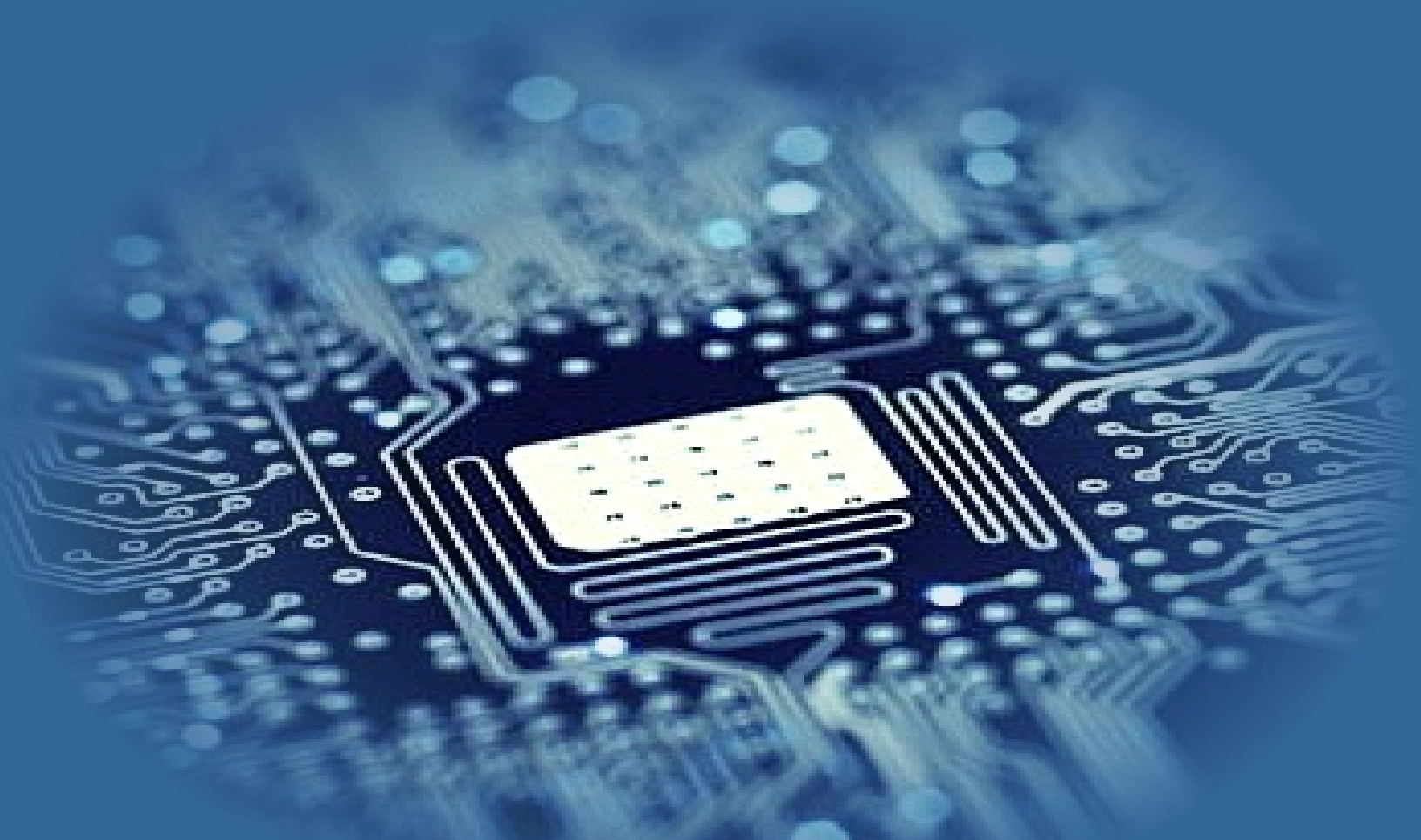
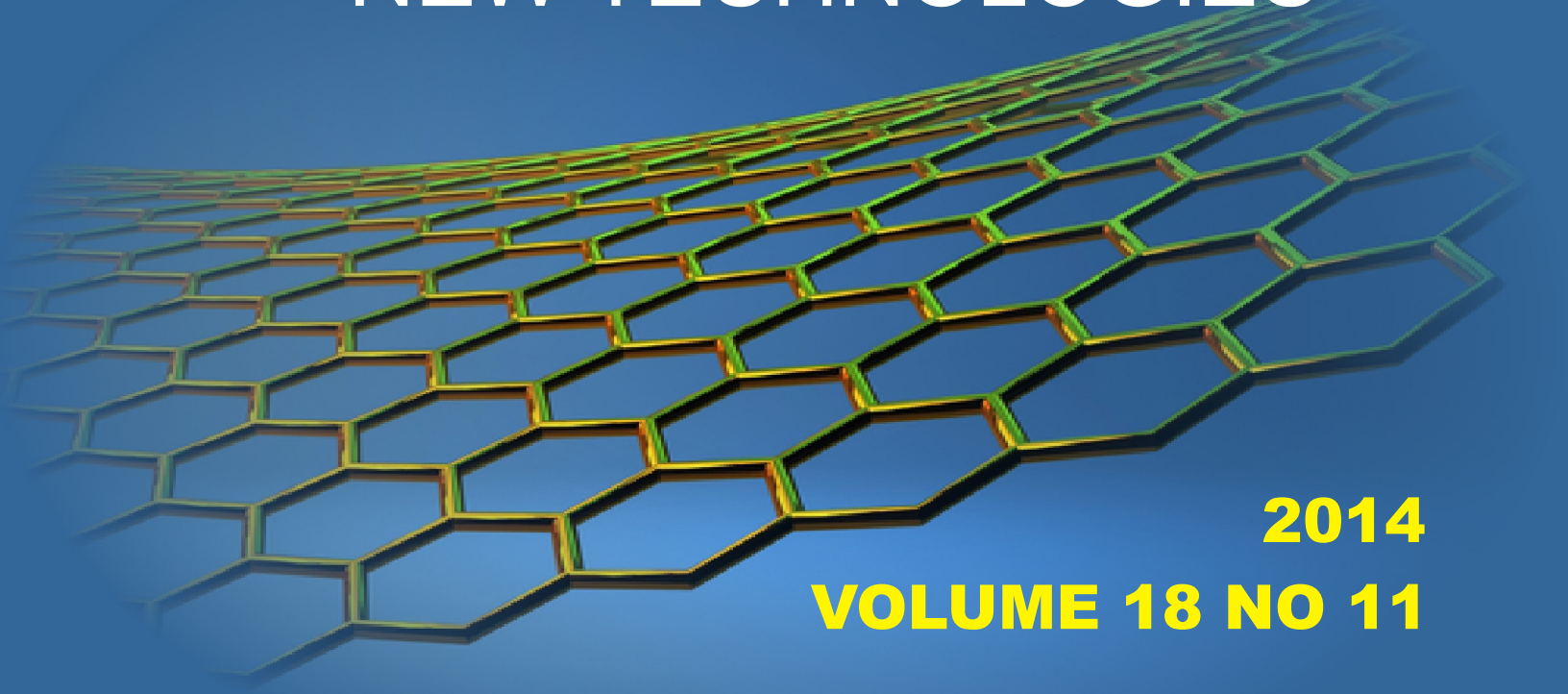


COMPUTER MODELLING AND NEW TECHNOLOGIES



2014
VOLUME 18 NO 11

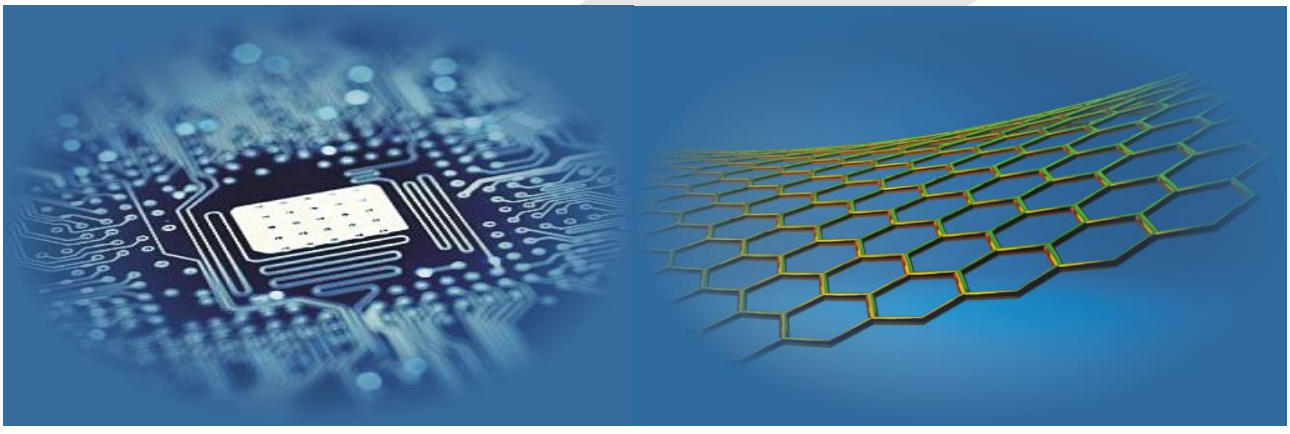
ISSN 1407-5806 ISSN 1407-5814 on-line

Latvian Transport Development and Education Association

Computer Modelling and New Technologies

2014 Volume 18 No 11

ISSN 1407-5806, ISSN 1407-5814 *(On-line: www.cmnt.lv)*



Riga – 2014

Journal topics:	**Publisher**	**Supporting Organizations**
• mathematical and computer modelling • computer and information technologies • natural and engineering sciences • operation research and decision making • nanoscience and nanotechnologies • innovative education	Latvian Transport Development and Education Association	Latvian Academy of Sciences Latvian Operations Research Society Transport and Telecommunication Institute, Latvia Fraunhofer Institute for Factory Operation and Automation IFF, Germany International IT University , Kazakhstan

Articles should be submitted in **English**. All articles are reviewed.

EDITORIAL CORRESPONDENCE	**COMPUTER MODELLING AND NEW TECHNOLOGIES, 2014, Vol. 18, No.11** **ISSN 1407-5806, ISSN 1407-5814 (on-line: www.cmnt.lv)**
Latvian Transport Development and Education Association	**Scientific and research journal** **The journal is being published since 1996**
68 Graudu, office C105, LV-1058 Riga, Latvia **Phone: (+371) 29411640** E-mail: yu_shunin@inbox.lv http://www.cmnt.lv	The papers published in Journal 'Computer Modelling and New Technologies' are included in: **INSPEC,** www.theiet.org/resources/inspec/ **VINITI ,** http://www2.viniti.ru/ **CAS Database** http://www.cas.org/ **EI Compendex**

Content

Editors' Remarks

****

THE ANSWER

by Rudyard Kipling

A Rose, in tatters on the garden path,
Cried out to God and murmured 'gainst His Wrath,
Because a sudden wind at twilight's hush
Had snapped her stem alone of all the bush.
And God, Who hears both sun-dried dust and sun,
Had pity, whispering to that luckless one,
"Sister, in that thou sayest We did not well --
What voices heardst thou when thy petals fell?"
And the Rose answered, "In that evil hour
A voice said, `Father, wherefore falls the flower?
For lo, the very gossamers are still.'
And a voice answered, `Son, by Allah's will!'"

Then softly as a rain-mist on the sward,
Came to the Rose the Answer of the Lord:
"Sister, before We smote the dark in twain,
Ere yet the stars saw one another plain,
Time, Tide, and Space, We bound unto the task
That thou shouldst fall, and such an one should ask."
Whereat the withered flower, all content,
Died as they die whose days are innocent;
While he who questioned why the flower fell
Caught hold of God and saved his soul from Hell.

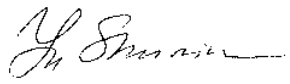
Rudyard Kipling (**1809-1849**) ♠

****

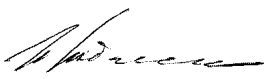
This 18th volume No.11 presents actual papers on main topics of Journal specialization, namely, **Mathematical and Computer Modelling, Computer and Information Technologies, Operation Research and Decision Making and Nature Phenomena and Innovative Engineering.**

Our journal policy is directed on the fundamental and applied sciences researches, which are the basement of a full-scale modelling in practice. This edition is the continuation of our publishing activities. We hope our journal will be interesting for research community, and we are open for collaboration both in research and publishing. We hope that journal's contributors will consider the collaboration with the Editorial Board as useful and constructive.

EDITORS



Yuri Shunin



Igor Kabashkin

♠ **Joseph Rudyard Kipling** (30 December 1865 – 18 January 1936) was an English short-story writer, poet, and novelist. He is chiefly remembered for his tales and poems of British soldiers in India and his tales for children. He was born in Bombay, in the Bombay Presidency of British India, and was taken by his family to England when he was five years old. Kipling is best known for his works of fiction, including The Jungle Book (a collection of stories, which includes and his poems, including "Mandalay" (1890), "Gunga Din" (1890), "The Gods of the Copybook Headings" (1919), "The White Man's Burden" (1899), and "If—" (1910). He is regarded as a major "innovator in the art of the short story"; his children's books are enduring classics of children's literature; and his best works are said to exhibit "a versatile and luminous narrative gift".

Modelling and simulation for the network-locomotive coupling of the co-phase continuous power supply and high speed railway

Shilong Chen*, Xingwang Li, Guihong Bi, Junxiang Rong, Ruirui Cao

School of Electric Power Engineering, Kunming University of Science and Technology, Kunming 650500, China

Received 1 October 2014, www.cmnt.lv

Abstract

Co-phase continuous power supply system (CCPSS) can completely eliminate electric phase break to realize the co-phase continuous power supply in the traction substation, thus the impact on the power quality of the utility grid can be greatly reduced. In this paper, double-loop control strategy, which involves the load current feedback for three-phase pulse width modification (PWM) rectifier and single-phase inverter in the traction substation were proposed. Also, the control strategy with neutral-point-potential balance for three-level rectifier as well as the constant V/f control for asynchronous motor of electric locomotive was proposed. A network-locomotive coupling model of an AC-DC-AC traction substation and electric locomotive was built in this paper using PSCAD/EMTDC. Thorough simulations were conducted to demonstrate the effectiveness of the proposed control strategies.

Keywords: CCPSS, AC-DC-AC electric locomotive of CRH2, control strategy, PSCAD/EMTDC

1 Introduction

In order to reduce the impact of traction power supply system on utility grid imbalance, the traction substation needs to be connected to the utility grid in turns; therefore an electric phase break section should be set at the output interface of a traction substation. Because of the existing of the electric phase break, the locomotive speed and traction power supply capacity were restricted seriously. To solve these technical problems, literature [1-3] put forward the scheme of co-phase power supply which based on active power filter and different connection forms of transformer. The method balanced the transform from three-phase AC to single-phase AC, and could compensate the harmonics and reactive power. However, these schemes could not realize the inter-connection of the power grid in a traction substation. Thus, it cannot cancel the electric phase break completely.

Literature [4-11] proposed a full transformation structure of the three-phase AC to DC to single-phase AC of the traction substation. In this way, there is no need of the electric phase break, so the electrified railway can be achieved to CCPSS. This structure reduced the impact of traction power supply system on utility grid power quality greatly. At the same time, the electric locomotives can be driven stable and safe high-speed.

In this paper, the network-locomotive coupling model of AC-DC-AC traction substation and electric locomotive was established using PSCAD/EMTDC.

Research the control strategies of traction substations and AC-DC-AC electric locomotives based on PWM.

2 The main structure of CCPSS

Compared with the original traction power supply system, the voltage magnitude, phase and frequency of traction network side of the CCPSS were the same, so the electrical phase break could be abolished completely, and it was suitable for high-speed rail operation. Since the substation used PWM symmetric transformation technology of three-phase rectifier to single inverter [12-16], this method could eliminate the interference between the utility grid and traction power supply system in the power quality. When the circuit breaker of the section post was closed, the substations could reach co-phase continuous power supply. The main structure was showed in Figure 1.

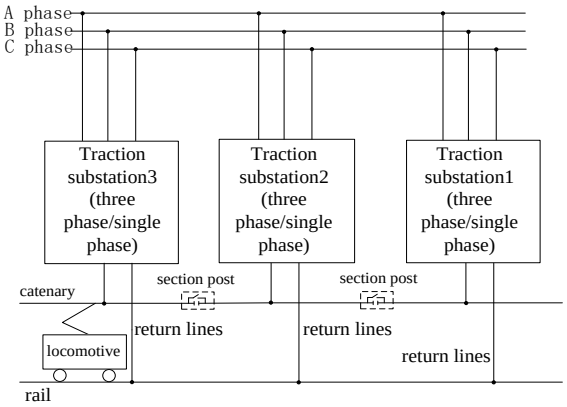


FIGURE 1 The main structure of the CCPSS

*Corresponding author e-mail: chenshilong3@126.com

3 The power converter's main circuit structure of the traction substation

This traction substation was based on the power converter of three-phase PWM rectifier and single-phase PWM inverter. Three-phase PWM rectifier was connected with the step-down transformer. When the DC voltage of rectifier was stable, the unity power factor should be achieved at the same time. Single-phase PWM inverter connected step-up transformer, on the one hand, the output voltage magnitude and frequency were stable; on the other hand, the reactive power, harmonics of the traction load should be compensated real-time. Due to the power supply mode of the traction substations was the AC-DC-AC symmetrical full transformation, traction power supply system and the utility grid were independent, and the three-phase load was balanced, so there was no negative sequence problems.

4 The power converter control strategy of the traction substation

4.1 THE CONTROL STRATEGY OF THE THREE-PHASE PWM RECTIFIER

Since the voltage mode of three-phase PWM rectifier [17] has several advantages, such as: adjustable power factor, low harmonics injected of AC side, DC side voltage stability and to work with the four-quadrant and two-way flow of energy, this control strategy was used in this article.

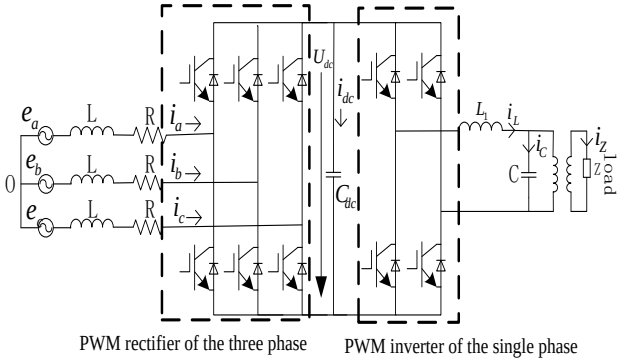


FIGURE 2 The main circuit configuration diagram of the power converter

This control of rectifier mainly included hysteresis current control, predictive current control, resonance control and linear PI controller [18]. Its control method had fast current response and robustness; by comparison, the voltage and current double closed loop PI controller [19] were used in this paper. Meanwhile, in order to improve the anti-jamming capability of rectifier side DC voltage, the paper increased the load current feedback compensation control, so that the power of three-phase PWM rectifier input and output could be balanced, the modulation method using SPWM. Figure 2 shown the equations of three-phase PWM rectifier circuit in ABC coordinates:

$$\begin{cases} C_{dc}\dfrac{du_{dc}}{dt} = i_a S_a + i_b S_b + i_c S_c - i_{dc} \\ L\dfrac{di_a}{dt} + Ri_a = e_a - u_{dc}S_a - u_{NO} \\ L\dfrac{di_b}{dt} + Ri_b = e_b - u_{dc}S_b - u_{NO} \\ L\dfrac{di_c}{dt} + Ri_c = e_c - u_{dc}S_c - u_{NO} \\ u_{NO} = -\dfrac{1}{3}u_{dc}(S_a + S_b + S_c) \end{cases} . \quad (1)$$

(S_a, S_b, S_c - Continuity and shutdown of the upper and lower bridge arm, u_{dc} - the rectifier side DC voltage i_a, i_b, i_c, e_a, e_b, e_c - the current and voltage of AC side; u_{NO} - the neutral point voltage). Through the stationary coordinate ABC to dq0 rotating coordinate transform, the equation of three-phase PWM rectifier under dq0 coordinates could be achieved as follows:

$$\begin{cases} L\dfrac{di_d}{dt} + Ri_d = e_d - U_{dc}S_d + \omega L i_q \\ L\dfrac{di_q}{dt} + Ri_q = e_q - U_{dc}S_q - \omega L i_d \\ C_{dc}\dfrac{du_{dc}}{dt} = \dfrac{3}{2}\left(i_d S_d + i_q S_q\right) - i_{dc} \end{cases} . \quad (2)$$

Let $U_d = U_{dc}S_d, U_q = U_{dc}S_q$, the following equation:

$$\begin{cases} U_d = e_d - L\dfrac{di_d}{dt} - Ri_d + \omega L i_q \\ U_q = e_q - L\dfrac{di_q}{dt} - Ri_q - \omega L i_d \end{cases} \quad (3)$$

In the Equations (2) and (3), i_d, i_q, e_d, e_q, respectively represented the current and voltage under the dq0 coordinate; where ω was the Angular frequency; S_d, S_q respectively represented continuity and shutdown function under the dq0 coordinate. i_{dc} isyhe rectifier side current.

The Equations (2) and (3) show that, when the current i_{dc} changed, the u_{dc} deviated firstly, then the system was adjusted, because the PI regulator delayed, the adjustment process of DC voltage u_{dc} resulted in large errors. In order to strengthen the stability of the DC voltage of three-phase PWM rectifier, a load current feedback compensation link was added. If we ignored line losses and switching losses, the Equation was:

$$\frac{3}{2}\left(e_d i_d + e_q i_q\right) = U_{dc}\left(C_{dc}\frac{du_{dc}}{dt} + i_{dc}\right), \quad (4)$$

when the system was stable, $i_q = 0$, $C_{dc}\dfrac{d_{uc}}{dt} = 0$ the equation was:

$$\frac{3}{2}e_d i_d = U_{dc} i_{dc}. \tag{5}$$

Because of the PI regulator, the current loop without static error tracking could be achieved, if we ignored the adjustment process of the current loop, front feed control variables of the load current was:

$$i_d^* = \frac{2}{3} \times \frac{U_{dc}}{e_d} \times i_{dc}. \tag{6}$$

Figure 3 showed the three-phase PWM rectifier joined the double-loop control system of the load current feedback loop:

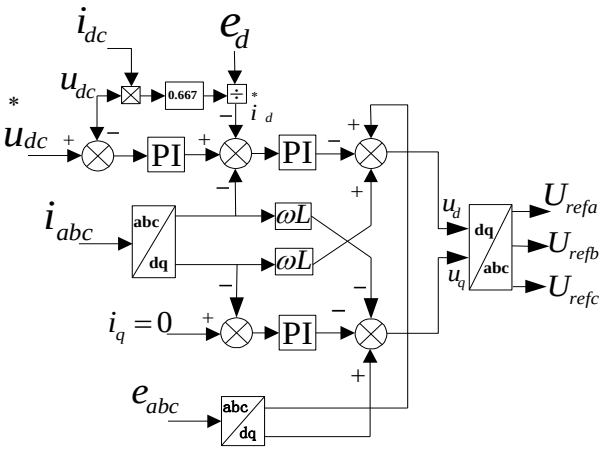


FIGURE 3 Joined the load current feedback loop of the double-loop control system

Double-loop PI control of three-phase PWM rectifier generated three-phase voltage as the reference signal was compared with the triangular carrier, and each switch action pulses to drive each switch-off and switch-on were got.

4.2 THE CONTROL STRATEGY OF SINGLE-PHASE PWM INVERTER

Because the double-loop controller had several characteristics [20, 21], such as: design simple, the output voltage waveform distortion small, fast dynamic response, and it required the power converter accurate tracking given traction load current, this control method was used in this article for the control of single-phase PWM inverter. The control system was shown in Figure 4.

The mathematical model of single-phase PWM inverter circuit was:

$$\begin{pmatrix} L_1 \dfrac{di_L}{dt} \\ L_1 \dfrac{du_c}{dt} \end{pmatrix} = \begin{pmatrix} u_{dc} \\ i_L \end{pmatrix} - \begin{pmatrix} u_C \\ i_Z \end{pmatrix}, \tag{7}$$

where i_L was the current of filter inductor; u_{dc} was the voltage of filter capacitor; i_z was the load current.

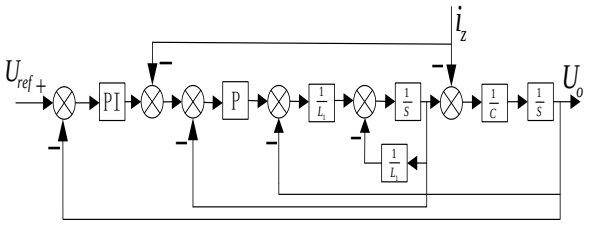


FIGURE 4 Single phase inverter control block

5 The converter main circuit structure of the AC-DC-AC electric locomotive traction

Currently, there are mainly two types of the traction converters，the one is the AC-AC converter, the other one is the AC-DC-AC converter.

The AC traction drive system generally adopted the following basic structure of AC-DC-AC electric locomotive:

The traction supply system of asynchronous motor power supply used voltage type AC-DC-AC converter; the traction supply system of synchronous motor power supply used current type AC-DC-AC converter, the traction supply system of synchronous motor power supply used AC-AC converter [22].

Compared with current type converter, the output current waveform of the voltage type converter contained less harmonic components [23]. Its performance was superior, the current waveform were closer sine. In summary, this paper adopted voltage type AC-DC-AC converter as the electric locomotive traction converters [24]. The main circuit configuration was shown in Figure 5.

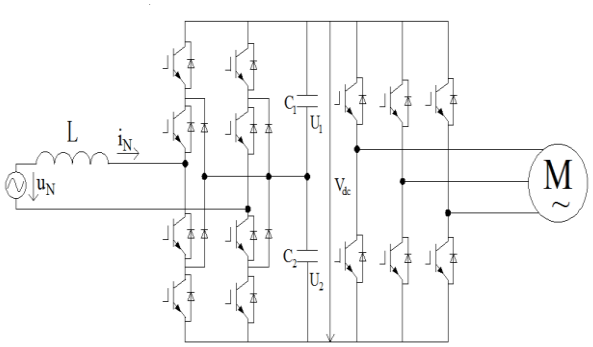


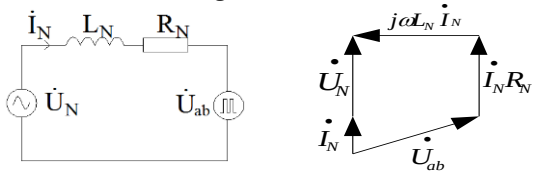
FIGURE 5 Traction converter main circuit diagram

The traction substation of decreasing from single-phase AC 27.5KV to 1.5KV single-phase AC; through three-level PWM rectifier to 2.6kV~3kV DC for the three-phase PWM inverter, three-phase PWM inverter and motor control as a whole, using the control mode of VVVF (variable voltage variable frequency).

5.1 THE CONTROL STRATEGY OF TRACTION CONVERTER

5.1.1 The control strategy of the single phase three-level PWM rectifier

Analysing rectifier side of Figure 5，the equivalent circuit diagram of the AC side of the rectifier and its control vector were shown in Figure 6.



a) Equivalent circuit diagram b) Control vector

FIGURE 6 Equivalent circuit and control vector of rectifier

In the Figure 6, where $\dot{U}_N$ was the voltage of the secondary winding of vehicle transformer, R_N , L_N ,were converted to resistance and leakage inductance of secondary side of the transformer windings respectively; $\dot{U}_{ab}$ was the fundamental vector of modulation voltage. Figure 6b showed the voltage vector equation of vehicle transformer secondary AC circuit:

$$\dot{U}_N = \dot{U}_{ab} + (R_N + j\omega L_N)\dot{I}_N . \tag{8}$$

The literature [25, 26] proposed an improved balance control with a midpoint voltage of transient current control method was used in the rectifier of the CRH2 locomotive. On the basis of Equation (8), when the compensated was increased the control Equation as follows:

$$\overset{*}{I}_{N1} = K_P(\overset{*}{U}_{dc} - U_{dc}) + \frac{1}{T_i}\int(\overset{*}{U}_{dc} - U_{dc})dt , \tag{9}$$

$$\overset{*}{I}_{N2} = \frac{I_{dc}U_{dc}}{U_N} , \tag{10}$$

$$\overset{*}{I}_N = \overset{*}{I}_{N1} + \overset{*}{I}_{N2} , \tag{11}$$

$$\overset{*}{U}_{ab}(t) = u_N(t) - (\overset{*}{I}_N R_N \sin\omega t + \overset{*}{I}_N \omega L_N \cos\omega t) - K[\overset{*}{I}_N \sin\omega t - i_N(t)] . \tag{12}$$

K_p, T_i were the proportional integral constant of PI regulator; K was the constant of PI regulator, i_N was the input current of AC side, other variables could be seen from Figure 7.

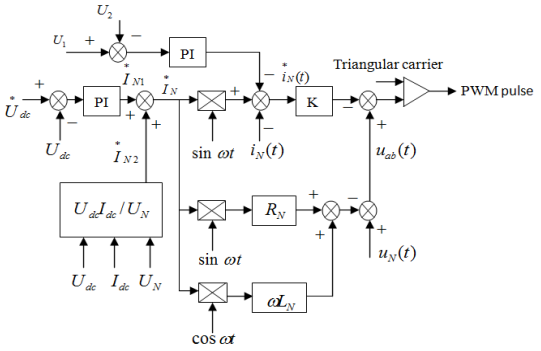


FIGURE 7 The control system block diagram of rectifier

The AC side current for a given value:

$$\overset{*}{i}_N(t) = \overset{*}{I}_N \sin\omega t - [K_{p1}(u_1 - u_2) + \frac{1}{T_{i1}}\int(u_1 - u_2)dt] - i_N(t) . \tag{13}$$

The traction converter rectifier control system block diagram was shown in Figure7.

5.1.2 The control strategies of the three-phase PWM inverter and motor

According to the knowledge about motor, electromagnetic torque *M* and *U/f* had a positive correlation.If the frequency *f* was adjusted, the voltage *U* should also be adjusted. So that the ratio $(U/f)^2$ could remained constant, and the flux ϕ remained unchanged, as well as the value of the torque *M* [27]. Control block diagram was shown in Figure 8.

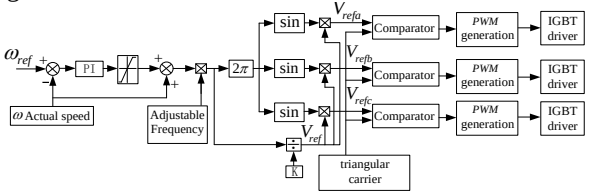


FIGURE 8 The control block diagram of the VVVF motor speed

6 Simulation results and analysis

Based on the PSCAD/EMTDC, the network-locomotive coupling model of AC-DC-AC traction substation and electric locomotive was established.

The main parameters were shown in Table 1.The other parameters of CRH2 according to the literature [28].

TABLE 1 The main parameters of the model

Three-phase power system	Three-phase rectifier side	Single-phase inverter side	Electric locomotive rectifier side
Fundamental frequency $f = 50Hz$	DC voltage set point $U_{dc} = 5KV$	Filter inductance $L_1 = 0.25mH$	Filter capacitance $C_1 = C_2 = 2200uF$
Source voltage $V_S = 110KV$	DC capacitor capacitance $C_{dc} = 9000uF$	filter capacitance $C = 130uF$	DC voltage set point $V_{dc} = 3KV$
three-phase step-down transformer ratio $110KV/3.5KV$	triangular carrier frequency $f_C = 1000Hz$	triangular carrier frequency $f_{C1} = 1200Hz$	triangular carrier frequency $f_{C2} = 1000Hz$

In this study, the pole-placement method was utilised [20], adjustment them through a lot of simulation experiments, to achieve the desired requirements.
The simulation results were as follows:

1) In this model, the co-phase continuous power supply system consisted of three traction substation, and three traction substations acquire power in the same utility grid together.

2) Traction substation output interface and rectifier of the waveform was shown in Figure 9 to Figure 13.

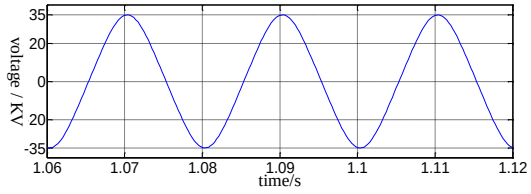


FIGURE 9 The voltage of traction substation 1

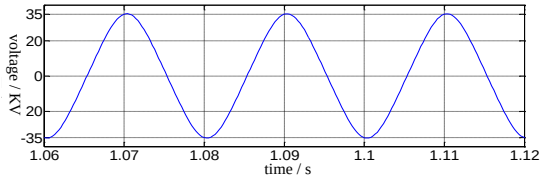


FIGURE 10 The voltage of traction substation 2

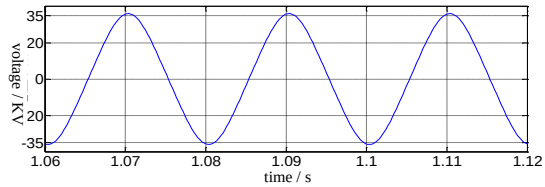


FIGURE 11 The voltage of traction substation 3

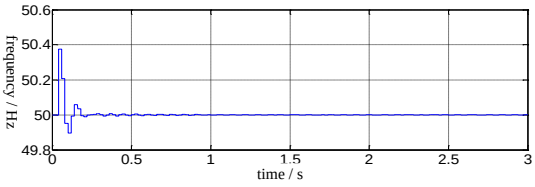


FIGURE 12 The frequency of traction substation

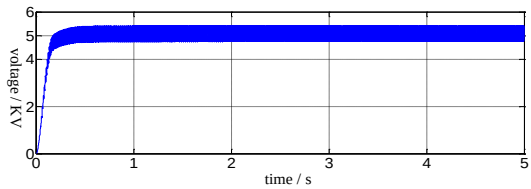


FIGURE 13 The DC voltage of traction substation rectifier side

It could be seen from Figure 9 to Figure 13 that in the case of electric locomotive load conditions, the voltage of traction substation rectifier side and the voltage of the output interface, frequency were stabilized at predetermined 5kV, 27.5kV, 50Hz,and harmonics injected and negative sequence were very low, have a high quality of voltage.

The rectifier side voltage of AC-DC-AC electric locomotive, current of the stator side and the stator flux trajectory waveform were shown in Figures 14-16.

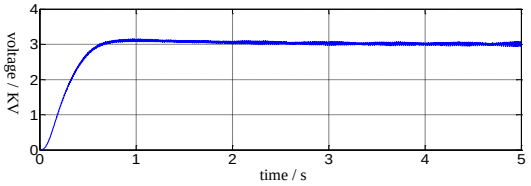


FIGURE 14 the voltage of electric locomotive rectifier side

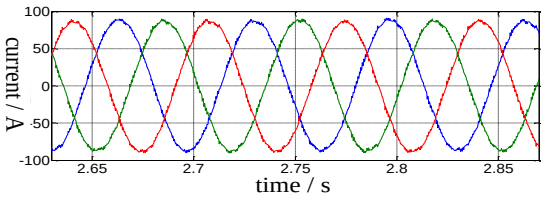


FIGURE 15 The current of the stator side

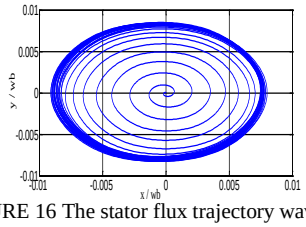


FIGURE 16 The stator flux trajectory waveform

It could be seen from Figure 14 to Figure 16 that, the voltage of electric locomotive rectifier side was stabilized at predetermined 3KV; the three-phase current of the stator side was symmetry. The stator flux trajectory waveform was approximate round.

7 Conclusions

This paper mainly studies the main circuit structure and control strategy of traction substation and AC-DC-AC electric locomotive. Application of electromagnetic transient simulation software PSCAD/EMTDC, it established the network-locomotive coupling model of AC-DC-AC traction substation and electric locomotive.

Through the analysis of this paper and the simulation results, the conclusions were as follows:

1) The voltage magnitude, phase and frequency were the same of three traction substation output interface. The electric phase break is completely cancelled, so the traction substation between the CCPSS is achieved. It reduces the impact of traction substation on the utility grid power quality greatly.

2) The double-loop control strategy of the loaded was suitable, not only the anti-interference ability be improved of DC voltage of three-phase PWM rectifier, but also make the DC voltage to track the command voltage fast.

Acknowledgements

This research is supported by National Natural Science Foundation of China (51267008), the personnel training fund of Kunming University of Science and Technology (kkz3201304019).

References

[1] Liu H 2012 Research on co-phase power supply system based on three phase transformer and DN power supply mode *Cheng du Southwest Jiaotong University* (*in Chinese*)
[2] Lv X, Zhang X 2006 Co-phase traction power supply system based on active filter and V/x connection *Proceedings of the CSU-EPSA* **18**(6) 73-8
[3] Zhang X, Gao S, Qian Q 2006 A novel co-phase traction power supply system based on impedance matching balance transformer and at power supply mode *Journal of the China Railway Society* **28**(4) 32-7 (*in Chinese*)
[4] Li Q 2010 Several key technical issues of the development of high-speed railway traction power supply *Journal of the China Railway Society* **32**(4) 119-24 (*in Chinese*)
[5] Liu X 2010 Control strategy of co-phase traction device of co-phase continuous traction power supply system *Cheng du Southwest Jiaotong University* (*in Chinese*)
[6] Li M 2008 The Simulation of a novel co-phase traction supply system D*aLian Da Lian Jiaotong University* (*in Chinese*)
[7] Wang Q 2012 Research on main circuit scheme of co-phase continuous traction power supply system *Cheng du: Southwest Jiaotong University* (*in Chinese*)
[8] Wang Q, Liu Z, Bai W 2009 Research on the simulation model of traction power supply system based on PSCAD/EMTDC *Power system protection and control* **37**(16) 35-45
[9] Hill R J. 1994 Electric railway traction. Part3: Traction power supplies *Power Engineering Journal* **8**(6) 275-86
[10] Dai N Y, Lao K W, Wong M C, et al. 2012 Hybrid power quality conditioner for co-phase power supply system in electrified railway *Power Electronics IET* **5**(7) 1084 -94
[11] Wei G, Li Q, Huang J 2008 A new co-phase traction power supply system *Automation of electric power systems* **32**(10) 80-3
[12] Xu M 2013 Design of three-phase to single-phase ac-dc-ac converter in co-phase-connected power supply system *Cheng du: Southwest Jiaotong University* (*in Chinese*)
[13] Wang Y 2013 Study of advanced co-phase traction power supply system based on three-phase to single-phase converter *Cheng du: Southwest Jiaotong University* (*in Chinese*)
[14] Lu H 2008 A novel traction supply system based on ac-dc-ac inverter *Da Lian Jiaotong University* (*in Chinese*)
[15] Chang F, Feng J, Zhao L 2012 Design of comprehensive power flow controller used in co-phase continuous traction power supply system [D] *Proceedings of the CSU-EPSA* **24**(1) 54-8 (*in Chinese*)
[16] Zhang R 2012 Research through in-phase inverter power supply system *Electric Railway* **23**(4) 19-22
[17] Huang S, Zhang T, Chen Y 2005 Investigation of an control strategy for three phase PWM voltage rectifier *Electrotechnical Application* **24**(3) 106-9
[18] Zhu Y, Ma H, Zhang Z 2006 Three-phase High Power Factor PWM Rectifier double-loop control system design *Electric Power Automation Equipment* **26**(11) 87-91 (*in Chinese*)
[19] Guo W, Lin F, Zheng Q 2006 The cascaded nonlinear PI control for three-phase voltage source PWM rectifier *Proceedings of the CSEE* **26**(2) 138-42
[20] Wang S, Peng L, Kang Y 2008 The research of PI digital dual-loop control method for PWM inverter based on pole assignment *Power Supply Technologies And Applications* **11**(9) 34-9 (*in Chinese*)
[21] Yang H, Song J 2008 Modeling and simulation of a single-phase voltage PWM inverter based on dual-loop control *Electrical Engineering* 5 72-5 (*in Chinese*)
[22] Guo M 2011 Mixed running simulation of ac-dc locomotive and ac-dc-ac locomotive *Cheng du: Southwest Jiaotong University* (*in Chinese*)
[23] Li G 2008 Research and realization on convertor control arithmetic of high-speed emu at traction state *Cheng du: Southwest Jiaotong University* (*in Chinese*)
[24] He G, Wang X 2010 Three-level PWM rectifier simulation and harmonic analysis of CRH2 electric multiple units *Journal of electric power* **25**(6) 455-7 (*in Chinese*)
[25] Xiao X 2012 Research on neutral point voltage of emu three-level traction converter *Nan Chang: East China Jiaotong University*
[26] Zhang Z, Zhang Z 2008 Study of control strategies and neutral point voltage balancing of single phase three-level rectifier *Electric Drive for Locomotives* 4 34-8 (*in Chinese*)
[27] Yu X, Peng C, Shi T 2012 Modeling and analysis of harmonic currents of type CRH2 electric locomotive *Engineer-ing Journal of Wuhan University* **45** (001) 107-10
[28] Deng X 2008 Traction drive system of 200 km/h crh2 emus *Electric Drive For Locomotives* 4 1-8 (*in Chinese*)

Authors

Shilong Chen, born in 1973, Si Chuan, China

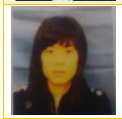
Current position, grades: associate professor, PhD.
Scientific interest: research of HVDC transmission and electrified railway traction power supply system.
Publications: 10 papers.

Xingwang Li, born in 1988, Tian Jin, China

Current position, grades: graduate student in the Kunming University of Science and Technology.
Scientific interest: electrified railway traction power supply system.
Publications : 1 paper, 1 patent.

Guihong Bi, born in 1968, Yun Nan, China

Current position, grades: professor, postgraduate, PhD
Scientific interest: research of power quality signal processing and control.



Jun Xiang Rong, born in 1985, Shan Dong, China

Current position, grades: graduate student in the Kunming University of Science and Technology.
Scientific interest: HVDC control in Kunming University of Science and Technology.

Rui Rui Cao, born in 1986, Shan Dong, China

Current position, grades: graduate student in the Kunming University of Science and Technology.
Scientific interest: HVDC control in Kunming University of Science and Technology.

Research on aircraft burst fault diagnosis based on T-S fuzzy neural network

Hui Yang, Lijing Wang*, Zheng Deng

College of Computer Science and Technology, Civil Aviation University of China, No. 2898 JINBEI Road, Dongli District, Tianjin, China

Received 1 October, www.cmnt.lv

Abstract

Aircraft burst fault is uncertainty and ambiguity. Considering QAR data as the research object, the fault diagnosis system based on the T-S fuzzy neural network combined with aircraft maintenance processes is built. First, the system designs the network performance oversight function to improve genetic neural network program. Then the fuzzy logic is used to deal with fuzzy rules, which can determine the location and severity of fault. And the result proves that the system has strong ability to deal with the questions.

Keywords: aircraft burst fault, QAR data, fuzzy rules, genetic neural network algorithm, T-S fuzzy neural network

1 Introduction

QAR is used for recording flight parameters and working conditions of the various components of airborne equipment [1]. It is streaming data acquired from the sensor, and its sampling frequency is once every second. QAR record details of the true state about aircraft in flight, and the data provide a strong basis to the real-time status monitoring and fault diagnosis of the aircraft.

U.S. National Aeronautics and Space Administration (NASA) have developed APMS (Aviation Performance Measuring System) software to build a model to identify abnormal QAR data, which can found the problems in the mass flight data [2]. UK CAA research centre devote to the Insight FDM system for airlines flight data analysis [3]. However, QAR data are only invoked when the faults happen and need to be removed, resulting in the gradual deterioration of some minor malfunction and causing serious consequences.

This paper uses T-S fuzzy neural network for fault diagnosis, then processes the QAR data modularly according to the structure characteristics of T-S fuzzy neural network system. Firstly, the BP neural network is improved to solve its existing problems for processing data, then making use of the expert experience and knowledge to turn the expert's experience into fuzzy logic rules, and adding fuzzy logic system to the improved BP neural network. Due to the excessive number of the input fuzzy rules, which caused by the size of input attribute values, the parallel network structure is raised in the fault diagnosis of T-S fuzzy neural network. Thus, faults could be detected and repaired according to the discrimination of fault types and assessment of failure severity detected by the output data.

2 The structure of fault diagnosis system

2.1 RELATED PROPERTIES

QAR data has the following characteristics: the sampling frequency is once per second, and the amount of data is very large; meanwhile, the QAR data is strictly in accordance with time, so the relationship between data is strong; besides, the QAR data has many uncertain factors such as some interference. QAR records lots of attribute values, but only parts have close relationship when particular fault occurs.

Related properties of air turbulence and air parking are analysed according to attribute reduction method based on rough set [4] and the maintenance experience provided by airlines, and eight related attributes are statistical: EPR (engine pressure ratio), ALV (vertical acceleration), N1 (low pressure rotor speed), N2 (high pressure rotor speed), EGT (exhaust gas temperature), TAT (total temperature probe signal), V1 (N1 vibration) and V2 (N2 vibration).

2.2 SYSTEM STRUCTURE

Neural network has the advantages of parallel computing, the ability of self-learning and nonlinear mapping ability. The complex relationship between a given input to the desired output can be realized through the network. It has been widely used in processing time series data such as fund price predicting [5], stock trend prediction [6] and so on. But the neural network is not suitable to process expert knowledge based on rules, while ignoring the system of expert experience knowledge often have a lot of defects in engineering; and fuzzy logic is suitable for processing some fuzzy or uncertain knowledge, but it does not have the ability of adaptive learning [7].

***Corresponding author** e-mail: wljroom@163.com

Aircraft quick access recorder (QAR) stores complex and a large number of data. Simply using the BP neural network to detect the abnormal data presents some problems, such as network convergence instability, slow and low accuracy. Therefore, on the basis of BP neural network, we propose an improved genetic neural network. Because of the defect of improved neural network model in lacking of using the expertise knowledge to handle aircraft emergencies troubleshooting process, we introduce the T-S fuzzy neural network exploration to research aircraft fault diagnosis. Figure 1 shows the whole procedures.

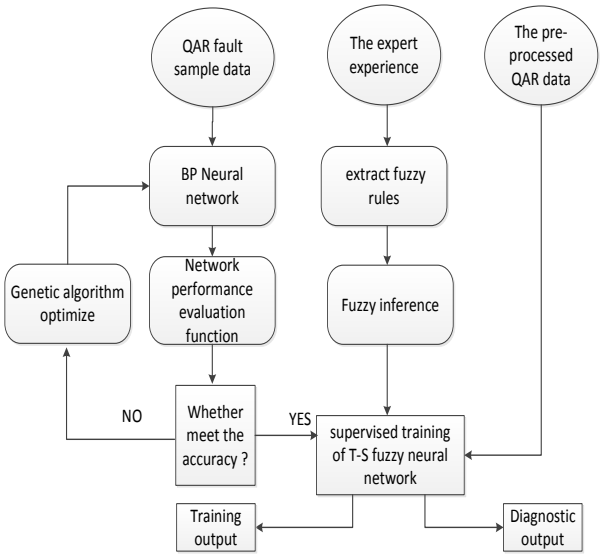


FIGURE 1 Fault diagnosis flow chart of T-S fuzzy neural network

3 Improved BP neural network algorithm

Back Propagation neural network is nonlinear mapping, parallel computing and self-learning [8]. However, simply using BP neural network to process fault QAR data has many problems such as the poor stability of convergence and low speed. As global search optimization algorithms, Genetic Algorithms is always used to improve BP neural network [9]. According to the characteristics of QAR data, we design a function to evaluate network convergence performance.

The network training times are divided into some parts with the same number, then calculating and recording the average error of each section, comparing the current average error with the last average error and the error setting before. If the current is greater than the set and less than the last, it shows that the internet has not yet reached saturation, continue training; if the current is greater than the set and equal to the last, it indicates that the network has been plunged into saturation, and genetic algorithms should be introduced. Terminate network training, extract weights and thresholds in BP neural network and encode as chromosome, which is used to form population with randomly generated chromosomes for genetic operation. The improved genetic neural network flow chart is shown in Figure 2.

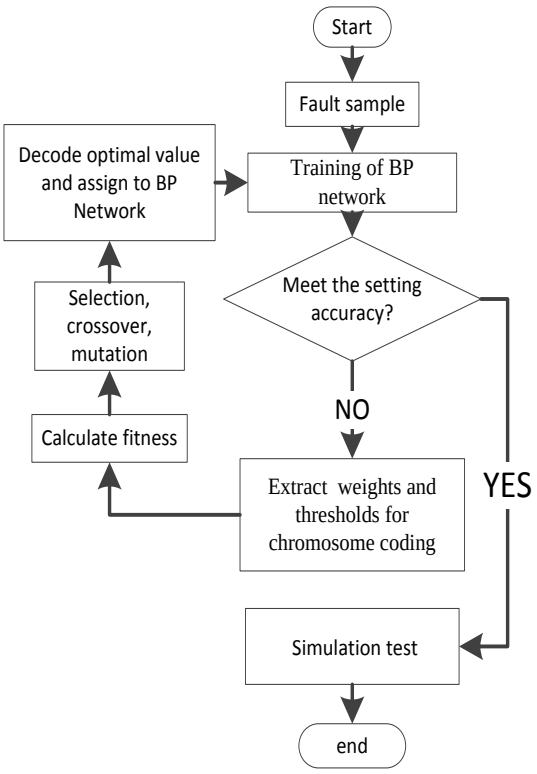


FIGURE 2 The process of improved genetic neural network

Based on the flow chart of improved genetic neural network, the steps of the algorithm can be designed for the diagnosis of QAR abnormal data.

1) Enter the eight values of air turbulence and air parking $X=\{x_1,x_2,\cdots x_8\}$, named EPR, ALV, N1, N2, TAT, EGT, V1 and V2, and the two kinds of output fault types are air turbulence and air parking. The appropriate number of neurons in the hidden layer based on empirical values and experiment is eight.

2) The input feature value is passed to the hidden layer neurons through the input layer neurons, and the transfer function is S-shaped function:

$$f(x)=1/1+e^{-ax}, \tag{1}$$

where a is the slope factor of S-function which can be adjusted freely. After calculation of hidden layer neurons function, we can get the output:

$$H_j=f\left(\sum_{i=1}^{8}\sum_{j=1}^{8}w_{ji}*X+\theta_j\right), \tag{2}$$

where w_{ji} is random initialization weights matrix between input layer and hidden layer; θ_j is random initialization threshold matrix of hidden layer neurons function, and the linear transfer function of output layer is:

$$\phi(x)=ax+b. \tag{3}$$

And the final output is:

$$O_k = \phi\left(\sum_{j=1}^{8}\sum_{k=1}^{2} \upsilon_{kj} * H_j + \theta_k\right), \tag{4}$$

where υ_{kj} is the random initialization weights matrix between hidden layer and output layer; θ_k is random initialization threshold matrix of output layer, and the total error is:

$$E = \frac{1}{2}\sum_{s=1}^{S}\left(O_s - T_s\right)^2, \tag{5}$$

where E is the sum of squared errors between the supervise values and the actual output value after sample data is trained, S is total number of the sample data, and the error back propagation algorithm named train is used.

3) Determine whether to use genetic algorithm to optimize BP neural network by the network performance evaluation function. Set the training times to n, $[1,n]$ represents the interval of training times, the value of n is set based on the sample size combined with the experiment, while the principle is to make the network speed as fast as possible. Training times will be divided as the same interval, and the quantity is $k = L/n$. $\overline{E}$ is the mean square error for each interval training of network, so

$$\overline{E} = E_1 + E_2 + \cdots + E_n / n. \tag{6}$$

Within such a range of k, we can obtain a row vector of error: $\left[\overline{E_1} \quad \overline{E_2} \quad \cdots \quad \overline{E_k}\right]$, for any two elements of the vector $\overline{E}_i$ and $\overline{E}_{i+1}$ $\left(i \in 1,2,\cdots,k\right)$, set $E_{trainerror}$ as error precision of network, if BP neural network training error satisfies the following Equation:

$$\begin{cases} \overline{E}_{i+1} > E_{trainerror} \\ \overline{E}_i > \overline{E}_{i+1} \end{cases}. \tag{7}$$

It means the BP neural network has not reached saturation point, continue training; if BP neural network training process error satisfies the following Equation:

$$\begin{cases} \overline{E}_{i+1} > E_{trainerror} \\ \overline{E}_i = \overline{E}_{i+1} \end{cases}. \tag{8}$$

It means the network reaches saturation and the error does not converge to the range of setting accuracy, then turn into the genetic algorithm optimization.

4) Terminate the train of BP neural network and extract the neural network weights and thresholds matrix when the neural network reached saturation. Do real number code according to the order of weight matrix of input layer to the hidden layer, the threshold matrix of hidden layer, weight matrix of hidden layer to the output layer and the threshold matrix of output layer, and the code can be expressed as

$$chrom = (w_{ji}\ \theta_j\ v_{kj}\ \theta_k). \tag{9}$$

Randomly generate the rest of the chromosomes to form the initial population.

5) Each chromosome is decoded as the BP neural network weights and thresholds, and the actual output of the network can be drawn from the sample data entered. The corresponding chromosome fitness is the reciprocal value of the sum of squares of the error between the actual output and the supervising:

$$f(i) = 1\Big/\frac{1}{2}\sum_{s=1}^{S}\left(O_s - T_s\right)^2, \tag{10}$$

where s is the number of samples. Crossover and mutate on population until genetic algorithm optimization process achieve the iteration times. Find out the chromosome with maximum fitness value, and decode the optimal chromosome as weights and thresholds matrix according to the inverse operation of encoding, then assigned the values to BP neural network, turn to steps 2 to continue training;

6) If the network converges, finishing the training, and the network can be used to QAR data detection; if not, turn to step 4).

4 Fuzzy neural network model

T-S model uses the consequent of fuzzy rules as function of the input linguistic variables [10]. In the fuzzy reasoning part of T-S network, fuzzy rules will exponentially grow with the input fuzzy partition number increasing. If fuzzy partition number is too many, it will result in the dimensionality curse of rules [5]. In order to reduce the difficulties, we set two T-S fuzzy neural networks apart for fault types. And the network framework is as follows:

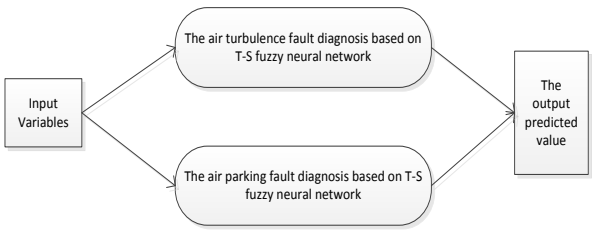


FIGURE 3 The overall design of the T-S network diagnosis system

We set air turbulence T-S fuzzy neural network as example to build network structure. The network structure is shown in Figure 4. Air parking T-S fuzzy neural network can be structured with the same approach of air turbulence, and then constitute a parallel network.

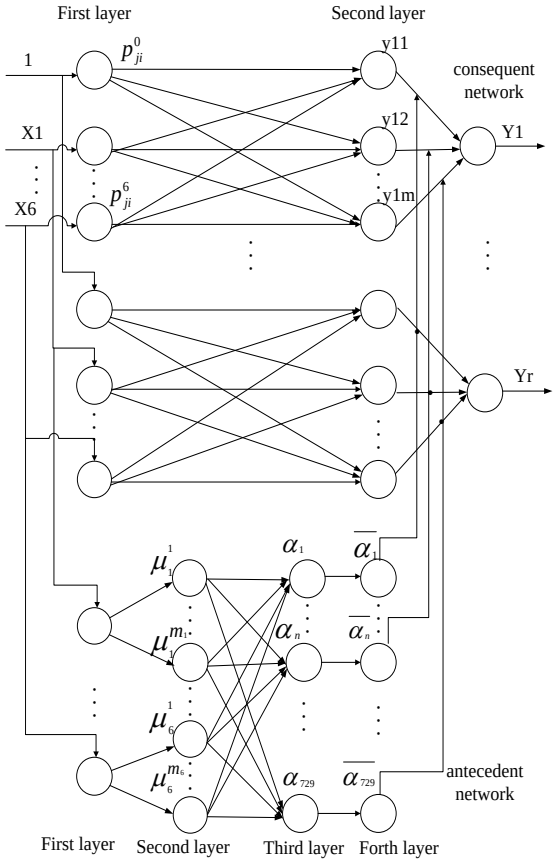


FIGURE 4 T-S network diagnosis structures

4.1 ANTECEDENT NETWORK

Antecedent network consists of four layers. The first layer is the input layer, and the input values $x=[x_1, x_2 \ldots x_6]^T$ represent 6 attribute values associated with air turbulence. The number of nodes in first layer is $N_1 = n$. At the second layer, each node represents a linguistic variable values, such as positive, zero, negative. Its role is to calculate the membership degree of the input component part $\mu_i^{S_i}$, where $\mu_i^{S_i} = \mu_{A_i^{S_i}}(x_i)$ $(i=1,2,\cdots,6; S_i=1,2,3)$. Input dimension number is 6, and fuzzy division number of x_i is 3. Membership function is Bell-shaped represented by Gaussian functions.

$$\mu_i^{S_i} = \exp - \left\{ \frac{\left(x_i - c_{i_{S_i}}\right)^2}{\sigma_{i_{S_i}}^2} \right\}, \tag{11}$$

where C_{iS_i} and σ_{iS_i} denote the centre and the width of the membership function. The number of nodes in second layer is $N_2 = \sum_{i=1}^{n} m_i = 6 \times 3 = 18$. Each node of the third layer represents a fuzzy rule, and its role is to match the antecedent of the fuzzy rules and calculate the applicable of each rules. The value of a_j is

$$a_j = \min\left\{\mu_1^{S_{1j}}, \mu_2^{S_{2j}}, ..., \mu_n^{S_{nj}}\right\} \tag{12}$$

and the necessary value to calculate a_j is

$$\begin{cases} S_{1j} \in \{1,2,...,m_1\} \\ S_{2j} \in \{1,2,...,m_2\} \\ ... \\ S_{nj} \in \{1,2,...,m_n\} \\ j = 1,2,\cdots,m \\ m = \prod_{i=1}^{n} m_i = 3^6 = 729 \end{cases}. \tag{13}$$

The number of nodes in third layer is $N_3 = m = 729$. For the given input, only linguistic variables near it could have greater membership grade, and those far away from it have smaller values. Therefore only some nodes have greater output, which is equal to a local approximation network. The fourth layer has the same number of nodes as the third layer, $N_4 = N_3 = m = 729$. It mainly do normalized operation

$$\overline{\alpha}_j = \alpha_j \Big/ \sum_{i=1}^{m} \alpha, \; j = 1,2,\cdots,729. \tag{14}$$

4.2 CONSEQUENT NETWORK

This part is composed of parallel sub-networks, and the output is calculated through the network.

The first layer of sub-network is the input layer; it transfers the input variables to the second layer. The node No.0 has value $x_0 = 1$, and its role is to provide the constant term for consequent fuzzy rules. The nodes in the second layer represent fuzzy rules, and its role is to calculate consequents for each rule. The Equation is as follows:

$$y_{kj} = p_j^{k_0} + p_j^{k_1} x_1 + \cdots + p_j^{k_n} x_n = \sum_{l=0}^{n} p_j^{k_l} x_l, \tag{15}$$

where $k = 1,2,\cdots,r$; $j = 1,2,\cdots,m$; $x_0 = 1$.The third layer of sub-network is to compute system's overall output:

$$y_k = \sum_{j=1}^{m} \overline{\alpha}_j y_{kj}, \; (k = 1,2,\cdots r), \tag{16}$$

where y_k is the weighted sum of consequent rules, and the weighting coefficients is the connection weights of third layers of the antecedent network.

4.3 ALGORITHMS OF T-S NETWORK

Weights of the consequent network p_{ji}^k $(j=1,2,\cdots,729; i=1,2,\cdots,6; k=1,2,3)$, central value of membership function c_{ij} and width σ_{ij} $(i=1,2,\cdots,729; \; j=1,2,3)$ should be installed if the T-S fuzzy neural network is used, which could make the linear

mapping from input to output complicated. Set the error cost function as follows:

$$E=\frac{1}{2}\sum_{k=1}^{3}\left(y_{dk}-y_k\right)^2, \tag{17}$$

y_{dk} and y_k are expected output value and actual output value. Learning algorithm of p_{ji}^k, c_{ij} and σ_{ij} is as follows:

1) Set weights between nodes in the network p_{ji}^k, central value of membership function c_{ij} and width σ_{ij} randomly.

2) Input the sample data of air turbulence, then train the network according to the following formulas:

$$\frac{\partial E}{\partial p_{jl}^k}=\frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial y_{kj}}\frac{\partial y_{kj}}{\partial p_{jl}^k}=-\left(y_{dk}-y_k\right)\overline{\alpha_j}x_l. \tag{18}$$

$$p_{jl}^k(t+1)=p_{jl}^k(t)-\beta\frac{\partial E}{\partial p_{jl}^k}=p_{jl}^k(t)+\beta\left(y_{dk}-y_k\right)\overline{\alpha_j}x_l. \tag{19}$$

$$j=1,2,\cdots,m;\ \ l=0,1,\cdots,n;\ \ k=1,2,\cdots,r.$$

$$c_{is_i}(k+1)=c_{is_k}(k)-\beta\frac{\partial E}{\partial c_{is_i}}, \tag{20}$$

$$\sigma_{is_i}(k+1)=\sigma_{is_i}(k)-\beta\frac{\partial E}{\partial \sigma_{is_i}}, \tag{21}$$

$$i=1,2,\cdots,n;\ \ s_i=1,2,\cdots,m^i\ \ ;\beta>0,$$

and β is learning rate. Regulate the values of p_{ji}^k, c_{ij} and σ_{ij} with the formulas above. And set up a network error oversight function (Specific constructor of the function is described in Section 2.2.1). Monitor the error of output value.

3) If the error is reduced continuously in the setting times of training, then go on using the Equations in step 2 to calculate the error convergence. If the error keeps constant or has minor change in the setting times of training, suspend the network training process temporarily and introduce genetic algorithm to find the optimal values of p_{ji}^k, c_{ij} and σ_{ij}, then assign the values to the network node. Turn to step 2 and continue to train. If the error can reach to the setting precision, it means the network has finished training, stop the training process; if not, go to step 2 and continue training.

5 Experiments and analysis

According to maintenance experience and analysis of failure data, we sort out the six properties and the severity of fault associated with the air turbulence in Table 1. The severity of fault can be inferred through the level of air turbulence fault.

This experiment is designed with the MATLAB R2009a system. 800 groups of sample data collected from the B737-800 aircraft are used, while 750 groups are randomly selected as the training samples and the remaining 50 groups are testing samples. The training times is =1000. According to the repeated experiment, when the learning rate is 0.006, the network has the best effect, and the error precision of training is 0.01. Considering the influence of the uncontrollable factors, the simulation output values representing fault classification should be allowed to exist reasonable error, and the error range in this article is as shown in the Table 2.

TABLE 1 Division of property values and fault level

	Normal	Relatively normal	Slight	Relatively heavy	Heavy
Engine pressure ratio	1.45	1.3	1.2	1.1	1
Vertical acceleration (m/s2)	0.95	0.8	0.7	0.6	<0.5
VIB Vibration ()	0.5	0.4	0.3	0.2	<0.1
Low pressure rotor speed (r/s)	120	110	95	80	<65
High pressure rotor speed (r/min)	90	80	65	45	<25
Exhaust temperature (℃)	650	550	450	400	<400
Output value	1	2	3	4	5

TABLE2 The error range of simulation output

	Normal	Relatively normal	Slight	Relatively heavy	Heavy
Simulation output value (x)	0.5<x<1.5	1.5<x<2.5	2.5<x<3.5	3.5<x<4.5	4.5<x<5.5

Air turbulence fault experiment is done according to the given data and T-S fuzzy neural network system. After experiments, we find that when the training times is 33~56, the error can reach the setting precision. Select the data which could reach the precision within 40 times, and draw an error graph as shown below:

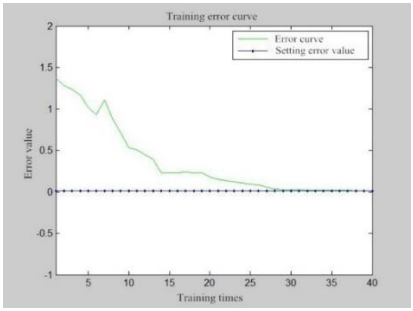


FIGURE 5 Training error curve

The picture is taken from an intermediate value after many training times. As shown in the figure, there is a smooth curve at 15-20 times in the process of training, and the network turns into the genetic algorithm optimization process through the network supervision function, then continue to train after the parameters are optimized until reach the setting error precision value. In order to test the rationality and validity of fault diagnosis system, 50 groups of sample data are used, and the results are as shown below:

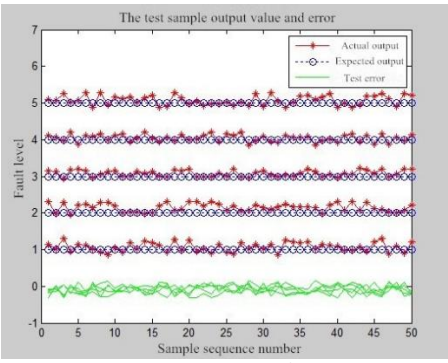


FIGURE 6 The output value of the test sample

As we can see, the result is in the permitted error, and the system can predict fault severity well, which can provide decision support and guidance for the repair. It also reaches our expectation.

6 Conclusions

In this paper, the aircraft faults warning system based on QAR data is realized. The system includes two fault diagnosis subsystems, air turbulence and air parking. Fault models based on QAR data and QAR outlier detection algorithm has been researched and realized. The parallel double network thought has been referred to separate the experimental data and import them to different fault diagnosis sub-system, which greatly simplifies the increasing quantity of fuzzy rules with the input. At the same time, the network training algorithm based on improved genetic algorithm is introduced into the antecedent and consequent parameters in the network training adjustment. The experimental results show that the algorithm has a good effect on sudden fault. It can diagnose the fault type and severity of aircraft effectively.

Because of the complexity of QAR data and information, only two fault types are built, and the fault models combine technical data with expert experience provided from airlines. Whether we need to build more models should be further studied.

Acknowledgments

The research work was supported by The National Natural Science Foundation of China and Chinese aviation jointly funded project No. 61179063.

References

[1] Lin C 2011 QAR principle analysis and application in aircraft maintenance *Jiangsu Aviation* **1** 36-7 (*in Chinese*)
[2] McCown G D, Patricia M, Conway, T J, Jessen K M 1991 *Methods and Apparatus for Monitoring System Performance* USA 5067099
[3] Knotts R M H, Exeter M A 1999 Civil aircraft maintenance and support Fault diagnosis from a business perspective *Journal of Quality in Maintenance Engineering* **5**(4) 335-48
[4] Li Y, Zhu S, Chen X, Zhang D, Han Z 1999 Data mining model based on rough set theory *Tsinghua Univ (Sci &Tech)* **39**(1) 110-3 (*in Chinese*)
[5] Xiao G 2011 Application Research of Fund Net Value Prediction Based on BP Neural Network *Computer simulation* **28**(3) 373-6
[6] Wang W, Chen L 2012 Modeling and Simulation of Stock Price Prediction *Computer simulation* **29**(1) 344-7
[7] Wang L, Guo Y 2010 T-S Fuzzy Neural Network Based Control of Accelerating Process of Turbo-fan Engine *Computer simulation* **27**(2) 26-9
[8] Zhang P, Wang Y, Fan S 2009 Civil Aviation Aircraft Fault Diagnosis Research Based GA and BP Net *Microcomputer Information* **25**(7) 110-2
[9] Li S, Liu L, Xie Y 2011 Chaotic prediction for short-term traffic flow of optimized BP neural network based on genetic algorithm *Control and Decision* **26**(10) 1581-5
[10] Ke Y, Xie B, Wu Qing 2012 The Vehicle Fault Diagnosis Research Based on T-S Model Fuzzy Neural Network *Journal of Hangzhou Dianzi University* **32**(2) 41-4 (*in Chinese*)

Authors

Hui Yang, born in 1957, Tianjin, China

Current position: professor.
Research direction: artificial intelligence, data mining and fault diagnosis.
Publications: many articles and a book.

Lijing Wang, born in 1990, Hebei, China

Current position: graduate student in the Civil Aviation University of China.
University studies: artificial intelligence, data mining and fault diagnosis.
Scientific interest: fault diagnosis.
Publications: 1 paper.

Zheng Deng, born in 1987, Henan, China

Current position: graduate student.
University studies: artificial intelligence, data mining and fault diagnosis.
Scientific interest: fault diagnosis.
Publications: 1 paper.

Design and implementation of a parallel algorithm based on Hadoop platform

Qingnian Zhang*, Zhao Chen, Zihui Wang

Wuhan University of Technology, China

Received 1 October 2014, www.cmnt.lv

Abstract

Existing clustering algorithm is transplanted into the Hadoop cloud computing platform, through the low price on the computer cluster nodes dynamically allocate huge amounts of data distributed task, solve the enterprise needs a large amount of data storage and the problem of real time analysis results. Graphs programming model can help developers to quickly realize the parallel clustering, and do not need too much to understand the specific underlying communication realization. This article will improve the clustering algorithm, which is transplanted into graphs on the programming model, realize the parallel design, and through the error sum of squares criteria such as function test and verify the reliability of the parallel algorithm. Under the Hadoop cluster composed of four machines respectively samples of different sizes of data clustering analysis, proves that the parallel algorithm of Hadoop platform on the large data applications better speedup and scalability.

Keywords: hadoop platform, mapreduce, clustering, k-mean

1 Introduction

Hadoop open source distributed cloud computing platform by the Hadoop distributed file system and the graphs of programming model. Existing clustering algorithm is transplanted into the Hadoop cloud computing platform, through the low price on the computer cluster nodes dynamically allocate huge amounts of data distributed task, solve the enterprise needs a large amount of data storage and the problem of real time analysis results [1, 2]. Graphs programming model can help developers to quickly realize the parallel clustering, and do not need too much to understand the specific underlying communication realization [3, 4].

Through learning clustering algorithm in data mining, this paper implemented the Hadoop distributed platform parallel Kmeans algorithm, effectively solves the massive user data classification problem.

2 Data clustering application framework design

This system adopts the Hadoop platform Hadoop distributed file system to store user data, with the clustering algorithm into graphs programming for data clustering. Figure 1 is the flow chart for the application of this system framework, which mainly consists of data collection module, data preprocessing module, data storage module and data mining analysis module [5, 6].

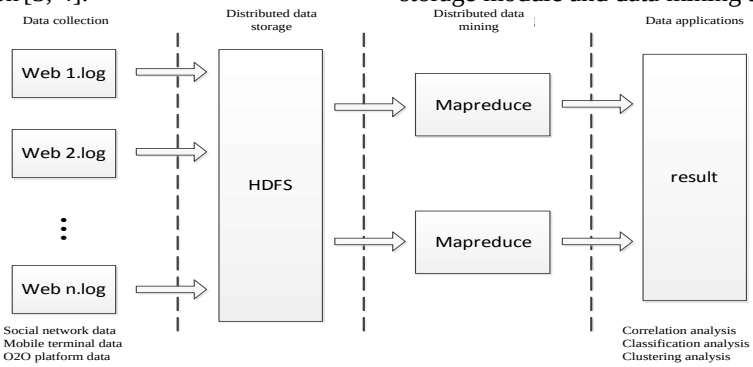


FIGURE 1 System frame

Mapreduce programming model can effectively solve the parallel processing data in the distributed storage and fault tolerance, how could the Hadoop platform to realize the big data mining work, the key is the transfer of the mining algorithm to graphs programming model. Under the framework of the k-means algorithm is transplanted into graphs, the concrete implementation process programming model as shown in Figure 2.

*Corresponding author e-mail: jieyang509@163.com

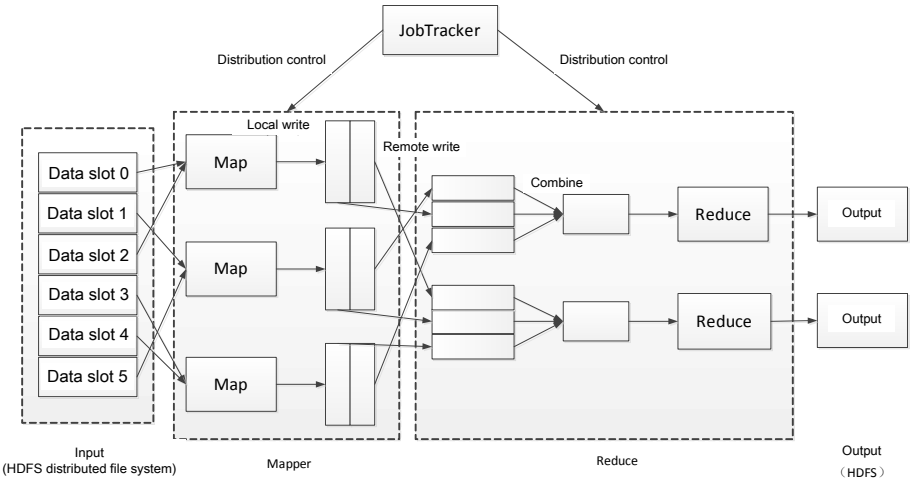


FIGURE 2 Mapreduce programming model specific implementation process

3 The site data clustering application framework

3.1 HARDWARE AND SOFTWARE CONFIGURATION

Website user behaviour data not only are there more big quantity, type, and for real-time analysis of the results, according to the characteristics of this system USES Hadoop platform Hadoop distributed file system to store user data, with transplanted into graphs programming under the clustering algorithm to cluster the user consumption behaviour [7, 8]. By the results of clustering, the difference of customers to provide high quality service, in view of the current marketing activities of the problem, provide favourable data support.

Using 4 PC computer in ubuntu12.04 platform structures, a small cluster with four nodes. Choose one of the highest frequencies of CPU a master, as the main nodes at the same time set up the master and the other three as a slave node. Master node configuration is 1 GB memory size, from the node configuration memory size is 512 m, all 20 GB hard drive size distribution. Every node in the cluster need to install the same version of Hadoop and under the same path JDK version, through a router connection to the machine, every day in the same local area network (LAN). Cluster structures, different from the pseudo distributed mainly in two aspects: the static IP Settings and SSH communications. After launch the Hadoop in a small cluster environment, on the master there should be the JobTracker, SecondaryNameNode, the NameNode, DataNode and TaskTracker five daemons, while slave1, slave2 and slave3 should have DataNode and TaskTracker two daemons. Table 1 below is the Hadoop cluster software and hardware configuration.

TABLE 1 hardware and software configuration information

	master	**slave1**	**slave2**	**slave3**
cpu	2.5GHz	1.79GHz	1.83GHz	1.93GHz
memory	1G	512M	512M	512M
hard disk	20G	20G	20G	20G
operating system		Ubuntu12.04, Windows XP		
open tools		Hadoop-1.0.2, Eclipse, JDK1.6		

3.2 THE HADOOP NODE CONFIGURATION

Hadoop is mainly composed of two basic modules: graphs programming model and Hadoop distributed file system. Graphs programming model is mainly used in distributed applications, this model is a JobTracker and several TaskTracker. After the client send a job application to the JobTracker, JobTracker is responsible for the operations of initialization, job scheduling and monitoring TaskTacker task. As long as find task failure, JobTracker will restart it, guarantee the stability of the cluster. TaskTracker is running on the slave service, responsible for executing the JobTracker distribution of parallel tasks [9, 10]. HDFS distributed file system is used for distributed storage, as the underlying support of graphs, the nodes in the cluster consists of one NameNode and more DataNode. The NameNode's role is to manage the file system namespace, the function of the DataNode is according to the need to store and retrieve data block. This experiment is configured with one NameNode and four DataNode, Table 2 for the specific planning of Hadoop environment.

TABLE 2 graphs and HDFS node configuration information

	master	**JobTracker/TaskTracker**
Mapreduce	slave1	TaskTracker
	slave2	TaskTracker
	slave3	TaskTracker
	master	**NameNode/DataNode**
HDFS	slave1	DataNode
	slave2	DataNode
	slave3	DataNode

3.3 SIMILARITY MEASURE METHOD

This paper uses the modified method of similarity measure in the map function to achieve the effect of normalized processing. Traditional k-means algorithm is used in the graphs, the map function defined in the two samples x_a, x_b similarity is the use of Euclidean distance, the smaller the $s(x_a, x_b) = \sqrt{\sum_{j=1}^{n}(x_a - x_b)^2}$, smaller $s = (x_a, x_b)$ shows that the greater the similarity between samples. Sample calculation and $s = (x_a, x_b)$ size of each centre, and clustering to the smallest clusters of $s = (x_a, x_b)$ distance class. In order to achieve the result of the normalized, modify the judgment criterion of the map function, to redefine the sample similarity distance is:

$$s'(x_a, x_b) = \sum_{j=1}^{n} \frac{1}{1 + |x_a - x_b|}. \quad (1)$$

Similarity between each attribute was defined as the reciprocal of the Manhattan distance plus 1, the greater the value of $s = (x_a, x_b)$ shows the greater the similarity between the sample data. By improving the similarity measure of $s = (x_a, x_b)$ calculated data sample and each centre distance, the result value, the greater the explain data samples is more similar with the centre. Due to the improved similarity measure $s = (x_a, x_b)$ to map samples of each attribute to between 0 and 1, guarantee the contribution of each attribute function in the same frame of reference, have the effect of the normalized. Such clustering result also has more explanatory.

4 Parallel algorithm design

4.1 CONSTRUCTION OF THE CLUSTER

Cluster is the master node and slave nodes distribution on different computers. In the practical application of Hadoop cluster scale is compared commonly big, but the main purpose of this paper is to validate the Hadoop in the usability of the website user data clustering, this simulation environment for the four nodes Hadoop cluster, transplanted clustering algorithm into Hadoop platform.

4.2 DESIGN AND IMPLEMENTATION OF PARALLEL ALGORITHM

In the Hadoop platform the key step for realizing the parallel Kmeans algorithm is to design the Map function and a Reduce function. Every process of Mapreduce is equivalent to a serial clustering algorithm in an iterative process. The Map function is to get the data object distance to the centre node, and distributing the data object to the nearest cluster class, and marking the category ID. The Reduce function completes new centre in the same cluster class value calculation. Relatively new clustering centre, compared with the previous clustering centre, when the error is less than the convergence value, the clustering process is over, otherwise the circulation process of Mapreduce. For the convenience of the recording and processing the data and Mapreduce calculation model uses the line form for storage, as there is no correlation, piece of data to be processed data line can be shard. To ensure the independence of the clustering process each iteration the data can be distributed to store the data node.

4.2.1 The map function design.

The main content of the Map function is calculated each data to all the selected centre distance, and put each data tag to the centre of the shortest distance, belong to the same data centre is a bunch of classes. The Map function is < key, value > way to input and output. The input data is the centre of the previous iteration (or the centre of the random selection) and all the sample data. Input function < key, value > object relative to the corresponding is the key data in the data file offset of the starting point and the value corresponding to the data object coordinate values of each dimension. Output data is each data object belongs to the ID, the output function < key, value > value or represent data objects in the coordinates of each dimension, and the key value indicates that the clustering centre of the data object. The Map function to realize pseudo code:

```
void map(LongWritable key,Text value){
mis_distance=getEuclideanDistance(point,cluster[0]);
for(int i=0; i<k; i++){
If(getEuclideanDistance(point, cluster[i])<mis_distance){
mis_distance=getEuclideanDistance(point, cluster[i]);
  currentCluster_ID=i;
  }
}intermediate_output(currentCluster_ID, point);
}
```

4.2.2 Reduce function design

The main content of the Reduce function is to the same cluster the data averaged class, get a new centre, if the centre of two adjacent results don't happen deviation, clustering results are produced. If the result deviation, this time the Reduce output centre as the initial centre of the next iteration. Reduce function of the input is the output of the Map function, the results of the same key value is assigned to the same Reduce calculation (i.e., the same cluster class assigned to a Reduce in). Corresponding form input function < key, value > for < cluster ID of the list (belong to the centre of the data objects) >, calculate the same in each Reduce the number of the class and each component and cluster computing the mean of each component of the clustering centre as a new file. Output function < key, value > for < > cluster ID, sample vector.The Reduce function implementation of pseudo code:

```
void reduce(Writable key, Iterator<PointWritable> points){
long num=0;
Float sum=0.0f;
While(points.hasNext()){
   PointWritable current_point=points.next();
   num+=current_point.getNum();
   for(int i=0; i<dimension; i++)
   sum[i]+=current_point[i];
   for(int i=0; i<dimension; i++)
   mean[i]=sum[i]/num;
}result_output(key,mean);
}
```

5 Similarity measure experiment and analysis

5.1 CLUSTERING CRITERION

Clustering analysis of the target is to belong to the similar research object into a class, belong to the same object of a class as similar as possible, and not the same kind object is different, as far as possible, the basic clustering criterion contains three functions.

5.1.1 The error sum of squares criterion function.

Data sets $R=\{r_1,r_2,r_3,...,r_n\}$, by clustering into *k* clusters, namely $R_1,R_2,R_3,...R_k$, each class contains data $m_1,m_2,m_3,...,m_k$ for a number of clusters, defined error sum of squares criterion function is:

$$j_a=\sum_{j=1}^{k}\sum_{n=1}^{m_j}\left\|r_n^{\ j}-c_j\right\|^2, \tag{2}$$

r_n^j of them belong to the same cluster class *j* of sample data. While c_j said *j*-th clusters mean all the samples in the class, the class cluster centre, its representation is:

$$c_j=\frac{1}{m_j}\sum_{j=1}^{m_j}r_j,\ j=1,2,3,...,k. \tag{3}$$

Error variance criterion function mainly calculation belong to the same cluster samples of a class and the class of cluster to Euclidean distance to the centre of the square, when the smaller the value of j_a, suggests that cluster-heads data between difference is smaller, the clustering effect is good. In the process of clustering iterations, similar data will be allocated according to the judgment of the distance to the same cluster, j_a function present a descending trend. However, only use this a judgment standard is not enough, the error sum of squares criterion suitable for sample data density is bigger, small differences between the clusters of judgment, in the case of within the cluster sample difference is very big, the error sum of squares criteria often cannot render good effect [11].

5.1.2 The weighted average of the sum of squares of the criterion function

This function is shown by the following:

$$j_b=\sum_{j=1}^{k}p_j f_j^{\ *}. \tag{4}$$

The p_j is the prior probability, of the total number of samples for *m*, assigned to the *j*-th class cluster of number for m_j, is the representation of a prior probability method is:

$$p_j=\frac{m_j}{m},\quad j=1,2,3,...,k \tag{5}$$

and $f_j^{\ *}$ said *j*-th cluster-heads mean value of the square of the distance between two objects, because m_j number within the cluster is, the combination of two number is $m_j\left(m_j-1\right)$, expression is:

$$f_j^{\ *}=\frac{2}{m_j(m_j-1)}\sum_{r\in R_j}\sum_{r'\in R_j}\left\|r-r'\right\|^2. \tag{6}$$

In view of the error variance criterion function cannot very good judgment sample data of markedly different defects, weighted average of the sum of squares of good make up for it. Rule of the main computing the sum of squared distance between classes in the same cluster sample. As the clustering iterations, similar samples gathered in the same cluster, cluster sample within the distance between the smaller and smaller, the weighted average of the sum of squares of the criterion function said local area density, will present a downward trend.

5.1.3 The distance between the class and the criterion function

The above two criteria are judgment within the cluster, the similarity between data, and the distance between the class and the criterion function is more focused on different types of separation degree, the rule of the well of the distance between each cluster are described, the expression is as follows:

$$j_c=\sum_{j=1}^{k}\left(c_j-c_{all}\right)^T\left(c_j-c_{all}\right). \tag{7}$$

Expression of c_j said all the samples in the class is the first j cluster averages, c_{all} said average of all the sample data. In the process of clustering iterations, each cluster degree of dissimilarity between more and more obvious, the distance between the class criterion function present a rising trend.

Taken together with the increase of clustering iterations, the error sum of squares criterion function and

weighted average of the sum of squares of the criterion function as class data cluster distance more and more close, property is more and more similar, function values can present a downward trend, show the same kind object difference as small as possible. And the distance between the class criterion function with data between clusters of separation, function values can present a rising trend, show differences between different objects as large as possible.

5.2 THE RESULT OF THE EXPERIMENT

This experiment from UCI Machine Learning Repository site selection of the iris data set, wine data set, the data set and the vehicle data sets upload the four sample data to hadoop distributed file system. Then the graphs programming under the framework of Kmeans algorithm to cluster analysis, and use and square error criterion function, weighted average of the sum of squares of the criterion function. The distance between the class and criterion function on the process of clustering analysis, observation is related to the changes in the process of iteration function, verify the reliability of the parallel clustering algorithm.

The iris sample set a total of 150 data, each data contains 4 kind of attribute, the whole data set can be divided into 3 clusters; Wine sample set a total of 178 data, each data contains 13 kinds of attributes, the whole data set can be divided into 3 clusters; Seed data samples with 210, each sample has eight kinds of attributes, samples can be classified into 3 clusters; Vehicle data samples with 94, each sample has 18 kinds of attributes, samples can be classified into four clusters. Because the Kmeans algorithm is randomly selected from the initial value, different initial value selection may get different clustering results, so this paper 20 times for each criterion function clustering operation, a randomly selected as the experimental results. At the beginning of the iterative function change is more obvious, therefore this thesis excerpts from the first five iterations as trend analysis. Table 3-6 of iris respectively, wine, seed, vehicle data aggregation class effect.

TABLE 3 Iris data aggregation class effect

Number of iterations	Error sum of squares	Weighted average sum	Distance between the class sum
1	3.143×10^{3}	1.278	10.354
2	2.653×10^{3}	1.169	10.803
3	2.487×10^{3}	1.147	11.028
4	2.067×10^{3}	1.137	11.286
5	2.048×10^{3}	1.120	11.652

TABLE 4 Wine data aggregation class effect

Number of iterations	Error sum of squares	Weighted average sum	Distance between the class sum
1	8.580×10^{7}	3.155×10^{4}	2.068×10^{5}
2	6.815×10^{7}	2.821×10^{4}	2.832×10^{5}
3	7.233×10^{7}	2.821×10^{4}	2.847×10^{5}
4	7.045×10^{7}	2.773×10^{4}	2.823×10^{5}
5	6.814×10^{7}	2.729×10^{4}	2.854×10^{5}

TABLE 5 Seed data aggregation class effect

Number of iterations	Error sum of squares	Weighted average sum	Distance between the class sum
1	6.045×10^{4}	9.971	30.261
2	4.672×10^{4}	8.874	30.921
3	3.438×10^{4}	7.162	31.992
4	2.467×10^{4}	6.448	31.968
5	2.243×10^{4}	5.973	31.594

TABLE 6 Vehicle data aggregation class effect

Number of iterations	Error sum of squares	Weighted average sum	Distance between the class sum
1	1.051×10^{7}	2.047×10^{4}	1.019×10^{5}
2	6.991×10^{6}	1.324×10^{4}	1.140×10^{5}
3	7.115×10^{6}	1.184×10^{4}	1.309×10^{5}
4	6.475×10^{6}	1.137×10^{4}	1.440×10^{5}
5	5.798×10^{6}	1.132×10^{4}	1.468×10^{5}

By the form can be found on the four sample data shows that with the increase of the number of iterations. Clustering similar sample data set in the same class makes the error sum of squares criterion function. The weighted

average of the sum of squares of the function presented a decreasing trend, and differences between class and class increases gradually, the distance between the classes and functions presented an increasing trend. But the three functions is not monotone changing, in the process of clustering iterative search criterion function of the uncertainty of the optimum path, the criterion function results may fluctuate, but the overall trend in line with the desired results. At the same time can also be observed in three previous iterations, the change of the function value is relatively obvious, to late function value change is small, the clustering results. The experiment proves that the graphs under the programming model of parallel Kmean algorithm, has good clustering effect. Under the Hadoop cluster composed of four machines respectively samples of different sizes of data clustering analysis, proves that the parallel algorithm of Hadoop platform on the large data applications better speedup and scalability.

References

[1] Lu H, Hu T-t 2012 Research on Hadoop Cloud Computing Model and its Applications *Networking and Distributed Computing(ICNDC)* 59-63

[2] Yu Hong, Wang D 2012 Mass Log Data Processing and Ming Based on Hadoop and Cloud Computing *Computer Science & Education (ICCSE)* 197-202

[3] Singh B, Singh H K 2010 Web Data Mining Research: a Survey *Computational Intelligence and Computing Research (ICCIC)* 1-10

[4] Kala Karun A, Chitharanjan. K A 2013 Review on hadoop -HDFS infrastructure extensions *IEEE Conference on Information and Communication Technologies* 132-7

[5] Dean J, Ghemawat S 2004 MapReduce: symplified data processing on large clusters O*SDI* 1-12

[6] Zhou J, Liu Z 2008 Distributed clustering based on k-means and CPGA *Fuzzy Systems and Knowledge Discovery* **2** 444-7

[7] Hai M, Zhang S, Zhu L, Wang Y 2012 A survey of distributed clustering algorithms *Industrial Control and Electronics Engineering* 1142-5

[8] Hirzel M, Andrade H, Gedik B 2013 IBM Streams Processing Language: Analyzing Big Data in motion *International Business Machines Corporation* 1-11

[9] Garlasu D, Sandulescu V, Halcu I, Neculoiu G A 2013 Big data implementation based on Grid computing *Roedunet International Conference (RoEduNet)* 17-9

[10] Kala Karun. A, Chitharanjan K A 2013 Review on hadoop -HDFS infrastructure extensions *IEEE Conference on Information and Communication Technologies* 132-7

[11] Pakhira M K 2009 Clustering large databases in distributed environment *IEEE International Advance Computing Conference* 351-8

Authors

Qing Nian Zhang, China

Current position, grades: professor of School of Transportation at Wuhan University of Technology since 2002.
University studies: PhD degree in Machinery design and theories from the Wuhan University of Technology, China, in 2002.
Scientific interest: traffic and transportation planning, optimization and decision making of transportation system, transportation safety management.

Zhao Chen, born in 1969, Shaanxi Province, China

Current position, grades: PhD degree in Logistics management at Wuhan University of Technology, China.
University studies: MS degree in Transportation management engineering from Wuhan University of Technology, China, in 2002.
Scientific interest: parallel computing on hadoop platform, optimization decision.

Zihui Wang, born in 1990, Wuhan, China

Current position, grades: M.S. degree in Electronics and Communication Engineering at Wuhan University of Technology, Wuhan, China.
University studies: B.S. degree in Electronic Communication Engineering from Wuhan University of Technology, Wuhan, China, in 2012.
Scientific interest: signal processing on Hadoop platform, pattern recognition.

Research on the scope of capacity for different EDAs

Caichang Ding[1, 2, 3], Wenxiu Peng[2*]

[1]*State Key Lab of Software Engineering, Wuhan University, Wuhan 430072, China*

[2]*School of Computer, Wuhan University, Wuhan 430072, China*

[3]*School of Computer Science, Yangtze University, Jingzhou 434023, China*

Received 1 March 2014, www.cmnt.lv

Abstract

In this paper we investigation the scope of capacity for different EDAs to successfully solve problems, which concern to the mutual effects among the variables. More specifically, we study the learning restrictions that different EDAs confront to solve problems, which can be expressed by some ADFs. The research is conducted in the worst situation. The sub-functions in the ADFs are the same deceptive functions. We think that the capacity for this kind of algorithm are primarily influenced by the probabilistic model they depend on. We employ three different kind of EDAs so as to investigate the effect that the complexity of the probabilistic model has on the behave of the algorithm. Because the population size is crucial for EDAs, we use different population sizes in the experiments. Nevertheless, the results indicate that, in general, enlarge population size is not useful to solve more complex problems.

Keywords: EDAs, capacity, complexity, problem structure, probabilistic model

1 Introduction

Estimation of distribution algorithms (EDAs) [1-4] are a new kind of evolutionary algorithms (EAs) that use probabilistic models instead of the typical genetic operators used by genetic algorithms (GAs) [5, 6]. The features of the search space in EDAs are extracted by machine learning methods. In EDAs, employing a probabilistic model to represent the collected information which is used to generate new individuals later. By this way, probabilistic models can lead the search to hopeful areas of the search space.

Mathematically, an optimization problem can be seen as the minimization or maximization of a given function. Thus, optimization problems can be formulated as,

$$x^* = \arg\max_x f(x), \tag{1}$$

where $f: S \to \mathrm{R}$ is called the objective function or fitness function, $x = (x_1, \ldots, x_n) \in S$ is a candidate solution of the problem and S is named the problem space. In most cases, the optimum x^* is not unique. In this paper, the problem space S is an n dimensional discrete space.

Because the EDAs [2, 4] can capture the structure of the problem, EDAs are considered to be more efficient than GAs. In EDAs, the specific interactions among the variables of solutions in the problem are taken into mind. In evolutionary algorithms, the interactions are display implicitly in mind; whereas in EDAs, the interrelations are showed explicitly through the joint probability distribution associated with the individuals of variables selected from each generation. The probability distribution is calculated according to a population of selected candidate solutions of previous generation. Then, offsprings are sampled from this probability distribution generate. Neither crossover nor mutation has been used in EDAs. Figure 1 displays the Pseudo-code of Estimation of Distribution Algorithms.

1 $D_{t=0} \leftarrow$ gerate N individuals randomly
2 **Do**
3 $D_t \leftarrow$ evaluate individuals
4 $D_t^{Se} \leftarrow$ select $M < N$ individuals from D_t according to a selection method
5 $p_t(x) = p(x \mid D_t^{Se}) \leftarrow$ estimate the joint probability distribution by means of a probabilistic model
6 $D_{t+1} \leftarrow$ sample M individuals from $p_t(x)$ and creat the new population
7 $t = t + 1$
8 **until** stopping criterion is met

FIGURE 1 Pseudo-code of Estimation of Distribution Algorithms

An EDA has such basic elements [2]: encoding of candidate solutions, objective function, selection of parents, building of a structure, generation of offspring, selection mechanism, and algorithm parameters like population size, selection size, etc.

Different EDAs [2] mainly differ in the kind of probabilistic models employed and the approaches used to learn, then sample from the obtained models. Bayesian

[*]***Corresponding author*** e-mail: hamigua_ping@hotmail.com

networks is one of the models that has been widely used in EDAs. One of the advantages of EDAs that employ these kinds of models is that the complexity of the learned structure relies on the characteristics of the selected individuals. Moreover, the analysis of the models learned during the search can provide useful information about the problem structure.

How the characteristic of the search space are reflected in the learned probabilistic models is another related and important issue in EDAs. This issue has received special attention from the EDA community, and is essential to understand the mechanisms that enable EDAs to efficiently sample from the problem space during the search process. However, the question of analysing the relationship between the problem space and the structure of the learned probabilistic models becomes more and more difficult because of the next two main reasons: the search process is random in the EDAs, and the methods used when learning the models can only detect approximate, suboptimal, structures.

To investigate the effect that the complexity of the probabilistic model has on the behave of the algorithm, we employ three different implementations. Firstly, an EDA which is called univariate marginal distribution algorithm (UMDA) [7] that is considered independent among the variables, secondly, an EDA which is called EDAdt [8] uses a dependency-tree model every generation and last, an EDA which is called estimation of Bayesian networks algorithm (EBNA) [9,10] uses Bayesian networks. Because the population size is crucial [11] for EDAs and particularly when the used probabilistic model is Bayesian networks, we employ varying population sizes.

The experimental results indicate that the ability of EDAs to solve the testing functions is lost quickly when the number of sub-functions exceeds a certain value in the objective function. This threshold of behave indicates a obvious phase-transition phenomenon that clearly delimits the frontiers of effectiveness. The experimental results also indicate that the ability to learn structures is very important to extend the restrictions of successful EDA implementations. However, EBNA displays a tremendous decreasing of performance because the learning method cannot build more sophisticated models. Moreover, according to the experimental results, the learning of unrestricted Bayesian networks requires a huge computational cost for solving the problems. Therefore, in order to solve the problems, the complexity of the models should grow exponentially along with growth of the number of sub-functions.

2 Probabilistic models

The probabilistic model associates to the qualitative and quantitative structure determined by the probability function. Though the EDAs could employ any kind of probabilistic model, the widely used models are those that show their qualitative component by a graph. Especially, one kind of models that has been widely used in EDAs are Bayesian networks.

Bayesian networks [12-14], which is called belief networks are a kind of probabilistic graphical model. This kind of probabilistic models is one of very popular paradigms that can deal with probability distributions efficiently in modelling uncertain information. The domain of expert systems is one of the most important sources for the development of Bayesian networks. Moreover, in the past few years, Bayesian networks have obtained considerable attention in domain of the machine learning. As a result of this attention, more and more papers and tutorials have appeared. Thus, besides expert systems, Bayesian networks are also applied in classification problems, bioinformatics and optimization.

Bayesian networks are the product of associating probability and graph theory [15], similarly with any other probabilistic graphical model. The graphical consist of the model encodes a number of conditional independences related to the probability distribution. Let $X=(X_1,\ldots,X_n)$ be an n dimensional discrete random variable. A Bayesian network is a graphical expression of the factorization of the joint probability distribution for X, $p(X=x)$, where $x=(x_1,\ldots x_n)$ is an assignment of the random variable X. More specifically, a Bayesian network can be expressed as a pair (s,θ_s)s where s is a directed acyclic graph that is model structure and θ_s is the set of parameters associated to the structure that is model parameters. The structure s determines the set of conditional dependences among the random variables of X. According to the structure s, the joint probability distribution $p(x)$ can be factorized by means of marginal and conditional probability functions. More specifically, the probability distribution factorizes according to the graph as,

$$p(x)=\prod_{i=1}^{n} p(x_i \mid pa_i), \tag{2}$$

where pa_i denotes a value of the variables Pa_i that is the parent set of X_i in the graph s.

The local probability distributions in the factorization are those which is induced by means of the product that appears in Equation (2). We suppose that these local probability distributions depend on the parameters $\theta_s=(\theta_1,\ldots,\theta_n)$. So Equation (2) could be rewritten by specifying the parameters:

$$p(x \mid \theta_s)=\prod_{i=1}^{n} p(x_i \mid pa_i,\theta_i) \tag{3}$$

Suppose that the variable X_i has r_i possible values, thus, the local probability distribution $p(x_i \mid pa_i^j,\theta_i)$ is an unbounded discrete distribution:

$$p(x_i^k \mid pa_i^j,\theta_i)=\theta_{ijk}, \tag{4}$$

where $pa_i^1,\ldots,pa_i^{q_i}$ represent the q_i possible values of the parent set Pa_i. The parameter θ_{ijk} denotes the probability of variable X_i which takes in its *k*-th value, at the same time, the set of its parents' variables takes in its *j*-th value. Therefore, the local parameters are determined by $\theta_i = ((\theta_{ijk})_{k=1}^{r_i})_{j=1}^{q_i}$.

In order to look for a Bayesian network [16, 17], which can make us to represent and deal with the uncertain knowledge of a specific field, setting both the structure and the parameters is very necessary. The structure and conditional probabilities that is necessary for describing the Bayesian network can be provided either externally by experts, by machine learning from datasets or by mixing both of these methods. In this paper. We mainly focus on the second method. Besides, when the structure has been automatically learned, it can provide us with perceptions into the interactions between the variables of the field.

The learning step can be separated into two subtasks that are structural learning and parameter learning [18-20]. Although there are different approaches to learn the structure of a Bayesian network, we mainly focus on the so-called score plus search method. This kind of methods copes with the structure learning as an optimization problem. Thus, the steps of learning a Bayesian network can be expressed as follows. Given a data set D containing N cases, $D=\{x_1,\ldots,x_N\}$, finding the structure s^* such that,

$$s^* = \arg\max_{s\in S^n} g(s,D), \tag{5}$$

where $g(s,D)$ is the score which measures the quality of any given structure s related to the data set D, and S^n is the set of all possible directed acyclic graphs (DAG) which have *n* nodes. A number of relevant and used heuristic techniques such as greedy search, simulated annealing, particle swarm optimization, genetic algorithms and ant colony optimization have been used in this task.

If score can be decomposed in presence of complete data sets, it is the one of the desirable character. These scores can be decomposed in sub-scores related to every node X_i and its parents Pa_i in the structure s. Formally, we can express a decomposable score as:

$$g(s,D) = \sum_{i=1}^{n} g_D(X_i, Pa_i), \tag{6}$$

where the g_D is the sub-score function. As a result of the decomposability, it is computationally more efficient when the local search is carried out, because when we add an arc into the network, it is only necessary to evaluate the set of nodes involved by this change.

When we have defined a score to assess Bayesian networks, we have to run a search process to look for the Bayesian network, which can return the best score given the data.

In practical applications, we must to look for an suitable model structure [21-25] as soon as possible. Thus, a simple learning method, which can find a relatively good structure, even though it is not best, is preferred. There are a number of learning methods, which can be employed for this task. However, a specific search algorithm which is called Algorithm B [19], which is typically used by most of Bayesian network based EDAs.

Algorithm B is a greedy search algorithm and its pseudo-code is presented in Figure 2, where D is a data structure which contains the information needed to deal with the candidate arcs which should be added into the network. At the beginning, Algorithm B starts from an arc-less structure, which represents independently among the variables, and at each iteration, an arc is added into network that can increase the score greatly. The algorithm stops when no arc that can increase the score any more, can be added into the network.

1 begin with an arcless strcture
2 compute $D[X_j \to X_i] = g_D(X_i, X_j) - g_D(X_i)$ for all distic X_i, X_j
3 **do**
4 find the largest $D[X_j \to X_i]$, and add an arc from X_j to X_i in the structure
5 $D[X_j \to X_i] = g_D(X_i, Pa_i \cup X_j) - g_D(X_i)$ for all disctinct X_i, X_j not belonging to Pa_i
6 $D[X_j \to X_i] = -\infty$
7 **until** every $D[X_j \to X_i] < 0$

FIGURE 2 Pseudo-code for Algorithm B

When we have got a Bayesian network by learning, this model could provide us with detailed probabilistic information which we are interested in. Usually, the information, which the researcher wants to know is the probability of some events on the basic of special observations. Generally speaking, the probabilities which we are interested in, are not reposted in the Bayesian network obviously. It is necessary to compute in order to obtain them,. This course is called probabilistic inference and it is usually an NP-complete problem.

Emulation of Bayesian networks, which is also named stochastic sampling, can be regarded as an option to the exact inference. The Emulation of a given probabilistic graphical model requires to get a sample from the probability distribution for X which the model encodes. Next, the marginal and conditional probabilities involved can be calculated from the sample.

For our goals about EDAs, the intension of the emulation of Bayesian networks is to get a new population in which the probability relations among the random variables of the network are potential. Specifically, for the purpose of sampling from the Bayesian network, the sampling method which we employ is forward. The variable must be sampled after all its parent variables have

been obtained. This approach is named probabilistic logic sampling (PLS). Figure 3 presents a pseudo-code of this approach.

```
1   π ← ancestral ordering of the nodes in the baysian network
2   for j = 1 to N
3     for i = 1 to n
4       x_{j,π(i)} ← randomly generate a value form p(x_{π(i)} | pa_{π(i)})
5     done
6   done
```

FIGURE 3 Pseudo-code of the probabilistic logic sampling method

3 Testing functions

In order to study the performance of EDAs when the complexity of the function grows, we handle additively decomposable functions (ADFs). This kind of functions are regularly employed by researchers [26-28]. As everyone knows, a number of optimization problems investigated recent years could be modelled by ADFs. The model of function employed in this paper, in which new sub-functions are one by one added, could be seen as a system that grows its complexity with the time, because new interactions appear among the variables [29].

We define ADFs in a general form. Let $S=\{0,1\}^n$ be the problem space, a fitness function $f:S\to R$ is decomposable if it can be expressed as a sum of sub-functions of lower dimension,

$$f(x)=\sum_{c_i\in C} f_i(c_i), \tag{7}$$

where $x=(x_1,\ldots,x_n)\in\{0,1\}^n$ and $C=\{c_1,\ldots,c_l\}$ is a set composed of distinct sub-sets $c_i\subseteq\{x_1,\ldots,x_n\}$. Moreover, we suppose that set c_i and c_j are distinct and not include each other. This kind of functions is also featured by its order *k*, which is determined by the size of the largest subset in C.

In this paper we employ detailed cases of this general kind of functions. Firstly, all the sub-sets in C have three variables, that is to say $k=3$. Thus, C includes all the sets of distinct sub-sets, which are selected from all the $C(n,k)$ possible sub-sets of variables. Secondly, all the sub-functions f_i are the same deceptive function f_{3dec} which is proposed by Goldberg. Given $u(y)=\sum_{i=1}^{k} y_i$, where $y\in\{0,1\}^k$, the function can be formulated as,

$$f_{3dec}(c_i)=\begin{cases}0.9 & for \quad u(c_i)=0\\ 0.8 & for \quad u(c_i)=1\\ 0.0 & for \quad u(c_i)=2\\ 1.0 & for \quad u(c_i)=3\end{cases}. \tag{8}$$

From standpoint of the EDA, the advantages of the fitness functions that we use are evident [30-32]. First of all, independently of the number of sub-functions, we all time know the globally optimal solution, which is with all ones. Secondly, the deceptive method generates strong interactions among the variables which belong to every sub-function.

About the functions f_{3dec}, when they are managed without overlapping among the set of variables, we get the function Deceptive3. This function was designed in the background of genetic algorithms with the purpose of analysing their restrictions. Thus, in this paper, the function will be a helpful reference so as to analyse the restrictions of behave in EDAs. At present, deceptive separable functions are widely employed to analyse evolutionary algorithms.

For the sake of raising the complexity of the functions step by step, we use a simple method. First of all, we create a series of objective functions in which every new function puts one more sub-function into the previous one. This series of functions is determined by the ordered set $C=\{c_1,\ldots,c_l\}$. This ordered set is composed of l different sub-sets of variables which are randomly taken from all the $C(n,k)$ possible combinations following a uniform distribution. Though we could add all the $C(n,k)$ different subsets in C, it will be not essential to get learning restrictions. Therefore, the *s*-th objective function in the list adds *s* sub-functions which are the first *s* sub-sets of variables in C. The *s*-th function can be formulated as,

$$f_s=\sum_{i=1}^{s\le l} f_{3dec}(c_i). \tag{9}$$

However, the ordered set C has a constraint. The join of the first n/k sub-sets equal to the whole set of n variables, constituting the function Deceptive3. Be aware that in the functions from $s=1$ to $s=n/k$, a number of the variables do not appear in any sub-function. In order to make these functions intact, we directly use the previously mentioned function u to include the set of variables $\{x\setminus\bigcup_{i=1}^{s} c_i\}$, which do not appear in the sub-function. This measurement is very helpful to investigate the univariate and bivariate EDA. Moreover, this separable function is helpful to indicate difficulty of a problem.

Finally, because of the random nature of the set C, we have produced for the experiments 100 different random instances of this kind of sets and the results presented are the average value from them. The entire set of experiments include three different problem sizes ($n\in\{24,48,72\}$) and the maximum number of sub-functions which is

determined by C, is $l = 200$. Moreover, for every possible function, we conduct ten independent EDA runs. The number of runs per instance is limited because of the high computational cost.

The intensity of interaction can be understood as a notion about the interdependences that appear among the variables of a problem. Though there are a number of different methods to estimate this notion, in the background of the present work, we suppose that the intensity of interaction is simply determined by the number of sub-functions contained in the objective function [33-35].

Moreover, in order to give a more direct measure of the intensity of interaction, we consider the frequency that every variable appears in sub-functions. For instance, in some ADFs, each variable appears in only one sub-function. Generally speaking, given s sub-functions with size k including n variables, we compute the expected number of variable appearing in sub-functions. It is formulated as,

$$\langle s \rangle = s\frac{k}{n}. \tag{10}$$

In order to demonstrate how the landscapes of the ADFs change along with the number of sub-functions increasing, we give a simple example in which the number of variables is 9 in Figures 4-7.

These figures shows changes of function value with four objective functions, which differ only in the number of variables. The solutions are sorted by the number of ones so as to supplying more intuitive graphs. For instance, the region with the number 4 is comprised of all the $C(9,4)$ solutions, which have 4 ones. Therefore, the region with the number 0 and number 9 only contains one solution. The two solutions play an important role in the kind of functions we employ, and thus, the solution with all zeros is emphasized with a circle and the solution with all ones, which is optimum is emphasized with two concentric circles.

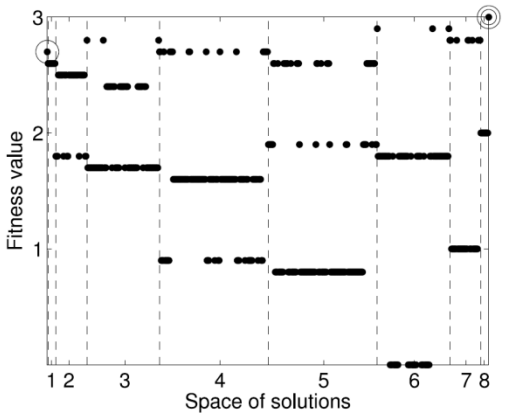


FIGURE 4 f_3, 3 sub-functions, $\langle s \rangle = 1$

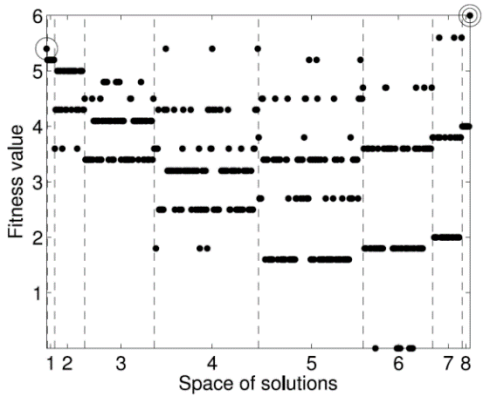


FIGURE 5 f_6, 6 sub-functions, $\langle s \rangle = 2$

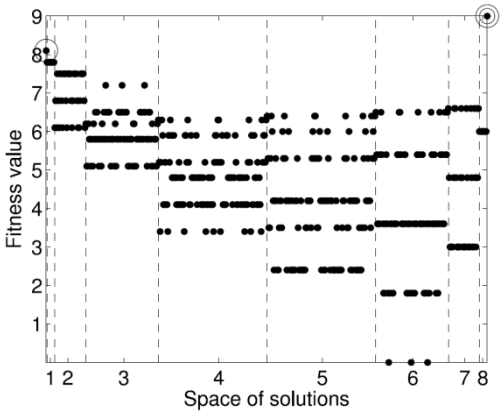


FIGURE 6 f_9, 9 sub-functions, $\langle s \rangle = 3$

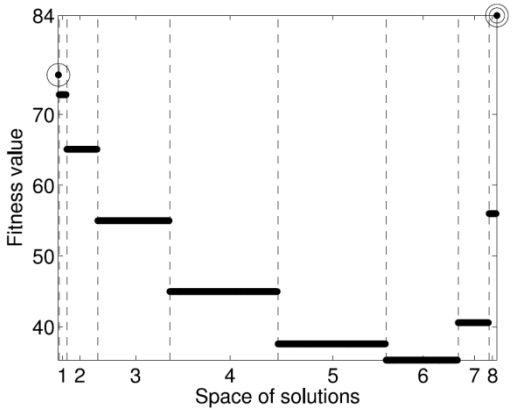


FIGURE 7 f_{84}, 84 sub-functions, $\langle s \rangle = 28$

In this part, for each function created, we compute the order of a possible related exact factorization. This computational process is a proximity to the exact factorization, which has minimum complexity. However, this process provides useful information about the complexity of the probabilistic models, and it is essential for an EDA to solve the created problems in the worst conditions. The orders of the used functions change as shown in Figure 8. This Figure presents when problem sizes is 3, the order change with the average number of sub-functions in which each variable appear. We begin from $\langle s \rangle = 1$, in this case it is the Deceptive3 function. The results show an exponential growth of the number of parameters related to exact factorizations and thus, the

complexity of these models rapidly becomes too high to manage. Even if having enough knowledge concerning the function, looking for this kind of probabilistic models is confronted with considerable computational restrictions. Moreover, if these factorizations were employed in an EDA, the population size should also grow with the complexity of the factorizations so as to get a robust behave of EDAs [36].

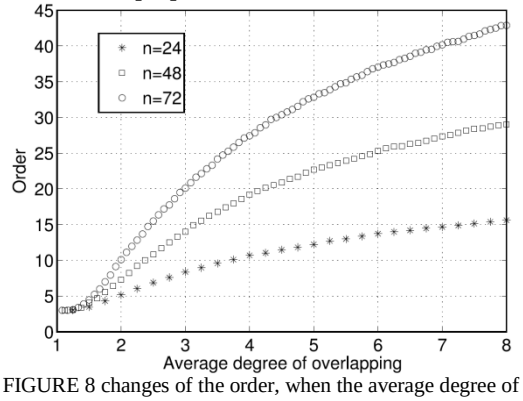


FIGURE 8 changes of the order, when the average degree of overlapping grows

4 Experimental results

In this part, we present experimental results. The relationship between the ratio of success and the number of sub-functions for each kind of algorithm gives in Figure 9, Figure 11 and Figure 13. The relationship between hamming distance to the best solution and the number of sub-functions for each kind of algorithm gives in Figure 10, Figure 12 and Figure 14. We only show the experimental results for the problems with $n = 72$ variables. The performance of the algorithms is alike for the three different problem sizes ($n \in \{24, 48, 72\}$) that we have used. Nevertheless, when the number of variables grows, the patterns are more obvious and clearer.

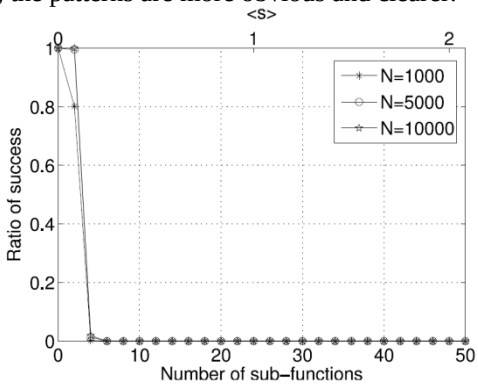


FIGURE 9 Ratio of success in UMDA

Generally speaking, by means of the descriptors employed in Figure 9 - Figure 14, we can obviously see the behave that different EDA implementations fails. The graphs exhibit a phase-transition effect when the interaction in the created problems exceeds a certain degree. However, this effect is especially remarkable in the graphs of ratio of successful runs, which fall off from 1 to 0 steeply after only adding very few sub-functions. Therefore, in UMDA and EDAdt, the number of sub-functions between total success and complete failure is within 2 sub-functions. For the EDA based Bayesian networks, though this algorithm also suffers a sudden fail, the conversion from 1 to 0 in the ratio of successful runs is relatively more gradual. concerning the hamming distance, it exhibits a more gradual change, which supplies complementary information concerning the quality of the solutions. For instance, though in Figure 9 the ratio of success can be equal to 0 when the number of sub-functions is 4, UMDA returns solutions which are near the optimum in hamming distance in Figure 10.

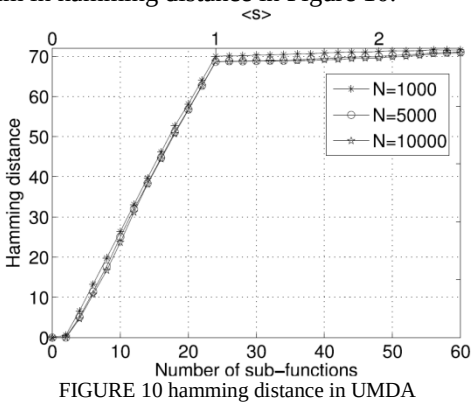


FIGURE 10 hamming distance in UMDA

From the results obtained from the experiment, we also achieve the understanding of the effects that the probabilistic models and the population size have on the EDAs when we solve problems with growing interaction. As anticipated, the probabilistic model employed in the EDAs has a determined influence on the scope of problems that it can solve. Therefore, UMDA begins to fail when the number of sub-functions is 2 in the objective functions, EDAdt can attain the equivalent level of Deceptive3 function and EBNA collapses between the ADFs and the functions with $\langle s \rangle = 2$.

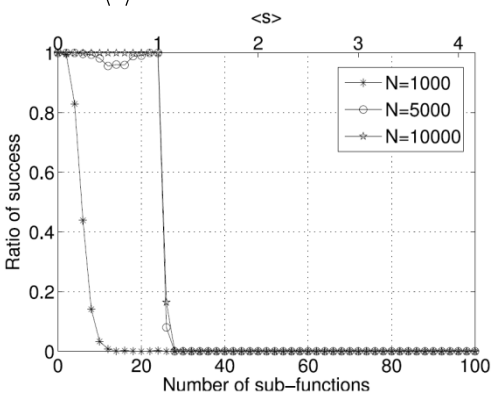


FIGURE 11 ratio of success in EDAdt

In addition, according to experimental results, if an EDA to can deal with more complex structural models, the population size has a more effect on the performance of the EDA. It is decisive to achieve a robust behaviour of EDAdt and EBNA, while UMDA is hardly affected by this parameter. As presented in Figure 11 and Figure 12, the smallest population size ($N = 1000$) is clearly inadequate,

therefore these algorithms cannot achieve a ideal performance.

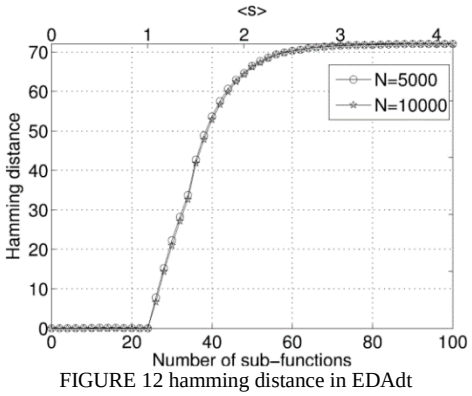


FIGURE 12 hamming distance in EDAdt

The greatest influence of the population size takes place in EBNA and it is reflected in the graphs shown in Figure 13 and Figure 14. However, even for EDA based Bayesian networks, this parameter exhibits a limited effect to overcome influence of threshold. This suggests that, though the population size is important to obtain a ideal behavior, growing this parameter is not an efficient way to solve the problems when the degree of interaction grows. In this respect, we can see in Figure 13 and Figure 14 that the different graphs tend to be closer as the size of the population grows.

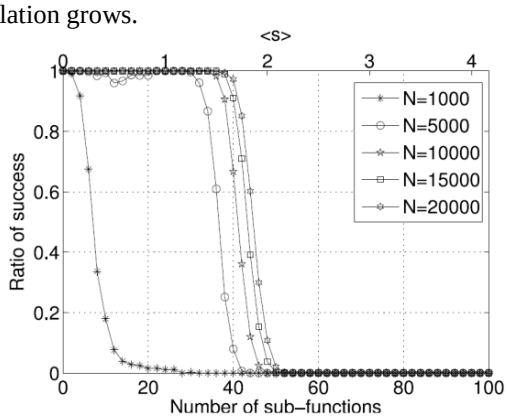


FIGURE 13 ratio of success in EBNA

From the graphs of hamming distance, we can see another phenomenon. when the interaction in the problem exceeds a certain extent, all EDAs only can find solution with all zeros. Thus, we could conclude that, from the aspect of efficiency, UMDA is the best choice to deal with the problems which exceed this critical threshold of difficulty. According to this viewpoint, and considering the whole scope of functions that can be created from $s=0$ to $s=C(n,k)$, EBNA exhibits better performance in a reduced sub-space. Be aware that Figure 9 - Figure 14 only presents the behaviour of the EDAs at the first ranks of difficulty. When n equal to 72 and *k* equal to 3, we could obtain in the fitness function up to $C(72,3)=59640$ sub-functions. Thus, the graphs only depict a small fraction of this number.

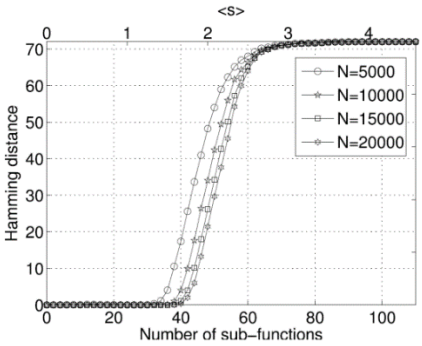


FIGURE 14 hamming distance in EBNA

From the experimental results given in Figure 9- Figure 14, we have observed that how the behave of different EDAs suddenly fails when the interaction among the problem variables exceeds a certain extent. The reason for this phenomenon is that UMDA and EDAdt are deficient in ability to learn models. Nevertheless, for EBNA, it is deserving to carry out a more thorough analysis of the reason of its failure. In order to do this, we consider the complexity of the Bayesian models learned from the experiment. We employ the order of the factorizations which is determined by these probabilistic models to assess their complexity. In Figure 15, we present the average maximum orders of the models learned from each execution.

Be aware that the influence of phase transition seen in Figure 13 and Figure 14 is associated with Figure 15. Therefore, EBNA starts to collapse quickly after the peaks in Figure 15, that is to say when the algorithm could not construct more complex models. Figure 15 indicates that the complexity of the Bayesian networks grows exponentially with the number of sub-functions so as to solve the problems. Be aware that when the learning method cannot construct the appropriate structures to solve more problems, the algorithm still spends a large number of computational resources on learning false models.

It is a question that what extent the behaviour mentioned above depends on the learning method, which need an thorough analysis. We argue that, in this worst situation, we will likely get a resembling performance for other approximate learning methods.

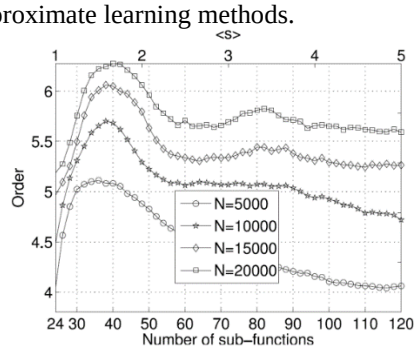


FIGURE 15 maximum orders of the models learned by EBNA

We compare the complexity of the learned structures with the complexity of the exact factorizations in Figure 16 and Figure 17. The curve shown in Figure 16 corresponds to the order of the exact factorizations, which were presented in Figure 8. However, in the graph, the number of

variables is only for $n = 72$ and curve is exhibited in relation to the number of sub-functions. We have used a circular mark so as to approximately indicate the region where EBNA fails. In Figure 17, we put the order of the exact factorizations and the order of the Bayesian networks obtained by the EDA together. The dashed line indicates the graph of the exact factorizations. In this Figure, we can observe that, when the EDA can run successfully, the maximum complexity of the models obtained is not less than the complexity of the exact factorizations. That is to say, when the EDA cannot learn the complexity of the models, which equals to the exact factorizations, it no longer solves more complex problems. The dashed line of the factorizations exactly separates the problems solved by EDA from the problems unsolved by EDA for every population size. Population size, which is insufficient may be the main reason to explain the bad behave in looking for the model structure.

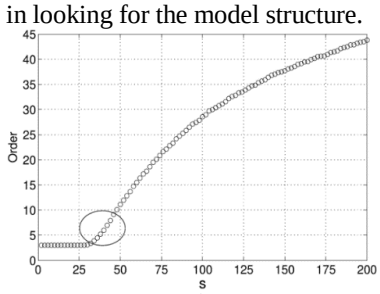


FIGURE 16 maximum orders of the models learned by EBNA

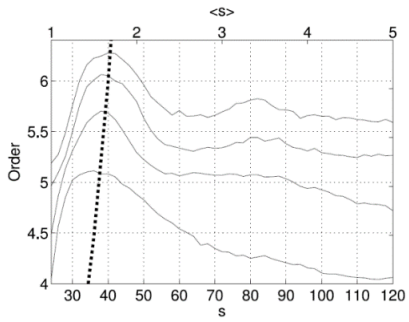


FIGURE 17 maximum orders of the models learned by EBNA

5 Conclusions

In this paper we have investigated the restrictions of behave that different EDA implementations confront when the interaction among the variables of the problem grows. We conduct the study by the employment of ADFs in which new sub-functions are added one by one. Therefore, the interaction can be measured by the number of sub-functions, which the objective function contains. In addition, we employ the separable deceptive function as a indication of problem difficulty so as to supply more perspicuous results. In the experiments, we have carried out three different type of EDA implementations. Because these algorithms only distinguish in the probabilistic model employed, the results indicate the influence that using more complex models has so as to solve a broader scope of problems. We have also employed different population sizes, which has been crucial so as to realize a potent performance in EDAs that base on Bayesian networks. Nevertheless, the results show that, generally speaking, growing this parameter is not useful to solve more difficult problems.

We have found that the EDAs fail with a phenomenon of phase transition when the number of sub-functions in the objective function grows in the worst situation,. The area in which EBNA collapses is between the separable deceptive functions and the objective functions with $2n/k$ sub-functions. The reason for the breakdown of the algorithm is that the EDA cannot obtain correct models. Once the EDA cannot to learn more complex models which is essential to solve more difficult problems, the algorithm fails. The complexity of the networks tends to grow exponentially so as to find the optimum needed for the algorithm. Nevertheless, exceeding a certain degree, the learning method cannot structure the suitable models to solve the problem and then, the algorithm collapses quickly. It implies that, when the degree of interaction exceeds a critical point, the learning of Bayesian networks may not be able to obtain the information needed to find the optimum from the population. Moreover, the relationship between the models learned by the EDA and the exact factorization indicates forceful computational restrictions because of the exponential increasing complexity of the structural essential to solve the problems.

The restrictions of effectiveness displayed in this paper are straight associated to the learning procedure of the algorithm. Nevertheless, these restrictions do not always have only one cause. We can confirm three different standpoints from which the learning restrictions in EDAs could be investigated:

1) restrictions of the learning either because the model needs a priori knowledge or because they learn approximate models with bounded complexity.

2) Even if we employ a desired learning method, there are efficiency restrictions because the complexity of models grows exponentially, which is disadvantage to solve the problems when the number of interactions grows.

3) Restrictions for the population either because of the deficiency of information that it includes to solve the problem or because this parameter should grow exponentially to supply an effective learning.

In general, we have investigated the concept of borders of EDA effectiveness concerning the extent of interaction of the problem. The main objective of this study is not to look for the best algorithm or negate any method but to comprehend which algorithms are the suitable for which problems.

Acknowledgments

This paper is supported in part by National Natural Science Foundation of China (Grant no.60975050), Research Fund for the Doctoral Program of Higher Education (Grant no.20070486081) and Fundamental Research Funds for the Central Universities (Grant no.6081014). The authors are grateful to the anonymous referees for their insightful and constructive comments, which greatly improved the quality of the paper.

References

[1] Mühlenbein H, Paaß G 1996 From recombination of genes to the estimation of distributions I. Binary parameters Lectures Notes in Computer Science: Parallel Problem Solving from Nature (PPSN IV) **1141** *Springer Verlag* 178-87

[2] Larrañaga P, Lozano J A 2002 Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation *Kluwer Academic Publishers*

[3] Armañanzas R, Inza I, Santana R, Saeys Y and Larrañaga P 2008 A review of estimation of distribution algorithms in bioinformatics BioData Mining **1**(6) 1–12

[4] Höhfeld M, Rudolph G 1997 Towards a theory of population based incremental learning *Proceedings of the 4th International Conference on Evolutionary Computation IEEE Press*

[5] Goldberg D E 1989 Genetic Algorithms in Search, Optimization, and Machine Learning *Addison: Wesley*

[6] Eiben E A and Smith J E 2003 Introduction to Evolutionary Computing (Natural Computing Series) *Springer*.

[7] Mühlenbein H 1998 The equation for response to selection and its use for prediction Evolutionary Computation **5**(3) 303–46

[8] Baluja S, Davies S 1997 Using optimal dependency-trees for combinatorial optimization: Learning the structure of the search space *Proceedings of the 14th International Conference on Machine Learning Morgan Kaufmann* 30-8

[9] Etxeberria R and Larrañaga P 1999 Global optimization using Bayesian networks *Proceedings of the Second Symposium on Artificial Intelligence Habana* 151-73

[10] Santana R, Larrañaga P, Lozano J A 2005 Interactions and dependencies in estimation of distribution algorithms *Proceedings of the 2005 Congress on Evolutionary Computation IEEE Press* 1418–25

[11] Wu H, Shapiro J L 2006 Does over-fitting affect performance in estimation of distribution algorithms *Proceedings of the 8th annual conference on Genetic and evolutionary computation ACM Press* 433–4

[12] Meek C 1995 Strong completeness and faithfulness in Bayesian networks *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers* 411-8

[13] Neapolitan R E 2003 Learning Bayesian Networks *Upper Saddle River: Prentice Hall*

[14] Jensen F V, Nielsen T D 2007 Bayesian Networks and Decision Graphs *Springer*

[15] Koller D, Friedman N 2009 Probabilistic Graphical Models: Principles and Techniques *Cambridge: MIT Press*

[16] Friedman N, Geiger D, Goldszmidt M 1997 Bayesian network classifiers *Machine Learning* **29**(2) 131–63

[17] Hauschild M, Pelikan M, Sastry K, Goldberg D E 2012 Using previous models to bias structural learning in the hierarchical *BOA Evolutionary Computation* **20(**1) 135–60

[18] Friedman N, Linial M, Nachman I. 2000 Using Bayesian networks to analyze expression data *Journal of Computational Biology* 7 601–20

[19] Buntine W 1991 Theory refinement on Bayesian networks *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence San Mateo Morgan Kaufmann* 52-60

[20] Chickering D M, Geiger D, Heckerman D 1995 Learning Bayesian networks: Search methods and experimental results *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics* 112-28

[21] Roberto S 2005 Estimation of distribution algorithms with Kikuchi approximations *Evolutionary Computation* **13**(1) 67-97

[22] Ocenasek J 2006 Entropy-based convergence measurement in discrete estimation of distribution algorithms Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms *Springer* 39–50

[23] Pelikan M, Sastry K and Goldberg D E 2005 Multi-objective hBOA, clustering, and scalability *Proceedings of Conference on Genetic and Evolutionary Computation, ACM Press* 663-70

[24] IInza P, Larrañaga P, Etxeberria B S 2000 Feature subset selection by Bayesian network-based optimization *Artificial Intelligence* **123**(1) 157-84

[25] Saeys Y, Degroeve S, Aeyels D, van de Peer Y and Rouze P 2003 Fast feature selection using a simple estimation of distribution algorithm: a case study on splice site prediction *Bioinformatics* **19**(s2) 179-88

[26] Joaquin R and Roberto S 1999 Improving the discovery component of classier systems by the application of estimation of distribution algorithms *Proceedings of the Students Sessions*

[27] Santana R 2002 An analysis of the performance of the mixture of trees factorized distribution algorithm when priors and adaptive learning are used *Institute of Cybernetics, Mathematics and Physics*

[28] Coffin D J and Smith R E 2007 The limitations of distribution sampling for linkage learning *Proceedings of the 2007 Congress on Evolutionary Computation IEEE Press* 364–9

[29] Naudts B and Kallel L 2000 *IEEE Transactions on Evolutionary Computation* **4**(1) 1–15

[30] Li X, Mabu S and Hirasawa K 2014 *IEEE transactions on evolutionary computation* **18**(1) 98-113

[31] Robles V, Peña J M, Pérez M S and Herves V 2006 GA-EDA: A new hybrid cooperative search evolutionary algorithm Towards a New Evolutionary Computation: Advances in Estimation of Distribution Algorithms *Springer* 187–220

[32] Armañanzas R, Saeys Y, Inza I, Garca-Torres M, Bielza C, van de Peer Y and Larrañaga P 2011 Peakbin selection in mass spectrometry data using a consensus approach with estimation of distribution algorithms *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **8**(3) 760–74

[33] Hauschild M, Pelikan M, Sastry K, Lima C 2009 *IEEE Transactions on Evolutionary Computation* **13**(6) 1199–217

[34] Abdollahzadeh A, Reynolds A, Christie M 2012 Bayesian optimization algorithm applied to uncertainty quantification *SPE Journal* **17**(03), 865-73

[35] Soto M R, González-Fernández Y and Ochoa A 2012 Vine estimation of distribution algorithm *Proceedings of the VIII Congreso Español sobre Metaheurísticas Algoritmos Evolutivos y Bioinspirados*

[36] Mühlenbein H 2012 Convergence of Estimation of Distribution Algorithms Markov Networks in Evolutionary Computation *Springer* 91–108

Authors

Caichang Ding, born on October 10, 1980, Hubei, China

Current position, grades: Ph.D. student at State Key Lab of Software Engineering, Wuhan University, Wuhan, China.
University studies: M.Sc. degree from the School of Computer, Wuhan University, Wuhan, China, in 2006.
Scientific interest: computational learning theory, statistical learning, basic theory of evolutionary computation and optimization theory.
Publications: 5 papers.
Experience: Caichang Ding received the B.Sc. degree from the School of Mechanical & Electronic Information, China University of Geosciences, Wuhan, China, in 2003, and the M.Sc. degree from the School of Computer, Wuhan University, Wuhan, China, in 2006. He is currently a Ph.D. student at State Key Lab of Software Engineering, Wuhan University, Wuhan, China, and a lecturer in the School of Computer Science, Yangtze University, Jingzhou, China.

Wenxiu Peng, born on February 14, 1981, Hubei, China

Current position, grades: lecturer at School of Computer Science, Yangtze University, Jingzhou, China.
University studies: M.Sc. degrees from Hubei University, Wuhan, China in 2006.
Scientific interest: computational learning theory, statistical learning, basic theory of evolutionary computation and optimization theory.
Publications: 3.
Experience: Wenxiu Peng received the B.Sc. and M.Sc. degrees from Hubei University, Wuhan, China, in 2003 and 2006. She is currently a lecturer in the School of Computer Science, Yangtze University, Jingzhou, China. Her main research interests include computational learning theory, statistical learning, basic theory of evolutionary computation and optimization theory.

An image threshold segmentation method based on multi-behaviour global artificial fish swarm algorithm

Jing Zeng*

School of Physics and Electronic Information, China West Normal University, Nanchong 637002, Sichuan, China

Received 1 July 2014, www.cmnt.lv

Abstract

Firstly, this paper describes how the histogram analysis method pre-processes the images to be segmented. Then is makes a detailed analysis of the working principles and behaviour pattern of basic artificial fish swarm algorithm (AFSA); dissects the defects of AFSA in principle and proposes an improved AFSA with global convergence. Finally, it presents the main steps of image threshold segmentation method based on AFSA; compares the performances of AFSA with those of other intelligent algorithms and proves that this improved AFSA makes all-around improvements on image segmentation.

Keywords: Image Threshold Segmentation, Artificial Fish Swarm Algorithm

1 Introduction

Image segmentation divides an image into several mutually-disjoint areas according to certain rules and every area meets the consistency under the rules for convenience of the subsequent target measurement as well as classification and identification. Image segmentation needs to divide the input image into two or more sub-regions, which is exactly the top priority encountered in the design and realization of image analysis, the text character identification and automatic target acquisition. Nowadays, there have emerged several image segmentation methods, including threshold segmentation method, statistic segmentation method and clustering segmentation, however, none of them is a universal image segmentation method. Therefore, it plays a significant influence on the entire performance of the segmentation method whether the appropriate image segmentation method is chosen.

Swarm intelligence is the intelligent behaviour of a group, which is formed by the individuals with no or simple intelligence in certain way and the individuals in this group show complicated intelligent behaviours through collaboration. For example, when the small fish with limited capacities are attacked by big fish, they usually hold together to form a fish swarm quickly and move around the swarm center continuously to withstand the attack and minimize the loss. In case of lack of centralized control and understanding of global situation, swarm intelligence can accomplish complicated and difficult tasks, which provides a clue for people to solve large-scale and complexly-distributed problems [1]. In image-processing applications, swarm intelligent optimization algorithm is relatively weak in mathematical theoretical foundation and it lacks the theoretical analysis in a general sense. When different swarm intelligent optimization algorithms are used in specific image processing, the parameters involved in these algorithms are usually determined by experience without any specific theoretical basis.

In order to improve the various shortcomings of AFSA, this paper has come up with an improved multi-behaviour AFSA with excellent global convergence after analysing and investigating the research results of other scholars. This algorithm has absorbed the advantages of other improved algorithms [2]. For instance, it optimizes the step size and horizon by using self-adaptive theory; increases the artificial fish behaviours such as jump and devour and makes distinct and all-around improvements over the basic AFSA. Therefore, it enhances the self-adaptive ability and convergence precision, accelerates the convergence time and avoids local extremum in searching global optimum [3].

2 The image pre-processing

Due to digital image acquisition and channel transmission, the images we obtain usually have noises, which may interfere the ornamental value of the images and even directly affects the subsequent image characteristics extraction, target measurement and target identification. Considering the influence of image noise, we reconstruct the images by using the characteristic of wavelet domain after multi-level wavelet transform, namely "the highest low-frequency coefficients and high-frequency coefficients have fewer noises", and so we can obtain enhanced noise-suppressing images of the original images. Then, the segmentation threshold to be determined shall be located between the two peaks of the

* ***Corresponding author*** e-mail 648357216 @qq.com

enhanced image histogram and its value shall be similar to the segmentation threshold of the enhanced image.

Assuming that the two peak values of the enhanced image histogram are *a* and *b*, narrow the distribution range of the threshold and design the objective function of threshold segmentation according to the following steps.

(1) Decompose the image to be segmented with three-level sym5 wavelet and proceed wavelet reconstruction by the decomposed 3rd-level low-frequency and high-frequency coefficients to obtain the noise-suppressing enhanced image.

(2) Analyze the histogram of the enhanced image; calculate the average grey level $\overline{c}$ and get the grey level *a* with the biggest possibility within $[0,\overline{c}]$ and the grey level *b* with the biggest possibility within $[\overline{c}+1,L-1]$. Then the threshold distribution range is $[a,b]\in[0,L-1]$ and $L=256$ is the grey level of the grey-level image[4].

(3) Segment the objective function by using OTSU according to the two-dimensional histogram information of the enhanced image and neighbourhood mean-value image.

3 The improved mechanism of fish swarm algorithm

Although AFSA has the merits like undemanding objective function value, greatly random parameter determination and quick convergence rate, improvements need to be made somewhere. When the optimization region of the algorithm is big or the change is relatively flat, the algorithm searching speed is becoming slow and the convergence performance decreases; the convergence speed is quick in the initial stage and it becomes slow in the later phase; optimization precision is subject to the randomness of horizon and step size so that it is difficult to get precise solutions [5].

3.1 THE IMPROVEMENTS OF ARTIFICIAL FISH SWARM BEHAVIORS

In the basic AFSA, assuming that there are *N* artificial fish in the search region and the *i*-th represents a feasible solution vector, $X_i=\left(x_1^i,x_2^i,\ldots,x_D^i\right)$ (D is the dimension). Then assuming that $X_i(t)$ is the current state of the artificial fish and (*t*+1) is the next status, the artificial fish will implement one of the four behaviours according to the current status in every iteration; update its ego state and add an added vector $\Delta X_i(t+1)$ on the basis of the original status X_i(t); therefore, the position update formula of the artificial fish in AFSA is described by using the following formula:

$$\Delta X_i(\mathrm{t+1})=\mathrm{Rand}()\bullet\mathrm{Step}\bullet\left[X_{next}(\mathrm{t+1})-X_i(\mathrm{t})\right], \quad (1)$$

$$X_i(\mathrm{t+1})=X_i(\mathrm{t})+\Delta X_i(\mathrm{t+1}). \quad (2)$$

In the above formula, Rand() is a random number within 0 and 1 and Step is the step size of the artificial fish. After every iteration, the artificial fish update the ego status according to the above formula [6].

3.1.1 The improvements of feeding

Assuming that (*t*) is the current status, $Y_i(t)$ is the function value of the current status $X_i(t)$; the next status is X_i(t+1) and X_j(t) is a randomly chosen status of its horizon. And the expression formula is as follows:

$$X_j(\mathrm{t})=X_i(\mathrm{t})+Visual\bullet Rand(). \quad (3)$$

In searching the minimal value, if $Y_i(t)>Y_j(t)$, then the artificial fish moves towards the vector sum direction of X_j and the global optimal position and the expression formula is as follows:

$$X_i(\mathrm{t+1})=X_i(\mathrm{t})+Visual\bullet Rand(). \quad (4)$$

3.1.2 The improvements of clustering

Assuming that (*t*) is the current status of the artificial fish, search the number of its companions n_f and the central position $X_c(t)$ within its horizon. If $Y_c(t)n_f<\delta Y_i(t)$, it is indicated that its neighbourhood centre has higher food concentration and low congestion degree and the artificial fish moves a step size towards the vector sum direction of the global optimal position X_{best} and the central position. If (*t*) $n_f>\delta Y_i(t)$, then the artificial fish implements the feeding [7].

3.1.3 The improvements of rear-end

Assuming that X_i is the current status of the artificial fish, search the smallest companion X_k(t) among the companions Y_k(t) within horizon. If $Y_K(t)\bullet n_f<\delta\bullet Y_i(t)$, it is shown that the status of the companion X_k(t) has higher food concentration and low congestion degree and then the artificial fish moves one step towards the vector sum of X_k(t) and the global optimal position X_{best} with its expression as follows. If not, the artificial fish implements feeding.

$$X_i(\mathrm{t+1})=X_i(\mathrm{t})+\left(\frac{\left(X_k(\mathrm{t})-X_i(\mathrm{t})\right)+\left(X_{best}-X_i(\mathrm{t})\right)}{\left\|\left(X_k(\mathrm{t})-X_i(\mathrm{t})\right)+\left(X_{best}-X_i(\mathrm{t})\right)\right\|}\right)\bullet Step\bullet Rand(). \quad (5)$$

3.1.4 The introduction of jumping

After several continuous iteration, if the objective function value of the optimal artificial fish is smaller than the pre-set value eps, then randomly choose some artificial fish and set their parameters with the expression formula as follows.

$$X_i(\mathrm{t+1})=X_i(\mathrm{t})+\beta\bullet visual\bullet Rand(). \quad (6)$$

In this formula, β can be either a parameter of a mutation function.

3.1.5 The introduction of devouring

We have increased a new fish swarm behaviour, devouring. In this new fish swarm movement mode, the artificial fish with low objective function value are usually considered as weak artificial fish because they have little influence on the algorithm performance. These weak artificial fish will be eliminated by the system after some iteration, which is just like they are devoured by big fish. Because these artificial fish which are eliminated are weak, they have little influent on the optimization capacity of the algorithm; thus reducing its complexity.

The specific behaviour description is as follows: in calculating the minimum, if the objective function value of a certain artificial fish is bigger than the set function value T_value after n continuous iteration, then the system will automatically release the system space occupied by this artificial fish and reduce the total amount of the artificial fish. Therefore, this artificial fish will not involve the subsequent iteration. Likewise, the maximum can be calculated [8].

3.2 THE IMPROVEMENTS OF PARAMETERS

The horizon and step size of the artificial fish in the basic AFSA are always fixed, which may lead to the slow convergence speed due to the undersized step size in the initial phase and the oscillation phenomenon caused by the oversized step size in the precise convergence in the later phase. In order to solve this problem, we have introduced self-adaptive step size with the specific behaviour description as follows.

Set X_i(t) as the current status of the artificial fish, then Y_i(t) is the function value of the current status X_i(t); set X_i(t+1) as the next status of the artificial fish and X_v(t) is the next status searched by the artificial, then Y_v(t) is the function value with the formula as follows:

$$X_v(\mathrm{t}) = X_i(\mathrm{t}) + Visual \bullet Rand(), \tag{7}$$

$$X_i(\mathrm{t+1}) = X_i(\mathrm{t}) + \frac{X_v(\mathrm{t}) - X_i(\mathrm{t})}{\| X_v(\mathrm{t}) - X_i(\mathrm{t}) \|} \bullet \left| 1 - \frac{Y_v(\mathrm{t})}{Y_i(\mathrm{t})} \right| \bullet Step \ (\text{Minimization problem}). \tag{8}$$

From the above formula, it can be seen that the movable step size in the algorithm is determined by the current status and the sensing status within horizon.

4 The noisy image threshold segmentation based on afsa

The main idea of the noisy image segmentation method based on AFSA is: consider the threshold of the noisy image as the individual fish in the fish swarm; the threshold distribution range $[a,b]$ is the feeding space of the fish swarm and find the optimal segmentation threshold through the chemotaxis, reproduction and elimination-dispersal with Formula 1 as the behaviour guide of fish feeding, namely the fitness function of AFSA.

The main steps are:

(1) Read in the image to be segmented;

(2) Decompose the image to be segmented with three-level sym5 wavelet and proceed wavelet reconstruction by the decomposed 3rd-level low-frequency and high-frequency coefficients to obtain the noise-suppressing enhanced image;

(3) Determine the threshold distribution range based on the histogram of the enhanced image;

(4) Get the fitness function of AFSA through Formula (1) by using the two-dimensional histogram of the enhanced image and its neighbourhood mean-value image;

(5) Set the control parameters in AFSA;

(6) Calculate the fitness of every individual fish and record the fish with the biggest fitness as well as its fitness;

(7) Chemotaxis operation: firstly, the fish overturns according to the following formula:

$$P(i, j+1, k, l) = P(i, j, k, l) + C(i) \times \Delta(i). \tag{9}$$

In this formula, $P(i, j, k, l)$ is the position of i-th fish in the l-th elimination-dispersal, the k-th reproduction and the j-th chemo taxis; $P(i, j+1, k, l)$ is the overturn new position of the i-th fish in the l-th elimination-dispersal, the k-th reproduction and the j-th chemotaxis; $C(i)$ is the step size of the i-th fish and $\Delta(i)$ is a random vector among [-1,1];

Then, compare whether the fitness of the fish after the overturn has been improved. If the fitness is improved, then the fish moves forward according to the below formula:

$$P(i, j+1, k, l) = P(i, j, k, l) + C(i) \times \Delta(i). \tag{10}$$

Otherwise, add 1 to the number of chemotaxis j; when $j < N_c$, the fish repeats the chemotaxis operation in this step. If not, the fish turns to Step (8).

Finally, compare the fitness after the fish moves forward and the biggest fitness recorded in this bulletin board to make the biggest fitness and the corresponding position are always kept in the board;

(8) Reproduction operation: calculate the fitness cumulative sum in the chemotaxis:

$$SumFitness(i) = \sum_{j=1}^{Nc+1} Fitness(i, j, k, l) . \quad (11)$$

Rank the fitness accumulative sum of the fish in the descending order. Make the weak fish to get the position and step size of the optimal fish and add 1 to the number of reproduction k. If $k < N_{re}$, turn to Step (7); otherwise, turn to Step (9);

(9) Elimination-dispersal Operation: the fish is dispersed to any random position in the solution space with the possible P_{ed} and add 1 to the number of elimination-dispersal. If $l < N_{ed}$, implement Step (7); otherwise, turn to Step (10);

(10) Output the optimum in the bulletin board and end the entire feeding process;

(11) Obtain the segmented image according to the following formula:

$$bw(x, y) = \begin{cases} 0, \text{当} 0 \le f(x, y) + g(x, y) \le s + t \\ 1, \text{当} f(x, y) + g(x, y) > s + t \end{cases} . \quad (12)$$

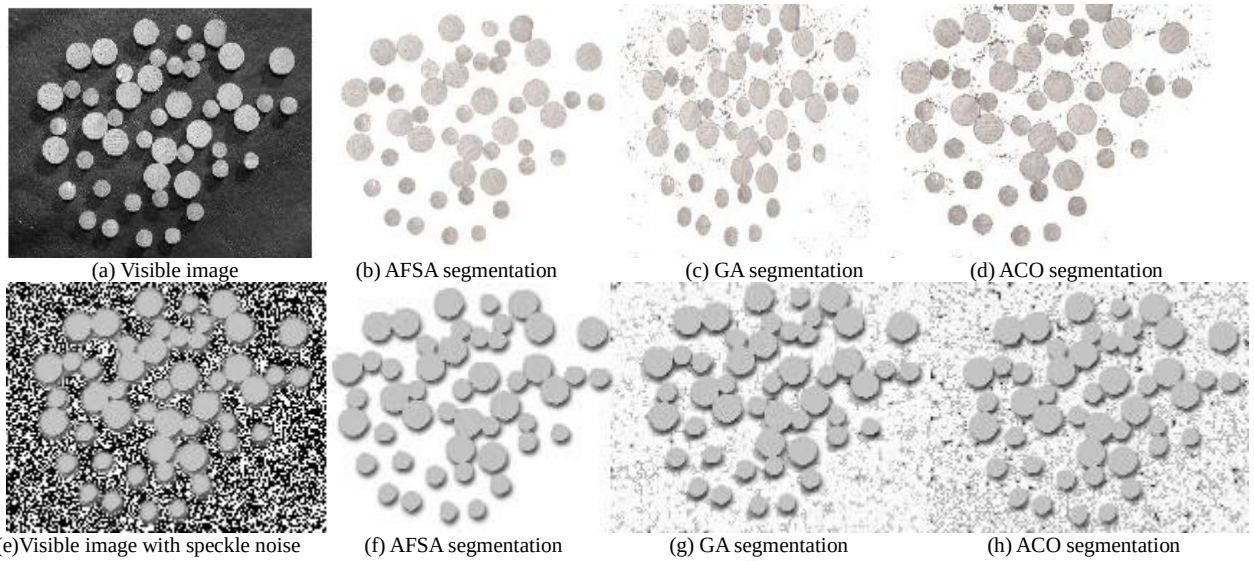
In this formula, $f(x, y)$ is the grey level of pixel (x, y); $g(x, y)$ is the average grey level of its 3×3 small neighbourhood pixel and $bw(x, y)$ is the grey value of pixel (x, y) after segmentation.

5 The experimental result and performance analysis

The following experiments are conducted in order to verify the effectiveness of the segmentation methods of AFSA.

5.1 THE COMPARISON OF SEGMENTATION EFFECTS

Segment the visible image and the visible image with speckle noise by using AFSA, genetic algorithm and ant colony optimization algorithm and the results are indicated as Fig.1. After conducting histogram analysis pre-processing on the images to be segmented, the threshold distribution range of Fig.1(a) is [103, 221] and that of Fig. 1(e) is [106, 218].



(a) Visible image (b) AFSA segmentation (c) GA segmentation (d) ACO segmentation

(e)Visible image with speckle noise (f) AFSA segmentation (g) GA segmentation (h) ACO segmentation

FIGURE 1 Image segmentation result comparison with different optimization algorithms

When using the segmentation method of AFSA, the segmentation result is clear-cut and intact with optimal performance.

5.2 THE COMPARISON OF SEGMENTATION SPEED

In order to test the segmentation speed of this method, the following table has presented the segmentation time and segmentation threshold by using AFSA segmentation methods, GA, ACO and PSO to segment Fig.1(a), as indicated as Table 1.

From Table 1, it can be seen that the segmentation time order of the above methods is: the method in this chapter<ACO<PSO<GA. The reason is that the new method adopted in this paper, on one hand, can reduce the search range of the threshold by conducting histogram pre-processing on the image to be segmented and on the other hand, quickly find the optimal threshold of the image to be segmented by introducing the parallel search mechanism of AFSA.

TABLE 1 The results by using different segmentation methods to segment fig.1

Optimization Algorithm	Fitness Function	Threshold	Run Time/S
AFSA	Two-Dimensional Maximum Between-Cluster Variance	68	5.047
GA	Two-Dimensional Conditional Entropy	131	8.372
ACO	Two-Dimensional OTSU	63	6.846
PSO	Improved Two-Dimensional Grey Entropy	153	7.478

6 Conclusion

This paper has proposed an improved AFSA image threshold segmentation method. Use histogram analysis techniques; design the trace of two-dimensional histogram scatter matrix of OTSU as the fitness function and instruct the search direction of the fish swarm in AFSA. Then, narrow down the threshold distribution range by enhancing the histogram information of the enhanced image and reduce the search range of the fish swarm in AFSA. Finally, with AFSA as the parallel threshold search strategy, accelerate the segmentation process and find the optimal threshold of the image to be segmented. The experimental results demonstrate the effectiveness of the new method, which can search the optimal threshold quickly and exactly.

References

[1] Na Fang, Jianzhong Zhou 2014 A hybrid of real coded genetic algorithm and artificial fish swarm algorithm for short-term optimal hydrothermal scheduling *International Journal of Electrical Power & Energy Systems* **62**(11) 617-29
[2] Md Abul Kalam Azad, Ana Maria A C Rocha 2014 Improved binary artificial fish swarm algorithm for the 0–1 multidimensional knapsack problems *Swarm and Evolutionary Computation* **14**(2) 66-75
[3] Weiling Zhu, Jingqing Jiang 2012 Clustering Algorithm Based on Fuzzy C-means and Artificial Fish Swarm *Procedia Engineering* **29** 3307-11
[4] Saka M P, Doğan E, Aydogdu I 2013 Analysis of Swarm Intelligence–Based Algorithms for Constrained Optimization *Swarm Intelligence and Bio-Inspired Computation* **26**(5) 25-48
[5] Fernanda M, Costa P, Ana Maria A C Rocha, Edite M G P Fernandes 2014 An artificial fish swarm algorithm based hyperbolic augmented Lagrangian method *Journal of Computational and Applied Mathematics* **259**(15) 868-76
[6] Yanbin Gao, Lianwu Guan 2014 Optimal artificial fish swarm algorithm for the field calibration on marine navigation *Measurement* **50**(4) 297-304
[7] Genrang Zheng, Zheng Chun Lin 2012 A Winner Determination Algorithm for Combinatorial Auctions Based on Hybrid Artificial Fish Swarm Algorithm *Physics Procedia* **25** 1666-70
[8] Jonas Krause, Jelson Cordeiro, Rafael Stubs Parpinelli, Heitor Silvério Lopes 2013 A Survey of Swarm Algorithms Applied to Discrete Optimization Problems *Swarm Intelligence and Bio-Inspired Computation* **67**(8) 169-91

Authors

Jing Zeng, born on January 1, 1982, China

Current position, grades: researcher at School of Physics and Electronic Information, China West Normal University, Nanchong, Sichuan, China.
Scientific interests: communication, signal processing and image processing.

A modified BFGS method and its convergence

Ganzhou Wu*, Haiyan Liang

School of Science, Guangdong University of Petrochemical Technology, Maoming 525000, Guangdong, China

Received 1 July 2014, www.cmnt.lv

Abstract

In this paper, a new modified BFGS method for unconstrained optimization problems is presented. The algorithm preserves the convergence properties of the famous BFGS algorithm. The relation between the new algorithm and a self-scaling quasi-Newton algorithm is revealed. If we assume the objective function is twice continuously differentiable and uniformly convex, we prove the iteration converge globally to the solution. And under some additional conditions, the superlinear convergence is given. Finally, the experimental results show that the proposed algorithm performs very well, which indicate that the numerical performance of the new algorithm is somewhat like the self-scaling quasi-Newton algorithm.

Keywords: modified BFGS method, convergence, unconstrained problems

1 Introduction

In this paper, we consider the following unconstrained optimization problems. See Equation (1).

$$\min f(x), x \in \mathbf{R}^n, \tag{1}$$

where $f: R^n \to R$ is continuously differentiable. The BFGS method is a well-known quasi-Newton method for solving unconstrained problems. During the past two decades, the study on global convergence of the quasi-Newton method has received growing interests. When f is convex, it was shown that Broyden's class of quasi-Newton methods converge globally if exact line search is used [1].When inexact line search is used, Powell [2] first obtained the global convergence of BFGS method. Byrd, Nocedal and Yuan [3] extended Powell's results to the restricted Broyden's class of quasi-Newton methods with Wolfe-Type line search except for DFP method. Yuan [4] and Wei [5] proposed modified secant equations which approximate the curvature of the objective function more accurate than the standard secant equation. On the other hand, in order to solve the nonconvex problems, Li and Fukushima [6] made a slight modification to the standard BFGS method.

Recently, Lewis and Overton [7] have shown in numerical experiments that the standard BFGS method works very well when applied directly without modifications to nonsmooth test problems as long as a weak Wolfe line search is used. All the mentioned methods in [3, 4, 5, 6] use both the gradient and function values available at the current iteration. In the cases that the Hessian matrix has some special structures and is partially available. For example, in the nonlinear least squares problems, the above mentioned secant equations have been developed by Amini and Ghorbani Rizi [8]. Xiao [9] proposed a modified limited memory BFGS method for solving the unconstrained. Saman [10] proposed two effective hybrid conjugate gradient algorithms based on modified BFGS updates. In the paper, we will analysis how to form the approximate matrix and how to use the formed approximate matrix to replace the inverse of the Hessian matrix in the BFGS method. Compared to the traditional BFGS method, the modified BFGS is also well-defined.

BFGS requires only matrix-vector multiplications which brings the computational cost at each iteration from $O(n^3)$ for Newton's method down to $O(n^2)$. However, if the number of variables is very large, even $O(n^2)$ per iteration is too expensive - both in terms of CPU time and sometimes also in terms of memory usage (a large matrix must be kept in memory at all times). According to the idea of Li and Fukushima. In the paper a modified BFGS method for unconstrained problems is applied even if the number of variables is very large. The modified BFGS method can save computing time.

The organization of this paper is as follows. In the next section we shall describe the new modified BFGS algorithm. If we assume the objective function is twice continuously differentiable and uniformly convex, according to the idea of Byrd, in section 3 we prove the iteration converge globally to the solution, and in section 4 under some additional conditions we show the method is superlinearly convergent. In section 5 the numerical results of the algorithm is shown. Finally, we make conclusions in section 6.

2 The new modified BFGS method

For brevity, we first introduce some notations used in this paper: $g_k = g(x_k), g = \nabla f$ gradient of f ; $G = \nabla^2 f$

* ***Corresponding author*** e-mail 330387742@qq.com

Hessian matrix of f ; d_k, λ_k, search vector/step length; $s_k = x_k - x_{k-1}, y_k = g_k - g_{k-1}$.

With $\|.\|$ we will denote the Euclidean vector norm in R^n, as well as the corresponding operator norm of matrices in $\mathbf{R}^{n\times n}$.

Algorithm 2.1(the new modified BFGS method)

Step1: Choose an initial point $x_1 \in \mathbf{R}^n$ and an initial positive definite matrix $\mathbf{B}_1$, and positive constants α, β with $0 < \alpha < 0.5, \alpha < \beta < 1$. Let $k = 0$.

Step2: If $g_k = 0$, stop; otherwise, go to Step3.

Step3: For given $x_k, \mathbf{B}_k$, solve the system of linear equation $\mathbf{B}_k d_k + g_k = 0$ to obtain a unique optimal solution d_k.

Step4: Compute λ_k, which satisfies

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k s_k s_k^{\mathrm{T}} \mathbf{B}_k}{s_k^{\mathrm{T}} \mathbf{B}_k s_k} + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}, \tag{4}$$

where $y_k = \begin{cases} \bar{y}_k \\ \bar{y}_k + (c\|g_k\| - \dfrac{\bar{y}_k s_k}{\|s_k\|^2}) \end{cases}$,

$$\text{if} \quad y_k^{\mathrm{T}} s_k \geq c\|g_k\|\|s_k\|^2, \tag{5}$$

else

$$\bar{y}_k = g_{k+1} - g_k + r_k s_k, \tag{6}$$

and

$$r_k = \frac{3(g_{k+1} + g_k)^{\mathrm{T}} s_k - 6(f_{k+1} - f_k)}{\|s_k\|^2} \tag{7}$$

then set $k = k+1$, go to Step2.

3 Global convergence

In this section, we study the global convergence of algorithm2.1.First we make the following assumptions:
(1): The level set $D = \{x \mid f(x) \leq f(x_0)\}$ is contained in a bounded convex set D.
(2): The objective function f is continuously differentiable on D and there is a constant $L > 0$ such that for any $x, y \in D$,

$$\|g(x) - g(y)\| \leq L\|x - y\|. \tag{8}$$

(3): The function f is uniformly convex, i.e., there exist two positive constants m and M such that

$$m\|z\|^2 \leq z^{\mathrm{T}} G(x) z \leq M\|z\|^2 \tag{9}$$

for all $z \in \mathbf{R}^n$ and all $x \in D$.

Lemma 3.1 Suppose the sequence $\{x_k\}$ is generated by Alorithm 2.1.Then we have for every $k = 0,1,\cdots$, $f(x_{k+1}) - f(x_k) < 0$.

Proof: For any positive definite matrix $\mathbf{B}_{k-1}$, from (2) we get $f(x_{k+1}) - f(x_k) \leq \alpha\lambda_k g_k^{\mathrm{T}} d_k = -\alpha\lambda_k d_k^{\mathrm{T}} \mathbf{B}_{j-1} d_k < 0$, it implies that the new iteration point belongs to D. Due to (3) we also have

$$y_k^{\mathrm{T}} d_{k-1} = (g_k - g_{k-1}) d_{k-1} \geq -(\beta - 1) d_{k-1}^{\mathrm{T}} \mathbf{B}_{j-1} d_{k-1} > 0. \tag{10}$$

Since $s_k = \lambda_{k-1} d_{k-1}$, from (10) we get $y_k^{\mathrm{T}} s_k > 0$, this indicates that Alorithm 2.1 has the property if $\mathbf{B}_{k-1}$ is symmetric positive definite, then $\mathbf{B}_k$ is also symmetric positive definite.

Lemma 3.2 Let the sequence $\{x_k\}$ is generated by Alorithm 2.1.Suppose the assumption hold. Then we can get

$$\|y_k\| \leq \max\{7L, 7L + cC_1\}\|s_k\|, k = 1, 2, \ldots. \tag{11}$$

Proof: By using (5), if $\bar{y}_k^{\mathrm{T}} s_k \geq c\|g_k\|\|s_k\|^2$, we have $y_k = \bar{y}_k$. Therefore, $\|y_k\| \leq \|g_{k+1} - g_k\| + |r_k|\|s_k\|$, from the (8), we have

$$y_k \leq (L + |r_k|)\|s_k\|, \tag{12}$$

where $|r_k| = \dfrac{|3(g_{k+1} - g_k)^{\mathrm{T}} s_k - 6(f_{k+1} - f_k)|}{\|s_k\|^2}$.

From the Taylor's formula, we get

$$|r_k| \leq \frac{|3(g_k + \theta_1 s_k) - g_{k+1} + 3(g_k + \theta_1 s_k) - g_k|}{\|s_k\|^2} \leq 6L. \tag{13}$$

Substitute (13) into (12), we can get $\|y_k\| \leq 7L\|s_k\|, k = 1, 2, \ldots$.

In the other hand, if $y_k = g_{k+1} - g_k + r_k + c\|g_k\| - \dfrac{\bar{y}_k^{\mathrm{T}} s_k}{\|s_k\|^2} s_k$, using the Taylor's formula and (8), we also get

$$\|y_k\| \leq \|g_{k+1} - g_k\| + |r_k| + c(\|g_k\| + \frac{\|\bar{y}_k^{\mathrm{T}} s_k\|}{\|s_k\|^2})\|s_k\|$$

$$\leq (7L + c\|g_k\|)\|s_k\|.$$

According to the assumption (1), there exists a positive constant C_1, $\|g_k\| \leq C_1$, thus, we have $\|y_k\| \leq (7L + cC_1)\|s_k\|$.

According to the assumption (1), there exists a positive constant C_1, $\|g_k\| \leq C_1$, thus, we have $\|y_k\| \leq (7L + cC_1)\|s_k\|$.

Lemma 3.3 Let the sequence $\{x_k\}$ is generated by Alorithm 2.1. Suppose $\mathbf{B}_0$ is symmetric positive definite.

Then there is a constant $H \geq 0$ for all $k \geq 1$, y_k and s_k satisfy

$$\frac{\|y_k\|^2}{y_k^{\mathrm{T}} s_k} \leq H, \tag{14}$$

Consequently, for any $p \in (0,1)$ there exist positive constants $\beta_1, \beta_2, \beta_3$ such that, for $k \geq 1$ the following inequality

$$\beta_2 \leq \frac{\|\mathbf{B}_j s_j\|}{\|s_j\|} \leq \frac{\beta_3}{\beta_1} \equiv \beta. \tag{15}$$

Proof: Using Lemma 3.2, we can easily get $\frac{\|y_k\|^2}{y_k^{\mathrm{T}} s_k} \leq \frac{\max\{7L, 7L + cC_1\}\|s_k\|}{\|s_k\|}$.

Therefore, we have $\frac{\|y_k\|^2}{y_k^{\mathrm{T}} s_k} \leq H$, where $H = \max\{7L, 7L + cC_1\}$. From the theorem 2.1 in [11], we get (15).

Our proof is completed.

We defined set $K = \{k \mid k \text{ satisfies (15)}\}$.

Thus, from the right hand of (15), we have

$$\|\mathbf{B}_k s_k\| \leq \beta \|s_k\|, \quad k \in K. \tag{16}$$

Moreover, for all $k \in K$, $\|\mathbf{B}_k s_k\| = |\lambda_k| \|\mathbf{B}_k d_k\| = |\lambda_k| \|g_k\| \leq \beta \|s_k\| = \beta |\lambda_k| \|d_k\|$.

So we can get

$$\|g_k\| \leq \beta \|d_k\|. \tag{17}$$

By using $y_k = \mathbf{B}_k s_k$ and (9), we have

$$m\|s_k\|^2 \leq \left\|s_k^{\mathrm{T}} \mathbf{B}_k s_k\right\| \leq M\|s_k\|^2, \tag{18}$$

where m and M are positive constants.

There m 3.4 Let the sequence $\{x_k\}$ is generated by Alorithm 2.1.Suppose the assumption hold. Then we get

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{19}$$

Proof: We firstly assume $\|g_k\| \geq \varepsilon$ for all $k \in K$ with positive constant ε. Using Lemma 3.1, (16), (18) and $\mathbf{B}_k s_k = \lambda_k \mathbf{B}_k d_k = -\lambda_k g_k$, we can get

$$+\infty \geq \sum_{k=0}^{\infty} (-g_k^{\mathrm{T}} s_k) = \sum_{k=0}^{\infty} \frac{1}{\lambda_k} s_k^{\mathrm{T}} \mathbf{B}_k s_k = \sum_{k=0}^{\infty} \lambda_k \frac{\|g_k\|^2}{\|\mathbf{B}_k s_k\|^2} s_k^{\mathrm{T}} \mathbf{B}_k s_k$$
$$\geq \sum_{k=0}^{\infty} \frac{\|g_k\|^2}{\|\mathbf{B}_k s_k\|^2} s_k^{\mathrm{T}} m \|s_k\|^2 \geq \sum_{k=0}^{\infty} \beta^{-2} m \|g_k\|^2 \geq \sum_{k=0}^{\infty} \beta^{-2} m \varepsilon^2 = \infty.$$

It is obvious that we have derived a contradiction, so (19) holds.

The above theorem is established the global convergence of Alorithm 2.1.

4 Superlinear convergence

In order to give the superlinear convergence of Algorithm 2.1, we also need the following assumptions. Let x^* be the limit of the sequence $\{x_k\}$.

(4): f is twice continuously differentiable for all x in the neighbourhood of x^*.

(5): $\{x_k\}$ converges to x^*, where $g(x^*) = 0$ and $G(x^*)$ is positive definite.

(6): There exist two constants $M_2 \geq 0$ and $v \geq 0$ such that

$$\left\|G(x) - G(x^*)\right\| \leq M_2 \left\|x - x^*\right\|^v. \tag{20}$$

Lemma 4.1 Suppose assumptions (4)-(6) hold. Let us define θ_k as the angle between s_k and $\mathbf{B}_k s_k$. According to Byrd [3], it follows that $\cos\theta_k = \frac{s_k^{\mathrm{T}} \mathbf{B}_k s_k}{\|s_k\| \|\mathbf{B}_k s_k\|}$, so we get

$$a_1 \|g_k\| \cos\theta_k \leq \|s_k\|, \tag{21}$$

where a_1 is a positive constant.

Proof: From the assumption (2) and the definition y_k, we obtain

$$y_k s_k^{\mathrm{T}} \geq c\|g_k\|\|s_k\|^2 \geq c\|s_k\|(-g_k^{\mathrm{T}} s_k) \geq c\varepsilon_0(-g_k^{\mathrm{T}} s_k). \tag{22}$$

Combining assumption (4) and assumption (2), we have $\|G(x)\| \leq M_1$, $\|g(x)\| \leq M_2$, with M_1 and M_2 are a positive constants.

So we get

$$M_1 \|s_k\|^2 \geq \|G(\varepsilon_1) s_k\| \|s_k\| = \|g_{k+1} - g_k\| \|s_k\| \geq y_k s_k^{\mathrm{T}}$$
$$\geq c\varepsilon_0(-g_k^{\mathrm{T}} s_k) = c\varepsilon_0 \|g_k\| \|s_k\| \cos\theta_k, \tag{23}$$

where $\varepsilon_1 = x_k + \delta s_k, \delta \in (0,1)$.

(23) implies (22) holds, where $a_1 = \frac{c\varepsilon_0}{M}$.

Lemma 4.2 If the assumptions (1)-(6) hold, for an arbitrary $v > 0$, we can obtain

$$\sum_{k=0}^{\infty} \left\|x_{k+1} - x_k\right\|^v < +\infty, \tag{24}$$

and

$$\sum_{k=0}^{\infty} r_k < +\infty, \tag{25}$$

where $r_k = \max\{\left\|x_{k+1} - x^*\right\|^v, \left\|x_k - x^*\right\|^v\}$.

Proof: By the assumption (4) and (5), we have the following inequality

$$\|g(x)\| = \|g(x) - g(x^*)\| \geq m_1 \|x - x^*\|, \quad (26)$$

where m_1 is a positive constant.

From (15) and Lemma 4.1, we have

$$\cos\theta_k \geq 1/\beta. \quad (27)$$

Therefore,

$$\begin{aligned} -g_k^{\mathrm{T}} s_k &= \|g_k\|\|s_k\|\cos\theta_k \geq a_1 \|g_k\|\cos^2\theta_k \\ &\geq a_1 m_1^2 \|x_k - x^*\|\cos^2\theta_k \\ &\geq a_1 m^2 \frac{1}{\beta}\|x_k - x^*\|^2 \end{aligned} \quad . \quad (28)$$

By (2), (3) and (28), we can get (24).It is obvious that (25) hold.

Theorem 4.3 Let $\{\mathbf{B}_k\}$ and $\{x_k\}$ be generated by Alorithm 2.1, f satisfies assumption (4)-(6), then for any $k \geq 0$, we can get

$$(1):\ \lim_{k\to\infty} \frac{\|(\mathbf{B}_k - G(x^*))s_k\|}{\|s_k\|} = 0, \quad (29)$$

(2): $\{x_k\}$ generated by Alorithm 2.1 superlinearly converges to x^*.

Proof: By using Lemma 4.1 and Lemma 4.2, we can easily obtain (29).Similar to the arguments of Dennis [11], we get

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0. \quad (30)$$

5 Numerical experiments

In this section we present the results of our numerical experiments to compare the algorithm suppose by Broyden in [1] and the algorithm 2.1 in this paper. For brevity, we use BFGS and MBFGS to represent the algorithm suppose by Broyden in [1] and the algorithm 2.1 in this paper. All the numerical experiments will be programmed by matlab2012b. Let us choose the same starting point for both BFGS and MBFGS. In algorithm 2.1, while $c = 10^{-3}$, $\alpha = 0.3, \beta = 0.8$, $\mathbf{B}_0 = E$, and the experiment is stop when $\|g\| \leq 10^{-5}$.

(i) Problem 1. $f(x) = x_1^4 + x_2^4 + 2x_1^2x_2^2 - 4x_1 + 3$,

(ii) Problem 2. $f(x) = (x_1 + x_2)^2 + [2(x_1^2 + x_2^2 - 1) - \frac{1}{3}]^2$,

(iii). Problem 3. (Rosenbrook's function, $n = 2$) $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$

(iv). Problem 4 (Cube function, $n = 2$) $f(x) = 100(x_2 - x_1^3)^2 + (1 - x_1)^2$

Please see Table 1.

TABLE 1 Numerical results of BFGS and MBFGS methods

Test Functions	Initial points	n_i		n_f		x^*	f^*
		BFGS	MBFGS	BFGS	MBFGS		
	(1.25,1.25)	8	5	41	26		
(i)	(12,12)	8	6	41	31	(1,0)	0
	(45,45)	8	6	41	31		
	(-1.25,0.5)	8	6	41	31		
(ii)	(-1.25,-1)	13	9	62	42	(0.763756, 0.763679)	0
	(-1.25,1)	7	5	35	26		
	(-10,10)	30	12	151	61		
(iii)	(10,-10)	68	23	341	156	(1,1)	0
	(100,100)	167	25	836	235		
	(10,10)	52	18	46	31		
(iv)	(30,80)	75	23	61	42	(1,1)	0
	(-15,15)	28	19	72	32		

In Table 1, n_i and n_f indicates the number of iteration and the number of function evaluation respectively. Meanwhile, *x** and *f** indicates minimum points and minimum value respectively. It can be seen from the comparison given above that the algorithm 2.1 in this paper is more efficient than BFGS method for solving unconstrained optimization.

6 Conclusions

In this paper, a modified BFGS algorithm for unconstrained problems is proposed. If we assume the objective function is twice continuously differentiable and uniformly convex, we prove the iteration converge globally to the solution. And under some additional conditions the superlinear convergence is given. The method retains the scale-invariance property of the original BFGS method. We have preliminary numerical results to show its efficiency. As demonstrated in Section 5, the reported numerical results show that the modified BFGS performs better than the BFGS in [1].

7 Acknowledgements

This work was supported by the youth innovation fund of Guangdong University of Petrochemical Technology (513022).

References

[1] Broyden C G 1970 The convergence of a class of double-rank minimization algorithms *Inst Math Appl.* **3**(6) 222-31

[2] Powell M J D 1971 On the convergence of the variable metric algorithm *Journal of the institute of Mathematics and Applications* **8**(7) 21-36

[3] Byrd R, Nocedal J, Yuan Y X 1987 Global convergence of a class of quasi-Newton methods on convex problems *SIAM Journal on Numerical Analysis* **7**(24) 1171-90

[4] Yuan Y X 1991 A modified BFGS algorithm for unconstrained optimization *IMA J Numer Anal.* **7**(11) 325–32

[5] Wei Z, Li G, Qi L 2006 New quasi-Newton methods for unconstrained optimization problems *Appl Math Comput.* **175**(2) 1156-88

[6] Li D, Fukushima M 2001 A modified BFGS method and its global convergence in nonconvex minimization *Journal of Computational and Applied Mathematics* **129**(4) 15-35

[7] Lewis A S, Overton M L 2010 Nonsmooth optimization via BFGS *SIAM*. 8(23) 73-90

[8] Amini K, Ghorbani Rizi A 2010 A new structured quasi-Newton algorithm using partial information on Hessian *Appl Math Comput.* **234**(2) 805-11

[9] Yunhai Xiao, Tingfeng Li, Zengxin Wei 2013 Global convergence of a modified limited memory BFGS method for non-convex minimization *Acta Mathematica Applicatae Sinica* **29**(3) 555-66

[10] Saman Babaie-Kafaki, Masoud Fatemi 2011 Two effective hybrid conjugate gradient algorithms based on modified BFGS updates *Numer Algorithm* **58**(3) 315-31

[11] Dennis J E, More J 1974 A characterization of superlinear convergence and its application to quasi-Newton methods *Math and Computation* **28**(6) 549-60

Authors

Ganzhou Wu, born on August 10, 1980, China

Scientific interests: optimization theory

Haiyan Liang, born on August 16, 1984, China

Scientific interests: data mining

Performance analysis and side lobe suppression in radon-Fourier transform based on random pulse repetition interval

Qian Chen[1, 2*], Junhao Liu[2], Chaowei Fu[2], Haitao Wang[2]

[1]*School of Information and Electronics, Beijing Institute of Technology, Beijing, 100081, China*

[2]*Shanghai Radio Equipment Research Institute, Shanghai, 200090, China*

Received 1 June 2014, www.cmnt.lv

Abstract

In order to solve the Blind Speed Side Lobe (BSSL) appeared in Radon-Fourier Transform (RFT) method used for dim target detection, a novel method of BSSL suppression is proposed in this paper. It is based on Random Pulse Repetition Interval (RPRI). The process of RPRI-RFT and the BSSL properties are described, the performance of coherent integration and the modulation noise of RFT algorithm based on RPRI are analysed in detail. Both the theoretical analysis and the numerical experimental results show that RPRI-RFT can be used to improve signal-to-noise ratio (SNR) and suppress BSSL effectively, and the influence of modulation noise of RPRI can be suppressed by the long-time integration characteristic, thus significantly improve the ability of low pulse repetition frequency radar to detect and measure long-range weak high-speed multi-targets.

Keywords: Blind Velocity Side Lobe, Radon-Fourier Transform, Random Pulse Repetition Interval

1 Introduction

With the development of radar stealth technology, the radar cross section (RCS) of target decreases sharply, weakening the energy of radar echo and diminishing the range of radar detection. Long-time integration is an effective method to improve the detection performance of dim targets [1-3]. With the increase of integration time, however, the problem of across range unit (ARU) walk will occur inevitably. Performing directly conventional methods of moving target detection (MTD) on high-velocity targets, the energy of targets spread over multiple resolution units [4-6] and cannot be integrated effectively. In this case, the increase of integration time cannot improve the performance of radar detection.

Non-coherent integration and coherent integration are two basic methods for long-time integration. The former mainly accumulates the energy along with the motion trajectory of targets to solve the problem of ARU. As a typical algorithm, the Hough Transform (HT) proposed by Carlson does not need phase compensation, making it easy to implement [7-9]. However, without phase information, the energy of target cannot be accumulated completely and this method cannot work well in extremely low SNR scenario [10]. The latter has drawn extensive attention from academic community and has been deeply investigated in recently years, because of its better integrated gain in the field of dim target detection. Techniques for ARU compensation in the long-time coherent integration can be classified into two types. In the first type, ARU compensation can be implemented by shifting or expanding the envelope after pulse compression [11]. In order to guarantee the compensation precision, interpolation in range dimension should be adopted, so the memory cost is huge. In the second type, ARU compensation is realized in the transform domain of the range-pulse plane, such as Keystone transformation, achieving ARU compensation of constant-velocity targets [12-14]. This method is free of searching operation, but it cannot eliminate the ARU effect for multiple targets with different ambiguity number simultaneously.

In recent years, based on the coupling relationship between range walk and Doppler frequency, Radon-Fourier Transform (RFT) is proposed to realize the long-time coherent integration [15-17] by mapping the echo pulse to the range-Doppler plane. RFT concentrating the energy to a focused peak, improves the radar detection performance of dim targets. However, because of the discrete pulse sampling, finite range resolution and limited integration time, the BSSL still exists in the case of Low Pulse Repetition Frequency (LPRF). Since the BSSL would increase the ratio of false alarm and deteriorate the detection performance, a symmetrical weighting method has been proposed to suppress the BSSL in RFT integration plane. The locations of BSSLs can be controlled via different weighting functions, and the BSSLs can be eliminated by combing the different weighted RFT results [16]. However, BSSLs still cannot be eliminated completely when the BSSLs appear simultaneously in both RFT outputs. Another method based on the design of pulse repetition interval (PRI) has been proposed in Ref. [18]. By jointly processing the RFT outputs in two adjacent Coherent Processing Intervals (CPIs), minimum criterion is employed between

* ***Corresponding author*** e-mail 3241126@qq.com

the two different PRI to achieve BSSL suppression. It should be note that the algorithm needs at least two CPIs, so that the integration time is increased twice and the radar efficiency is halved. A BSSL suppression method based on CLEAN algorithm has been proposed in [19]. The peak position of echo envelope is first obtained after RFT operation, and according to the relationship among BSSL, velocity and range, the positions of BSSL are confirmed and removed. And then the RFT peak will be determined. Though the BSSL of one target can be well restrained, the target cannot be detected when its main lobe overlaps the BSSL of another target.

In this paper, a novel RFT algorithm based on Random Pulse Repetition Interval (RPRI) is proposed to suppress BSSL. And the paper is organized as follows: the principle of RFT algorithm is analysed, and the causes of BSSL of discrete RFT are presented in section 2. In section 3, the RFT algorithm based on RPRI is proposed, and the analyses about the modulation noise and BSSL suppression performance of RFT algorithm are presented. In section 4, the simulation results of the proposed method are shown. Finally, the conclusions are given out in section 5.

2 The characteristics of RFT

2.1 SIGNAL MODEL

Suppose a linear frequency modulation (LFM) signal $p(t)$ is used by the radar transmitting, i.e.,

$$p(t)=\mathrm{rect}\left(t/T_{\mathrm{P}}\right)\exp\left(\mathrm{j}\pi\gamma t^{2}\right), \tag{1}$$

where T_{p} is the pulse duration, γ is the modulation frequency rate, and $\mathrm{rect}(\bullet)$ is the rectangular function.

For a point target with the RCS of Swerlling-0 type, the received echo signal can be written as

$$s_{\mathrm{r}}(\hat{t},t_m)=A_{\mathrm{r}}\,p\left(\hat{t}-2r\left(t_m\right)/c\right)\exp\left(-\mathrm{j}4\pi f_{\mathrm{c}}\,r\left(t_m\right)/c\right), \tag{2}$$

where $\hat{t}$ and t_m ($t_m=mT_{\mathrm{r}}$, T_{r} is the PRI) denote the quick-time and slow-time respectively, f_{c} is the carrier frequency, c represents the light velocity and A_{r} is the amplitude of target. It is supposed that the radial velocity of the target is uniform. The instantaneous distance between radar and target can be expressed as $r\left(t_m\right)=r_0+v_0t_m$ where r_0 is the initial distance from radar platform to target, and v_0 is the radial velocity of target.

After range compression, the echo signal can be represented as

$$\begin{aligned}s_{\mathrm{r,M}}(\hat{t},t_m)=&A_0\mathrm{sinc}\left(B\left(\hat{t}-2\left(r_0+v_0t_m\right)/c\right)\right)\bullet\\&\exp\left(-\mathrm{j}4\pi f_{\mathrm{c}}\,r_0/c\right)\exp\left(-\mathrm{j}4\pi f_{\mathrm{c}}\,v_0t_m/c\right)\end{aligned}, \tag{3}$$

where A_0 is the amplitude of baseband signal, B is the bandwidth of LFM signal.

For the conventional MTD method，echo envelop is supposed to be in the same range unit, and the coherent integration results can be written as

$$S_{\mathrm{P}}(t,v)=\int_{T_{CPI}}s_{\mathrm{r,M}}\left(\hat{t},t_m\right)\exp\left(\mathrm{j}4\pi vt_m/\lambda\right)\mathrm{d}t_m, \tag{4}$$

where λ is the wavelength, $S_{\mathrm{P}}\left(t,v\right)$ is the coherent integration result in the range-velocity plane.

In equation (4), the pulse integration time should satisfy

$$T_{\mathrm{CPI}}\le\frac{\rho_{\mathrm{r}}}{v_{\max}}, \tag{5}$$

where ρ_r is the range resolution, and $\rho_{\mathrm{r}}=c/\left(2B\right)$.

However, in general cases, the maneuvering targets cannot satisfy (5) during the long integration time, so ARU will appear, thus affecting the coherent integration of energy. RFT algorithm is a solution way to deal with the effect of ARU. For uniform velocity motion, it can be expressed as

$$G\left(r,v\right)=\int s_{\mathrm{r,M}}\left(\frac{2\left(r+vt_m\right)}{c},t_m\right)\exp\left(\mathrm{j}\frac{4\pi vt_m}{\lambda}\right)\mathrm{d}t_m, \tag{6}$$

Equation (6) shows that the echo envelope of each slow-time is extracted along with the motion track $r\left(t_m\right)=r_0+v_0t_m$, and the phase terms are compensated based on the relationship between velocity and Doppler frequency. In contrast, equation (4) shows that the sample points are extracted in the same range cell for different pulse repetition interval. Therefore, RFT can integrate the energy of target completely, while MTD cannot. Since motion track of the target is unknown before detection, two-dimensional joint searching of range and velocity is required in the RFT algorithm.

In practical application, the discrete form of (6) can be rewritten as

$$\begin{aligned}G_D\left(r,v\right)=&\sum_{m=0}^{M-1}s_{\mathrm{r,M}}\left(\mathrm{round}\left(\frac{r+vmT_r}{\rho_r}\right),m\right)\bullet\\&\exp\left(\mathrm{j}4\pi\,vmT_{\mathrm{r}}/\lambda\right)=\\&A_0\sum_{m=0}^{M-1}\mathrm{sinc}\left(2B\left((r\text{-}r_0)+mT_r\left(v-v_0\right)\right)/\mathrm{c}\right)\bullet\\&\exp\left(\mathrm{j}4\pi\left(v-v_0\right)mT_{\mathrm{r}}/\lambda\right)\end{aligned}, \tag{7}$$

where M is the number of integrated pulses. The maximum of integration peak can be found by searching in distance-velocity parameters' plane, and the peak value can be written as

$$G\left(r_0,v_0\right)=A_0M. \tag{8}$$

2.2 RFT BSSL CHARACTERISTICS

For LPRI Pulse Doppler (PD) radar, the accumulating gain of MTD method (shown in (4)) at ambiguous frequency point $f_{\mathrm{d}}+k/T_{\mathrm{r}}$, corresponding to the ambiguous velocity $v(k)=v_0+kv_{\mathrm{b}}$, where $v_{\mathrm{b}}=\lambda/2T_{\mathrm{r}}$ is the blind velocity, is the same with true Doppler frequency point, the true velocity cannot be ascertained. For RFT method from equation (7), although phase fluctuation of ambiguous velocity $v(k)$ can still be compensated completely, only some sampling units can be accumulated effectively according to the false motion track of ambiguous velocity $v(k)$, if they are in the same range unit with the real motion track $r(t_m)=r_0+v_0t_m$. The valid integration number L of accumulated pulses, which satisfies the above condition, can be represented as

$$L=\begin{cases}\mathrm{round}\left(\left|\rho_r/(v-v_0)T_r\right|\right) & v\neq v_0\\ M & v=v_0\end{cases}. \tag{9}$$

In this case, RFT results can be rewritten as

$$G\left(r_0,v(k)\right)=A_0L_k=\frac{2A_0\rho_{\mathrm{r}}}{k\lambda}. \tag{10}$$

Thus, for LPRI PD radar, one main accumulating peak and several BSSLs are generated in RFT output, and the Primary Lobe-to-Side Lobe Ratio (PSLR) is

$$I_k=20\lg\left(\frac{Mk\lambda}{2\rho_{\mathrm{r}}}\right). \tag{11}$$

For constant wavelength, BSSL is relevant with the ambiguity number k, the number of accumulated pulse M and the distance resolution ρ_{r}. RFT processing can suppress the effect of MTD method on velocity ambiguity, and the suppression ratio is I_k. But RFT still cannot resolve the velocity ambiguity completely. In the case of multiple targets, the phenomena of false alarms will still exist when BSSLs exceed the threshold.

3 BSSL Suppression of RFT based on RPRI

3.1 PROCESS OF RPRI-RFT

PD radar with RPRI resolves the problem of velocity ambiguity by adding random jitter to uniform PRI, and it shows good performance in anti-interception and anti-electronic-jamming.

In RPRI case, slow-time t_m is random, and it can be represented as

$$t_m=mT_{\mathrm{r}}+d_m=(m+\delta_m)T_{\mathrm{r}}, \tag{12}$$

where δ_m is a jitter of uniform distribution added to the average repetition interval T_{r}, which satisfies $-0.5\le\delta_m\le0.5$. Then the discrete expression of RPRI-RFT can be written as

$$G_D(r,v)=\sum_{m=0}^{M-1}s_{\mathrm{r,M}}\left(\mathrm{round}\left((r+vmT_{\mathrm{r}}+vd_m)/\Delta R\right),m\right)\cdot\exp\left(4\pi\mathrm{j}v(mT_{\mathrm{r}}+d_m)/\lambda\right). \tag{13}$$

Comparing (13) with (7), there are two differences between RPRI-RFT and RFT. First, the envelope delay of the former increases with the random jitter, which is related to d_m instead of increasing linearly with slow-time t_m. When $vd_m\ll\Delta R$, the envelope delay jitter caused by RPRI is much less than the range resolution, and the RPRI can be considered to have no influence on the sampling of envelope. In fact, the condition usually can be satisfied in reality. Secondly, the phase term of RPRI-RFT is not only related to Doppler modulation, but also including a random term $\exp(4\pi\mathrm{j}v_0d_m/\lambda)$, which introduces random noise component to the processing result of RFT. Therefore, RPRI-RFT is equivalent to Non-Uniform Discrete Fourier Transform (NUDFT) along the track of range migration. The Mean and variance of $G_D(r,v)$ are analysed as follows.

The Mean of RPRI-RFT is expressed as

$$\begin{aligned}&E\left(G_D(r,v)\right)=\\ &E\left(\sum_{m=0}^{M-1}s_{\mathrm{r,M}}\left(\mathrm{round}\left((r+vmT_{\mathrm{r}})/\rho_r\right),m\right)\cdot\exp\left(\mathrm{j}4\pi\frac{v(mT_{\mathrm{r}}+d_m)}{\lambda}\right)\right)=\\ &A_0E\left\{\sum_{m=0}^{M-1}\mathrm{sinc}\left(2B(r-r_0)+(mT_r+d_m)(v-v_0)/c\right)\cdot\exp\left(\mathrm{j}4\pi(v-v_0)(mT_{\mathrm{r}}+d_m)/\lambda\right)\right\}\approx\\ &A_0E\left\{\left(\sum_{m=0}^{M-1}\mathrm{rect}(r-r_0)+(mT_r+d_m)(v-v_0)/\rho_r\right)\cdot\exp\left(\mathrm{j}4\pi(v-v_0)(mT_{\mathrm{r}}+d_m)/\lambda\right)\right\}\approx\\ &A_0E\left\{\sum_{m=0}^{L-1}\exp\left(\mathrm{j}4\pi(v-v_0)(mT_{\mathrm{r}}+d_m)/\lambda\right)\right\}\triangleq\\ &A_0x(r,v)\end{aligned}. \tag{14}$$

In (14), the condition supported the final approximate equation is $vd_m\ll\rho_{\mathrm{r}}$, that is, the envelope jitter caused by RPRI is far less than the range resolution. As a result, $x(r,v)$ in equation (14) is written as

$$
\begin{aligned}
&x(r,v)= \\
&E\left(\sum_{m=0}^{L-1}\exp\left(\mathrm{j}4\pi(v-v_0)(mT_\mathrm{r}+d_m)/\lambda)\right)\right)= \\
&\sum_{m=0}^{L-1}\exp\left[\mathrm{j}4\pi\left(v-v_0\right)mT_\mathrm{r}/\lambda\right]\bullet \\
&\mathrm{E}\left\{\exp\left[\mathrm{j}4\pi\left(v-v_0\right)d_m/\lambda\right]\right\}=\mathrm{sinc}\left[4\left(v_0-v\right)\alpha T_\mathrm{r}/\lambda\right]\bullet \\
&\sum_{m=0}^{L-1}\exp\left[\mathrm{j}4\pi\left(v-v_0\right)mT_\mathrm{r}/\lambda\right]=\mathrm{sinc}\left[4\left(v_0-v\right)\alpha T_\mathrm{r}/\lambda_\mathrm{r}\right]\bullet \\
&\frac{L\,\mathrm{sinc}\left[2\left(v_0-v\right)LT_\mathrm{r}/\lambda\right]}{\mathrm{sinc}\left[2\left(v_0-v\right)T_\mathrm{r}/\lambda\right]}\exp\left[\mathrm{j}2\pi\left(v_0-v\right)\left(L-1\right)T_\mathrm{r}/\lambda\right]
\end{aligned}
\quad . (15)
$$

The variance of RPRI-RFT is expressed as

$$
\begin{aligned}
\mathrm{D}\left(G_D\left(r,v\right)\right)=\mathrm{D}\Bigg\{&\sum_{m=0}^{M-1}\mathrm{s}_{\mathrm{r,M}}\left(\mathrm{round}\left(\left(r+vmT_\mathrm{r}\right)/\Delta R\right),m\right)\bullet \\
&\exp\left(\mathrm{j}\frac{4\pi v\left(mT_\mathrm{r}+d_m\right)}{\lambda}\right)\Bigg\} \\
=LA_0^{\,2}&\mathrm{sinc}^2\left(2B(r-r_0)\right)\left\{1-\mathrm{sinc}^2\left(4\left(v_0-v\right)\alpha T_\mathrm{r}/\lambda\right)\right\}
\end{aligned}
\quad . (16)
$$

In (16), when $v=v_0$, $D\left(G_D\left(r_0,v_0\right)\right)=0$, the main integration peak of RPRI-RFT has no noise component. When $v\neq v_0$, $L=\mathrm{round}\left(\left|\rho_\mathrm{r}/\left(v-v_0\right)T_\mathrm{r}\right|\right)$, however, $\mathrm{D}\left(G_D\left(r,v\right)\right)$ becomes smaller with increasing the deviate degree of v from v_0 according to equation (9). From (15) and (16), it can be seen that both the mean and variance of $G_D\left(r,v\right)$ have no relationship with the number of integrated pulses M. Therefore, the advantage of RPRI-RFT is that the amplitude of main peak corresponding to target parameter $\left(r_0,v_0\right)$ increases with the same proportion as M, while the variance of modulation noise corresponding to other range-velocity parameter points decreases when v deviates from v_0. Although noise is introduced by random modulation, it is still restrained after RFT processing so that target can be detected in low SNR when M reaches a certain value.

3.2 BSSL PROPERTIES OF RPRI-RFT

For the case of $v=v_0$, $L=M$, $E\left(G_D\left(r_0,v_0\right)\right)=A_0M$ and $D\left(G_D\left(r_0,v_0\right)\right)=0$, the random modulation has no influence on the integration amplitude of RFT Primary lobe.

The BSSL of RPRI-RFT output is random, and its mean is written as

$$
E\left(G_D\left(r_0,v(k)\right)\right)=\frac{2A_0\rho_r}{k\lambda}\mathrm{sinc}(2\alpha k)\,. \qquad (17)
$$

Compared with (10), the improvement factor of BSSL for RPRI-RFT can be written as

$$
I_a=\mathrm{sinc}(2\alpha k)\,. \qquad (18)
$$

It can be seen from Equation (18) that, the improvement of BSSL suppression is related to the jitter α and ambiguity number k. Specifically when $\alpha=0.5$, the mean of BSSL is 0.

The noise variance introduced after RPRI –RFT is computed to be

$$
\mathrm{D}\left(G_D\left(r_0,v(k)\right)\right)=\frac{2\rho_\mathrm{r}A_0^{\,2}\left(1-\mathrm{sinc}^2(2\alpha k)\right)}{k\lambda}\,. \qquad (19)
$$

In this case, PSLR of RPRI-RFT can be rewritten as

$$
\bar{I}_k=10\lg\left(\frac{M^2}{\left(\frac{4\rho_\mathrm{r}^{\,2}}{k^2\lambda^2}-\frac{2\rho_\mathrm{r}}{k\lambda}\right)\mathrm{sinc}^2\left(2\alpha k\right)+\frac{2\rho_\mathrm{r}}{k\lambda}}\right). \qquad (20)
$$

Equation (20) shows that, PSLR of RPRI-RFT can increase with the increase of M or the jitter α, so the performance of PSLR can be improved effectively by M and α. Formula (19) also shows that BSSL variance will reduce when ambiguity k increases, and however, BSSL variance cannot be completely zero even if α =0.5. Therefore, in order to eliminate BSSL, parameters α and M should be chosen reasonably.

4 Numerical experiments

Numerical simulations are done to verify the above analysis. The parameters of targets and radar are as follows: the radar carrier frequency is 2.5 GHz; the signal bandwidth is 5MHz; the complex sampling frequency is 10MHz; the pulse duration is 10μs; and average repetition frequency is 1 kHz. Three targets are assumed, and their distance and velocity are T_1(96 km, 680 m/s), T_2(96 km, 620 m/s) and T_3(94 km, 680 m/s) respectively.

4.1 UNIFORM PRI PROCESSING

The range walk does not occur when M = 32. For uniform PRI, the result of MTD is shown in Figure 1. Velocity ambiguity has emerged on three targets so that the real velocity cannot be obtained correctly. In addition, MTD processing cannot distinguish T_1 from T_2 in the same distance and Doppler unit in Figure 1.

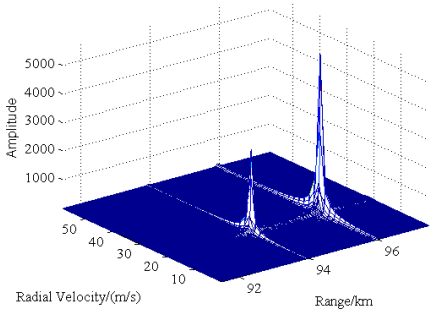
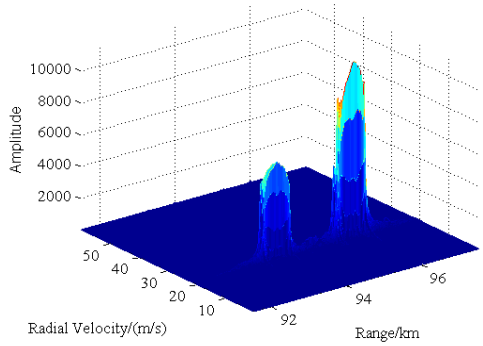
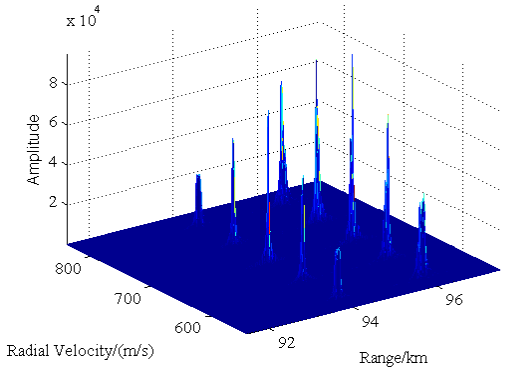


FIGURE 1 MTD results of uniform PRI when M=32

The ARU effect become much more evident when M =1024. The results of MTD and RFT are shown in Figure 2. Figure 2(a) shows that MTD processing cannot distinguish the three targets, because of ARU effect. Figure 2(b) shows that the targets cannot be distinguished because of BSSL, in this case, the PSLR is $I_k = 6.2\text{dB}$.



(a) MTD results



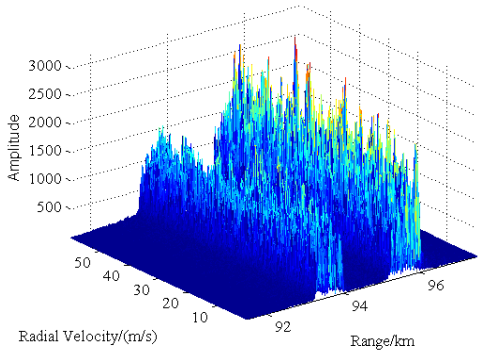
(b) RFT results

FIGURE 2 Results of uniform PRI when M =1024

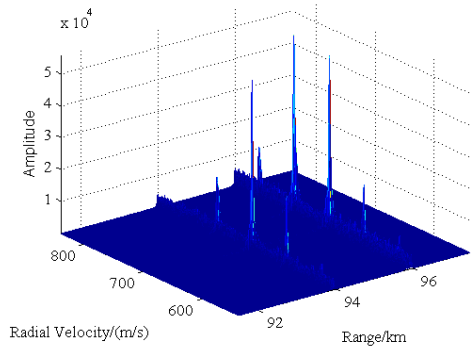
4.2 RESULT OF RPRI-RFT

On the basis of uniform PRI, RPRI status is presented, and the suppression of different jitters on blind side lobe is analysed, which is shown in Figure 3: (1) When $\alpha = 0.3$, MTD shown in Figure 3(a) is directly applied, and the velocity dimension accords with noise distribution; (2) When $\alpha = 0.3$, the RFT processing results are shown in Figure 3(b). The BSSL is further suppressed, and $\bar{I}_1$ =10.9dB, which is 4.7dB higher than Figure 2(b); (3) When $\alpha = 0.5$, the RFT processing results are shown in Figure 3(c). The BSSL of RFT is reduced sharply, and the PSLR is $\bar{I}_1$ =27.7dB.

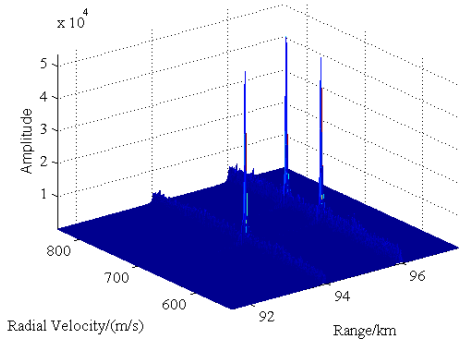
From the above simulation results, it can be observed that the noise does not affect distance resolution along the distance dimension, and the effect of modulation noise is small out of range resolution unit after RPRI-RFT processing; along the velocity dimension, the random modulation noise can be further suppressed, and the influence of modulation noise will be smaller when a certain pulse number M is selected.



(a) MTD results of $\alpha = 0.3$



(b) RPRI -RFT results of $\alpha = 0.3$



(c) RPRI-RFT results of $\alpha = 0.5$

FIGURE 3 Results of RPRI at M =1024

The following Figure 4 shows that PSLR values of different jitter α and the number of accumulated pulse M correspond to the position of the first side lobe.

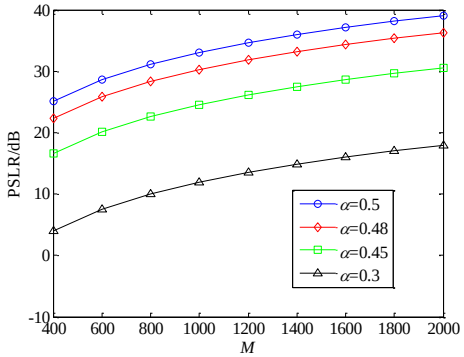
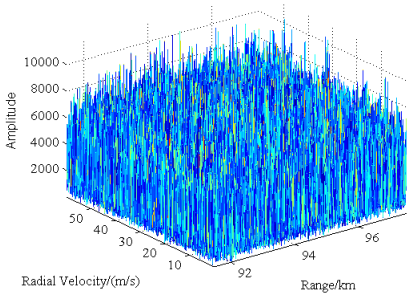


FIGURE 4 PSLR value of different α and M

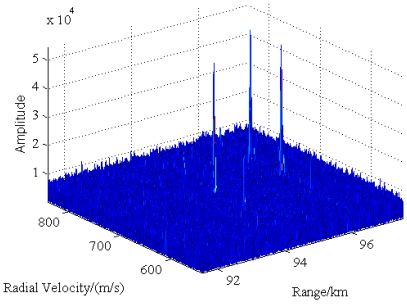
From Figure 4, it shows that the PSLR can be improved to 20 dB when M is more than 600 and $\alpha \geq 0.45$, and the PSLR can be improved to 30 dB when M is more than 1000 and $\alpha \geq 0.48$. These results show different ability to suppress of the BSLL at different parameters.

4.3 PERFORMANCE ANALYSIS OF RPRI-RFT IN LOW SNR

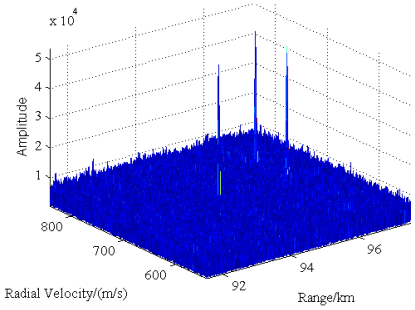
On the basis of section 4.2, Gaussian white noise is introduced, and the input SNR is -20dB. Experimental results are shown in Figure 5. Figure 5(a) shows that targets are submerged in noise after MTD processing. Figure 5(b) shows that signal energy can be integrated so that SNR can be improved substantially after RPRI-RFT processing. Figure 4(c) shows that after the increase of the PRI jitter, BSSL is submerged in noise, and the three pecks of the targets are all much higher than the noise background when the pulse number M is large enough (M =1024),in this case, the output SNR is 20.7dB.



(a) MTD results at $\alpha = 0.3$



(b) RPRI-RFT results at $\alpha = 0.3$



(c) RPRI- RFT results at $\alpha = 0.5$

FIGURE 5 Results of RPRI at M=1024 and SNR=-20dB

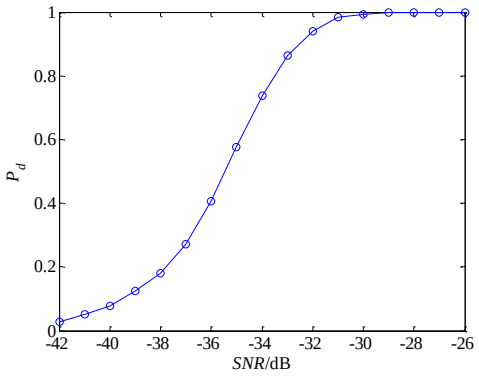


FIGURE 6 Detection performance at different input SNR

In order to analyse the detection performance at low input SNR, the results are given at different input SNR of 1000 Monte-Carlo simulation experiments in Figure 6. With detection probability P_d =80% and a given constant false alarm ratio P_f =10^{-6}, Figure 6 shows that the needed input SNR is -34.3dB at RPRI-RFT method.

Simulation results illustrate that RPRI-RFT can be well applied to resolve the problem of dim target detection and anti-velocity ambiguity, thereby achieving accurate velocity and distance of targets.

5 Conclusion

For the detection of the high-velocity dim target, the problem of ARU in long coherent integration time can be solved by RFT algorithm. However, BSSL of RFT algorithm may affect the detection when multiple targets exist in the observed scene. Therefore, a RFT algorithm with RPRI modulation has been proposed to restrain BSSL in this paper. The process of RPRI-RFT has been described and the BSSL properties of RPRI-RFT have been analysed in details. The experimental results have shown that RPRI-RFT can effectively restrain BSSL and the influence of modulation noise of RPRI can be suppressed by the characteristic of long-time integration of RFT. Therefore, the ability to detect long-range and high velocity dim targets has been improved effectively by RPRI-RFT, and proved by the simulation results.

Acknowledgment

This work has been supported in part by Shanghai Aerospace Technology Innovation Fund (Project No. SAST201224).

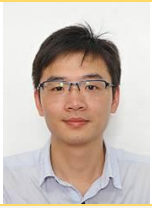
References

[1] Skolnik M 2002 Opportunities in radar-2002 *Electronics and Communication Engineering Journal* **14**(6) 263-72
[2] Cleetus G M 1976 Properties of staggered PRF Radar Spectral Components *IEEE Transaction of Aerospace and Electronic System* **12**(6) 800-3
[3] Vergara Dominguez L 2004 Analysis of the digital MTI filter with random PRI *IEE Proceedings-F* **140**(2) 129-37
[4] Barton D K 2004 *Radar System Analysis and Modeling* Beijing: Publishing House of Electronics Industry
[5] Skolnik M 2002 *Introduction to Radar System* (3nd rd) Columbus, OH: McGraw-Hill
[6] Yuxi Zhang, Jinping Sun, Bingchen Zhang, Wen Hong 2011 Doppler Ambiguity Resolution Based on Compressive Sensing Theory *Journal of Electronics & Information Technology* **33**(9) 2103-7
[7] Carlson B D, Evans E D, Wilson S L 1994 Search Radar Detection and Track with the Hough Transform Part I: System Conception *IEEE Transactions on Aerospace and Electronic Systems* **30**(1) 102-8
[8] Carlson B D, Evans E D, Wilson S L 1994 Search Radar Detection and Track with the Hough Transform Part II: Detection Statistics *IEEE Transactions on Aerospace and Electronic Systems* **30**(1) 109-15
[9] Carlson B D, Evans E D, Wilson S L 1994 Search Radar Detection and Track with the Hough Transform Part III: Detection Performance with Binary Integration *IEEE Transactions on Aerospace and Electronic Systems* **30**(1) 116-25
[10] Richards M A 2005 *Fundamentals of Radar Signal Processing* New York: McGraw-Hill
[11] Wang Y M, Ma J G, Fu Q, Zhuang Z W 2006 Research on integration detection of high-velocity moving target *Modern Radar* **28**(3) 24-7 *(in Chinese)*
[12] Perry R P, Dipietro R C, Fante R L 2007 Coherent integration with range migration using Keystone formatting *In Proceedings of IEEE Radar Conference*
[13] Su J, Xing M, Wang G, Bao Z 2010 High-velocity multi-target detection with narrowband radar *IET Radar Sonar and Navigation* **4**(4) 595-603
[14] Zhang S S, Zeng T 2005 Dim target detection based on Keystone transform In *Proceedings of IEEE International Radar Conference* May 889-94
[15] Xu J, Yu J, Peng Y, *et al.* 2011 Radon-Fourier transform for radar target detection I: generalized Doppler filter bank *IEEE Transactions on Aerospace and Electronic Systems* **47**(2) 1183-202
[16] Xu J, Yu J, Peng Y-N, *et al.* 2011 Radon-Fourier Transform for Radar Target Detection, II: Blind Velocity Side lobe Suppression *IEEE Transactions on Aerospace and Electronic Systems* **47**(4) 2473-89
[17] Xu J, Yu J, Peng Y-N, *et al.* 2012 Radon-Fourier transform for radar target detection III: Optimality and Fast Implementations *IEEE Transactions on Aerospace and Electronic Systems* **48**(2) 991-1004
[18] Li-chang Qian, Jia Xu, Wen-feng Sun, Ying-ning Peng 2012 Blind Speed Side Lobe Suppression in Radon-Fourier Transform Based on Radar Pulse Recurrence Interval Design *Journal of Electronics & Information Technology* **34**(11) 2608-14 *(in Chinese)*
[19] Lichang Qian, Jia Xu, Wenfeng Sun, *et al.* 2012 CLEAN based blind velocity side lobe (BSSL) suppression in the Radon Fourier Transform (RFT) for multi-target detection *Proc. of 2012 IEEE 12th International Conference on Computer and Information Technology* 490-5

Authors

Qian Chen, born on March 16, 1975, China

Current position, grades: pursuing his Ph.D. degree in Beijing University of Science and Technology.
University studies: B.S. degree from University of Electronic Science and Technology of China, in 1996, and the M.S. degree from Shanghai Radio Equipment institute, in 1999, China.
Experience: He joined in the Shanghai Radio Equipment Research Institute in 1999, and has been a vice director.
Scientific interest: are signal process, automobile anti-collision and radar system. He has been long engaged in short-range radar technology research.

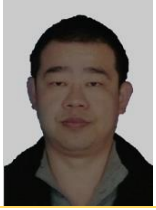


Junhao Liu, born on February 12, 1988, China

University studies: B.S. degree in communication engineering from Nanjing University of Science & Technology, Nanjing, China, in 2011. He received the M.S. degree in 2014 from Shanghai Academy of Spaceflight Technology, Shanghai, China.
Experience: He joined Shanghai Radio Equipment Research Institute, China, in 2014, where he has been engaged in research on weak signal processing, and short-range radar system.

Chaowei Fu, born on May 21, 1985,China

Current position, grades: a researcher at Shanghai Radio Equipment Research Institute, China.
University studies: B.S. and M.S. degrees in engineering of signal and information processing from Xidian University, Xi'an, China, in 2010.
Scientific interest: SAR Imaging and Signal processing.



Haitao Wang, born on August 13, 1978, China

Current position, grades: senior engineer in Shanghai Radio Equipment Research Institute.
University studies: B.S. degree in Shandong University, Jinan, Shandong, in 2000. He received the M.S. degree in 2003 from Shanghai Academy of Spaceflight Technology, Shanghai, China.
Scientific interest: radar signal processing, radar system and Terahertz detection technology.

Improvements of ant colony algorithm and its applications in artificial neural network

Kuangfeng Ning*, Xiliang Zeng

Hunan University of International Economics, Changsha 410205, Hunan, China

Received 1 June 2014, www.cmnt.lv

Abstract

Ant colony algorithm (ACA) is a bionic intelligent optimization algorithm with positive feedback, distributed computing and heuristic search. As an important branch of computational intelligence and swarm intelligence, ACA has been successfully applied in solving many combinatorial optimization problems. Artificial neural network is a large-scale distributed parallel processing system with the characteristics of self-organization, self-study, self-adaptation and non-linear dynamic processing and it has a broad prospect in settling the complicated non-linear problems. This paper has proposed an algorithm used to solve multi-objective optimization problems and the applications of ACA.

Keywords: Ant Colony Algorithm; Neural Network

1 Introduction

Among the numerous swarm intelligence algorithms, the most remarkable is a self-organizational behaviour of the ant colony, namely that the ant can always find the shortest path between the ant nest and the food source in the foraging. Without visual capacity in finding the food, the ants can only find the path by releasing pheromone on the way. The ant individuals convey information through pheromone and choose the passing path according to the pheromone concentration. The higher pheromone concentration the path has, the higher probability the ants will choose this path. In this way, the process where the ants choose the path has formed a positive-feedback mechanism. The pheromone concentration of the optimal path increases gradually with time while those of the other paths decrease relatively and the ants will finally find the shortest path [1].

Dorigo M, an Italian scholar had first put forward a heuristic swarm intelligent bionic optimization algorithm, namely Ant Colony Algorithm (ACA) by simulating ant's behaviour in finding the shortest path. This algorithm has strong robustness, global convergence ability, distributed computing and feedback mechanism. In a short span of ten years, ACA has been extensively applied in a great number of optimization fields, including combinatorial optimization, network route, function optimization and robot path planning and it has been doing well especially in lots of discrete or continuous combinatorial optimization problems. At present, ACA has been used to settle a lot of practical problems such as Traveling Salesman Problem (TSP), the blind inspection problem of the signals, the assignment problem and Job Shop Scheduling Problem (JSSP) and it has achieved excellent effects [2].

ACA has a short period of development and there are still plenty of problems to be further researched and settled. It has been proved by the theoretical research and the practical applications in many fields that this algorithm has a bright development future. With the development of information technology and the advancement of human cognition, swarm intelligent bionic optimization algorithm will gradually become a useful tool of scientific knowledge; therefore, ant colony optimization algorithm and its applications will eventually be a long-term research focus and a forefront topic and it will arouse more attention from the public and have a broader development prospect [3].

2 Basic ant colony algorithm

Traveling Salesman Problem (TSP) is a typical combinatorial optimization problem and it is also the most successful example of ACA; therefore, ACA is always described with TSP as the application background. TSP can be described as follows: a traveling salesman needs to go back to the departure city after going through n cities only once apart from the starting city and TSP is supposed to find the shortest path among all the cities.

The ants in the natural world can always find the optimal path from the food to the cave or bypassing the obstacles when searching for food or encountering obstacles. The main reason is that the ants will release pheromone on the way and the subsequent ants can choose the next path to go according to the pheromone. The more pheromone a path has, the more times the path

* ***Corresponding author*** e-mail: ssss@dddd.com

is selected, the more excellent performance the path has and the higher probability the subsequent ants choose this path is; therefore, it forms a positive feedback of learning information and gets gradually closer to the optimal solution [4, 5].

TSP can be interpreted by the directed diagram G=(V,E), where V=(1,2…n) is the collection of the nods, E={(ij)} is the collection of the sides and D=(d_{ij}) is the Euclidean distance between i and j. Before using ACA to solve TSPs, it needs to restrict every artificial ant to only choose one city once along a path. After all the ants have searched a complete and legal path, update the corresponding pheromone to every side according to the paths the ants have gone through. In the searching process, the ant computes the probability according to the pheromone and the heuristic information in every path, based on which, the probability that the artificial ant moves from i city to j city at t in the next city is:

$$\rho_{ij}^{k}(t)=\begin{cases}\dfrac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}}{\sum_{s\in D_k}[\tau_{is}(t)]^{\alpha}[\eta_{is}(t)]^{\beta}}, j\in D_k\\ 0, \text{Others}\end{cases}. \quad (1)$$

In this formula, $D_k=\{0,1\cdots n-1\}-tabu_k$ is the city collection to be chosen when the artificial ant arrives at kth city. With a memory function, the artificial ant records the cities it has passed and makes dynamic adjustment with the evolution via $tabu_k(k=1,2,\cdots m)$. As time passes, the information left before disappears gradually. $\eta_{ij}(t)$ is the visibility of side ij and $\eta_{ij}=\dfrac{1}{d_{ij}}$; $\tau_{ij}(t)$ is the pheromone of ij at the time of t; d is the relative importance degree of the pheromone and 1-p is the volatization of pheromone. After some time, the ant finishes one cycle and adjusts the pheromones in every path cording to Formula (2) and (3).

$$\tau_{ij}(t+1)=\rho\tau_{ij}(t)+\Delta\tau_{ij}(t), \quad (2)$$

$$\Delta\tau_{ij}(t)=\sum_{k=1}^{m}\Delta\tau_{ij}^{k}(t). \quad (3)$$

In the above formula, $\Delta\tau_{ij}^{k}(t)$ is the pheromone left on ij by k ants in this cycle and $\Delta\tau_{ij}$ is the pheromone increment left on ij in this cycle.

$$\Delta\tau_{ij}^{k}\begin{cases}\dfrac{Q}{L_k}, \text{If kth ant pass through the ij in this loop}\\ 0, \text{Else}\end{cases}. \quad (4)$$

If the kth ant passes ij in this cycle, otherwise.

In Formula (1)-(4), the pheromone intensity Q is a constant and L_k is the total length of the path the kth ant takes in this loop.

At the initial moment, $\tau_{ij}(0)=C$, $\Delta\tau_{ij}=0$ $(i,j=0,1\ldots,n-1)$ [6, 7].

3 The applications of ant colony optimization algorithm in artificial neural network

3.1 ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks (ANNs), also called Neural Networks (NNs), is a mathematical model by simulating Biological Neural Networks (BNNs) to conduct information processing. Based on the psychological research achievements of brain, it is aimed to simulate some mechanism of the brain and realize some specific functions. Currently, ANNs have been widely used in many fields.

The basic information processing unit of ANNs is artificial neuron, which is the design foundation of ANNs and its model is indicated in Fig.1.

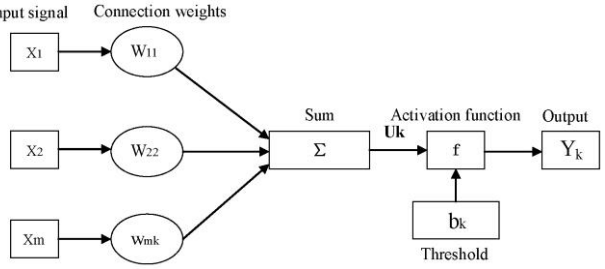


FIGURE 1 Artificial neuron model

An artificial neuron k can be indicated with the following formula.

$$u_k=\sum_{i=1}^{m}w_{ik}x_i \qquad y_k=f(u_k+b_k). \quad (5)$$

In this formula, m is the number of input signals; $x_i(i=1,\ldots,m)$ is the input signal; $w_{ij}(i=1,\ldots,m)$ is the synaptic weight of neuron k (the positive value is at the excited state and the negative value is at the depressed state); u_k is the output of linear combiner of the input signal; b_k is the bias (threshold) of the neuron unit; f is the activation function and y_k is the neuron output signal.

Organize numerous neurons with simple functions through certain topological structure and form the colony parallel processing computing structure. This is the artificial neural network.

According to different connection types, neural network can be divided into two types: layered and interconnected neural network. The layered neural network divides all the neurons into several layers, including input layer, hidden layer and output layer and every layer has some neurons. According to the feedback relationship between layers, the layered neural network can also be divided into the simple feed-forward network, the feedback feed-forward network and the inlayer interconnected feed-forward network. The topological

structure of the layered neural network is demonstrated in Fig.2.

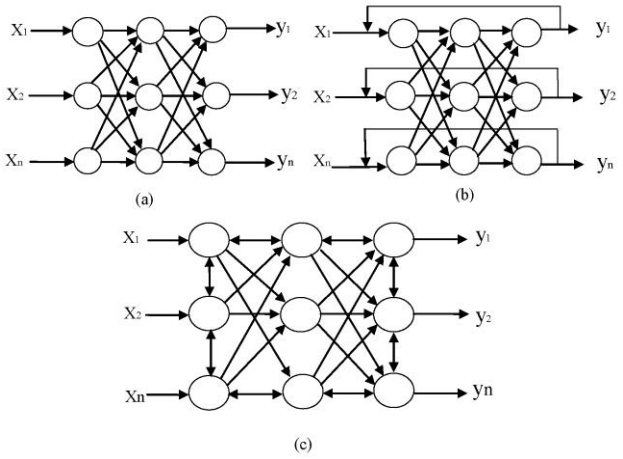


FIGURE 2 The topological structure of layered neural network

a) Simple Feed-Forward Network b)Feed-Forward Network with Feedback c) Inlayer Interconnected Feed-Forward Network

The difference between the interconnected neural network and the layered neural network is that any two units in the interconnected neural network is accessible. According to whether the output of every neuron is connected with other neurons, neural network can also be divided as: fully-connected network and locally-connected network. The topological structure of the interconnected neural network is shown in Fig.3.

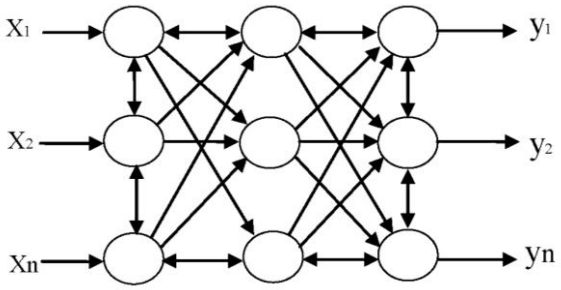


FIGURE 3 The topological structure of interconnected neural network

The basic attributes of neural network reflect the characteristics of neural network. The main characteristics of neural network include:

(1) It has strong robustness and fault tolerance because the information of neural network is distributed and saved in neurons with different network positions.

(2) It has strong information synthesization ability. Every neuron of neural network can process and save information; can handle quantitative and qualitative information and can better coordinate several input information relationships.

(3) It uses parallel processing method. Every neuron in the neural network is parallel in structure and similar processing can be conducted at the same time according to the information received to make faster computing rate.

(4) It has the advantages of self-study, self-organization and self-adaptation. The neurons of the neural network are connected in various ways and the connection strength changes by learning the training samples to make the neural network can handle uncertain systems.

(5) Neural network can fully get close to any complicated non-linear relationship.

3.2 THE NEURAL NETWORK LEARNING ALGORITHM BASED ON ANT COLONY OPTIMIZATION

The basic idea of ACA optimization neural network is to assume that there are N parameters in the neural network to be optimized, including all the weights and thresholds. Firstly, sort these parameters and form a collection I_{pi} of all the possible solutions for the parameter $p_i(1 \le i \le N)$. Then define that m ants go to find food from the ant nest. Every ant starts from the first collection and chooses one element from every collection I_{pi} according to the information status of every element and adjusts the corresponding pheromone to the selected element. When the ant has finished choosing the elements in all the collections, it has arrived at the food source and goes back to the nest in the previous path and adjusts the relevant pheromone. After continuous iteration of the ant, the optimal solution for the parameter can finally be found.

Because BP algorithm is easily trapped in local minimum, this paper has adopted the max-min ant system training neural network which can better solve problems and which has drawn most attention in the improved ant colony algorithm. The neural network training can be seen as an optimization problem, namely to find a group of optimal real number weight combination to minimize the error between the output result and the expected result in this weight. The main steps of the ant colony algorithm optimization neural network are classified as follows:

Step 1: Initialize the pheromone. Equally divide into spn the weight range $[W_{\min}, W_{\max}]$ and the point of every sub-region boundary is an alternative weight. In the initial moment, every point has the same pheromone τ_0. Set the pheromone volatilization coefficient as ρ; the pheromone increment intensity as Q; the maximum iterations N_{ACO} of ant colony optimization algorithm as and the learning rate of BP algorithm as η.

Set the maximum iterations of BP algorithm as N_{BP}; the training error exit criteria as E_0 and the retaining number of the optimal solutions as σ. The number of ants is m and all of them are in the ant nest.

Step 2: Any ant $k(k = 1, \cdots, m)$ starts from the first collection and moves from one point to another according to the probability computed by the following formula.

$$prob(\tau_i^k) = \tau(i) / \sum_{j=1}^{m} \tau(j) . \quad (6)$$

The ant records the marks in the passing points, namely to choose a numerical value for the weight and records them in $tabu_k$. When the ant has chosen values for all the weight parameters, it has completed one traversal and all the recorded values have become the parameters of this neural network. Input the training sample; get the corresponding output and compute the error E.

Step 3: Record the σ th group of weights with fewer errors and adjust the pheromone of every element according to the following formula. Assume that the above-mentioned ant has gone through n time units in its foraging.

$$\tau_j(t+n) = (1-\rho)\tau_j(t) + \Delta\tau_j(t), \tag{7}$$

$$\Delta\tau_j(t) = \sum_{k=1}^{m} \Delta\tau_j^k(t). \tag{8}$$

In the above formulas, $\Delta\tau_j^k(t)$ is the pheromone the k th ant has left on the j th element of this collection in this cycle and it can be computed according to the following formula.

$$\Delta\tau_j^k(t) = \begin{cases} \dfrac{Q}{e}, & \text{Select jth element of the set} \\ & \text{if the ant k in this loop} \\ 0, & \text{Else} \end{cases}. \tag{9}$$

In this formula, Q is a constant and is the pheromone increment intensity and e is the maximum output error when the element chosen by all the ants is seen as the weight of the neural network, which can be defined as $e = \max_{n=1}^{s} \left| O_n - O_p \right|$. In addition, s is the number of samples and O_n and O_q are the actual output and the expected output of the neural network respectively. Therefore, it can be seen that the fewer the error, the more pheromone increment [8].

Step 4: Repeat the above steps until all the ants have converged to a path, namely having found the optimal solution of the parameters and the cycle ends. The flow chart for the training neural network of ant colony algorithm is shown as Fig.4.

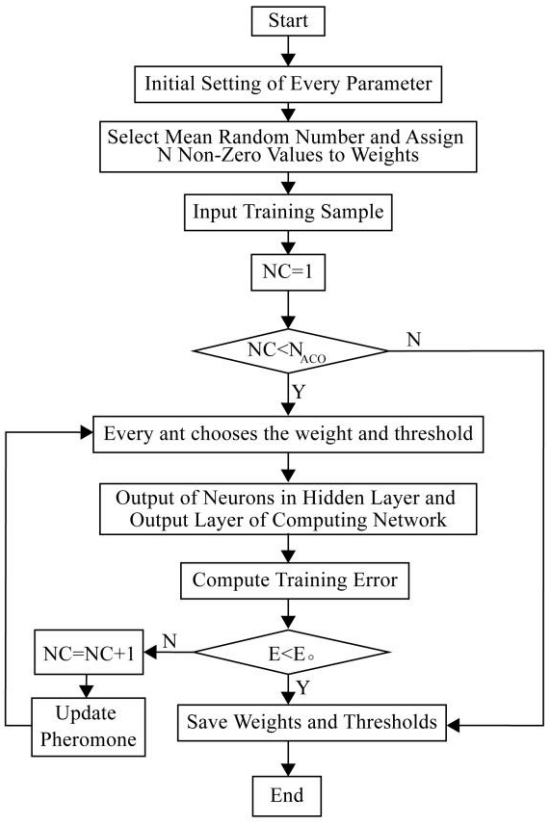


FIGURE 4 The flow chart for training neural network of ant colony algorithm

In this paper, we solve the classification problems with ant colony algorithm and the traditional BP algorithm separately and use the most commonly-used Iris data collection in the public database as the pattern classification data collection with UCI machine. Train the weights and thresholds of the neural networks with ant colony algorithm and the traditional BP algorithm and make comparisons of the results. The neural network in the simulation experiment is the three-layer neural network with a structure of 4-10-3. The algorithm parameters are set as follows: the number of ants S=50; the pheromone volatilization coefficient ρ=0.02 and the total information amount Q=30. The random value of the network weight parameter ranges from -5 to 5. Take the maximum iterations as the end condition of the algorithm. The test error, the classification accuracy and CPU running time in the training phase are displayed in the following table.

TABLE 1 Performance comparison of neural network training

	Test error	**Classification accuracy (%)**	**CPU running time (s)**
Ant colony optimization algorithm	1.353	96.25	20.28
Traditional BP algorithm	2.734	93.68	37.17

It can be seen from the table that ant colony optimization algorithm has fast convergence rate at the training phase and high error precision.

4 Conclusion

Although ant colony algorithm has made big breakthroughs in many application fields since its appearance, there are still some problems about ant colony algorithm to be settled in its practical applications. This paper explains the basic principles and typical characteristics of ant colony algorithm in detail; briefly introduces the model structure and learning algorithm of BP neural network; explores the establishment and optimization algorithm of the optimization neural networks of ant colony algorithm; finishes the network training and test data diagnosis with the extractive samples; makes comparative analysis with the training process and results of BP neural networks and proves the advantages of ant colony optimization neural network.

References

[1] Ghahreman Rezaei, Mohammad Hadi Afshar, etc. 2014 Layout optimization of looped networks by constrained ant colony optimisation algorithm Original Research Article *Advances in Engineering Software* **70**(4) 123-33
[2] Verdaguer M, Clara N, etc. 2013 Application of Ant-Colony-Optimization algorithm for improved management of first flush effects in urban wastewater systems *Science of The Total Environment* **485**(1) 143-52
[3] Zhaojun Zhang, Na Zhang, etc 2014 Multi-satellite control resource scheduling based on ant colony optimization *Expert Systems with Applications* **41**(6) 2816-23
[4] Peng-Yeng Yin, Ray-I Chang, etc. 2012 Niched ant colony optimization with colony guides for QoS multicast routing *Journal of Network and Computer Applications* **40**(4) 61-72
[5] Salabat Khan, Abdul Rauf Baig, etc. 2014 A novel ant colony optimization based single path hierarchical classification algorithm for predicting gene ontology *Applied Soft Computing* **16**(3) 34-49
[6] Zhang B, Qi H, Ren Y T, etc. 2013 Inverse transient radiation analysis in one-dimensional participating slab using improved Ant Colony Optimization algorithms *Journal of Quantitative Spectroscopy and Radiative Transfer* **133**(1) 351-63
[7] Michalis Mavrovouniotis, Shengxiang Yang 2014 Ant colony optimization with immigrants schemes for the dynamic travelling salesman problem with traffic factors *Applied Soft Computing* **13**(10) 4023-37
[8] Juan Rada-Vilela, Manuel Chica, etc. 2013 A comparative study of Multi-Objective Ant Colony Optimization algorithms for the Time and Space Assembly Line Balancing Problem *Applied Soft Computing* **13**(11) 4370-82

Authors

Kuangfeng Ning, born on November 16, 1980, China

Current position, grades: researcher at Hunan International Economics University, China.
University studies: B.S degree in computer science and technology from National University of Defense Technology, 2002, and received the M.S degree in computer application from Middle and Southern University, China in 2008.
Scientific interests: include Signal processing and multimedia technology.

Xiliang Zeng, born on February 7, 1979, China

Current position, grades: researcher at college of computer, Hunan University of International Economics, China.
University studies: B.S degree in computer science and technology from National University of Defense Technology, 2003, and received a M.S. degree in computer science and technology from Hunan University.
Scientific interests: image processing and information security.

An improved algorithm for multi-factor fuzzy correlation

Chang Liu, Zhenyu Na*, Xin Zhang

School of Information Science and Technology, Dalian Maritime University, Dalian 116026, Liaoning, China

Received 1 June 2014, www.cmnt.lv

Abstract

As a means for Vessel Traffic Service (VTS) to oversee the vessels, the traditional radar and the new navigation method of Automatic Identification System (AIS) are the two sources of getting the vessels' information. Tracks Fusion of the data received from these two sensors becomes the fundamental problem to be resolved in VTS. The tracks correlation is the premise and basis of the tracks fusion. This paper proposed an improved algorithm of multi-factor integrated fuzzy correlation based on the least square-time interpolation. We make generous correlation decision of distance and achieve the targets set in a fixed range, and then after time correction based on the least square-time interpolation we get the correlated tracks set and make fuzzy correlation used the membership function of normal distribution. The simulation experiment shows the proposed fuzzy correlation algorithm is more precise and the data are more close to the actual data of the vessel. The result of this effort can become an efficient method that impacts greatly on the vessel traffic management.

Keywords: Fuzzy Correlation, Tracks Fusion, Automatic Identification System (AIS)

1 Introduction

The VTS (Vessel Traffic Service) is a system of receiving and processing the traffic data, and providing vessels with service. Up to now, many VTS systems have been introduced in world-wide ports. In China, there have been 22 VTS systems covering most of the water area. VTS played an important role in traffic safety, efficiency and environmental protection [1-2]. In VTS, radar is the main tool to monitor and collect data [3]. With the establishment of the shore-based AIS (Automatic Identification System) and inter-fusion with VTS, VTS can get target data such as dynamic, static and navigation information [4-7]. For the VTS' of multiple radars, the cross section data can be collected by nearly radars. When the shore-based AIS is involved in the processing, the track data can be achieved by both radar and AIS [8]. For such a condition, we need a multiple sensors processing data fusion method (MSPDF) [9-12]. MSPDF can fuse the data from different sensors and establish the system tracking, which can increase the tracking performance of VTS.

Some research proposed the method of central clustering used in the tracks correlation. Comparing with the method of focal clustering, in this paper, we propose an improved algorithm multi-factor integrated fuzzy correlation based on the least square-time interpolation method. By experiment, we can conclude that the improved method is more precise and the data are more close to the real navigation data of the ship. Firstly, we get to make a correlation decision of distance from AIS and radar and achieve the correlated targets set [13-15]. And next we go for a multi-factor fuzzy correlation algorithm [16-18]. The fuzzy correlation algorithm design the membership functions of the distance and bearing parameters of the target ships from radar and AIS. Finally, by multi-factor integrated fuzzy correlation we can evaluate the correlation tracks and then make fusion of them. This approach is more effective and accurate especially in the busy water area with high density of ships. To do so, the remainder of this paper is outlined as follows. In the next section, we present the idea of generous correlation of distance and time correction. In Section 3, we have developed an approach to fuzzy correlation based on normal distribution membership function. In Section 4, the fusion method is offered. In section 5, we make the simulation and analyse the result. In Section 6, we close the paper and give some remarks.

2 Decision of distance correlation and time correction

First according to the data characteristic of radar and AIS we should make the distance correlation and time correction.

2.1 DISTANCE CORRELATION

Suppose the position of ship A detected by radar is ($X_{A(t)}$, $Y_{A(t)}$), and the speed is $v_{A(t)}$. Here we suppose ship A is of uniform linear motion. According to the speed at the time we can get the AIS report period T_a of ship A. If the tracks of AIS and radar are correlated at time *t*, the track of AIS should be within the circle of radius *R*, shown as in Figure 1.

* ***Corresponding author*** e-mail: nazhenyu@dlmu.edu.cn

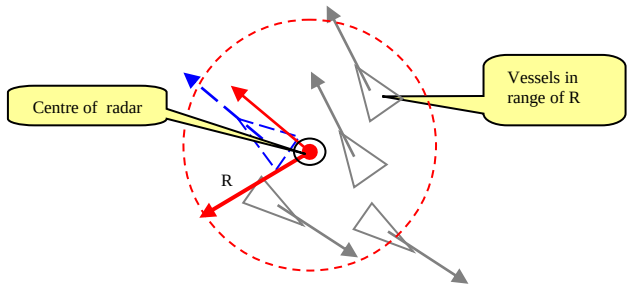


FIGURE 1 Distance correlation

We define *R* as in Equation (1).

$$R = v_{A(t)}T_a + \delta_r + \delta_a . \tag{1}$$

In Equation (1), δ_r and δ_a are the distance errors of radar and AIS. We correlate the tracks set *l* of radar with the tracks set l^* detected by AIS. So *S*(*l*) is defined as the correlation tracks set of the ship. See Equation (2).

$$S(l) = \{l^* \mid d\left(l(t), l^*(t)\right) \le R\}. \tag{2}$$

In Equation (2), $l(t)$ is the track-point of time *t*, beginning from 1. And $d(x, y)$ is the distance of *x* and *y*.

2.2 TIME CORRECTION BASED ON THE LEAST SQUARE- INTERPOLATION METHOD

Tracks of targets from radar and AIS may be not at the same sample time. So we define a sample time set, and the sample time of all the targets should be corrected by the method of interpolation according to the definition set in order to obtain the position set of them at the same time. The interpolation equation is illustrated in Equation (3).

$$\begin{cases} x_c = v_x \Delta T + x_0 \\ y_c = v_y \Delta T + y_0 \end{cases}. \tag{3}$$

In (3) x_c and y_c are the corrected coordinated value of distance (*x*, *y*); v_x and v_y are the speed component of (*x*, *y*); ΔT is the time difference of the original sample time; and x_o, y_o are the original distance value of direction (*x*, *y*). We select the cubic spline function to make the curve smooth. Suppose $(n+1)$ measurement in the time interval of $[a,b]$ for the targets from radar or AIS, and then divide the time interval of $[a,b]$ into $a \le x_0 \le x_1 \le x_2 \le x_3 \le \cdots \le x_n \le b$. The sample time is x_i $(i = 0,1,2,\cdots,n-1)$. The measurement value of $y_i = f(x_i)$ $(i = 0,1,2,\cdots,n-1)$ for each x_i can be the target speed or distance. We can construct the cubic spline interpolation function $S(x)$, satisfying the three conditions as follows.

(1) $S(x_i) = y_i, i = 0,1,2,\cdots,n-1$.

(2) in each $[x_i, x_{i+1}]$ $(i = 0,1,2,\cdots,n-1)$, $S(x)$ is a cubic polynomial.

(3) $S(x)$ has the continuous second derivative in $[a,b]$.

According to each measurement value $(x_i, y_i)(i = 0,1,\cdots,n-1)$ we make the $f(x_i) = y_i$ satisfying the Equation (4).

$$\sum_{i=0}^{n} (S^*(x_i) - y_i) = \min . \tag{4}$$

Comparing with the method of central clustering, the data of least square-interpolation method is more close to the actual data of the navigation ship.

3 Multi-factor fuzzy correlation algorithm

3.1 FUZZY FACTORS SET

We take the Euclidean distance of positions and bearings of vessels as the factors of a fuzzy set, as shown in Equation (5) and Equation (6).

$$U = \{u_1, u_2\}, \tag{5}$$

$$\begin{cases} u_1(t) = \left[\left(x_r(t) - x_a(t)\right)^2 + \left(y_r(l) - y_a(t)\right)^2\right]^{\frac{1}{2}} \\ u_2(t) = \left|\arctan\left[y_r(t)/x_r(t)\right] - \arctan\left[y_a(t)/x_a(t)\right]\right| \end{cases}. \tag{6}$$

$u_1(t)$ and $u_2(t)$ are respectively the differences of Euclidean distance and the bearing at the sample time of *t* from two targets. $x_r(t)$ and $y_r(t)$ are respectively the position components of (*x*, *y*) from radar at the time *t*. $x_a(t)$ and $y_a(t)$ are respectively the position components of (*x*, *y*) from AIS at the time *t*.

In order to simplify the calculation we set the evaluation set only two conditions as *V*.

$$V = \{\text{Correlation}, \text{Non-correlation}\}. \tag{7}$$

3.2 FUZZY CORRELATION OF SINGLE-FACTOR EVALUATION

We set the single parameter fuzzy array as in Equation (8).

$$R = (r_{kl})_{n\times m}(l = 1, 2). \tag{8}$$

In Equation (8) r_{kl} means the possibility of the correlation result *l* of the two tracks considering the parameter *k*. This is the single parameter correlation membership. r_{k1} is the correlation membership and r_{k2} is the non-correlation membership. We select Normal distribution membership in the algorithm as shown in Equation (9).

$$\begin{cases} r_{k1} = \exp\left\{-\tau_k \left(u_k^{\ 2} / \sigma_k^{\ 2}\right)\right\}, k = 1,2 \\ r_{k2} = 1 - \exp\left\{-\tau_k \left(u_k^{\ 2} / \sigma_k^{\ 2}\right)\right\}, k = 1,2 \end{cases}. \quad (9)$$

In equation (9) τ_k is correction modulus. That is, τ_1 and τ_2 are respectively the correction modulus of position and bearing. $\sigma_k^{\ 2}$ is the variance. $\sigma_1^{\ 2}$ and $\sigma_2^{\ 2}$ are respectively the precision variance of position and bearing.

The fuzzy evaluation array of single-factor is indicated by Equation (10).

$$R = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}. \quad (10)$$

3.3 FUZZY CORRELATION OF MULTI-FACTOR EVALUATION

The weighting fuzzy set is described as in (11).

$$A = (a_1, a_2). \quad (11)$$

Here considering the precision of position is higher than that of bearing we set a_1 as 0.6 and a_2 as 0.4 according to the experience by many experiments.

The fuzzy correlation evaluation of two tracks depends on the weighted fuzzy set A and the fuzzy evaluation array of single parameter R. So the intenerated fuzzy set is defined as in (12).

$$B = A \bullet R = (b_l) = (b_1, b_2). \quad (12)$$

In Equation (12) b_1 and b_2 are respectively the membership of correlation and non-correlation. We adopted the weighted average method to integrate the fuzzy sets, as showed in Equation (13).

$$\begin{cases} b_1 = a_1 r_{11} + a_2 r_{21} \\ b_2 = a_1 r_{12} + a_2 r_{22} \end{cases}. \quad (13)$$

We define a threshold Td_g to evaluate the membership b_1. When it satisfied with the Equation (14), we take the decision of the correlation.

$$b_1 > Td_g. \quad (14)$$

4 Tracks fusion

4.1 WEIGHTED MODULUS

Here we adopt the statistics weighted estimation algorithm. The data fusion is according to the precision of data from radar and AIS. The higher the precision is, the greater the weight is.

$$\begin{cases} W_{RL} = \dfrac{\sigma_{AL}^{\ 2}}{\sigma_{AL}^{\ 2} + \sigma_{RL}^{\ 2}} \\ W_{AL} = \dfrac{\sigma_{RL}^{\ 2}}{\sigma_{AL}^{\ 2} + \sigma_{RL}^{\ 2}} \end{cases}. \quad (15)$$

In Equation (15), $W_{AL} + W_{RL} = 1$. σ_{RL} and σ_{AL} are respectively the precision of radar and AIS. According to the experience we set σ_{RL} and σ_{AL} as 15 m and 8 m.

4.2 WEIGHTED FUSION

By equation (15) we can get the fusion track l in Cartesian coordinates. See Equation (16).

$$\begin{cases} L_x = W_{RL} L_{Rx} + W_{AL} L_{Ax} \\ L_y = W_{RL} L_{Ry} + W_{AL} L_{Ay} \end{cases}. \quad (16)$$

$L_{R(x,y)}$ and $L_{A\,(x,y)}$ are the position values of radar and AIS.

5 Experiment and analysis

5.1 PARAMETERS OF THE TRACKS FROM RADAR AND AIS

We suppose three tracks of targets A, B and C from radar, and one track from AIS.

(1) The speed of ship A in direction (x, y) are respectively $v_{xA} = 10$, v_{yA} = 0. The track of ship A from radar is shown in Table 1. x_{RA} *and* y_{RA} are respectively the position data of ship A from radar.

TABLE 1 Tracks of ship A from radar

Time series (s)	x_{RA} (m)	y_{RA} (m)
3	80	60
6	120	70
9	130	80
12	200	90
15	170	80
18	200	70
21	210	60
24	230	50
27	20	30
30	290	30
33	330	10
36	370	10
39	410	30
42	460	40
45	500	30
48	520	20
51	590	20
54	590	10
57	620	50
60	620	30

(2) The speed of ship B in direction (x, y) are respectively $v_{xB} = 10$, v_{yB} =10. The track of ship B from radar is shown in Table 2. x_{RB} *and* y_{RB} are respectively the position data of ship B from radar.

TABLE 2 Tracks of ship B from radar

Time series (s)	x_{RB} (m)	y_{RB} (m)
1	30	60
4	70	30
7	110	50
10	150	100
13	190	100
16	130	170
19	180	220
22	220	180
25	230	290
28	240	310
31	330	350
34	390	350
37	400	360
40	400	380
43	400	430
36	450	470
39	530	520
52	580	560
55	530	590
58	540	630

(3) The speed of ship C in direction (x, y) are respectively $v_{xC} = 10$, $v_{yC} = 20$. The track of ship C from radar is shown in Table 3. x_{RC} *and* y_{RC} are respectively the position data of ship C from radar.

TABLE 3 Tracks of ship C from radar

Time series (s)	x_{RC} (m)	y_{RC} (m)
2	35	90
5	65	160
8	90	230
11	100	300
14	130	370
17	120	430
20	130	480
23	135	530
26	160	580
29	155	630
32	150	670
35	155	720
38	180	770
41	185	840
44	190	910
47	195	970
50	200	1020
53	215	1070
56	240	1120
59	260	1170

(4) Suppose the sample period of data from AIS is 6 second and the track data is shown in Table 4. A_x *and* A_y are respectively the position data of the ship from AIS.

TABLE 4 Tracks of ship from AIS

Time series (s)	A_x (m)	A_y (m)
2	35	90
5	----	----
8	65	210
11	----	----
14	85	330
17	----	----
20	110	445
23	----	----
26	140	565
29	----	----
32	155	670
35	----	----
38	185	790
41	----	----
44	215	910
47	----	----
50	240	1035
53	----	----
56	270	1155
59	----	----

5.2 FUZZY CORRELATION SIMULATION

Firstly, by the distance correlation we can see the tracks of A and B are out of the threshold, and the track of C is within the threshold, so we continue the fuzzy correlation decision of C with AIS track.

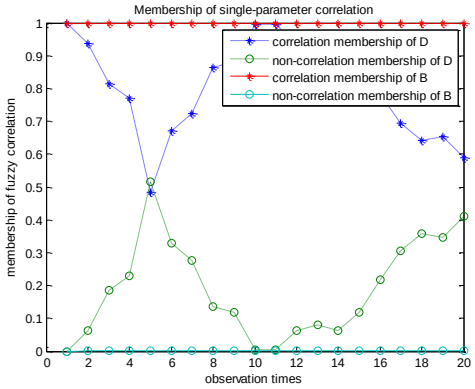


FIGURE 2 Single-factor correlation results

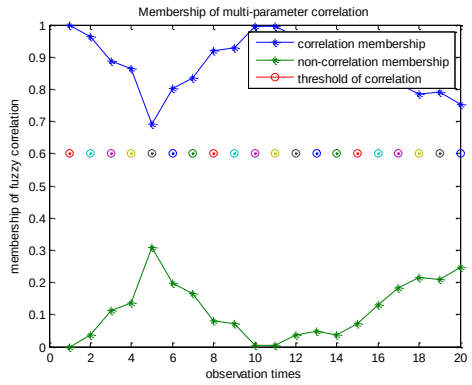


FIGURE 3 Multi-factor correlation results

After the time correction of track C and AIS, we got the correlation membership of a single parameter as showed in Figure 2 and the multi-factor correlation in Figure 3. From Figure 2 the correlation membership of D (distance) is greater than 50% and the correlation membership of B (bearing) is 100%. From Figure 3 fuzzy correlation membership of multi-factor exceeds the threshold of 0.6 as we set. It can be seen obviously that track C is related to the track of AIS.

The fusion track of ship C and the original tracks of A, B is shown in Figure 4. We can see the fusion track is the same as the track of AIS, because the precision of AIS is much higher than that of radar.

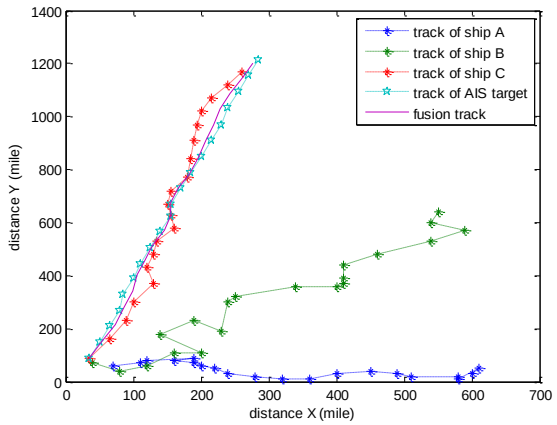


FIGURE 4 Fusion track of C and AIS

We make the experiments for the comparison of the two correlation methods, one is based on the least square-interpolation and the other is the central clustering, the Standard deviation is shown in table 5. From table 5 we can conclude that the correlation algorithm based on the least square- interpolation time corrected is more precise.

TABLE 5 Comparison of standard deviation of the two methods for the time corrected

Method	AIS	radar	fusion
central clustering	1.4308	7.8015	1.5908
least square- interpolation	1.4025	6.9446	1.3091

6 Conclusions

This paper described a fuzzy correlation algorithm to fuse redundant observations to multi-sensor coverage in order to provide an accuracy track of AIS and radar in the Vessel Traffic Service (VTS) and make the navigation safety. The experiment showed that the improved algorithm correctly fuses the redundant sensor observation on the same target and the result is a computationally efficiency and cost effective software solution to a deficiency that impacts greatly on overall waterway safety and protect the ocean environment. The correlation algorithm can be enhanced by adding other attributes such as speed, course, size and track quality to improve the accuracy and reliability of the method.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (3132013054).

References

[1] Sawano N 2008 Current Situation of Digitalized Ship Navigation System for Safety *22nd Intl. Conf. on Advanced Information Networking and Applications-Workshops* 1134-7
[2] Danu D, Sinha A, Kirubarajan T 2007 Fusion of Over-the-Horizon Radar and Automatic Identification Systems or Overall Maritime Picture *10th Intern. Conf. on Information Fusion*, Quebec, Canada, 1-8
[3] Barton D K 2004 *Radar System Analysis and Modeling* Artech House: London
[4] Lee A, Zetterberg S 2010 Establishing an IALA AIS Binary Message Register: Recommended Process *IALA Conference 2010*, Cairo, Egypt, 108–15
[5] Eriksen T, Høye G, Narheim B, Meland B J 2006 Maritime traffic monitoring using a space-based AIS receiver *Acta Astronautica* **78**(1) 537–49
[6] Norris A 2007 AIS implementation – Success or failure? *Journal of Navigation* **60**(1) 1−10
[7] Harre I 2000 AIS adding new quality to VTS systems *Journal of Navigation* **53**(1) 527−39
[8] Guerriero M, Willett P, Coraluppi S, Carthel C 2008 Radar/AIS data fusion and SAR tasking for Maritime Surveillance *Proc. of 11th Intl. Conf. on Information Fusion*, Cologne, Germany, 3−30
[9] Kaplan L M, Blair W D 2004 Assignment costs for multiple sensor track-to-track association *Proc. of the 7th Intl. Conf. on Information Fusion*, Stockholm, Sweden 1231−7
[10] Chong C Y, Mori S Barker W H, Chang K C 2000 Architectures and algrithms for track association and fusion *IEEE Aerospace Electronic System* **15**(1) 5−13
[11] Deb S, Yeddanapudi M, Pattipati K, Bar-Shalom Y 1997 A generalized S-D assignment algorithm for multisensor-multitarget state estimation *IEEE Trans Aerospace Electronic System* **33**(1) 523−38
[12] Chen H, Kirubarajan T, Bar-Shalomm Y 2003 Performance limits of track-to-track fusion versus centralized estimation: Theory and application *IEEE Trans Aerospace Electronic System* **39**(1) 386-400
[13] Waltz E, Llinas J 1990 *Multi Sensor Data Fusion* Artech House Radar Library
[14] Bar-Shalom Y, Blair W D 2000 *Multitarget/Multisensor Tracking: Applications and advances-Volume III* Artech House Radar Library
[15] Chaudhuri B B, Bhattacharya A 2001 On correlation between two fuzzy sets *Fuzzy Sets System* **118**(1) 447−56
[16] Lin N P, Chueh H E 2007 Fuzzy Correlation Rules Mining *Proc. of the 6th WSEAS Intl. Conf. on Applied Computer Science, Hangzhou, China*, 13–8
[17] De Cock M, Cornelis C, Kerre E E 2005 Elicitation of fuzzy association rules from positive and negative examples *Fuzzy Sets System* **149**(1) 73–85
[18] Delgado M, Marín N, Sánchez D, Vila M 2003 Fuzzy Association Rules: General Model and Applications *IEEE Trans Fuzzy System* **11**(1) 214−25

Authors

Chang Liu, born on March 14, 1976, China

Current position, grades: lecturer at School of Information Science and Technology of Dalian Maritime University (DMU).
University studies: B.S. degree in communication engineering, M.S. degree in communication and information system from Dalian Maritime University in 1999 and 2002, respectively. In 2013, she received the Ph.D. degree in information and communication engineering at Dalian Maritime University.
Scientific interests: include communication and information system, Automatic Identification System and vessel traffic system.
Publications: more than 10 papers in international journals and conferences.

Zhenyu Na, born in January, 1981, China

Current position, grades: lecturer at School of Information Science and Technology of Dalian Maritime University (DMU) with research interests including satellite communications and networks, wireless networks, MIMO-OFDM communications.
University studies: B.S. degree in communication engineering, M.S. degree in information and communication engineering from Harbin Institute of Technology (HIT) in 2004 and 2007, respectively. In 2010, he received doctoral degree in information and communication engineering at Communication Research Center of HIT.
Publications: more than 20 papers in international journals and conferences.

Xin Zhang, born on January 27, 1988, China

Current position, grades: pursuing the Ph.D. degree in information and communication engineering at Dalian Maritime University
University studies: B.S. degree in communication engineering, M.S. degree in information and communication engineering from Harbin Institute of Liaoning Technical University in 2010 and 2013, respectively.

Kinematics analysis and simulation of a novel metamorphic parallel cutting mechanism

Shenghai Hu[1], Xiulian Liu[1, 2*], Manhui Zhang[1]

[1]*School of Mechanics Engineering, Harbin Engineering University, Harbin 150001, China*

[2]*School of Mechanical Engineering, Heilongjiang University of Science and Technology, Harbin 150027, China*

Received 1 June 2014, www.cmnt.lv

Abstract

Based on the metamorphic mechanisms, we proposed a new type of metamorphic kinematic pair, with which a novel metamorphic parallel cutting mechanism was designed to achieve changing pose and trajectory movement. Using the screw theory and the product of exponentials formula, we constructed a nonlinear system of equations for each configuration framework of the metamorphic parallel cutting mechanism, and the Sylvester resultant elimination method was employed to simplify equations of motion for the mechanism, thereby completing the research about the forward and inverse kinematics solution of the original and sub-configuration framework and verifying the feasibility of this novel mechanism via MATLAB and ADAMS simulation.

Keywords: Metamorphic Mechanism; Screw Theory; Product of Exponentials Formula; Forward and Inverse Kinematics Solution

1 Introduction

The manufacturing of ship products is often accompanied by large-diameter hole cutting and processing on the hull surfaces, which are pieced together with complex surface patches in most cases. And then square groove or bevel preparation of these large size holes are often needed [1-2]. As a critical technical problem in shipbuilding industry, large-diameter-hole and changing-angle bevel structure processing is characterized by high difficulty, low efficiency and low accuracy, etc., and the processing quality of which will directly influence the quality and efficiency of the subsequent industrial processes.

Seamus Gordon, et al., from University of Limerick, designed a high-speed CNC cutting machine realizing spatial movement of the cutting tool along three axial directions. Controlled by computer numerical control programming, the high speed of this cutting device reaches 40m/min [3]. The Omnimat cutting machine developed and produced by Messer, is a high-performance, multi-functional, large CNC cutting machine applicable for high speed and precision cutting of large size flat plate [4]. The model RV-016 cutting robot, developed and produced by Panasonic Corporation, adopts a serial mechanism, and its end infinite rotating cutting gun can perform high speed and precision cutting task on complex surfaces.

In China, many experts and engineering technicians in machinery manufacturing field have made a series of researches on numerical control cutting technology. The multi-functional NC-fair incision machine with a gantry-bridge structure, designed by Ming, et al., can cut large 5000mm × 3000mm work-piece and perform arbitrary plane curve cutting tasks [5]. The cantilever structured CNC large intersecting circle flame cutting machine with five-bar cooperating, designed by Sun, et al., can process a work-piece with a diameter of 4000-8000mm (pipe); and the diameter of the cutting hole lies in a range of 50-550mm, and the bevel angle is between 30° and 40° [6]. Zhang, from Harbin Engineering University, designed a large complex surface cutting mechanism with the end cutting torch being a five-bar metamorphic mechanism to achieve square groove and bevel groove of large diameter hole cutting on complex surfaces; controlled by electric motor, it can perform spatial movements with changing pose and trajectory [7].

Based on the analysis of literatures at home and abroad, we know there are two types of cutting techniques and equipment: one is round hole cutting on a plate with its vertical displacement almost unchanged in the cutting process. And the bevel angle remain unchanged in bevel cutting; the other is CNC pipe-cutting machine, which is applicable for round pipe (whose diameter is less than 1.4m) intersection hole cutting.

The parallel mechanism has the merits of simple high-rigidity structure, high processing speed and accuracy, and is less affected by the inertia, etc. For these advantages over the serial mechanism, it has been widely employed in various processing and manufacturing industries. Furthermore, its end moving platform can many complex spatial motions via multi-axis coordination [8-13]. Studies about metamorphic mechanism starts late, but this mechanism has stimulated the research interests of many scholars for its flexible topology, based on which many tasks can be completed [14-21].

* Corresponding author e-mail: lxl--2002@163.com

According to the principles of metamorphic mechanism, a parallel cutting mechanism was designed to solve the problem of changing pose and trajectory in square groove and bevel groove cutting. Since there are different requirements for the cutting torch motion in the process of cutting different shapes of holes on complex surfaces, we reasonably designed different conFigureuration frameworks of the metamorphic mechanism based on their metamorphic characteristics, which can meet the technology requirements of changing spatial pose and trajectory cutting on large complex surface.

2 Metamorphic Parallel Cutting Mechanism Design

2.1 A NOVEL DESIGN OF SPATIAL METAMORPHIC KINEMATIC PAIR

A conventional spherical pair has three spatial degrees of freedom (DOF). And there are 3 degrees of freedom of rotating between the two components linked by the spherical pair. In actual analysis, the motions of the two components can be simplified as the rotation around the three spatial axes that are perpendicular to each other and also meet the right-hand rule. Therefore, a spatial spherical pair is equivalent to a compound kinematic pair consisting of a Hooke's joint and a revolute pair whose axis is perpendicular to the Hooke's joint axial plane. For a Hooke's joint, the shaft connected with the base is the main shaft, and that connected with the linkage is a counter shaft.

According to the principles of metamorphic mechanism, we designed a slot where the two ends of the counter shaft can slide on the main shaft of Hooke's joint. The counter shaft can be fixed in a certain position as required with a pin; therefore, we can reduce the degree of freedom of the kinematic pair by putting pin into the pinhole located on the revolute pair. Based on the framework before and after metamorphic operation, the novel metamorphic kinematic pair can be classified into four configurations:

(1) The original configuration S_{m3}

Fix the counter shaft in a position perpendicular to the main shaft, and remove the pin on the revolute pair. The kinematic pair has three degrees of freedom and is equivalent to a spatial spherical pair, as shown in Figure. 5(a);

(2) The a-type 2-DOF sub-configuration S_{m2}^{a}

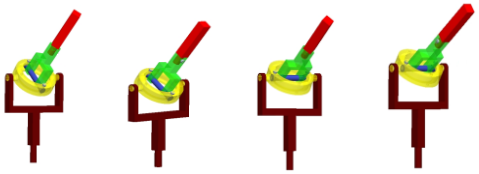
Fix the counter shaft in a position perpendicular to the main shaft, and put the pin into the pin hole on the revolute pair. Now the kinematic pair has two degrees of freedom and is equivalent to a traditional Hooke's joint, as shown in Figure. 5(b);

(3) The b-type 2-DOF sub-configuration S_{m2}^{b}

Fix the counter shaft in a co-axial position with the main shaft, and remove the pin on the revolute pair. Now the kinematic pair has two degrees of freedom, as shown in Figure. 5(c);

(4) Sub-configuration S_{ml} with single degree of freedom

Fix the counter shaft in a co-axial position with the main shaft, and remove the pin on the revolute pair. Now the kinematic pair has only one degrees of freedom and is equivalent to a traditional revolute pair, as shown in Figure 1 (d);



(a) S_{m3} (b) S_{m2}^{a} (c) S_{m2}^{b} (d) S_{m1}

FIGURE 1 The Structure of a Novel Metamorphic Kinematic

2.2 THE METAMORPHIC PARALLEL CUTTING MECHANISM'S ORIGINAL CONFIGURATION FRAMEWORK

Based on the above designed metamorphic kinematic pair unit, a limb kinematic chain where we can switch the degree of freedom is constructed. Assign the position of each kinematic pair, and form the metamorphic parallel mechanism using three same limb chains (see Figure 2). The end effectors $M_1M_2M_3$ are connected with the fixed base $B_1B_2B_3$ via the three same kinematic chains $S_{m2}^{a}S_{m2}^{b}U$. Under this framework, it is 6-DOF mechanism that can achieve upper and lower bevel cutting. Therefore, it is regarded as the original configuration of the cutting machine. When the mechanical properties of the mechanism are considered, the driving input should be as near as possible to the fixed platform to reduce the influence of inertia and to improve load performance. So we choose three Hooke's joints on the fixed platform as driving input.

2.3 THE NOVEL METAMORPHIC PARALLEL MECHANISM'S SUB-CONFIGURATION FRAMEWORK

Based on the original configuration, the 3-DOF sub-configuration framework of metamorphic parallel mechanism is obtained by transforming the metamorphic kinematic pair S_{m2}^{a} into S_{m1} framework, and the metamorphic kinematic pair S_{m2}^{b} into S_{m2}^{a}. Then the limb chain structure changes after these metamorphic operations. The kinematic pair distribution in the limb chain turns into $S_{m1}S_{m2}^{a}U$, which means the sub-configuration framework of the metamorphic parallel cutting mechanism is $3S_{m1}S_{m2}^{a}U$, as shown in Figure 3. With three revolute pairs on the fixed platform chosen

as driving inputs, the square groove cutting on the surface can be achieved.

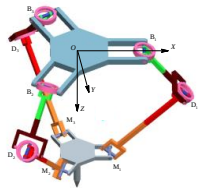


FIGURE 2 The Metamorphic Parallel Mechanism's Original Configuration Framework $3S^a_{m2}S^b_{m2}U$

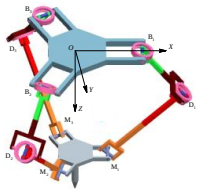


FIGURE 3 3-DOF Sub-configuration Framework $3S_{m1}S^a_{m2}U$

3 The kinematics position analysis of the metamorphic parallel cutting mechanism

3.1 THE ANALYSIS OF THE ORIGINAL CONFIGURATION FORWARD AND INVERSE KINEMATICS SOLUTION

3.1.1 The analysis of original configuration forward kinematics solution

Set the absolute coordinate system $OXYZ$, as shown in Figure 4; the XOY plane is parallel to the plane $B_1B_2B_3$ of the fixed base; axis Z is downward perpendicular to $B_1B_2B_3$; and suppose the origin of coordinates and the geometric centre of $B_1B_2B_3$ overlap. With the structures and parameters of the three limb chains being identical, we assume that: the radius of the circumcircle of $B_1B_2B_3$ is R; the length of $B_iD_i(i=1,2,3)$ is l, and the length of $D_iM_i(i=1,2,3)$ is m. Then suppose the angles of rotation about e_{i1}, e_{i2}, e_{i3}, e_{i4}, e_{i5}, e_{i6} ($i=1\sim3$) in each limb chain are θ_{i1}, θ_{i2}, θ_{i3}, θ_{i4}, θ_{i5} $(i=1\sim3)$ respectively. Let the Hooke's joint of the metamorphic kinematic pair B_i be the driving joint and the rest be driven joint, and $\eta_i(i=1\sim3)$ be the azimuth at point B_i, as shown in Figure 4, where $\eta_1=0^\circ$, $\eta_2=120^\circ$, $\eta_3=$-120°.

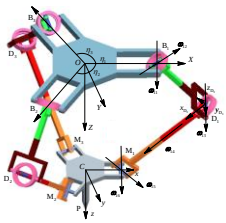


FIGURE 4 The Moving Coordinate System and the Kinematic Screw of the Metamorphic Parallel Cutting Mechanism's Original Configuration Framework

According to rigidity constraint,

$$\begin{aligned}&(M_2-M_1)^T(M_2-M_1)=(M_3-M_1)^T(M_3-M_1)\\&=(M_2-M_3)^T(M_2-M_3)=3r^2\end{aligned}. \quad (1)$$

The coordinates of point M_i in the equation can be obtained through the following methods

$$g_{D_i}(\theta)=e^{\theta_{i1}\hat{\xi}_{i1}^{\omega}}e^{\theta_{i2}\hat{\xi}_{i2}^{\omega}}e^{\theta_{i3}\hat{\xi}_{i3}^{\omega}}g_{D_i}(0), \quad (2)$$

where $e^{\theta\hat{\xi}}=\begin{bmatrix}e^{\theta\hat{e}} & (I-e^{\theta\hat{e}})(e\times v)+\theta ee^Tv\\ 0 & 1\end{bmatrix}$ $(e\neq0)$,

$$\xi=\begin{bmatrix}\omega\\ r\times\omega\end{bmatrix},\quad \hat{\xi}=\begin{bmatrix}\hat{\omega} & r\times\omega\\ 0 & 0\end{bmatrix},\quad \omega=\begin{bmatrix}\omega_1\\ \omega_2\\ \omega_3\end{bmatrix},$$

$$\hat{\omega}=\begin{bmatrix}0 & -\omega_3 & \omega_2\\ \omega_3 & 0 & -\omega_1\\ -\omega_2 & \omega_1 & 0\end{bmatrix}.$$

$$M_i=R_{D_i}(\theta)M_i^{D_i}+t_{D_i}. \quad (3)$$

Substituting (2) and (3) into (1), we have three equations of $\theta_{i3}(i=1\sim3)$:

$$\begin{cases}L_{11}c\theta_{13}c\theta_{23}+L_{12}s\theta_{13}c\theta_{23}+L_{13}c\theta_{13}s\theta_{23}+L_{14}s\theta_{13}s\theta_{23}\\+L_{15}c\theta_{13}+L_{16}c\theta_{23}+L_{17}s\theta_{13}+L_{18}s\theta_{23}+L_{19}=0\\L_{21}c\theta_{13}c\theta_{33}+L_{22}s\theta_{13}c\theta_{23}+L_{23}c\theta_{13}s\theta_{33}+L_{24}s\theta_{13}s\theta_{33}\\+L_{25}c\theta_{13}+L_{26}c\theta_{33}+L_{27}s\theta_{13}+L_{28}s\theta_{33}+L_{29}=0\\L_{31}c\theta_{23}c\theta_{33}+L_{32}s\theta_{23}c\theta_{23}+L_{33}c\theta_{23}s\theta_{33}+L_{34}s\theta_{23}s\theta_{33}\\+L_{35}c\theta_{23}+L_{36}c\theta_{33}+L_{37}s\theta_{23}+L_{38}s\theta_{33}+L_{39}=0\end{cases}, \quad (4)$$

where $L_{mn}(m=1\sim3,n=1\sim9)$ is a polynomial containing $\theta_{i1},\theta_{i2}(i=1\sim3)$.

Based on half-angle formula: $\sin\theta=2\tan(\theta/2)/[1+\tan^2(\theta/2)]$, $\cos\theta=[1-\tan^2(\theta/2))]/[1+\tan^2(\theta/2)]$, supposing $x=\tan(\theta_{13}/2)$, $y=\tan(\theta_{23}/2)$, $z=\tan(\theta_{33}/2)$, and manipulating the above three equations, we have:

$$P_{1y}x^2+P_{2y}x+P_{3y}=0, \quad (5)$$

$$P_{4x}z^2+P_{5x}z+P_{6x}=0, \quad (6)$$

$$P_{4y}z^2+P_{5y}z+P_{6y}=0, \quad (7)$$

where, $P_{1y}=Q_{11}y^2+Q_{13}y+Q_{15}$,

$P_{2y}=Q_{12}y^2+Q_{16}y+Q_{18}$, $P_{3y}=Q_{14}y^2+Q_{17}y+Q_{19}$,

$P_{4x}=Q_{21}x^2+Q_{23}x+Q_{25}$, $P_{5x}=Q_{22}x^2+Q_{26}x+Q_{28}$,

$P_{6x}=Q_{24}x^2+Q_{27}x+Q_{29}$, $P_{4y}=Q_{31}y^2+Q_{33}y+Q_{35}$,

$P_{5y}=Q_{32}y^2+Q_{36}y+Q_{38}$, $P_{6y}=Q_{34}y^2+Q_{37}y+Q_{39}$.

Eliminate the variable z using Sylvester resultant; then regard z and the constant terms as new variables; multiply z with equation (6) and (7), combed with which a matrix is constructed. Since only non-zero solutions are considered, the resultant is zero:

$$\begin{vmatrix} P_{4x} & P_{5x} & P_{6x} & 0 \\ 0 & P_{4x} & P_{5x} & P_{6x} \\ P_{4x} & P_{5x} & P_{6x} & 0 \\ 0 & P_{4x} & P_{5x} & P_{6x} \end{vmatrix} = 0 . \tag{8}$$

Then we have the equation, in which the highest exponent degree of x is not more than 2:

$$F_{1y}x^2 + F_{2y}x + F_{3y} = 0 , \tag{9}$$

where $F_{iy}(i = 1 \sim 3)$ are expressions in which the highest exponent degree of y is 2.

Eliminate the variable x using Sylvester resultant; regard x and the constant terms as new variables; multiply x with equation (5) and (10), combed with which a matrix is constructed. Since only non-zero solutions are considered, the resultant is zero:

$$\begin{vmatrix} P_{1y} & P_{2y} & P_{3y} & 0 \\ 0 & P_{1y} & P_{2y} & P_{3y} \\ F_{1y} & F_{2y} & F_{3y} & 0 \\ 0 & F_{1y} & F_{2y} & F_{3y} \end{vmatrix} = 0 . \tag{10}$$

Then we have the equation in which the highest exponent degree of y is not more than 4:

$$K_4 y^4 + K_3 y^3 + K_2 y^2 + K_1 y + K_0 = 0 , \tag{11}$$

where $K_l(l = 0 \sim 4)$ is a polynomial containing $\theta_{i1}, \theta_{i2}(i = 1 \sim 3)$. θ_{i1}

Because the driving kinematic pair rotation angles θ_{i1}, θ_{i2} are known quantities, the value of y can be calculated out from equation (11). Substituting the y value into equation (5) and (6), we yield twos sets of values of x and z, and then substituting, which into equation (7) to determine the end values of x and z.

After that, the coordinates of M_1, M_2 and M_3 on the end moving platform can be determined, and then we can further determine the pose of the moving platform in the absolute coordinate system.

3.1.2 The analysis of the original configuration inverse kinematics solution

Supposing that the geometric center of the moving platform is *C*, the cutting gun endpoint is *P*, and the length of the cutting gun is n, we set up the moving coordinate system *Cxyz* of the moving platform $M_1M_2M_3$. The x-axis passes through point M_1; z-axis is perpendicular to the moving platform in an outward direction; and y-axis is determined according to right-hand rule. The analysis of original configuration inverse kinematics solution is to determine the driving kinematic pair rotation angle θ_{i1} and $\theta_{i2}(i = 1 \sim 3)$, based on the known pose of the end moving platform $M_1M_2M_3$. Supposing that the moving coordinate system *Cxyz* is formed by rotating the absolute coordinate system along the X-axis, Y-axis and Z-axis by the angles of γ, β, α respectively, the coordinates of C are (x_C, y_C, z_C), and the coordinates of P is $(x_P, y_{P,} z_P)$, the pose of centre of mass (PCM) of the end moving platform is represented as:

$$g_C = \begin{bmatrix} R & t_c \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} e^{Z\alpha} e^{Y\beta} e^{X\gamma} & t_c \\ 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} c\alpha c\beta & c\alpha s\beta s\gamma - s\alpha c\gamma & c\alpha c\beta c\gamma + s\alpha s\gamma & x_C \\ s\alpha c\beta & s\alpha s\beta s\gamma + c\alpha c\gamma & s\alpha s\beta c\gamma - c\alpha s\gamma & y_C \\ -s\beta & c\beta s\gamma & c\beta c\gamma & z_C \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{12}$$

We calculate out the coordinates of M_1, M_2, M_3, P in the moving coordinate system *Cxyz*: $M_1^C(r,0,0)$, $M_2^C(-\frac{1}{2}r, \frac{\sqrt{3}}{2}r, 0)$, $M_3^C(-\frac{1}{2}r, -\frac{\sqrt{3}}{2}r, 0)$, $P^C(0,0,n)$. Therefore, their coordinates in the absolute coordinate system are:

$$M_i = RM_i^C + t_c , \tag{13}$$

$$P = RP^C + t_c . \tag{14}$$

In solving forward kinematics in the above section, we know point M_1, M_2 and M_3 can be represented by θ_{i1}, θ_{i2}, θ_{i3} $(i = 1 \sim 3)$. Based on the corresponding relationship, we have three square matrix equations:

$$\begin{bmatrix} (c\theta_{11}c\theta_{12}\theta_{13} - s\theta_{11}s\theta_{13})m + lc\theta_{11}c\theta_{12} + R \\ -msc\theta_{12}c\theta_{13} - lsc\theta_{12} \\ -(s\theta_{12}c\theta_{13}c\theta_{13} + c\theta_{11}s\theta_{13})m - ls\theta_{11}c\theta_{12} \end{bmatrix} = \begin{bmatrix} rc\alpha c\beta + x_C \\ rs\alpha c\beta + y_C \\ -rs\beta + z_C \end{bmatrix} . \tag{15}$$

$$\begin{bmatrix} p_1 & q_1 & k_1 \end{bmatrix}^T = \begin{bmatrix} d_1 & e_1 & f_1 \end{bmatrix}^T , \tag{16}$$

$$\begin{bmatrix} p_2 & q_2 & k_2 \end{bmatrix}^T = \begin{bmatrix} d_2 & e_2 & f_2 \end{bmatrix}^T , \tag{17}$$

where $p_1 = -\left[(c\theta_{21}c\theta_{22}c\theta_{23} - s\theta_{21}s\theta_{23})m + lc\theta_{21}c\theta_{22} + R\right]/2 + \sqrt{3}(ms\theta_{22}c\theta_{23} + ls\theta_{22})/2$,

$q_1 = \sqrt{3}\left[(c\theta_{21}c\theta_{22}c\theta_{23} - s\theta_{21}s\theta_{23})m + lc\theta_{21}c\theta_{22} + R\right]/2 + (ms\theta_{22}c\theta_{23} + ls\theta_{22})/2$

$k_1 = -(s\theta_{21}c\theta_{22}c\theta_{23} + c\theta_{21}s\theta_{23})m - ls\theta_{21}c\theta_{22}$,

$$d_1 = -\gamma c\alpha c\beta/2 + \sqrt{3}\gamma(c\alpha s\beta s\gamma - s\alpha c\gamma)/2 + x_C$$
$$e_1 = -\gamma s\alpha c\beta/2 + \sqrt{3}\gamma(s\alpha s\beta s\gamma + c\alpha c\gamma)/2 + y_C \quad ,$$
$$f_1 = rs\beta/2 + \sqrt{3}rc\beta s\gamma/2 + z_C \quad ,$$
$$f_2 = rs\beta/2 - \sqrt{3}rc\beta s\gamma/2 + z_C$$
$$k_2 = -(s\theta_{31}c\theta_{32}c\theta_{33} + c\theta_{31}s\theta_{33})m - ls\theta_{31}c\theta_{32}$$
$$d_2 = -rc\alpha c\beta/2 - \sqrt{3}r(c\alpha s\beta s\gamma - s\alpha c\gamma)/2 + x_C$$
$$p_2 = -[(c\theta_{31}c\theta_{32}c\theta_{33} - s\theta_{31}s\theta_{33})m + lc\theta_{31}c\theta_{32} + R]/2$$
$$-\sqrt{3}(ms\theta_{32}c\theta_{33} + ls\theta_{32})/2$$
$$q_2 = -\sqrt{3}[(c\theta_{31}c\theta_{32}c\theta_{33} - s\theta_{31}s\theta_{33})m + lc\theta_{31}c\theta_{32} + R]/2$$
$$+(ms\theta_{32}c\theta_{33} + ls\theta_{32})/2$$
$$e_2 = -rs\alpha c\beta/2 - \sqrt{3}r(s\alpha s\beta s\gamma + c\alpha c\gamma)/2 + y_C$$

Based on the coordinates of P in the coordinate system, the coordinates of the centre of mass C of the moving platform can be calculated out. And based on equation (15), (16) and (17), values of θ_{i1}, θ_{i2} and $\theta_{i3}\,(i=1\sim3)$ can be calculated out, thereby deriving the expressions of θ_{i1} and θ_{i2} in terms of α, β, γ, x_P, y_P, z_P. So the inverse kinematics solution is completed.

3.1.3 A numerical example of forward and inverse kinematics solution

Supposing that the geometric parameters of the original configuration framework of the metamorphic parallel cutting mechanism are: $R=1000$mm, $r=800$mm, $l=1000$mm, $m=1400$mm, $n=200$mm, the coordinates of the cutting gun endpoint P in the absolute coordinate system are (1000,0,1973), and the Euler rotation angles (γ,β,α) are (0°, 30°, 0°), we can calculate out the coordinates of the centre of mass C of the moving platform (900,0,1800). Substituting all these known quantity into the matrix equations of (15), (16) and (17), we obtain the values of input crank rotation angles θ_{i1}, $\theta_{i2}\,(i=1\sim3)$, and for more detailed calculation results, see Table 1. It can be seen that the input crank rotation angles $(\theta_{11},\theta_{12},\theta_{21},\theta_{22},\theta_{31},\theta_{32})$=(3.445°, 0°, -95.963°, 23.286°, -95.963°, -23.286°) are reasonable solutions of the inverse kinematics.

TABLE 1 Inverse kinematics solution of the metamorphic parallel cutting mechanism's original configuration framework

C (x_c,y_c,z_c) (γ,β,α)	$(\theta_{11}\quad\theta_{12})$	$(\theta_{21}\quad\theta_{22})$	$(\theta_{31}\quad\theta_{32})$
(900,0,1800) (0°,30°, 0°)	(3.445° 0°)	(-95.963°23.286°)	(-95.963°-23.286°)
	(-130.724° 0°)	(-133.334°23.286°)	(-133.334°-23.286°)
	(-49.337°180°)	(95.963° 156.714°)	(95.963° -156.714°)
	(176.557°180°)	(133.334° 156.714°)	(133.334° -156.714°)

Taken $(\theta_{11},\theta_{12},\theta_{21},\theta_{22},\theta_{31},\theta_{32})$
$=(3.445^\circ,0^\circ,-95.963^\circ,23.286^\circ,-95.963^\circ,-23.286^\circ)$
as the input crank rotation angles of the mechanism, we obtain eight sets of forward kinematics solutions by solving the forward kinematics equation of the mechanism. For more detailed calculation results, see Table 2.

It can be seen that the two sets of x and z values calculated out based on the four sets of y values of equation (11) all satisfy equation (7) and thus are all reasonable solutions. Hence, we obtain eight sets of forward kinematics solutions.

TABLE 2 Forward Kinematics Solution of the Metamorphic Parallel Cutting Mechanism's Sub- configuration Framework

No	θ_{13}	θ_{23}	θ_{33}
1	2.5956	3.0572	3.0572
2	1.1422	3.0572	3.0572
3	1.9517+1.2135i	-2.2954+1.0379i	-2.2954+1.0379i
4	-2.1642+0.7489i	-2.2954+1.0379i	-2.2954+1.0379i
5	1.9516-0.2133i	-2.2954-1.0379i	-2.2954-1.0379i
6	-2.1642-0.7484i	-2.2954-1.0379i	-2.2954-1.0379i
7	2.5728	0.6784	0.6784
8	1.8045	0.6784	0.6784

After calculation, the obtained coordinates of M_1, M_2 and M_3 in the absolute coordinate system are shown in Table 3.

TABLE 3 The Coordinates of M_1, M_2 and M_3 in the Absolute Coordinate System

No	M_1			M_2			M_3		
	x_{M_1}	y_{M_1}	z_{M_1}	x_{M2}	y_{M2}	z_{M2}	x_{M3}	y_{M3}	z_{M3}
1	301.79	0	282.205	301.79	0	282.205	301.79	0	282.205
2	-453.439	0	-1418.2	-453.439	0	-1418.2	-453.439	0	-1418.2
3	272.887	0	268.508	272.887	0	268.508	272.887	0	268.508
4	1592.8	0	1400	1592.8	0	1400	1592.8	0	1400

Based on Table 3 and the equation (15), (16) and (17), we can calculate out the coordinates of the geometric center C and the z-y-x Euler angles of the moving platform, see Table 4.

TABLE 4 The coordinates of the geometric centre C and the z-y-x Euler angles of the moving platform

No	C			α	β x_C	γ y_C
	x_C	y_C	z_C			
1	-411.729	0	-79.67	0°	-411.729	0
2	-663.473	0	-646.41	0°	-663.473	0
3	486.443	0	1387.103	0°	486.443	0
4	899.993	0	1800	0°	899.993	0

Since the phase spaces of the first and the third set of solutions will cause interference in the mechanism, and the phase space of the second set is outside the range of maximum rotation angle $\theta_{i3}(i=1\sim3)$, we give all these three sets of solutions up. Errors resulting from using approximate solutions of inverse kinematics to solve the forward kinematics of the mechanism are small between the fourth set of solutions and the inverse kinematics solutions. Figure 5 is the abbreviated drawing of the mechanism pose of the fourth solutions set, and the corresponding model pose is shown in Figure 6.

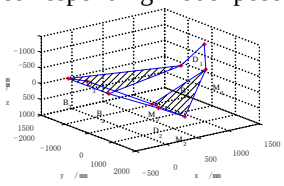


FIGURE 5 Mechanism pose diagram of position positive solutions

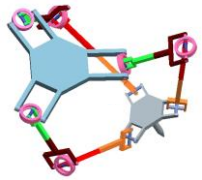


FIGURE 6 Model pose of position positive solutions

3.2 FORWARD AND INVERSE KINEMATIC ANALYSIS OF THE METAMORPHIC PARALLEL CUTTING MECHANISM'S SUB-CONFIGURATION FRAMEWORK

3.2.1 Forward Kinematic Analysis of the Sub-configuration Framework

Figure 7 shows the moving coordinate system and kinematic screw of metamorphic parallel cutting mechanism's the sub-configuration framework. Establish the absolute coordinate system *OXYZ*, and the sub-configuration framework has only 3 degrees of freedom.

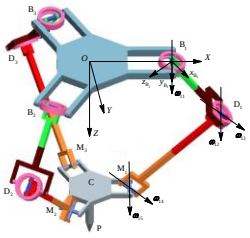


FIGURE 7 Coordinate of motion and twist of sub-configuration

Suppose the angles of rotation about e_{i1}, e_{i2}, e_{i3}, e_{i4}, e_{i5} $(i=1\sim3)$ in each limb chain are θ_{i1}, θ_{i2}, θ_{i3}, θ_{i4}, θ_{i5} $(i=1\sim3)$ respectively. Let metamorphic kinematic pair B_i be the driving joint and the rest be driven joints. Suppose the coordinates of the geometric center C of the moving platform $\mathrm{M_1M_2M_3}$ in the absolute coordinate system are $(x_\mathrm{C}, y_\mathrm{C}, z_\mathrm{C})$, based on which the coordinates of M_1, M_2, M_3 in the absolute coordinate system are :

$$\begin{cases} \mathrm{M}_1=\begin{bmatrix} x_C+r & y_C & z_C \end{bmatrix}^{\mathrm{T}} \\ \mathrm{M}_2=\begin{bmatrix} x_C-r/2 & y_C+\sqrt{3}r/2 & z_C \end{bmatrix}^{\mathrm{T}} \\ \mathrm{M}_3=\begin{bmatrix} x_C-r/2 & y_C-\sqrt{3}r/2 & z_C \end{bmatrix}^{\mathrm{T}} \end{cases}. \tag{18}$$

To calculate the coordinates of D_i in the absolute coordinate system, let fully expanded limb chain $\mathrm{B_iD_iM_i}$ be the initial phase space, at which the conversion between $\mathrm{B}_i x_{Bi} y_{Bi} z_{Bi}$ and the absolute coordinate system is:

$$g_{\mathrm{B}_i}(\theta)=e^{\theta_{i1}\hat{\xi}_{i1}^{\omega}}g_{\mathrm{B}_i}(0). \tag{19}$$

And the coordinates of D_i in the absolute coordinate system are:

$$\mathrm{D}_i=R_{\mathrm{B}i}\mathrm{D}_i^{\mathrm{B}_i}+t_{\mathrm{B}_i}. \tag{20}$$

According to the rigidity constraint equation $\left|\mathrm{D}_i\mathrm{M}_i\right|^2=m^2(i=1\sim3)$, we have the position kinematics equation of the metamorphic parallel mechanism's sub-configuration framework:

$$\begin{cases} (x_c+r-l\cos\theta_{11}-R)^2+y_c^{\,2}+(z_c+l\sin\theta_{11})^2-m^2=0 \\ \left[x_c-\frac{1}{2}r+\frac{1}{2}(l\cos\theta_{21}+R)\right]^2+\left[y_c+\frac{\sqrt{3}}{2}r-\frac{\sqrt{3}}{2}(l\cos\theta_{21}+R)\right]^2+(z_c+l\sin\theta_{21})^2-m^2=0 \\ \left[x_c-\frac{1}{2}r+\frac{1}{2}(l\cos\theta_{31}+R)\right]^2+\left[y_c-\frac{\sqrt{3}}{2}r+\frac{\sqrt{3}}{2}(l\cos\theta_{31}+R)\right]^2+(z_c+l\sin\theta_{31})^2-m^2=0 \end{cases}. \tag{21}$$

Let $a=r-l\cos\theta_{11}-R$, $b=l\sin\theta_{11}$, $e=l\sin\theta_{21}$, $h=l\sin\theta_{31}$, $c=-r/2+(l\cos\theta_{21}+R)/2$, $d=\sqrt{3}r/2-\sqrt{3}(l\cos\theta_{21}+R)/2$, $f=-r/2+(l\cos\theta_{31}+R)/2$, $g=-\sqrt{3}r/2+\sqrt{3}(l\cos\theta_{31}+R)/2$.

So that the simultaneous equations (21) can be written as:

$$\begin{cases} (x_c+a)^2+y_c^{\,2}+(z_c+b)^2-m^2=0 \\ (x_c+c)^2+(y_c+d)^2+(z_c+e)^2-m^2=0 \\ (x_c+f)^2+(y_c+g)^2+(z_c+h)^2-m^2=0 \end{cases}, \tag{22}$$

which can be simplified to a matrix form:

$$2\begin{bmatrix} a-c & -d & b-e \\ a-f & -g & b-h \end{bmatrix}\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}=\begin{bmatrix} d^2-a^2+c^2-b^2+c^2 \\ g^2-a^2+f^2-b^2+h^2 \end{bmatrix}. \tag{23}$$

Based on equation (23), let z_c represent x_c and y_c, and substitute it into equation (24), thereby deriving the

analytic expression of z_c relative to θ_{11}, θ_{21} and θ_{31}. Then x_c and y_c can be calculated out, and the forward kinematics solution is achieved.

3.2.2 Inverse kinematic analysis of the sub-configuration framework

Substituting the half angle formula $\sin\theta = 2\tan(\theta/2)/[1+\tan^2(\theta/2)]$, $\cos\theta = [1-\tan^2(\theta/2)]/[1+\tan^2(\theta/2)]$ into equation (21), we have the expressions of $\tan(\theta_{11}/2)$, $\tan(\theta_{21}/2)$ and $\tan(\theta_{31}/2)$ about x_c, y_c and z_c, thereby obtaining the input crank rotation angles θ_{11}, θ_{21} and θ_{31}. Then the inverse kinematics solution of the sub-configuration framework is achieved.

3.2.3 A numerical example of forward and inverse kinematics solution

Suppose the coordinates of the cutting gun endpoint P are (1000,0,1000), and the coordinates of the moving platform geometric center C are (1000,0,800). Substituting these two known quantities into equation (21), we have the input crank rotation angles θ_{11}, θ_{21} and θ_{31}. See Table 5 for detailed calculation results. The results $\theta_{11} = -156.033^\circ$ and $\theta_{21} = \theta_{31} = -167.665^\circ$ will cause interference or unreasonably distribution of the components in the mechanism, so we give them up. Then supposing the coordinates of the moving platform geometric center C is (-1000,0,200), $\theta_{11} = -159.101^\circ$ and $\theta_{21} = \theta_{31} = -96.178^\circ$ among the results obtained are also unreasonable and are therefore casted.

TABLE 5 Inverse kinematics solution of the metamorphic parallel mechanism's sub-configuration framework

C(x_c, y_c, z_c)	(1000,0,800)	(-1000.0.200)
θ_{11}	23.174°, -156.033°	-159.101°, -146.005°
θ_{21}	-116.654°, 167.665°	-96.178°, 32.956°
θ_{31}	-116.654°, 167.665°	-96.178°, 32.956°

Taken $(\theta_{11}, \theta_{21}, \theta_{31}) = (23.174^\circ, -116.654^\circ, -116.654^\circ)$ in Table.5 as input crank rotation angles, we obtain the forward kinematics solution by solving the forward position equation of the mechanism. See Table 6 for detailed calculation results.

TABLE 6 Forward Kinematics Solution of Metamorphic Parallel Mechanism's Sub-conFigureuration Framework

$(\theta_{11}, \theta_{21}, \theta_{31})$	(23.174°,-116.654°, -116.654°)
C (x_c, y_c, z_c)	(1006.427,0,995.852) (-280.248,0,-429.140)

Obviously, (1006.427,0,995.852) are reasonable solutions; Figure 8 is the corresponding pose of end moving platform at the model mechanism; Figure 9 is the corresponding pose of the model. Taking the inverse kinematics solution results calculated to the thousandth as the input of forward kinematics solution may result in errors between the forward and the inverse kinematics solution results.

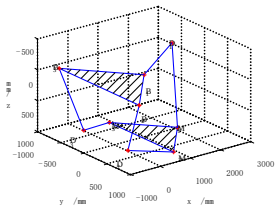


FIGURE 8 Mechanism pose diagram of position positive solutions

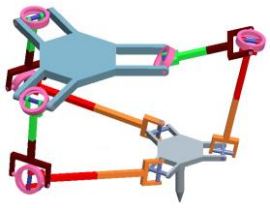


FIGURE 9 Model pose of position positive solutions

4 Study on kinematics simulation of the metamorphic parallel cutting mechanism

4.1 KINEMATICS SIMULATION OF THE METAMORPHIC PARALLEL CUTTING MECHANISM'S METAMORPHIC ORIGINAL CONFIGURATION FRAMEWORK

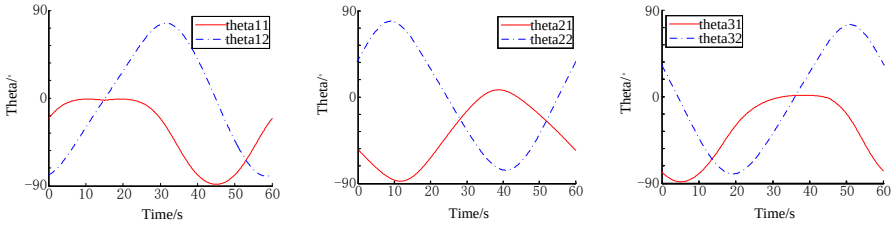
4.1.1 MATLAB-based kinematics numerical simulation of the original configuration framework

The sub-configuration framework of metamorphic parallel cutting mechanism should be able to perform upper and lower bevel cutting on a surface, which means the end moving platform can move with changing spatial pose and trajectory. Now we take upper bevel cutting as an example, and the space motion requirement for the cutting gun endpoint P are:

$$\begin{cases} x = 1000\sin(2\pi t/2) \\ y = 1000\cos(2\pi t/2) \\ z = -400\sin(2\pi t/2) + 1400 \\ \alpha = 0 \\ \beta = -(\pi/6)\cdot\sin(2\pi t/2) \\ \gamma = (\pi/6)\cdot\cos(2\pi t/2) \end{cases}, \tag{24}$$

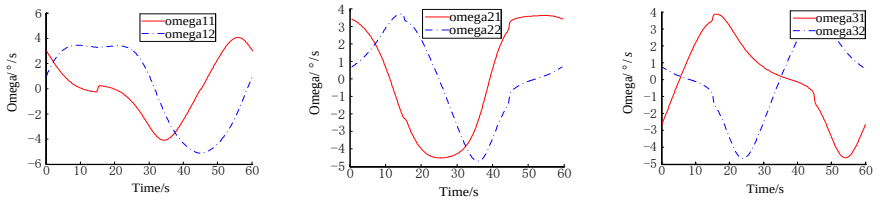
where T is a time cycle with the value of 60s.

Set 300 evenly spaced numerical points t from 0 to 60s. Based on the above inverse kinematics equation of metamorphic parallel cutting mechanism's original configuration framework, using Matlab, we calculate out the 300 numerical values of the 6 driving kinematics pairs rotation angles in the original configuration framework. Then draw the curve of driving kinematic pair rotation angles to time based on the numerical values obtained. As shown in Figure 10, compute the derivatives of the analytic expressions of the six input crank rotation angles with respect to t, and obtain 300values of inputs angular velocities. Then draw the angular velocity curve, as shown in Figure 11.



(a) Angle changing curve of the drive pair in the branched chain $B_1D_1M_1$ (b) angle changing curve of drive pair in the branched chain $B_2D_2M_2$ (c) angle changing curve of drive pair in the branched chain $B_3D_3M_3$

FIGURE 10 Input angle curve of source configuration of metamorphic parallel cutting mechanism



(a) Angular velocity changing curve of drive pair in the branched chain $B_1D_1M_1$ (b) Angular velocity changing curve of drive pair in the branched chain $B_2D_2M_2$ (c) Angular velocity changing curve of drive pair in the branched chain $B_3D_3M_3$

FIGURE 11 Input angular velocity curve of source configuration of metamorphic parallel cutting mechanism

4.1.2 ADAMS-based kinematics simulation of the original configuration framework

Figure 12 shows the simulation model. Add six driving force on the three Hooke's joints; and draw spline curve using the 300 numerical values of driving kinematic pair rotation angles calculated out by MATLAB and add it as driven spline curve to the AKISPL function. Figure 13 is the spline curve of the input crank rotation angle θ_{11}.

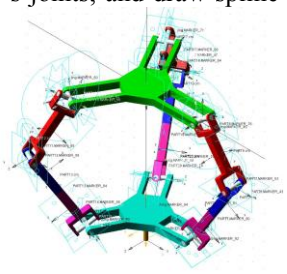


FIGURE 12 Simulation models in ADAMS

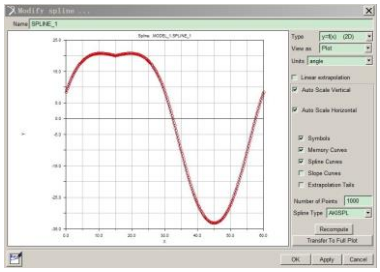
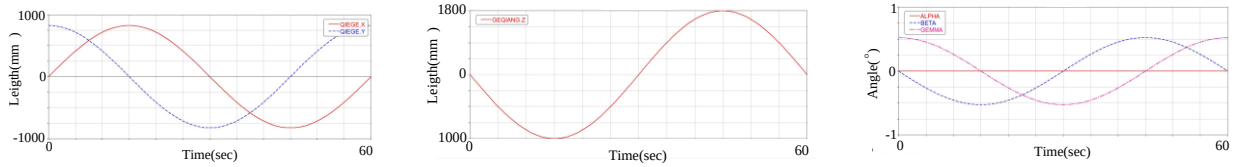


FIGURE 13 The spline curve of input angles

Perform kinematics simulation of the above model by setting the simulation time 60s and the number of time steps 300. And measure the displacement and angle changes of the coordinate system of the end moving platform's center of mass relative to that of the fixed platforms, thereby obtaining the angular displacement and the angular velocity change curves of the end moving platform's center of mass, as shown in Figure 14.



(a) displacement curve on X, Y axis at the end point of cutting gun (b) displacement curve on Z axis at the end point of cutting gun (c) angle curve of the end point of cutting gun relative to system coordinate axis

FIGURE 14 Change curves of displacement and angle of the end moving platform's center relative to the static platform

4.1.3 The simulation output analysis

The curve derived from the simulation shows the ADAMS-based simulation output coincides well with the motion characteristics as well as the changing trend of the center of mass of the mechanism end moving platform with initial settings. To find the numerical errors between the simulation output and the theoretical motion characteristics, the simulation values and theoretical values of the each motion parameters at the time of t=0,10,20,30,40,50,60s are collected, thereby obtaining the maximum relative errors. See Table 7 for more detail.

TABLE 7 The Simulation-theoretical Relative Errors of the Mechanism Motion Parameter

Motion Parameters	X(mm)	Y(mm)	Z(mm)	X-Angle(°)	Y-Angle(°)	Z-Angle(°)
Relative Errors	0.0048	0.0059	0.0005	0.006	0.0021	0.0044

Comparing the ADAMS simulation values and theoretical values of motion parameters of the metamorphic parallel cutting mechanism's original configuration framework, as shown in Table 7, we know the maximum relative errors at each time point are all restricted on a level of 0.01. It means the kinematics method adopted in this paper for the study of the metamorphic parallel cutting mechanism's original configuration framework is correct and reasonable, and the bevel cutting on the surface can be achieved via this mechanism.

4.2 KINEMATICS SIMULATION OF THE METAMORPHIC PARALLEL CUTTING MECHANISM'S SUB-CONFIGURATION FRAMEWORK

4.2.1 MATLAB-based kinematics numerical simulation of the sub-configuration framework

The metamorphic parallel cutting mechanism's sub-configuration framework has 3 degrees of freedom, and the motion of the end moving platform is 3D satisfying the requirement of square groove cutting on a surface.

Now we give the specific requirements for trajectory of the cutting gun endpoint:

$$\begin{cases} x = 1000\sin\dfrac{2\pi}{\mathrm{T}}t \\ y = 1000\cos\dfrac{2\pi}{\mathrm{T}}t \\ z = -400\sin\dfrac{2\pi}{\mathrm{T}}t + 1400 \end{cases}, \tag{25}$$

where T is a time cycle with value of 60s.

Set 300 evenly spaced numerical points t from 0 to 60s. Based on the inverse kinematics equation of the sub-configuration framework, using Matlab, we calculate out the input crank rotation angles and angular velocities at the 300 numerical points, and then draw the curve of input rotation angle and angular velocity to time, as shown in Figure 15 and Figure 16.

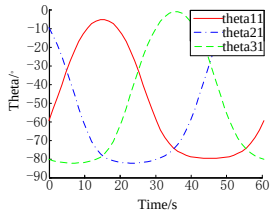


FIGURE 15 The input angle's changing curve of sub configuration

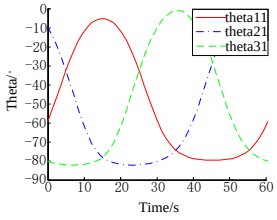


FIGURE 16 The input angular's velocity curve of sub configuration

4.2.2 ADAMS-based kinematics simulation of the sub-configuration framework

The kinematics simulation method applied to the study of the metamorphic parallel cutting mechanism's sub-configuration framework is the same with that applied to the original configuration. And Figure 17 shows the simulation model. Using the same method in section 3.1.2, import the values of the three driving pair rotation angles obtained from MATLAB into ADAMS, and use Spline curve to fit the data points of driving pair rotation angles. Then, set spline-driven pattern for the sub-configuration framework.

Perform kinematics simulation of the above model by setting the simulation time 60s and the number of time steps 300. And measure the displacement and angle changes of the coordinate system of the center of mass of the end moving platform relative to that of the fixed platforms, thereby obtaining the angular displacement and the angular velocity change curves of the center of mass of the end moving platform, as shown in Figure 18.

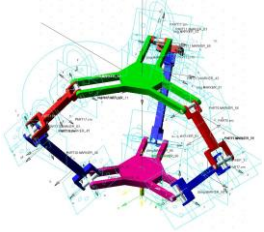
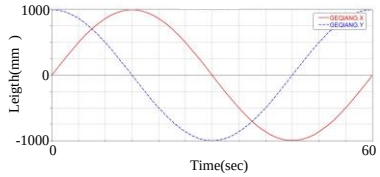
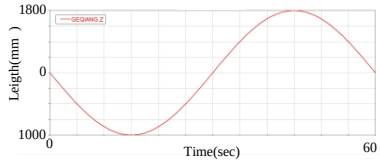


FIGURE 17 The ADAMS Simulation Model



(a) X, Y axis displacement curve of cutting gun end point



(b) Z axis displacement curve of cutting gun end point

FIGURE 18 Each axis displacement curve of the end moving platform's centroid

4.2.3 The simulation output analysis

The analysing method is the same as that applied to the original configuration framework. Similarly, the ADAMS simulation values and theoretical values (motion values of the end moving platform's center of mass in the mechanism with initial settings) of the each motion parameters at the time of t=0,10,20,30,40,50,60s are collected, thereby obtaining the maximum relative errors as shown in Table 8.

TABLE 8 The Simulation-theoretical Relative Errors of the Mechanism's Motion Parameters

Motion Parameters	X(mm)	Y(mm)	Z(mm)
Relative Errors	0.0028	0.0045	0.0013

Comparing the ADAMS simulation value and theoretical value (the mechanism with initial settings) of the motion parameters of the metamorphic parallel cutting mechanism's sub-configuration framework, as

shown in Table 8, we know the maximum relative errors at each time point are all restricted to a level of 0.01. It means the kinematics method adopted in this paper for the study of sub-configuration framework is correct and reasonable, and the bevel cutting on the surface can be achieved via this mechanism.

5 Conclusions

(1) Based on the traditional spherical pair, a novel metamorphic kinematic pair with four different frameworks was designed according to the principles of metamorphic mechanism. And then we devised a metamorphic parallel cutting mechanism structure using this kinematic pair.

(2) We analysed the forward and the inverse kinematics solution for the original and the sub-configuration of the metamorphic parallel cutting mechanism respectively, and made numerical example verification using MATLAB. Finally, the ADAMS simulation showed that this novel metamorphic parallel cutting mechanism is capable of square groove and bevel cutting on complex surfaces satisfying the technical requirements of changing spatial pose and trajectory cutting on complex large-size surfaces.

Acknowledgements

This work is supported by National Natural Science Foundation of China (51175099).

References

[1] Jinli Lv 2009 Design and Simulation on Series-parallel Mechanisms of cutting machines for complex surface. Harbin Engineering University 1-3
[2] Zongyi Wang, Shenghai Hu, Shijun Zhao 2003 Design of big intersecting circle flame cutting machine *Journal of Harbin Engineering University* **24**(3) 258-62
[3] Xiaodong Zhang 2012 *Design and dynamic characteristics study on metamorphic mechanisms of cutting machines for complex surface* Harbin Engineering University 1-3
[4] Seamus Gordon, Hillery M T 2005 Development of a high-speed CNC cutting machine using linear motors *Journal of Materials Processing Technology* **166** 321-9
[5] http://www.messercs.cn/pt/asiapacific/systems/machines/metalmaster/
[6] Xingzu Ming, Jinhua Liu 2005 Design of large-scale NC-fair incision machine with multi-functions *Modern Manufacturing engineering* 107-9
[7] Meina Sun, Yongchun Zheng, Zhenrong Gang 2007 Design of CNC.big intersecting circle flame cutting machine *Machinery Design & Manufacture* (10) 139-40
[8] Kumar Vijay 1990 Characterization of Workspaces of Parallel Manipulators *Mechanism Synthesis and Analysis* (25) 321-9
[9] Boudreau R, Gosselin V 1999 The Systhesis of Planar Parallel Manipulators with a Genetic Algorithm *ASME Journal of Mechanical Design* **12**(4) 533-7
[10] Feng Gao 2005 Reflection on the current status and development strategy of mechanism research *Chinese Journal of mechanical engineering* **41**(8) 3-17
[11] Shu Zhang, Heisel U 2003 *Parallel Kinematics Machine Tool* 1st ed. Beijing: Machinery Industry Press, China
[12] Bhattacharya S, Hatwal H, Ghosh A 1995 On the Optimum Design of a Stewart Platform Type Parallel Manipulators *Robotica*, **13** 133-40
[13] Yang Zhao 2014 Study on Predictive Control for Trajectory Tracking of Robotic Manipulator *Journal of Engineering Science and Technology Review* **7**(1) 45-51
[14] Shu Xu, etc. 2014 Local fractional Laplace variational iteration method for non-homogeneous heat equations arising in fractal heat flow *Mathematical Problems in Engineering* **2014** 1-10
[15] Dai J S, Zhang Q 2000 Metamorphic mechanisms and their configuration models *Scientific Journal of Mechanical Engineering the Scientific Journal of Chinese Mechanical Engineering Society* **13** 212-8
[16] Zongh Guo, Lvzhong Ma, Qizhi Yang 2005 Topological type analysis of the variable freedom mechanism based on the metamorphic principle *China Mechanical Engineering* **16** 1-3
[17] Zhang P, Dai J S 2009 Metamorphic techniques and geometric reconfiguration principles *Proceedings of the 2009 ASME/IFToMM International Conference on Reconfigurable Mechanisms and Robots* 32-40
[18] Shujun Li, Dai J S 2011 Topological representation of planar mechanisms based on Assur group elements *Journal of mechanical engineering* **47** 8-12
[19] Xianjin Wang, Yang Zhao, Xiaojun Yang 2014 Local fractional variational iteration method for inhomogeneous Helmholtz equation within local fractional derivative operator M*athematical Problems in Engineering* **2014** 1-10
[20] Dai J S, Jones J R 1999 Mobility in metamorphic mechanism of foldable/ereetable kinds *Transaction of the ASME, Journal of Mechanical Design* **121**(3) 375-82
[21] Dai J S, Zhang Q X 2000 Metamorphic mechanism and their configuration models *Chinese Journal of Mechanical Engineering (English Edition)* **13**(3) 212-8

Authors

Shenghai Hu, born on October 5, 1954, China

Current position, grades: Professor, College of mechanical engineering of Harbin Engineering University, a member of the Institute of China shipbuilding Surface weapon Society. He is engaged in the study of naval gun weapon system, NC machining technology, mechanical design theory and so on. He has won first prize of progress of science and technology of Heilongjiang Province, and has awarded the second prize of state scientific and technological progress.
Publications: edited three books and published more than 40 Academic papers.

Xiulian Liu, born on March 19, 1979, China

Current position, grades: lecturer, ph.D. degree in mechanical design from Liaoning technical university, China in 2005.
Scientific interests: the application of variable topology mechanism, NC machining technology.

Manhui Zhang, born on May 24, 1991, China

Current position, grades: ph. degree in mechanical design from Harbin Engineering University, China in 2012.
Scientific interests: the application of variable topology mechanism, NC machining technology.

Test paper composition based on fish swarm algorithm

Wenfa Wang*, Shiyao Wang

Department of Computer Science, Yanan University, Shaanxi, 716000, China

Received 1 May 2014, www.cmnt.lv

Abstract

Fish swarm algorithm was applied to seek optimization solution of test paper composition. Imitation results showed that fish swarm algorithm had better performance than random drawing algorithm in both composition accuracy and running time. The validity and superiority of the algorithm in this work, therefore, were verified.

Keywords: fish swarm algorithm, test paper composition, mathematical model, running time

1 Introduction

Development of intelligent item bank is an important aspect of computer application in educational field. By taking advantage of computer and technology and combining modern educational theory, questions are selected from current bank to automatically generate test questions satisfying teaching and examination requirements.

Currently, there are two main algorithms to solve test paper composition:

1) random search in current question bank for questions satisfying conditions with required amount. With great randomness and disconfirmation, this method is poorly intelligent, which cannot satisfy with education demands;

2) questions extracted for the whole test paper by random drawing. This method, which is relatively inflexible, cannot satisfy the different requirements of question types in papers.

2 Artificial fish swarm algorithm

Proposed by Li Xiaolei, et al. in 2002, Artificial Fish Swarm Algorithm (AFSA) is a new-type biotic optimization algorithm based on research into animal swarm intelligent behaviour. "Generally, the place with the most existing fish in water area is also the place with the most nutrients". According to this characteristic, this algorithm enables optimization to be realized by imitating foraging behaviour of fish swarm.

As a kind of optimizing method based on autonomous animats behaviour, AFSA is a new-type intelligent biotic algorithm built upon activity characteristics of fish swarm.

Fish has following behaviours by observation: foraging behaviour, swarm behaviour, following behaviour, and random behaviour. Because fish behaviours are tightly related to problem solution in this part, a major problem of algorithm implementation will be the adoption of simple and effective methods for constructing and realizing the behaviours.

2.1 BASIC CONCEPT

It is assumed that there is a swarm composed of *N* artificial fish in an n-dimensional target searching space. Status of every artificial fish per day can be expressed as vector quantity $\mathrm{X}=(x_1,x_2,\cdots,x_n)$, in which $x_i\ (i=1,2,\cdots,n)$ is the variable for optimization. Food concentration in current location of artificial fish is expressed as $\mathrm{Y}=f(x)$, while Y is the target function. Figure 1 shows that δ is the factor of congestion degree, step is the step length of artificial fish movement, $d=\|x_i-x_j\|$ represents distance between individuals, try number is the maximum try number for foraging by artificial fish, and visual means perception range of artificial fish.

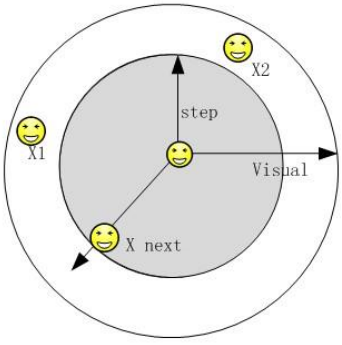


FIGURE 1 Visual field and step length of artificial fish

2.2 BEHAVIOUR DESCRIPTION OF ALGORITHM

1) Random behaviour: means that artificial fish move in visual field randomly. Once the food is found, it will lead the fish towards the direction with increasing foods rapidly.

***Corresponding author** e-mail: ydwangwenfa@126.com

2) Foraging behaviour: means that artificial fish move to a place with more foods. An artificial fish X_i will randomly select a status X_j in visual field, and then their target function values should be calculated, respectively. If Y_j is better than Y_i according to comparison, X_i will move one step to the direction of X_j. Otherwise, X_i should keep selecting status X_j in its visual field until satisfying the forward condition. After repeated try number times, a random step should be taken to give X_i a new status if the forward condition is not satisfied.

3) Swarm behaviour: means that artificial fish gather into a swarm during swimming to ensure their survival and avoid danger. The following three rules should be observed in a swarm. Division rule: is to avoid excessive congestion with neighbouring fish as much as possible. Alignment rule: is to keep consistent average direction to neighbouring fish as much as possible. And cohesion rule: is to move towards the centre of neighbouring fish as much as possible. An artificial fish X_i will search the fish number n_f and central location X_c. When $Y_c / n_f > \delta Y_i$, which represents the optimal and uncrowded status of partner's central location, X_i will move a step towards the central location, or, perform foraging behaviour.

4) Following behaviour: means that artificial fish move to the optimal direction in visual field. An artificial fish X_i will search the optimal partner X_j of function among all partners. When $Y_c / n_f > \delta Y_i$, which represents it is not too crowded around the optimal partner, X_i will move a step towards the central location, or, perform foraging behaviour.

5) Bulletin board: is a space for recording individual status of the optimal artificial fish. The current status of an artificial fish should be compared with recorded status on bulletin board after one time of iteration. When the comparison shows, that the current status is better, the status on bulletin board should be updated with it, or, keep unchanged. The required optimal value is the value on bulletin board outputted after all irritations in the whole algorithm.

2.3 ALGORITHM STEPS

Given that above-mentioned behaviours of artificial fish swarm, every artificial fish will select a behaviour for actual performance by exploring the condition of current environment and its partners. Eventually, artificial fish will gather around several local extremums in different numbers. In discussion of maximum problem, however, artificial fish with larger adaptive value will generally gather around larger extremum field, which is beneficial to obtain overall extremum field. Additionally, a greater number of artificial fish will gather around extremum field in larger value, which is beneficial to estimate and obtain overall extremum. Specific steps of artificial fish swarm algorithm are shown as follows:

Step 1. Swarm scale is determined as *N*, and *N* individuals are randomly generated in feasible region of variable. It is assumed that the visual range of artificial fish is Visual, the step length is Step, the factor of congestion degree is δ, and the try number is try number.

Step 2. Adaptive value of every individual in initial fish swarm is calculated, and the status and its value of the optimal artificial fish will be given to bulletin board.

Step 3. Individual will update itself by behaviours, such as foraging, swarm and following to generate a new fish swarm.

Step 4. All the individuals should be evaluated. The bulletin board should be updated with the individual, which is better.

Step 5. Algorithm will be finished when the optimal solution on bulletin board reaches the satisfying error range, or turn to step 3.

3 Fish swarm algorithm steps for test paper composition

During test paper composition, logic order must be followed to compose a test paper conforming to composition constraints. Specific steps are as Figure 2.

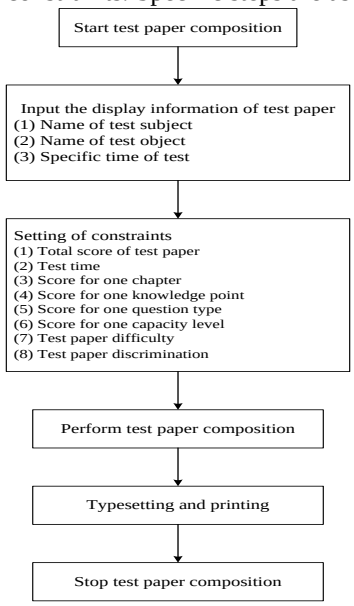


FIGURE 2 Fish swarm algorithm steps for test paper composition

4 Constraints for test paper composition

The objective of test paper composition is enabling test system to automatically select a certain amount of test questions from item bank according to constraints. These questions will compose a test paper satisfying examination requirement, which strives to sincerely and effectively reflect teaching level, and test examinees on mastery degree of relevant knowledge. Thus, test quality can be improved, and test objective can be reached.

Consequently, all the constraints affecting test paper problem should be determined according to examination requirement in composition to reach the objective. Generally, there are following constraints on test paper composition according to above analysis of composition steps:

1) Constraint of knowledge point. Knowledge points in teaching are always corresponding to section content in textbook, so this constraint is also the constraint of chapters. In occupational or other tests, however, constraint of knowledge point means the constraint of professional knowledge.

2) Constraint of question type. Constraint of question type means the question types included in the test paper, that is, question types for test paper composition in the examination.

3) Constraint of question amount. Constraint of question amount is the amount of test questions included in the test paper. Specifically, it is the question amount for one question type. This constraint is positively correlated with answer time

4) Constraint of answer time. Constraint of answer time means the longest time for finishing all the questions in the paper. In general, the longer the answer time is, the greater the question amount is.

5) Constraint of difficulty. Constraint of difficulty means the average difficulty of the test paper and the difficulty of test questions. Generally, test difficulty depends on different test targets. For example, in the test selecting minority of outstanding students, test difficulty should properly increase. In qualification test requiring majority of passers, however, the difficulty should properly reduce. Better effectiveness can be reached only if this constraint combines with the constraint of discrimination.

6) Constraint of discrimination. Constraint of discrimination is the discrimination capability of test paper. Total discrimination of test paper is the average discrimination of all test questions.

7) Constraint of exposure time. This means the test times of knowledge. The more test times, the more times the knowledge point will be repeated.

5 Mathematical model of test paper composition

The above analysis on constraints shows that the system is always limited by some constraints when selecting a test question from item bank in composition. If every constraint is a local constraint, all the constraints, which should be satisfied in the test paper, will constitute global constraint. Four vectors will be defined here:

Attribute set: $s=\{s_1,s_2,\cdots,s_n\}$, which is attribute variable of test question (question type, affiliated section, knowledge point, difficulty factor, score, etc.) or their quantization value;

Value set: $L=\{l_1,l_2,\cdots,l_n\}$, which corresponds to the value range of every attribute variable of test question;

Constraint set: $R=\{r_1,r_2,\cdots,r_n\}$, which corresponds to the constraint relation between different values for attribute variable of test question;

Question set: $I=\{i_1,i_2,\cdots,i_n\}$, which corresponds to the test question selected from item bank and satisfying with certain constraints.

It can be seen that test paper composition is to randomly select a group I from item bank, while the attribute variable *S* of this group should satisfy the constraint of *R* within value range *L*. Automatic test paper composition, therefore, can be described as a problem of multi-constraint objective satisfaction. In addition, the mathematical model of test paper composition can also be considered as a constraint satisfaction problem. The constraints can be divided into hard constraint and soft constraint in type.

5.1 MATHEMATICAL MODEL OF HARD CONSTRAINT

Actually, the test paper composition is to select test questions satisfying all the constraints from item bank. It is assumed that there are m test questions in a test paper, and each of these questions should satisfy n indicators. Therefore, the following *m*n* target matrix is established. (*n*-dimensional vectors include difficulty a_1, question type a_2, section a_3, teaching requirement a_4, time a_5, score a_6, etc. a_i means the indicator No. *i*. *m* refers to question amount in paper.) The matrix is as follows:

$$s=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

In this matrix every line represents one question in item bank, while every list is the attribute value of the question. The target matrix should satisfy the following constraints:

1) $\sum a_{i1} * a_{i6} / 100 = ND$ (difficulty of the whole paper, which is determined by the user of composition), the constraint of difficulty. The determination of question difficulty: scoring rate (d)=1- (average score/full score of the question);

2) $\sum a_{1i} * a_{i2} = m_d$ (m_d is the score of question type *d*), in which $C_{1i}=\begin{cases}1\\0\end{cases}$. If $a_{i2}=d$, take 1 for C_{1i}, otherwise, take 0;

3) $\sum c_{2i} * a_{i3} = Z_h$ (Z_h is the score for section *h*), in which $C_{2i}=\begin{cases}1\\0\end{cases}$. If $a_{i3}=h$, take 1 for C_{2i}, otherwise, take 0;

4) $\sum c_{3i} * a_{i4} = p_k$ (p_k is the score for teaching requirement k), in which $C_{3i}=\begin{cases}1\\0\end{cases}$. If $a_{i2}=k$, take 1 for C_{3i}, otherwise, take 0;

5) $\sum a_{i5} = t$ (t is the test time);

6) $\sum a_{i6} = G$ (G is the total score of test paper).

5.2 MATHEMATICAL MODEL OF SOFT CONSTRAINT

In test paper composition, "hard constraint" is to completely obtain the target value, while "soft constraint" is to incompletely obtain the target value. In practical process, above target functions (1)–(6) are difficult to be completely satisfied. In order to minimize the deviation between target function and target value, a priority should be established based on the importance of every target according to user requirements. All the targets should be ordered in this priority to achieve more targets as much as possible. The imported positive and negative deviation variables represent the parts more and less than target value, respectively. The process of test paper composition, therefore, becomes the solution process of "soft constraint" target value. The multi-target optimization model based on "soft constraint" is established as follows:

$$\min\ z = p_1\left|\sum_{i=1}^{p}\left(da_i^- + da_i^+\right)\right| + \ldots + \mathrm{p}_2\left|\sum_{j=1}^{q}\left(db_j^- + db_j^+\right)\right| .$$

Constraints:

$$\sum_{j=1}^{q}\sum_{k}^{n} f_k x_{ijk} + da_i^- + da_i^+ = a_i , \quad i = 1, 2K, p$$

$$\sum_{i=1}^{p}\sum_{k}^{n} f_k x_{ijk} + db_j^- + db_j^+ = b_j , \quad j = 1, 2K, \mathrm{q} .$$

$$\sum_{i=1}^{p}\sum_{j=1}^{q} x_{ijk} = S_k , \qquad k = 1, 2K, \mathrm{n}$$

In the constraints, p_i, the priority factor, represents relative importance of every target. p_i is always prior to p_i+1 to all values of i, in which i=1, 2,3… x_{ijk} represents the amount of questions with difficulty i, scope j and question type k in the test paper. da_i^- and da_i^+ represent the positive and negative deviation variables for total score of questions with difficulty i, respectively, which correspond to difficulty requirement a_i. db_j^- and db_j^+ represent the positive and negative deviation variables for total score of questions with scope j, respectively, which correspond to scope requirement b_j.

6 Algorithm simulation

Table 1 shows the parameter setting in algorithm simulation with MATLAB software according to algorithm steps and mathematical model of test paper composition:

TABLE 1 Parameter setting for fish swarm algorithm

Parameter	Value	Parameter	Value
Artificial fish swarm	50	Perception distance	1
The maximum iterations	100	Factor of congestion degree	0.618
The maximum try number of foraging	100	Step length of movement	0.1

The result of 100 iterations simulation results shows fish better convergence of the algorithm (see Figure 3).

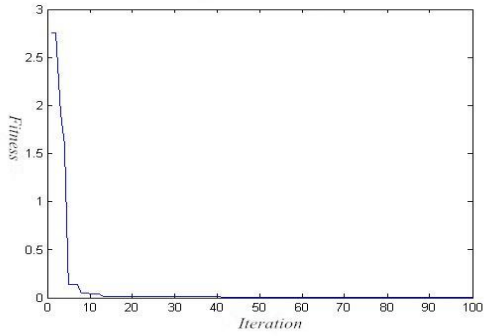


FIGURE 3 Convergence map of fish swarm algorithm

In order to verify the effectiveness and superiority of the algorithm in this work, the algorithm is compared with random drawing in composition accuracy and running time. Figure 4 and Figure 5 show the comparison results:

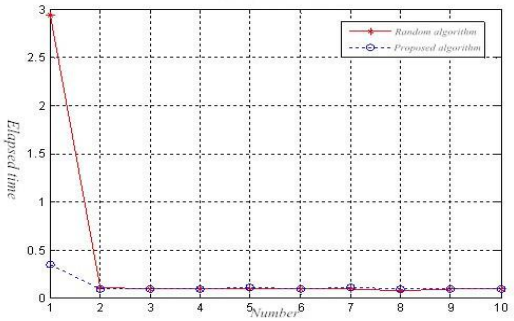


FIGURE 4 Comparison in running time

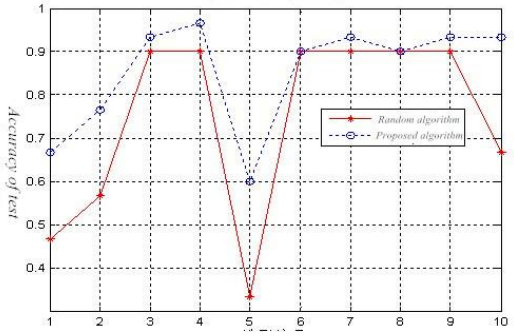


FIGURE 5 Comparison in accuracy

Simulation experiment is conducted by MATLAB software. By comparing with the algorithms in literature, the algorithm in this work is verified in both composition accuracy and running time. It is found in Figure 3 that the accuracy of fish swarm algorithm is higher than that of

random drawing. Moreover, it is found in Figure 4 that running time of fish swarm algorithm is also better than that of random drawing.

7 Conclusions

Fish swarm algorithm was applied to make optimization solution for test paper composition. Simulation result showed that fish swarm algorithm was better than random drawing in both composition accuracy and running time. The effectiveness and superiority of algorithm in this work, therefore, were verified.

Acknowledgements

The work was supported by Natural Science Fund of Yan'an (YA2013-12), "Related Models and Algorithms Improving Oil Production of Enterprises"; Natural Science Special Fund of Yan'an University (YDZ2013-02), "Oil Pump Scheduling Policy under Multi-Constrained Condition — A Case Study of Oil Plant in Qili Village"; National Innovation Training Program (201410719001), "Design and Implementation of Universal Questions System in Universities".

References

[1] Yuan X 2008 *Research on intelligent examination system based on adaptive genetic algorithm* Central South University Changsha
[2] Yin X, et al. 2009 Analysis and study of volume pattern in network test system *Computer Technology and Development* **19**(2) 206-8
[3] Lei Guoyu, Zhouyong 2006 *Analysis and design of strategy for test paper composition* Southwest University of Science and Technology, Sichuan
[4] Zhang Y, Li H 2007 Maneuver of composing examination papers from questions in view of genetic algorithms *Journal of Inner Mongolia Agricultural University* **27**(3) 103-5 (*in Chinese*)
[5] Zhang Q, Xing C 2008 Application of real-number encoding in intelligent test paper composition in view of genetic algorithms *Science & Technology Information* **1**(3) 44-5
[6] Zhou H 2009 Research on strategy for test paper composition in test system based on web *Northeast Normal University Jilin*
[7] Xia G 2008 *Research on general question bank management system of self-learning examination and its strategy for intelligent test paper composition* Tianjin Normal University, Tianjin
[8] Yun Q, Huang G 1997 *Genetic algorithm and genetic technology* Metallurgical Industry Press Beijing
[9] Wang Y, Hou S, Guo M 2003 Research on algorithm for automatic test paper generation in testing system *Journal of Harbin Institute of Technology* **35**(3) 343-8

Authors

Wenfa Wang, born in October, 1968, Tongshan, Shanxi Province, China

Current position, grades: an associate professor of Department of Computer Science, Yanan University, China.
Scientific interest: software technology and algorithm design.
Publications: 20 papers.

Shiyao Wang, born in November, 1989, Yanan, Shanxi Province, China

Current position, grades: a master of Department of Computer Science, Yanan University, China.
Scientific interest: algorithm design and analysis.
Publications: 3 papers.

Epidemic spreading in weighted homogeneous networks with community structure

Chang Zinan*, Shao Fei

[1]*Jiangsu Information Analysis Engineering Laboratory, Nanjing 211169, China*

[2]*School of Information Technology, Jinling Institute of Technology, Nanjing 211169, China*

Received 1 October 2014, www.cmnt.lv

Abstract

Community structure has been proven to have great impact on epidemic spread in weighted networks. To understand the epidemic propagation in weighted homogeneous networks with community structure, a model of pseudo-random network is presented with adjustable community structure. By changing the number of edges connecting to the nodes in the same community and the weight of edges connecting to the nodes in the same community, we investigate the epidemic spreading in weighted homogeneous networks with different community structure. Simulations show that both the number of within-community edge and the weight of within-community edge have great impact on epidemic spreading behaviour.

Keywords: epidemic spread, weighted network, homogeneous network, community structure

1 Introduction

In the past few years, the study of complex network has attracted the increasing interest since the small-world phenomenon introduced by Watts and Strogatz [1] and the scale-free phenomenon proposed by Barabási and Albert [2]. The ultimate goal of these studies on topology is to understand and explain the dynamic process up these networks, for instance, to understand how the topology of social network affects large-scale epidemics such as H1N1 [3] and so on. It is of great importance to control the epidemic spreading taking place in real world networks. A great deal of models has been proposed to investigate the feature of epidemic spreading where the node is classified in three states: susceptible (which will not infect others but may be infected), infected (which is infective) and recovered (which has recovered from the disease and has immunity). The SI [4-6], SIS [7-9], and SIR [10-12] models are proposed based on the discrete states of the nodes. To investigate the dynamical behaviours in the very early stage of epidemic outbreaks when the effects of recovery and death can be ignored, we shall focus on the susceptible-infected (SI) model in which individuals can be in two discrete states, either susceptible or infected. Each individual is represented by a node of the network and the edges are the links between individuals along which the infection may spread. An infected node can infect any of its neighbourhood nodes with a fixed probability λ at each time step and the infected nodes remain always infective. At the beginning time, I_0 nodes are randomly selected to infect the rest of the network, the dynamical process being affected by the topology of the network. The behaviours of the SI model are not only of theoretical interest, but also of practical significance.

The previous studies on networks have been principally focused on the unweighted network where edges between nodes are either present or not. However, lots of real world systems such as the scientific collaboration networks [13], the world-wide airport network [14], the mobile networks [15] and the Internet [16] have proved to be specified not only by the topology but also by the edge weight.

Accompany with the continuing study of complex networks, another common feature of many real world systems, the community structure, is founded [17-21]. Community structure is the tendency for nodes to divide into subsets within which node-node connections are dense, but between which connections are sparser. There have also been some studies to investigate the impact of community structure upon epidemic spreading in scale-free networks [22-24].

However, there are few studies to combine weight and community structure well to investigate the epidemic spreading in homogeneous networks. Indeed, the weight distribution of the edges would impact the community structure and the epidemic spreading in weighted homogeneous networks with community structure. In this paper, we proposed a model of pseudo-random network with adjustable community structure. By changing the number of edges connecting to the nodes in the same community and the within-community edge weight in the same community, we investigate the epidemic spreading in weighted homogeneous networks with different community structure.

**Corresponding author* e-mail: changzn@jit.edu.cn

This paper is organized as follows. In section 2 we describe the models, followed by the experimental evaluations in section 3. The conclusions are given in section 4.

2 Models

In order to study the dynamical behaviours in the very early state of epidemic outbreaks, we focus on the traditional SI model in which nodes are either susceptible or infected. If $s(t)$ and $i(t)$ are the density of susceptible and infected nodes at time t respectively, then $s(t)+i(t)=1$. Denote the spreading rate as λ at which each susceptible node acquires infection from an infected neighbour during one time step. In homogeneous networks, each node has approximately the same degree which makes it possible to use mean-field theory to obtain approximate results. In this case the system, we have

$$\frac{di(t)}{dt} = \lambda < k > i(t)[1-i(t)]. \tag{1}$$

Equation (1) states that the growth rate of infected nodes is proportional to the spreading rate λ, the density of susceptible nodes that may become infected $s(t)=1-i(t)$, and the number of infected nodes in contact with any susceptible one. The homogeneous mixing hypothesis considers that the last term is the product of node degree $<k>$ and the average density of infected nodes $i(t)$.

For our weighted SI model, We assume that transmit probability through the edge with weight w, λ_w, is equivalent to the infected probability that w infected nodes simultaneously influence the susceptible nodes [25,26], which is

$$\lambda_w = 1-(1-\lambda)^w. \tag{2}$$

We also employ the pseudo-random network model to investigate the epidemic spreading since all other properties such as average node degree will be equivalent to fully random networks except the controllable varying strength of community structure. These networks are comprised of n nodes which are split into *mods* communities of *n/mods* nodes each. Each node has on average Z_{in} edges connecting it to nodes of the same community and Z_{out} edges to nodes of other communities. While Z_{in} is varied, the value of Z_{out} is chosen to keep the total average degree constant, and set to $<k>$. And we assign different weights to the different kinds of edges: between-community edges are given a fixed weight of w_{out} (which is often set to 1 for simplicity), while within-community edges are given the weight $w_{in}=w$. As Z_{in} and w are increasing, the communities become better defined and easier to identify.

To know the influence of the accuracy of community structure identification on information transfer capacity, we employ the modularity measure. $\mathbf{A}$ is the adjacency matrix where $\mathbf{A}_{ij}=1$ if nodes i and j are connected and 0 otherwise. Let c_i be the community, which node i belongs to. The modularity measure, Q, is defined as follows [18]:

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i \cdot k_j}{2m}\right]\delta(c_i, c_j), \tag{3}$$

where $m = \frac{1}{2}\sum_{ij} A_{ij}$ is the number of edges in the network, $\delta(u,v)$ is 1, if $u=v$ and 0 otherwise. The higher the modularity Q_{max} is the stronger community structure the network has. In practice values for such networks typically fall in the range from about 0.3 to 0.7. Higher modularity values are very rare.

3 Simulations and analysis

At first, we check the impact of within-community edge weight w on the maximum modularity Q_{max} using pseudo-random networks with n=128 nodes which are divided into *mods*=4 communities with 32 nodes in each community. The average degree $<k>$ is set to 16. Simulation results are shown in Figure 1.

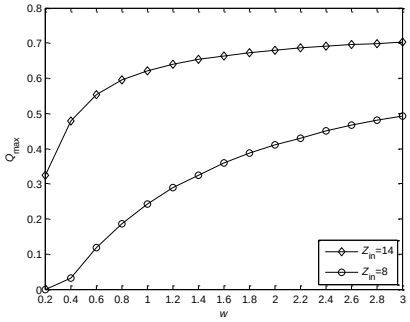


FIGURE 1 Q_{max} vs w, n=128, *mods*=4, $<k>$=16

From Figure 1 we can obtain that the increasing of edge weight will result in the increase of community structure especially in the traditional random network. In the pseudo-random network which has more with-community edges (Z_{in}=14 in Figure 1), the network has pronounced community structure even though the within-community edge weight is smaller. And the greater the number or the weight of within-community edge is, the higher the maximum modularity Q_{max} is.

Then we check the impact of the number of within-community edge Z_{in} on the maximum modularity Q_{max}. Simulation results are shown in Figure 2.

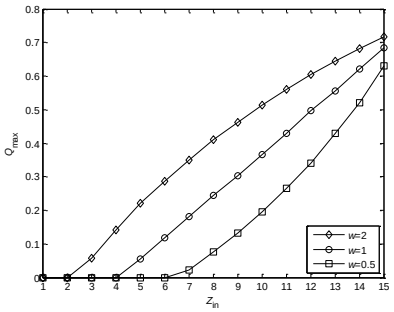
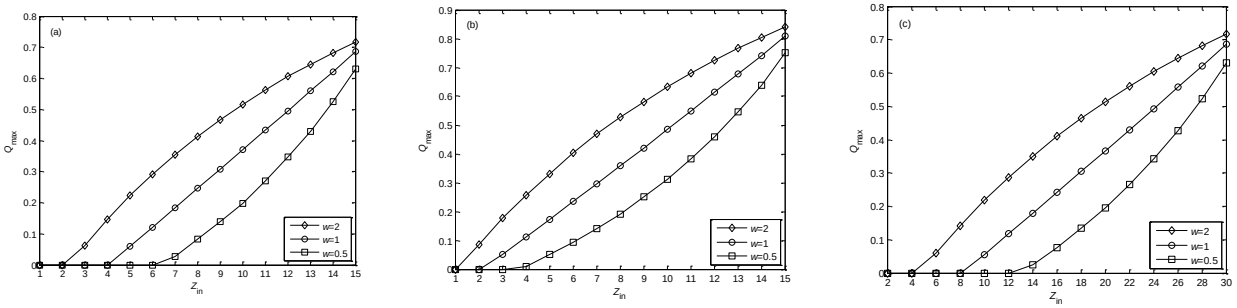


FIGURE 2 Q_{max} vs Z_{in}, n=128, *mods*=4, $<k>$=16

Figure 2 also proves that the greater the number or the weight of within-community edge is, the higher the maximum modularity Q_{max} is.

Then we check the impact of the total node number n, the community number *mods*, and the average node degree $<k>$ on the maximum modularity Q_{max}. Simulation results are shown in Figure 3.

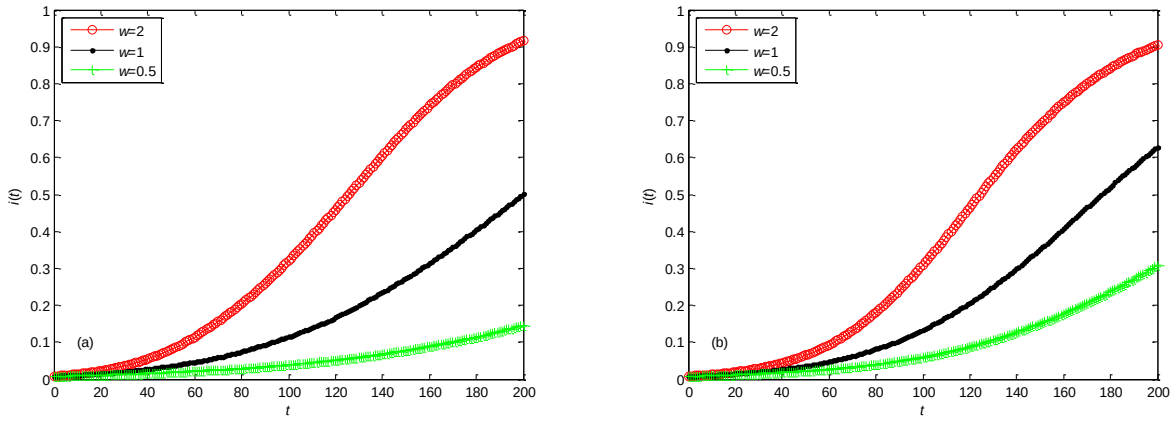


a) n=256, *mods*=4, $<k>$=16 b) n=128, *mods*=8, $<k>$=16 c) n=128, *mods*=4, $<k>$=32

FIGURE 3 Q_{max} vs Z_{in}

Comparing with Figure 2, we double the total node number, the community number and the average node degree to get the results shown in Figure 3a, 3b and 3c correspondingly. From Figure 3 we can obtain that the maximum modularity Q_{max} is also increasing accompany with the number and the weight of within-community edge. Only the changing of the community number affects the absolute value of maximum modularity. The changing of the total node number and the average node degree will barely affect the result of maximum modularity.

Now we focus on the epidemic spreading in weighted homogeneous networks. We set the the number of within-community edge Z_{in} as 14 for the network which has stronger community structure and 8 for the network which is a traditional random network. (In the scenario where average degree is 32, it changes to 28 and 16 accordingly.) At the initial age, we select a node randomly and make it an infected node. At each time step, the infected nodes will interact with their neighbours with probability λ_w which is defined in Equation (2). Simulations of different scenarios are shown in Figures 4, 5, 6 and 7.

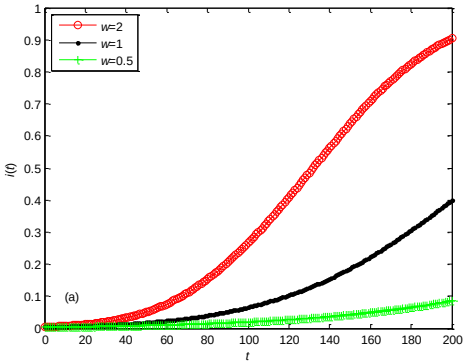


a) Z_{in}=14 b) Z_{in}=8

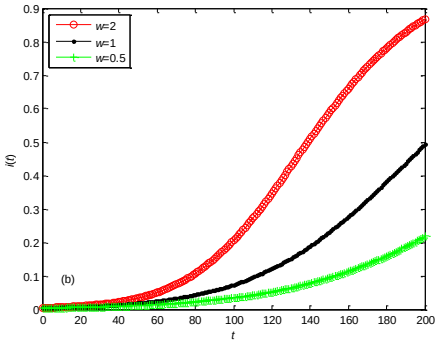
FIGURE 4 Epidemic spreading in weighted networks, n=128, *mods*=4, λ=0.002, $<k>$=16

In Figure 4, we report the density of infected nodes versus time in weighted homogeneous networks with different community structure. As shown in each figure, the epidemic propagation velocity is higher in networks with greater weight of within-community edge. However, when the weight of within-community edge is less than the weight of the between-community edge (w=0.5), less within-community edge will unexpectedly accelerate the epidemic propagation.

Then we also double the total node number, the community number and the average node degree to get the results shown in Figures 5, 6 and 7 correspondingly.

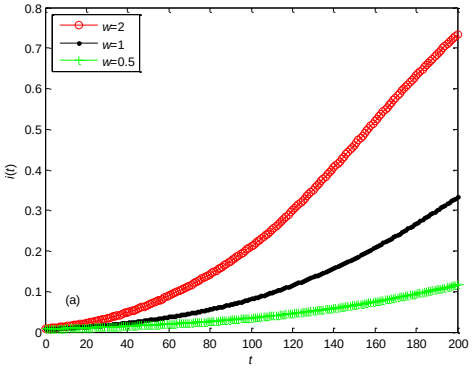


a) Z_{in}=14

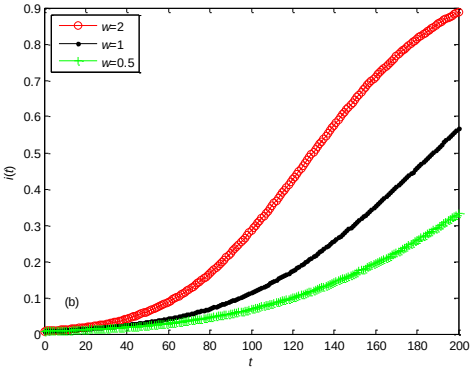


b) Z_{in}=8

FIGURE 5 Epidemic spreading in weighted networks, n=256, *mods*=4, λ=0.002, <k>=16

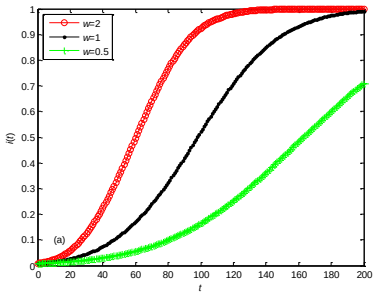


a) Z_{in}=14

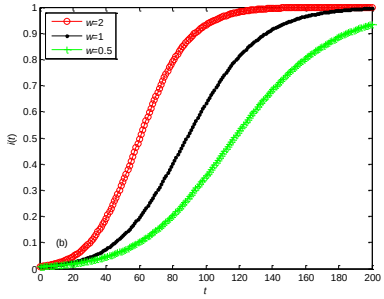


b) Z_{in}=8

FIGURE 6 Epidemic spreading in weighted networks, n=128, *mods*=8, λ=0.002, <k>=16



a) $3Z_{in}$=28

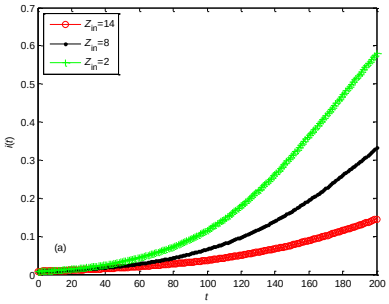


b) Z_{in}=16

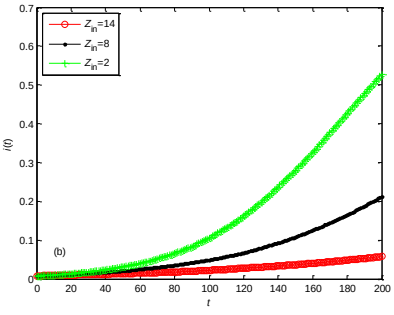
FIGURE 7 Epidemic spreading in weighted networks, n=128, *mods*=4, λ=0.002, <k>=32

As shown in each Figure, the increasing of weight of within-community edge will result in the acceleration of the epidemic propagation and less within-community edge will also accelerate the epidemic propagation when the weight of within-community edge is less than the weight of the between-community edge.

We utilize different within-community edge number Z_{in} to check the impact on epidemic spreading as shown in Figure 8.



a) w=0.5



b) w=0.25

FIGURE 8 Epidemic spreading in weighted networks, n=128, *mods*=4, λ=0.002, <k>=16

In both scenarios where the weight of within-community edge is less than the weight of the between-community edge, less within-community edge can accelerate the epidemic propagation.

4 Conclusions

How the community structure impact on the epidemic spreading in weighted homogeneous networks is studied in this paper. With our weighted SI model and the computer-generated pseudo-random networks model, the epidemic propagation velocity is studied in difference scenarios. Both increasing the number of within-community edge and increasing the weight of within-community edge can enhance the community structure. And increasing the weight of within-community edge will result in the acceleration of the epidemic propagation. Furthermore, we also proposed that less within-community edge will accelerate the epidemic propagation when the weight of within-community edge is less than the weight of the between-community edge. This study will shed light on how to restrain the epidemic spreading in weighted homogeneous network.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61373136), the Natural Science Foundation of Jiangsu Province, China (Grant No. BK2012082), the Research Foundation of Jinling Institute of Technology (Grant No. JIT-B-201406) and sponsored by Qing Lan Project. The author also gratefully acknowledges the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] Watts D J , Strogatz S H 1998 *Nature* **393** 440-2
[2] Barabási A L, Albert R 1999 *Science* **286** 509-12
[3] Fraser C, Donnelly C A, Cauchemez S, Hanage W P 2009 *Science* **324** 1557-61
[4] Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A 2004 *Phys Rev Let.* **92** 178701
[5] Barthelemy M, Barrat A, Pastor-Satorras R, Vespignani A 2005 *Journal of Theoretical Biology* **235** 275-88
[6] Zhou T, Liu J G, Bai W J, Chen G R, Wang B H 2006 *Phys Rev E* **74** 056109
[7] Pastor- Satorras R, Vespignani A 2001 *Phys Rev Let.* **86** 3200-3
[8] Pastor-Satorras R, Vespignani A 2001 *Phys Rev E* **63** 066117
[9] Hethcote H W 2000 *SIAM Review* **42** 599-653
[10] May R M, Lloyd A L 2001 *Phys Rev E* **64** 066112
[11] Moreno Y, Pastor-Satorras R, Vespignani A 2002 *European Physical Journal B* **26** 521-9
[12] Bailey N T J 1975 The mathematical theory of infectious diseases and its applications *Hafner Press: New York*
[13] Newman M E J 2001 *Phys. Rev. E* **64** 016132
[14] Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A 2004 *PNAS* **101** 3747-52
[15] Onnela J P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A L 2007 *PNAS* **104** 7332-6
[16] Pastor-Satorras R, Vespignani A 2007 Evolution and structure of the Internet: A statistical physics approach *Cambridge University Press*: *Cambridge*
[17] Girvan M, Newman M E J 2002 *PNAS* **99** 7821-6
[18] Newman M E J 2004 *Phys Rev E* **70** 056131
[19] Danon L, Díaz-Guilera A, Duch J, Arenas A 2005 *Journal of Statistical Mechanics: Theory and Experiment* P09008
[20] Newman M E J 2006 *PNAS* **103** 8577-82
[21] Fortunato S 2010 *Physics Reports* **486** 75-174
[22] Chu X W, Guan J H, Zhang Z Z , Zhou S G 2009 *Journal of Statistical Mechanics: Theory and Experiment* 2009 P07043
[23] Liu Z H, Hu B B 2005 *Europhysics Letters* **72** 315-21
[24] Huang W, Li C G 2007 *Journal of Statistical Mechanics: Theory and Experiment* P01014
[25] Tasgin M, Bingol H O 2012 *Advances In Complex Systems* **15** 1250061
[26] Sun Y, Liu C, Zhang C X, Zhang Z K 2014 *Phys Lett A* **378** 635-40

Authors

Zinan Chang, born in May, 1979, Jiangsu China

Current position, grades: a lecture at the School of Information Technology, Jinling Institute of Technology.
University studies: Master degree in Computational Mathematics from Nanjing Normal University in 2006.
Scientific interest: information security and complex dynamical networks.
Publications: 6 papers.

Fei Shao, born in December, 1978, Jiangsu China

Current position, grades: an associate professor at the School of Information Technology, Jinling Institute of Technology.
University studies: Ph. D Degree in Information Security from Nanjing University of Posts & Telecommunications in 2013.
Scientific interest: information security and complex dynamical networks.
Publications: 20 papers.

Design and realization of road geometry alignment simulation

Shengneng Hu[1*], Xiaoming Lu[2], Juan Han[2]

[1]*School of Civil Engineering and Communication, North China University of Water Resources and Electric Power, Zhengzhou, 450011, China*

[2]*University for Science and Technology Zhengzhou, Zhengzhou, 450007, China*

Received 1 October 2014, www.cmnt.lv

Abstract

Embarking from road's characteristic and road design's demand to the simulation system, road simulation system's characteristics have been discussed, and then road simulation system frame based on multi-agent has been constructed. The simulation agents have been divided into response agent, corporation agent and interface agent, and their structures have been analysed too. The car dynamics model, road ground view model and driver model have been built, the simulation results are validated by basic experiments. The simulation system is an efficient tool for road design evaluation.

Keywords: road simulation, road engineering, multi-agent, simulation model

1 Introduction

Road engineering investment is huge; it is not possible to verify rationality of design program through making "sample". Design is influenced by many factors; it is difficult to use a mathematical model to evaluate. Because of the lack of enough information, experts can only rely on experience to judge, it is difficult to make an objective and comprehensive evaluation [1]. The computer simulation technology is introduced into the evaluation of road design, through studying and establishing the road simulation system, takes the road geometry linear as the main body, repeatedly carries on the simulation experiment to the road design plan, scientific analysis and evaluation of the simulation results, may achieve the purpose of scientific evaluation of road design program.

The road simulation system is continuous variable dynamic system (CVDS) and discrete event dynamic system (DEDS) interaction promiscuous system (HS), has complexity, hybridity, interactivity, timeliness and modularity and other characteristics [2]. Therefore, in the road simulation system, how to describe the interaction of the discrete event and continuous variable, portray promiscuous system behaviour essence of road, so coordinate news transmissions of among various agent based on DEDS and CVDS in the process of simulation, is one of the key of simulation system [3].

Under these guides, the characteristics of road simulation system, system framework and development method are studied in this paper, which lays a theoretical foundation for the development of the simulation system, and has an important significance to realize signification, visualization, intelligent of road design.

2 Characteristics and framework of road simulation

2.1 ANALYSIS OF ROAD SIMULATION SYSTEM'S CHARACTERISTICS

Time-variable characteristic in road simulation process may describes with the following the time-variable characteristic model of person, car, road and environment.

$$\begin{cases} \Phi(t_i) = \{\alpha^i, r^i, w^i\} \\ \alpha^i = \{\alpha_v^i, \alpha_{av}^i, \alpha_{aH}^i, \alpha_\beta^i, \alpha_o^i, \alpha_p^i\} \\ r^i = \{r_H^i, r_V^i, r_C^i\} \\ w^i = \{w_E^i, w_C\} \end{cases}, \quad (1)$$

where $\Phi(t_i)$ is the time-variable characteristic model of person, car, road and environment. α^i is car's condition at t_i moment. $\alpha_v^i, \alpha_{av}^i, \alpha_{aH}^i, \alpha_\beta^i, \alpha_o^i, \alpha_p^i$ are car speed, longitudinal acceleration, lateral acceleration, front wheel corners, car conditions as well as positions describes respectively at t_i moment. r^i is road geometrical characteristic at t_i moment. r_H^i, r_V^i, r_C^i are road's horizontal, vertical and Cross-section geometric characteristics respectively at t_i moment. w^i is surrounding environment characteristic at t_i moment. w_E^i is environment characteristic at t_i moment. w_c expresses weather characteristic [4, 5].

With $Q_{i+1}(\Phi(t_{i+1}))$ described transport condition that is determined by environmental information at t_{i+1} moment:

$$Q_{i+1}(\Phi(t_{i+1})) = p_i(Q_i(\Phi(t_i))), \quad (2)$$

* ***Corresponding author*** e-mail: hushengneng2006@126.com

where p_i is an operator that has function to car, and non-correlated with road or environment, namely driver's operation to car. Type (2) also explained the transport condition $Q_{i+1}\left(\Phi\left(t_{i+1}\right)\right)$ that the process expected may achieves the travel goal, through adjusting process characteristic parameter α_i.

The above time-variable characteristic mode reveals, road simulation system has the following features:

1) Continued corresponding feedback character, the road simulation experiment advances continually along with extending of route. In this process, the driver model car model and road virtual environment model has the characteristic of persistent response.

2) Intelligent simulation characteristic, the road simulation system's pilot controls behavior is one kind of intelligent decision behavior, pilot model in the simulation system should have the auto-adapted characteristic, along with the change of person, car, road and environment time-variable condition, pilot model automatic adjust its control policy.

3) Hybridity, the road simulation system is promiscuous system that is unified of road and environment characteristic's discrete state, pilot manipulative separate behavior and car operational continual condition. The evolution of system is driven by the time (car movement) and event (pilot operation, road and environment state event) the two different mechanism, displays for the structure, algorithm hybridity.

2.2 MULTI-AGENT SYSTEM THEORY

Agent is a software entity that can realize the function under the specific environment continuously and spontaneously, and associates with the relevant agent and process. Continuous and spontaneous request originates from the change of environment, requests agent to real-time respond to user demand under nobody guides and interference by flexible, intelligent way. In addition, there is also hope that the agent could communicate and cooperate with other agent in the environment and the process [6].

The multi-agent system needs realizes with the aid of some correlation theories and technologies, mainly includes the knowledge expression and inference, agent communication language, multi-agent coordinates model, multi-agent to consult model:

1) Knowledge expression and inference. If a single agent in the system through a certain data structure and the corresponding operation to address the problem of modeling, must use the knowledge expression and inference method correlation theories.

2) Agent communication language. Communication and information exchange between agents need to a certain level of agreement as a basis, agent communication language can be used to accomplish this function.

3) Multi-agent coordinates model. In the multi-agent system category, reciprocity needs to be coordinated between agents. Coordination model is to provide a formal framework for the interaction among agent multi-agent coordination model can be divided into control actuation and data driven two types.

4) Multi-agent negotiation model. The multi-agent system is a system composed of a plurality of agent, which can solve the complex problem which single agent cannot solve, question solution network are formed by many agent coordinated cooperation. In multi-agent system, cooperation between agents is to be done through communication and consultation. Regarding multi-agent system model, has the different viewpoint, generally thinks that agent should include sensors, decision controller, mental state, knowledge base, communicator etc. several parts of compositions, Including the BDI model is a widely accepted model of agent negotiation model, it reveals the agent energetic factor including belief, desire, intention etc., as shown in Figure 1.

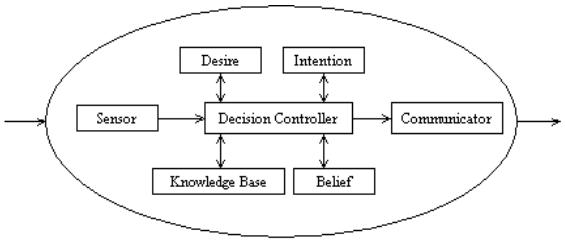


FIGURE 1 BDI model

2.3 DEFINITION OF ELEMENTAL AGENT

In order to facilitate the formal description of Agent, commonness of agent system can be abstracted, is defined as elemental agent. According to the characteristics of agent in the road simulation system and refers to the agent dynamics characteristic, defines agent as:

$$\begin{aligned} &Agent = \left\langle A \cdot id, T, I, O, S, IA, OA, R \right\rangle \\ &T \in R_0^+ \\ &I = EI \bigcup II \\ &\forall_i \in I^{\Rightarrow} i \cdot source \in IA \\ &\forall_o \in O^{\Rightarrow} o \cdot source \in OA \end{aligned} \quad ,$$

where $A \cdot id$ is identifier for the Agent. T is a logical clock. I, O are respectively for the input and output messages. EI is the external input news, II for internal input news. S is state set of elemental agent, the status in S is the limited dimensions real number vector. IA is all of the agent set that is possible to input message for this Agent. OA is all of the agent set that is possible to output message for this Agent. R is rule processing of elemental agent.

The message and the message flow in agent can be defined, according to the BNF form, the message is defined as:

*<system message>::=<message ID><message type>(:<parameters>)**

<*message ID*>::=<*digital*>*
<*message type*>::=<*identifier*>*
<*parameters*>::=<*parameter name*><*blank*>(:<*parameter value*>)*
<*parameter name*>::=<*identifier*>
<*parameter value*>::=<*identifier*>
<*identifier* >::=<*letter*>(<*letter*>|<*digutal*>|_)*
While the message flow can be defined as:
<*system message flow*>::=<*message flow aim*>
<*message flow source*>
<*send time*><*effective time*>
<*message format*><*system message*>

The status (*S*) of agent can be described by five vectors:

$$S = \langle A \cdot id, EI \cdot n, A \cdot e, EO \cdot n, t \rangle .$$

Type: $A \cdot id$ is identifier for the Agent; $EI \cdot n$ is name of input event for agent; $A \cdot e$ is activation state of agent; $EO \cdot n$ is name of output event for agent; t is the running time of agent after activation.

2.4 SIMULATION SYSTEM FRAME FOUND ON MULTI-AGENT

The above characteristic of road simulation system determines the software development method of traditional which is oriented to process oriented and object, is hard to process under the time-variable condition among the subsystems alternately, coordination, auto-adapted, self-feedback and other characteristics, simultaneously is also hard to solve the road simulation system hybridity problem. Statement of multi-agent system software development thought that to solve the software intelligence and cooperation has provided a new way.

According to person, car, road and environment simulation system's time-variable dynamic characteristic, the person, car, road and environment simulation system frame based on multi-agent as shown in Figure 2.

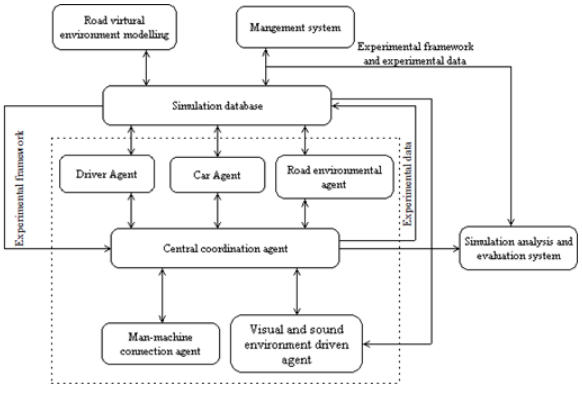


FIGURE 2 Person, car, road and environment simulation framework based on multi-agent

Here human, car, road and environment simulation system is composed of 6 interactions of agent, according to its duty and the difference of solution mechanism undertaking to duty, may be divided to respond agent, cooperation agent and interface agent 3 kinds. Driver agent, car agent and center coordinate agent to belong to cooperation agent, namely has internal behavior initiative agent, has knowledge representation, question solution expressed and so on. Road environmental agent and visual-audio driver agent belong to respond agent, but simply has the response to the exterior stimulation simply, is similar to customer/server architecture, agent is not only the customer, is the server, according to the procedure arrangement, makes to reply or send out the request. Man-machine connection agent belongs to interface Agent, can interactive with the user, carries on display and control [7].

In the road simulation system, various agent characteristics are different, its basic structure is also different, this article only describes the basic structure of central coordination agent, the center coordinates agent is at the core position in the multi-agent system, is responsible for message distribution, coordination, simulation clock coordination, its basic structure as shown in Figure 3.

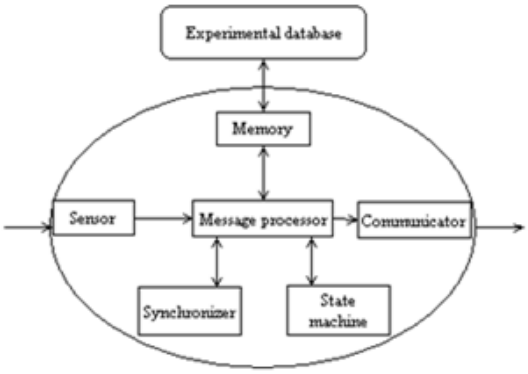


FIGURE 3 Center coordinates agent

Here the sensor is responsible for the receive of environmental information, distribute news through message processor to memory, synchronizer and state machine, the memory and experimental analysis, evaluation of relevant information store in the experiment database. The message processor has a message distribution function and agent internal state storage and adjustment function. The state machine memory system's time-variable environmental information, and distribute the previous time-variable environmental information by the message processor to the synchronize hand-in-hand processing row state renewal [8]. The synchronizer according to the definite rule, decided the message that in advance must distribute, uses the message processor and communicator distribution to other agent. For example, when central synchronizer receives driver agent operation information, the synchronizer compares the information with the information that state machine save, if there is no change, the information will no longer be distributed to the car agent, otherwise, then transmits [9].

3 Establishment of Road Simulation Model

3.1 CAR MULTI-BODY DYNAMICS SIMULATION MODEL

Traditional analyses of road alignment index is usually based on the classic car dynamic model, regards car as a particle, neglects various car part of relations, but does not tally with the reality. The car dynamics model established in this paper is based on multi-body dynamics system. In order to facilitate the modelling and the mathematical expression, simultaneously took into consideration the road simulation systems characteristic, has carried on the following simplification to the automotive system:

1) The car, frame and packing container as an object to be processed.

2) The front axle and rear axle of car may reciprocate to the car body and rotate around the longitudinal axis parallel to the axis of the car, around the shock absorber is simplified as a damping restraint, the frame and axles are connected by leaf springs and shock absorbers.

3) Steering system the steering wheel and column can be simplified to rotate objects around the car body, the steering unit is only considered that rotates around the body of the axle, the steering column and between steering units' axes with the universal joint.

4) The tire model uses the FIALA tire model.

3.2 DRIVER MODEL BASED ON PREVIEW THEORY

The car presents time-variable dynamic road environment to the driver in the process of travel, and the driving behavior the effect on the car displays for the driver to take operation behavior after the environment sensation. The basic principle of the driver control car is to cause car's movement as far as possible consistent with expected trajectory and speed. Therefore, according to person's behavior stimulation- organism - response classic mode, driving behavior can be divided into 3 stages, namely information sensation, judgment decision-making and operation adjustment. As shown in Figure 4:

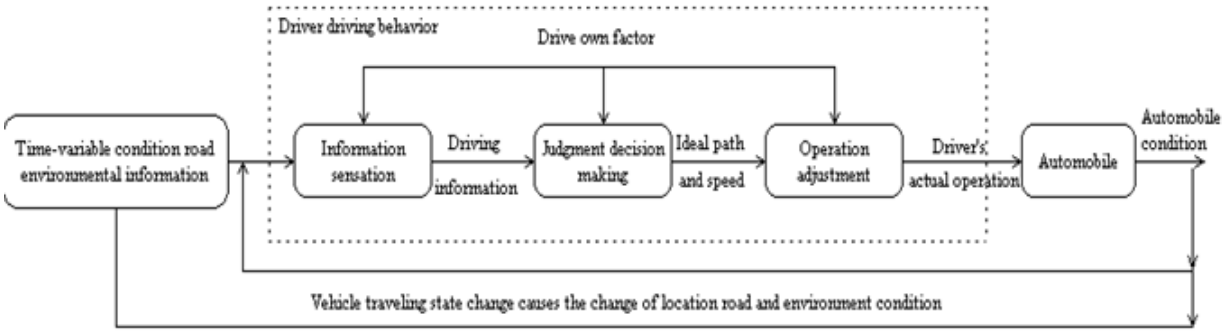


FIGURE 4 Driver operation behaviour flow char

3.3 ROAD SCENE MODEL BASED ON CONSTRAINED DELAUNAY TRIANGULATION

The request of road ground view model is expressed by the different resolution through constraint dlaunay triangulation (CDT), modelling plan of this system is used for road simulation model based on CDT, and the modelling process is described as follows:

1) Carries on the pretreatment to the topographical data, first carries on the data filters, the rejection wrong spot and closely coincide, next data grid, to enhance data retrieval speed.

2) Using the point by point insertion algorithm, establishes the basic Delaney triangulation.

3) Using "wears hat" principle of road cross-sectional design, obtains the intersection of road side slope and triangular net, forms the closed polygon.

4) Using the CDT algorithm, the closed polygon according to the line segment is inserted into the basic triangle net, forms the constrained Delaney triangulation.

5) Using triangular net virus algorithm, rejection closed polygon interior triangular net.

6) Using point by point insertion algorithm, builds the road three-dimensional model, the road three-dimensional odel and topographical model superimposition, forms ground view simulation model - line frame model.

7) Based on edge collapse algorithm,

Multi-level detail model of ground view simulation model has been built.

8) Enduing the model material and texture, finally builds the ground view simulation model.

4 Simulation Model Confirmations

Simulation model is set up for simulation experiments, to reveal the behaviour characteristics of the original system by simulation experiment. However, it is not possible to confirm whether a model is effective well at a modelling. When the system is quite huge and complex, in order not to make the cost of model establishing and running is too high, the complexity of the model is limited, has to make some assumptions or simplified, so it is necessary to verify the effectiveness of the model. Key field tests were conducted on vehicle dynamics. In model validation, using the method based on experimental data, in the same boundary condition and scene consistent situation, compared with actual experimental data simulation experiment data, the model was modified according to the comparison result.

4.1 EXPERIMENTAL DESIGN

Experimental road section is Dian Liang – Cai Guan Ling section of Bao Ji – Han Zhong road in China's Shanxi province. This road is second-class road of mountainous hilly area, design speed 40 km/h, asphalt pavement width 7 m, roadbed width 8.5 m. Region's altitude of the route is at 1 200m~1 350m, by the topographical constraint, this road section consecutively 5 km downhill, plane indices except sight distance of individual curves cannot meet the requirements, other indices conform to the standard stipulation.

Experiment vehicles and instrument: Test vehicle uses YaXin JS6820C32D1, maximum speed 110 km/h, carrying capacity 5 t; the car speed uses the fifth wheel instrument real-time collection and record.

4.2 SIMULATION EXPERIMENTAL PLAN

According to the uphill and downhill two cases are considered, in the downhill experiment the car is pressed 30,40,50 km/h to hang 3 grades of glides by the top of slope to the base of slope, in the glide process, loosens the accelerator pedal. In the uphill experiment the car is pressed 40, 50, 60 km/h to hang 3 grade and accelerator full hill climbing to go by the base of slope. In the simulation, road model is established according to the actual road, the car simulation model is built, which is consistent with 5 t experimental cars.

4.3 RESULT ANALYSIS

Figure 5 is the field test data and model test data of car with 50 km/h from top hanging 3 archives to glide. The test section longitudinal slope of 7%, there are two horizontal curve, radius is larger than 200m, In the simulation model, friction coefficient between tire and road surface takes 0.65, indicates that the simulated data and empirical datum have certain error, The error is mainly from the following two aspects: First, in simulation testing the road surface friction coefficient value and actual situation have the difference. This experimental road section puts into use for many years later, the different sector road surface attrition and destruction condition are different, cause the road surface friction coefficient to be different; Second, the local transportation department responsible for the work to prevent the vehicles exceeds the speed limit, has established the caution belt in the steep slope road section, has certain influence to the field test data. Even so, Figure 5 shows the simulation results and experimental data are in good agreement and the error is in the allowable range, so test data of 40 km/h and 30 km/h. therefore the simulation model is effective. In the later design proposal appraisal experiment, only need change parameter of path model, but does not transform the car model, may carry on to analyse and study to various parameters in road route design.

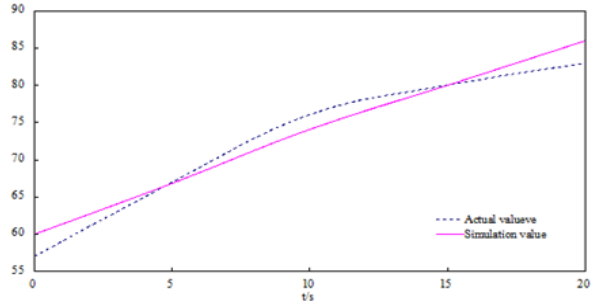


FIGURE 5 Actual data and simulated data contrast

5 Conclusions

The road simulation system has been established to evaluate the road design, on the basis of the analysis simulation characteristic, development and framework of road simulation system have been proposed, agents of road simulation have been defined, the center coordinates agent and driver agent have been discussed, and this system is confirmed through the project example, an important guiding significance for the development of road simulation system has been provided.

Acknowledgments

The research project was funded by a grant from the Ministry of Transport of China. The authors gratefully thank my professional brothers at Chang'an University in Xi'an. Finally, we would like to express our gratitude to the study subjects who uncomplainingly underwent the rather demanding trials.

References

[1] Chen L J, Wu F 2002 Modeling of urban traffic system simulation based on hybrid Petri net *Journal of Tong Ji University* **12** 1478-82 *(in Chinese)*
[2] Yang H Z, Xu J, Li J 2005 Evaluation of road route design based on computer simulation *China Journal of Road and Transport* **1** 14-7 *(in Chinese)*
[3] Zou Z J, Yang D 2001 Software design of urban road traffic simulation system *Journal of Traffic and Transportation Engineering* **3** 86-9 *(in Chinese)*
[4] Wei L, Zhou W, Li C, et al 2005 Driver perception model for road structure parameters *Journal of Traffic and Transportation Engineering* **4** 116-20 *(in Chinese)*
[5] Wang X J, Wen C Y 1997 One kinds of mix dynamic systems' simulation methods *Journal of System Simulation* **9** 51-5
[6] Fan Y S, Cao J W 2002 Multi-agent system theory, method and application *Qinghua University Press* Beijing 1-25 *(in Chinese)*
[7] Li H L 2001 Agent-based complicated system distributed simulation *National University of Defense Technology [Dissertation]* Changsha China *(in Chinese)*
[8] Liu L H, Song Z H 2001 Based on multi-agent to process promiscuous system's integration technology *Systems Engineering Theory and Practice* **9** 62-6 *(in Chinese)*
[9] Wang X 2000 The research of agent system communication theory and organizational structure *Chinese Academy of Science institute of software [Dissertation]* Beijing *(in Chinese)*

Authors

Shengneng Hu, born in October, 1979, Xinyang, China

Current position, grades: lecturer, North China University of Water Resources and Electric Power.
University studies: PhD degree in Highway & Railway Engineering from Chang'an University in 2011.
Scientific interest: road alignment design technology, traffic safety, road landscape planning.
Publications: 18.

Xiaoming Lu, born in October, 1986, Zhengzhou, China

Current position, grades: lecturer, University for Science & Technology Zhengzhou.
University studies: master's degree in Henan Polytechnic University.
Scientific interest: green building construction.
Publications: 6.

Han Juan, born in June, 1980, China

Current position, grades: lecturer, University for Science & Technology Zhengzhou.
University studies: master's degree of engineering of road & railway engineering from Changsha University.
Scientific interests: teaching and research of geotechnical engineering.
Publications: 15.

A multivariate analysis-based for range-free localization algorithm

Hao Yan[1], Xiaoyong Yan[2*], Rulin Dou[1]

[1]*School of Software Engineering, Jinling Institute of Technology, Nanjing 211169, China*

[2]*School of Intelligence Science and Control, Jinling Institute of Technology, Nanjing 211169, China*

Abstract

Proposed an improved DV-Hop localization algorithm (PLS-DVHop) based on partial least squares, which uses the partial least squares to model of hop-count and the Euclidean distances, along with the maximum covariance of input matrix and output matrix to estimate the location of unknown nodes. PLS-DVHop has strong adaptability for different deployment network, and overcomes the shortage of only suitable for isotropic networks in the original algorithm. Simulation results show that PLS-DVHop algorithm has high estimate precision and stable performance, can adapt to different network topologies, and is very suitable for large scale deployment network.

Keywords: range-free localization, wireless sensor network, partial least squares

1 Introduction

Wireless sensor network (WSN) [1, 2] is constituted in ways of self-organization and multi-hop by large volumes of sensor nodes with communication and computation capability. Nodes in the network are able to collaboratively perceive, collect, process, and transmit the information of perceived objects within the coverage area of the network, as well as to report the information to users. WSN has great potential application value in military, transportation, medical care, and environment monitoring [1]. Location estimation is a key issue for WSN [3, 4]. Different from traditional networks, WSN is a data-based network. Thus, in WSN localization researches, statistics and multivariate analysis methods are often applied for quantitative analysis.

In the application of WSN, nodes' location information can be acquired by adding global position system (GPS)/BeiDou Navigation Satellite System (BDS) devices on nodes. However, this way is only applicable with outdoor. Besides, GPS/BDS device is large in volume, and high in cost and energy consumption. Moreover, GPS/BDS device also needs stable base installations. These facts have made it difficult to realize the requirements of WSN, which is "low price, low cost and low energy consumption" [5]. As for this, in practice, only some of the nodes can be installed with GPS/BDS device. For the rest nodes, their location information can only be estimated via a certain algorithm or method. After several years' development, researchers have proposed many node localization approaches. According to whether the range information is used among the localization process, the localization techniques can be classified into range-based and range-free [3, 4, 6]. The range-based method exploits distance or angle information between neighbour nodes, and then uses the information to localize nodes. The range-based localization has higher location accuracy but requires additional hardware support and thus, is very expensive to be used in large scale sensor network. The range-free localization is being considered as a cost-effective alternative to range-based methods because of hardware limitation in large scale deployment. On the other hand, range-free schemes do not need additional hardware support and makes use of connectivity, multi-hop routing and other information between nodes to estimate nodes location. Therefore, range-free technique is considered to be most effective solution for the localization issues in WSN.

The DV-Hop localization algorithm proposed by Dragos Niculescuet et al. [7, 8] from Rutgers University is one of a series of distributed localization algorithms, it is a localization algorithm not related to the distance, it smartly uses the distance vector routing and the idea of GPS localization, and this algorithm has great distributive and expandability. DV-Hop method is an ideology based on distance vector routing and GPS, which makes use of hop distance to replace real distance between nodes. Eventually, least squares are applied to estimate the position. DV-hop algorithm assumes that the network is isotropic and uniformly distributed, that is, when the properties of the graph are the same in all directions, so that the corrections that are deployed reasonably estimate the distances between hops. Unfortunately, in practice, networks may be anisotropic and may contain com plex inner or outer boundaries, which make the least hop counts deviating the Euclidean distances. This paper integrates PLS in multivariate analysis, making use of the correlation between hop-counts and Euclidean distances between known nodes to establish an optimal linear conversion matrix. On this basis, the matrix is used to convert hop-

***Corresponding author** e-mail: xiaoyong_yan@jit.edu.cn

counts between unknown nodes and known nodes into distance, so as to realize compensation on distance estimation in networks with unevenly distributed nodes, reaching the effect of high localization accuracy.

2 Related works

In recent years, it has become a new research hot spot to make use of multivariate analysis technique in modelling and algorithm design of localization mechanism [9, 10]. The method puts to use the relation between distribution feature and measurement information of known nodes to establish a mapping function. On this basis, the function will be used to estimate the location of unknown nodes. Compared with previous methods, multivariate analysis is able to effectively discover network topology, correlation and other information hidden behind data. Lim et al. [11] proposed a PDM (Proximity Distance Map) algorithm based on TSVD (Truncated Singular Value Decomposition) technique. PDM describes the optimal linear transformations between the hop-count and the Euclidean distances under the least-squares metric. With the help of PDM, an unknown node is able to obtain more accurate distance translation, thus to get a better location estimation. Firstly, the PDM method uses matrices to express the collected Euclidean distances and the hop-count between known nodes; secondly, TSVD technique is used to conduct linear transformation of two matrices to obtain an optimum linear transformation model; lastly, the hop-counts from the unknown nodes to the known nodes will be applied to this model to estimate the Euclidean distances between the unknown nodes and the known node. In essence, TSVD [12] is a multivariable linear regularization learning method, the estimated Euclidean distances obtained through method is actually the weighted sum of the estimated values of other known nodes in the monitoring area, and therefore, the obtained estimated value is close to the actual value. In addition, the TSVD method has abandoned the small singular values, which can to a certain extent reduce the impact of noise during the transformation process, so the collinearity problem during the localization process can be avoided, and the stability of algorithm can be increased. All these have caused the algorithm to have a low requirement for the deployment of sensor nodes, connection and signal attenuation method, which more benefits its use in complex application environments. In a certain degree, TSVD can solve some problems of range-free method, but the literature and experiment show that the PDM method only works under certain conditions, and when the beacon nodes are sparse or various radio ranges have serious anisotropy, the performance of TSVD method will sharply decrease. The main drawback of PDM is that it need to set a threshold parameter k. TSVD technique directly sets the singular values smaller than the threshold parameter k as zero, and if k is properly chosen, the solution of TSVD is stable, otherwise, it will reduce the algorithm's performance. Moreover, the PDM method has not conducted standard processing to the hop-counts and Euclidean distances, and different dimensions have caused a certain degree of data submergence. In addition, TSVD modelling only takes into consideration hop-counts information, disregarding Euclidean distance information. As for this, the model built is unable to truly reflect the relationship between hop-counts and real distance. Inspired by PDM method, Lee et al. [13, 14] put forward SVR-based localization method – LSVR (Localization through Support Vector Regression). The localization method is fit for different networking environment. Moreover, under small sample condition, it still leads to good positioning accuracy. However, LSVR is a multi-input and single-output algorithm [15]. In practical localization practices, modelling is to be performed frequently, sharply increasing the complexity of the algorithm. Moreover, with the number of beacons increases, time and space resource required by positioning will grow geometrically.

In order to reduce the complexity of localization problem, to improve the generalization performance of positioning method, and to simplify localization model, it is quite necessary to perform feature extraction before model building. Researchers have found that, PLS (Partial Least Squares) [16, 17] is able to perfectly realize feature extraction from input to output. PLS is a standard multivariate regression method, which is to form components that capture most of the information in the explanatory variables that is useful for predicting dependent variables, while reducing the dimensionality of the regression problem by using fewer components than the number of explanatory variables. PLS technique is considered especially useful for constructing prediction equations when there are many explanatory variables, comparatively little sample data and the collinearity between independent variables. It also has strong anti-noise property and great generalization ability, it does not require obtaining the distribution model of the sample in advance, and it also has various characteristics such as a high predication precision, so it is also called the second-generation regression method. Inspired by the PLS and based on DV-Hop localization method, the paper makes use of PLS technique in multivariate regression to optimize DV-Hop algorithm, and proposes a PLS-based DV-Hop localization method (PLS-DVHop).

3 Brief Reviews of DV-Hop, PLS

Before the introduction of our algorithm, for completeness, we will briefly review DV-Hop and PLS in the next sub-sections.

3.1 DV-HOP

DV-Hop method does not require any hardware to measure ranges or angles to neighbours. It only relies on the connectivity of the underlying graph and it comprises three no overlapping stages:

1) First, each node estimates the least hop-counts to each beacon and maintains a table $[x_i, y_i, h_i]$ and exchanges updates only with its neighbours. Where $[x_i, y_i]^T$ denotes the physical location of beacon i, h_i is a counter to record the hop-counts to beacon. This phase is the classical Bellman-Ford distributed shortest path algorithm.

2) In the second stage, beacons cooperatively estimate the average distance of each hop in the network. Once a beacon j gets the hop count h_i to beacon i, it reports the value of h_i to beacon i. After collecting these values from all other beacons, beacon i (locating at $[x_i, y_i]^T$) calculates the average distance of each hop in the network, uses it as an adjusted value and broadcasts it to the network. The average distance of each hop can be expressed by the following equation:

$$HopSize_i = \frac{\sum_{i \neq j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{i \neq j} h_i}. \quad (1)$$

3) After receiving the correction, an arbitrary node may estimate distances to beacons. Suppose an unknown node receives the messages from three beacons, i.e. beacon i, j, and k. It uses the three distance estimates ($HopSize_i \times h_i$, $HopSize_j \times h_j$, and $HopSize_k \times h_k$) to determine its location by trilateration or maximum likelihood method.

3.2 PLS

Before regression, PLS makes use of covariance to guide feature extraction of input variable and output variable. In the process of extraction, information integration and screening technology is applied, so that PLS method overcomes correlation of input variable, eliminating the influence of co-linearity on regression. Moreover, in the process of regression, interpretation and prediction role of input on output is emphasized, which eliminates noise unfavourable to regression. As for this, PLS method leads to good robustness and prediction stability. PLS method always converts multivariate regression problem into several simple regression problems, so that it is also fit for small sample. Owing to such favourable natures, PLS is quite fit for WSN localization.

Linear PLS method comprehensively considers input variable **H** and output variable **D**, which solves component **t**, **u** (**t** is linear combination of variable **H**, and **u** is linear combination of variable **D**). The principle is shown in Figure 1:

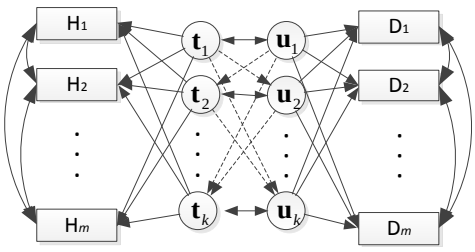


FIGURE 1 The schematic diagram of PLS

It may be seen from Figure 1 that, cross covariance between **H** and **D** is important information. Such cross covariance is hoped to be described through k. Thus, solution of PLS shall meet:

1) **t** and **u** shall try to carry variation information of their respective data sheet as much as possible.

2) The correlation between **t** and **u** may be maximized. PLS is described as solution of optimization problem:

$$\max \left\| \mathrm{cov}(\mathbf{t}, \mathbf{u}) \right\|^2 = \max[\mathbf{Hw}, \mathbf{Dc}], \quad (2)$$

where, **w**, **c** is weight vector.

Basic steps of PLS algorithm is shown below:

Firstly, input and output variables are to be standardized, so as to eliminate influence on final calculation result caused by inconsistent dimension of input and output variable;

The first principle component $\mathbf{t}_1$ and $\mathbf{u}_1$ are separately extracted from input and output variable H, D. In accordance with the demand of regression, the below conditions shall be satisfied:

- $\mathbf{t}_1$ and $\mathbf{u}_1$, shall try to carry variation information of variable **H**, **D** as much as possible;
- Correlation between $\mathbf{t}_1$ and $\mathbf{u}_1$ shall be able to be maximized.

After extraction of the first component $\mathbf{t}_1$ and $\mathbf{u}_1$, regression of **H**, **D** on $\mathbf{t}_1$ shall be separately performed. If the regression equation has reached satisfactory precision, PLS method will be terminated. Or else, residual of **H** after being interpreted by $\mathbf{t}_1$, as well as residual of **D** after being interpreted by $\mathbf{u}_1$ will be utilized to perform the second round of component extraction. Repeat the process, until a satisfactory precision is obtained.

4 The establishment of PLS-DVHop localization model

The localization process of PLS-DVHop work in two phases: offline training phase and an online localization phase. More specifically, in the offline training phase, we take two steps for model building. In the first step, we collect hop-counts and Euclidean distances between beacons as the training set. In the second step, we make use of PLS technique to obtain the mapping model between hop-counts and Euclidean distances. In the online localization phase, the real-time hop-counts samples received from the beacons by the unknown node, and then the unknown node uses the mapping model obtained through the training to conduct location estimation.

Consider a WSN which is comprised of n sensor nodes $\{S_i\}_{i=1}^n$ deployed in a 2D geographic region. Without loss of generality, let the first m nodes be beacons whose locations are known, for all $i = 1,\ldots,m$, where $m \ll n$. For every pair of sensors S_i and S_j, h_{ij} denotes shortest hop-count and d_{ij} denotes Euclidean distances that sensor S_i receives from sensor S_j. After running for a period of time, we can obtain two matrices, *i.e.*, the shortest hop-counts matrix $\mathbf{H}=[\mathbf{h}_1,\ldots,\mathbf{h}_m]$ and the Euclidean distances matrix $\mathbf{D}=[\mathbf{d}_1,\ldots,\mathbf{d}_m]$, where $\mathbf{h}_i=[h_{i1},\ldots,h_{im}]^T$, $\mathbf{d}_i=[d_{i1},\ldots,d_{im}]^T$.

Considering the relationship Euclidean distances and hop-counts, and according to the multiple regression theory. We can obtain an equation, which can be expressed as:

$$\mathbf{D} = \mathbf{H}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{3}$$

where, $\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_m)^T$ is the regression coefficient vector, $\boldsymbol{\varepsilon}$ is the errors vector.

In order to minimize the error, as well as for the convenience of computation, we often use quadratic sum of error as the judgment standard. In order to figure out the optimal relation between the hop-counts and Euclidean distances, we are to figure out the partial derivative of the quadratic sum of error, and assuming that it is 0, hereby:

$$\mathbf{H}^T\mathbf{H}\hat{\boldsymbol{\eta}} = \mathbf{H}^T\mathbf{D}. \tag{4}$$

It may be seen from Equation (2) that, there may be as well multiple correlations between variables in **H**, or the number of samples in **H** may be smaller than the number of variables. If so, forced calculation of Equation (2) may leads to invalid result. In addition, the precision of estimation value $\hat{\boldsymbol{\eta}}$ is not only related with input variable, but also related with output variable **D**. Input and target co-determines the prediction direction of $\hat{\boldsymbol{\eta}}$.

Algorithm1: PLS-DVHop Localization Algorithm

Input $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_m]$: hop-counts matrix of beacons; $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_m]$: Euclidean distances matrix of beacons; k : the number of principal element; $\{\mathbf{c}_i = (x_i, y_i)\}_{i=1}^{m}$: the location of beacons

output $\{\hat{\mathbf{c}}_i = (x_i, y_i)\}_{i=m+1}^{n}$: estimated location of the non-beacons

1 Centring matrix **H** and matrix **D**

2

for $j = 1, \cdots, k$

$\mathbf{u}_j$ = first column of $\mathbf{H}_j^T\mathbf{D}$

$\mathbf{u}_j = \mathbf{u}_j / \|\mathbf{u}_j\|$

repeat

$\mathbf{u}_j = \mathbf{H}_j^T\mathbf{D}\mathbf{D}^T\mathbf{H}_j\mathbf{u}_j$

$\mathbf{u}_j = \mathbf{u}_j / \|\mathbf{u}_j\|$

until convergence

$$\mathbf{w}_j = \frac{\mathbf{H}_j^T\mathbf{H}_j\mathbf{u}_j}{\mathbf{u}_j^T\mathbf{H}_j^T\mathbf{H}_j\mathbf{u}_j}$$

$$\mathbf{c}_j = \frac{\mathbf{D}_j^T\mathbf{H}_j\mathbf{u}_j}{\mathbf{u}_j^T\mathbf{H}_j^T\mathbf{H}_j\mathbf{u}_j}$$

$\hat{\mathbf{D}} = \hat{\mathbf{D}} + \mathbf{H}_j\mathbf{u}_j\mathbf{c}_j^T$

$\mathbf{H}_{j+1} = \mathbf{H}_j(\mathbf{I} - \mathbf{u}_j\mathbf{w}_j^T)$

end

$\hat{\boldsymbol{\eta}} = \mathbf{U}(\mathbf{W}^T\mathbf{U})^{-1}\mathbf{C}^T$

3 Putting the obtained estimation term $\hat{\boldsymbol{\eta}}$ into the original equation to figure out the prediction model; putting the hop-counts from unknown nodes to known nodes into the equation to further figure out corresponding estimation distance. At the moment, integrating coordinates and estimation distance of known nodes, least squares may be used to estimate unknown nodes, so as to obtain the estimation coordinates.

PLS is a multivariate data analysis method integrates multivariate regression, canonical correlation analysis and principle component analysis altogether. It makes use of covariance of input and output to guide feature selection, which perfect fits for modelling and prediction of real situation. Being applied in DV-Hop localization method, PLS is fit for different localization situation, and is able to obtain high localization accuracy. At the moment, hop-count and Euclidean distances modelling process based on PLS may be concluded as Algorithm 1.

5 Performance evaluations

One of the important features of range-free localization method is that, it is quite fit for large-scale deployment. This requires thousands of sensor nodes, while in labs; it is difficult to realize such large-scale real network. As for this, in researches on large-scale range-free node localization algorithm, software simulation is often applied to estimate the advantages and disadvantages of localization algorithm.

In this section, the performance of PLS-DVHop algorithm is to be analysed and assessed via simulation experiment. Matlab2013b software is employed to analyse and compare methods proposed in this paper. In the experiment, all nodes are evenly distributed in two-dimensional space. In order to reduce the one-sidedness of single experiment, each deployment goes through 50 simulations while nodes in each experiment are randomly re-distributed in the experiment area. Mean value of 50 RMS (Root Mean Squares) [18] is taken as the assessment basis.

In order to assess the performance of methods proposed in this paper, nodes are assumed to be randomly or regularly deployed in the monitoring area. In addition, in order to evaluate the adaptability of the proposed methods to network topology anisotropy, obstruction is added in the aforementioned two deployment strategies, i.e. assuming that there is a large obstruction in the deployment area, impeding the direct communication between nodes. Such area is of C-Shape. In allusion to different network topology structure, nodes are re-deployed in the same area for several times, assessing the average localization error. This experiment also compare our method with three previous methods:

1) The classic DV-Hop method proposed in [7];

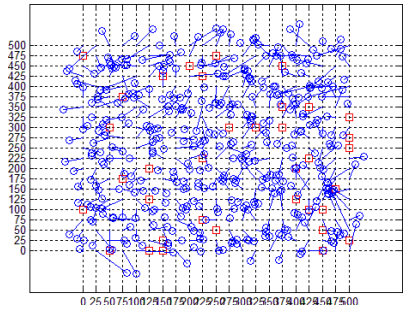
2) PDM proposed in [11];

3) LSVR proposed in [13] in two group experiments.

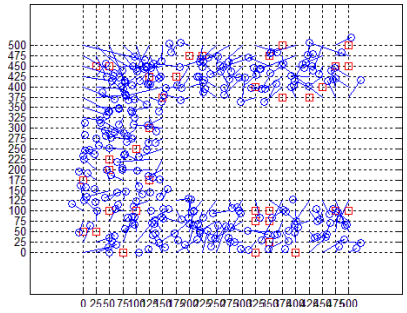
Furthermore, for fairness, in PDM localization, we denoted abandoning threshold in TSVD as 2, i.e. abandoning eigenvectors with eigenvalue less than or equal to 2. There is certain relationship between kernel parameter σ and the distance between training samples. In the experiment, we assume σ as 50 times of the average distance between sample nodes. Configuration of C and ε in SVR method uses for reference related reference [19], while C is also configured based on σ according to related reference [20]

5.1 REGULARLY DEPLOYED SENSORS

In this group of experiments, there are 441 and 315 nodes deployed in the area of square or C-Shape region. The number of beacons was gradually increased from 30 to 50, with a step size of 2.before analysing the performance of PLS-DVHop algorithm, let us first investigate the two final localization results (in figure the number of beacons is 36). In Figure 2, the circles denote the unknown node and the squares are the beacon node, the line connects the actual coordinate and estimated coordinate of unknown nodes, and the longer the line, the more the estimated value deviates from the actual location.



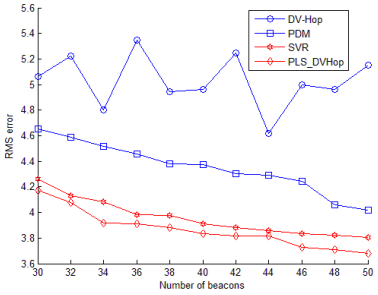
a) Square Regular Distribution



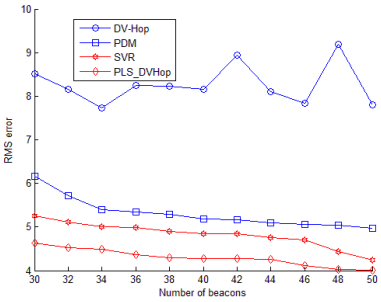
b) C-shaped Regular Distribution

FIGURE 2 Localization Results Under Regular Distribution

Figures 3a and 3b describes the influence of beacons quantity on localization accuracy in regular deployment network. Theoretically, more beacons in monitoring area leads to smaller localization error. According to the result shown in Figures 3a and 3b, RMS error of DV-Hop is the largest, and the curve fluctuates up and down. RMS error of the other three methods simply decreases with the number of beacons grows. PLS-DVHop algorithm proposed in this paper takes on highest localization accuracy. In addition, RMS error of PLS-DVHop and LSVR method are similar in the two deployment area. By contrast, the rest methods take on significant difference. This also shows that PLS-DVHop and LSVR methods have high environment adaptability, and are fit for anisotropic networks. As PLS-DVHop is multivariate regression method, in calculation process, it considers correlation between nodes. Thus, its localization accuracy is the highest. It may as well be seen from Figure 3 that, PLS-DVHop monotonically decreases with the number of beacon increases. Yet, such tendency is weak. This also indicates that, the number of beacons has quite slight influence on precision of the algorithm. PLS-DVHop method is also fit for environment with fewer beacons.



a) RMS error under Square Regular Distribution

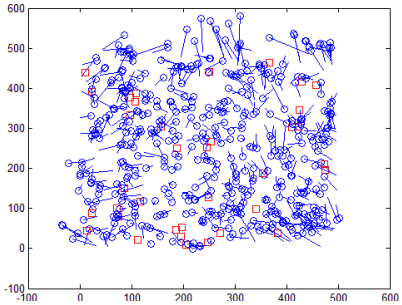


(b) RMS error under C-Shape Regular Distribution

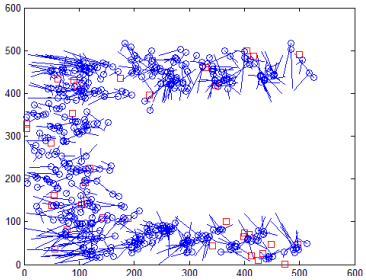
FIGURE 3 RMS error Under Regular Distribution

5.2 RANDOM DEPLOYED SENSORS

In this group of experiments, 500 nodes were randomly deployed in a 500×500 two-dimensional square area, and 30 to 50 nodes were chosen from the 500 as the beacon nodes. Like the regular deployment, in order to investigate the impact of non-line-of-sight on the localization algorithm, the experiment scenario with obstacle was added to the random deployment experiment. Similarly, the two final localization results were analysed first. As shown in Figure 4, in these two experiments, the number of beacon nodes is still 36.



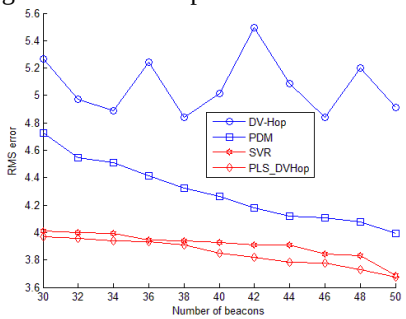
a) Square Random Distribution



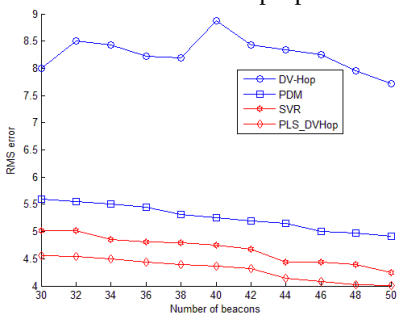
b) C-Shape Random Distribution

FIGURE 4 Localization Results Under Random Distribution

Figures 5a and 5b describes the relation between beacon quantity and RMS error in random deployment network. Similar to regular deployment, RMS error of DV-Hop is the highest, and the curve fluctuates up and down. Moreover, in C-shaped region, the localization error is larger than that in square region. This proves that, DV-Hop method is instable, and is sensitive to anisotropy of network topology. PDM method eliminates some data of small eigenvalue. As the elimination is set manually, disregarding factors of hop count and real distance dimension, its localization accuracy is only slightly improved than DV-Hop method. The localization accuracy of LSVR method is quite close to the method proposed in this paper. Yet, LSVR is a multi-input and single-output method, and is powerless when establishing inter-matrix relation like hop-counts and real distances. LSVR method has to perform frequent modelling. Although being optimized, it still does not considered correlation between data. Thus, the localization accuracy of SVR-based method is lower than the method proposed in this paper.



a) RMS error under Square random Distribution



b) RMS error under C-Shape random Distribution

FIGURE 5 RMS error Under random Distribution

6 Conclusions

In this paper, a range-free localization method (PLS-DVHop) is proposed. The method establishes the relation between hop-counts and Euclidean distances, so as to build measurement distance (hop-counts) and Euclidean distance model, and to effectively solve complicated topological structure problem in WSN. Compared with other similar methods, PLS-DVHop method inherits the advantages of DV-Hop method. Shown by simulation test, under different deployment environment, PLS-DVHop algorithm is able to present high positioning accuracy, and is only slightly affected by the number of beacon.

Acknowledgements

The paper is sponsored by Provincial University Natural Science Research Foundation of Jiangsu Education Department (12KJD510006, 13KJD520004); Natural Science Foundation of Jiangsu(BK2012082); Doctoral Scientific Research Startup Foundation of Jinling Institute of Technology (JIT-B-201411); Educational Reform Foundation of Jinling Institute of Technology(2010JGXM-02-2).

References

[1] Emary I M M E, Ramakrishnan S 2013 Wireless Sensor Networks: From Theory to Applications CRC Press
[2] Khan S, Pathan A-S K, Alrajeh N A 2013 Wireless Sensor Networks Current Status and Future Trends CRC Press
[3] Liu Y, Yang Z 2011 Location, Localization, and Localizability Location-awareness Technology for Wireless Networks *Springer*
[4] Mao G, Fidan B 2009 Localization Algorithms and Strategies for Wireless Sensor Networks: Monitoring and Surveillance Techniques for Target Tracking *Information Science Reference*
[5] Zheng J, Jamalipour A 2009 Wireless Sensor Networks: A Networking Perspective *A John & Sons Inc*
[6] Zekavat R, Buehrer R M. 2011 Handbook of Position Location: Theory, Practice and Advances *Wiley-IEEE Press*
[7] Niculescu D, Nath B D V 2003 Based Positioning in Ad Hoc Networks *Telecommunication Systems* **22**(1-4) 267-80
[8] Wu G, Wang S, Wang B, Dong Y, Yan S 2012 A novel range-free localization based on regulated neighborhood distance for wireless ad hoc and sensor networks *Computer Networks* **56** 3581-93
[9] Gu J, Chen S, Sun T 2011 *IEEE Transactions on Wireless Communications* **10**(9) 2841-9
[10] Pan J J, Pan S J, Yin J, Ni L M 2012 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(3) 587-600
[11] Lim H, Hou J C 2006 Distributed localization for anisotropic sensor networks ACM Transactions on Sensor Networks (TOSN) **5**(2) 1-26
[12] Hogben L 2007 Handbook of Linear Algebra *CRC Press*
[13] Lee J, Chung W, Kim E 2013 A new kernelized approach to wireless sensor network localization *Information Sciences* **243** 20-38
[14] Lee J, Choi B, Kim E 2013 *IEEE Transactions on Neural Networks and Learning Systems* **24**(7) 1099-113
[15] Xu S, An X, Qiao X, Zhu L, Li L 2011 Multi-output least-squares support vector regression machines *Pattern Recognition Letters* **34**(9) 1078-84
[16] Li B, He Y, Guo F, Zuo L 2013 *IEEE Transactions on Instrumentation and Measurement* **62**(2) 304-14
[17] Geladi P, Kowalski B R 1986 Partial least-squares regression: a tutorial *Analytica Chimica Acta* **185** 1-17
[18] Chen J, Wang C, Sun Y, Shen X 2011 Semi-supervised Laplacian regularized least squares algorithm for localization in wireless sensor networks *Computer Networks* **55**(10) 2481-91
[19] Cherkassky V, Ma Y 2004 Practical selection of SVM parameters and noise estimation for SVM regression *Neural Networks* **17** 113-26
[20] Keerthi S S, Lin C-J 2003 Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel *Neural computation* **15**(7) 1667-89

Authors

Hao Yan, born in October, 1980, Jiangsu, China

Current position, grades: PhD student at Hohai University since 2013. Lecturer at the School of Information Technology, Jinling Institute of Technology.
University studies: Bachelor and Master Degree at Nanjing University of Science and Technology in 2003 and 2006 respectively.
Scientific interests: information security and the technology of the Internet of things.
Publications: about 12 papers.

Xiaoyong Yan, born in February, 1977, Jiangsu, China

Current position, grades: lecture at the School of Intelligence Science and Control, Jinling Institute of Technology, China. Postdoctoral Fellow at School of Instrument Science and Engineering, Southeast University, China.
University studies: MS and PhD Degrees in computer application technology at Nanjing University of Science and Technology (NUST) in 2006 and 2013, respectively.
Scientific interests: wireless sensor networks, pattern recognition and machine learning.

Rulin Dou, born in October, 1979, Jiangsu, China

Current position, grades: technician at the School of Information Technology, Jinling Insitute of Technology, China.
University studies: MS Degree in computer application technology from Nanjing University of Science and Technology (NUST) in 2011.
Scientific interests: wireless sensor networks, machine learning, pattern recognition and image processing.

The dynamic characteristic analysis of spindle based on ANSYS

Hualong Xie[1*], Xiaopeng Mei[1], Yongxian Liu[1]

School of Mechanical Engineering & Automation, Northeastern University, Shenyang, Liaoning, 110819, P.R. China

Received 12 September 2014, www.cmnt.lv

Abstract

The basic concept of dynamic characteristics of spindle is introduced. The first ten order vibration mode and natural frequency of shaft are obtained by modal analysis and harmonic response analysis using software ANSYS. The deformation trend of spindle under resonance state is analysed and the contrast between theory and simulation critical speed is done. The frequency response curve of spindle is given. The first order, second order dynamic stiffness and dynamic stiffness under general conditions are calculated. The influence of bearing damping to spindle dynamics characteristics is discussed. In the end, the measures of improving spindle performance are summarized.

Keywords: Shaft, Dynamic characteristic, ANSYS, modal analysis, harmonic response analysis

1 Introduction

In actual manufacturing, cutting force and other exciting forces are applied in the form of dynamic load onto the structure. In order to ensure the manufacturing accuracy and efficiency of work piece, not only the spindle system of machine tool should have some static rigidity, but also the dynamic property of spindle structure should be considered. The dynamic property of spindle structure refers to its resistance to forced and self-excited vibration. Notably in high-speed precise machine tool, the dynamic property of spindle usually has great influence on the quality of manufactured goods [1]. For a long time, researches on the dynamic property of spindle have mainly adopted analogies of experience or other methods, which brings about some negative effects, such as prolonged product development cycle, relatively higher cost and blindness. In addition, its theoretic calculation method is found to be extremely difficult for relatively complex structures. Virtual electrodynamics research is therefore proposed. It mainly constructs the model of finite element analysis of digital mock-up through some large-size finite element analysis software to make dynamics analysis accordingly, which can be used to know the dynamics property of digital mock-up. In this way, product design cycle and cost incurred would be reduced with specific target. This paper focuses on the spindle of HTC3250μn and HTC2550hs CNC turning centre, and uses ANSYS software to research the dynamic properties of this machine tool spindle. This greatly simplifies the performance analysis process of spindle unit, lessens the intervention of manual operation and avoids repetitive work.

At present, the researches on dynamic property can be mainly classified into two respects: natural vibration characteristics and response characteristics. Wherein, the research of natural vibration characteristics focuses on undamped free vibration, which can be used to get the natural property of vibration system, namely natural frequency and vibration mode; while the research of response characteristics focuses on the steady-state response of structure that is sinusoidal-varying with time, which can be used to calculate the dynamic response of structure and get the displacement response and stress response [2].

2 Spindle modal analysis

2.1 SUMMARY OF MODAL ANALYSIS

Modal analysis is a kind of technology used for determining the natural frequency and vibration mode of the structure (Natural frequency and vibration mode of the structure are two extremely important parameters in designing the structure on dynamic loads) [3]. The natural frequency reflects the degree of structure rigidity, namely bigger natural frequency means bigger rigidity. Through modal analysis, the structure can be designed to avoid resonance or vibration at specific frequency, designers would know how the structure responds to the dynamic loads of different types, and control parameter will be provided for other dynamic analysis as well. As the structure vibration characteristics determines its response status at different dynamic loads, modal analysis is the basis for further analysing structure dynamics, such as transient dynamics, harmonic response and spectral analysis.

Machine tool vibration forces the change on relative position between cutting tool and work piece as well as relative speed, which changes the cutting process and limits the product manufacturing precision and efficiency. Most vibrations of machines tool come from the spindle.

[*] ***Corresponding author*** e-mail: 52425362@qq.com

Therefore, the natural frequency and vibration mode of spindle are important metrics for analysing and evaluating the dynamic performance of machine tool. Although the spindle of machine tool is an infinite multi-degree of freedom system with infinite multiple natural frequencies and their corresponding vibration modes, the low-order mode of the spindle is actually enough to express its dynamic property. Therefore, a few of low-order modes starting from the first-order mode are enough to make a finite element modal analysis of the machine tool.

The steps of making a modal analysis by ANSYS software are more or less the same with other analyses, but it should be emphasized that the application of constraint and boundary condition on the modal analysis will affect the result to a larger degree. In this concern, a proper application of boundary condition is necessary to accurately get the natural frequency of structure. Normally the mode extraction methods are described as below:

1) Block Lanczos Method: it is applicable to the solution of the eigenvalue of large-size symmetric matrix, which has the characteristics of lower quality requirement of model unit and faster solution availability;

2) Power Dynamics Method: it is applicable to multi-degree of freedom problems, especially for only solving the first several orders of structure;

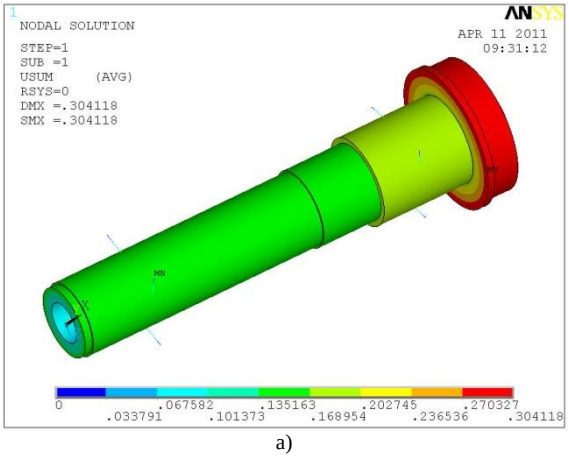
3) Unsymmetric Method: it is used for the problems when the system matrix is unsymmetric;

4) Damped Method: it is used for problems considering damp effect; this paper uses this method as the bearing damp effect cannot be ignored.

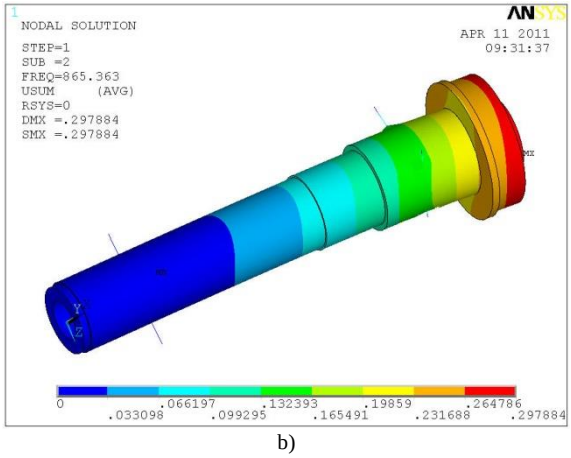
In typical modal analysis, the only effective load is zero-displacement constraint. If one non-zero displacement constraint is appointed at some point of DOF, the program will replace the position of this DOF with zero-displacement constraint. Although other types of loads can be appointed in modal analysis, such as force, stress, temperature and acceleration, they are invariably ignored in modal extraction. All loads vector will be calculated by the program and saved in vibration mode documents, which can be used in harmonic response analysis or transient analysis by the mode superposition method.

2.2 SPINDLE MODAL ANALYSIS BASED ON ANSYS

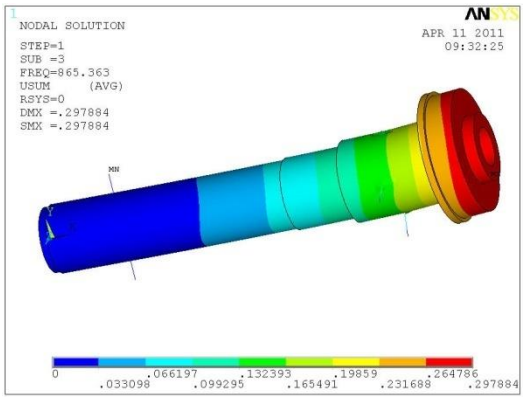
In order to get an accurate result of the spindle modal analysis, this paper properly modifies the constructed three-dimensional static analysis model, which would be used as the finite element model for spindle modal analysis. The researcher alters the analysis type to modal analysis in the solver, and sets the frequency range of extracted mode from the minimum "0" to the maximum "3000". Spring damped unit is used to simulate the boundary condition and supporting form, the pitch point of inner ring is noted to restrict the axial degree of freedom , partial useless natural frequency and vibration mode are filtered out, and the spindle modal analysis and modal extension are finally made. Calculated by ANSYS software, the first 10-order modes of spindle are extracted to get their vibration characteristics. Vibration modes diagram are shown as Fig.1. From Fig.1, the vibration modes of spindle are mainly torsion, oscillation and bend. First-order displays the vibration mode of rigid body; the natural frequencies for two-order and three-order, four-order and five-order, six-order and seven-order are the same and vibration modes are in orthogonal, which can be mathematically understood as different characteristic vectors for one same characteristic value or multiple root, but the deformation trends are differently from one another.



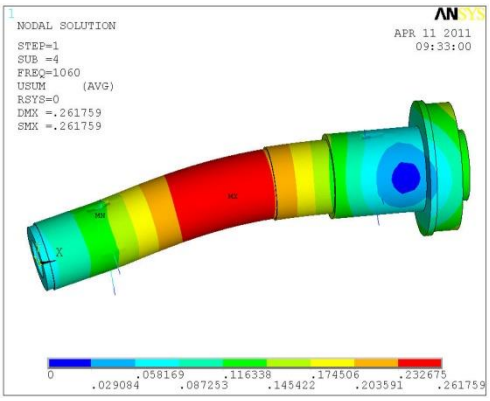
a)



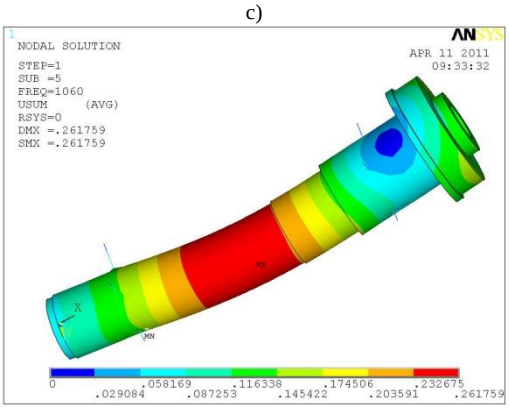
b)



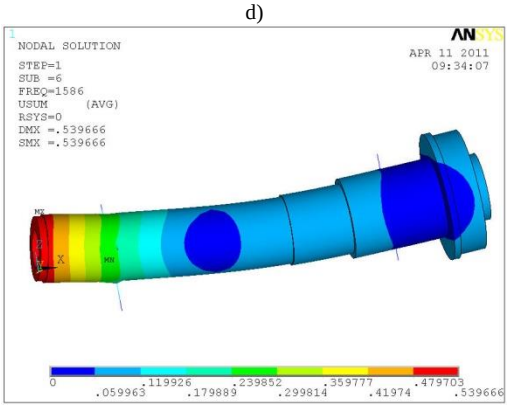
c)



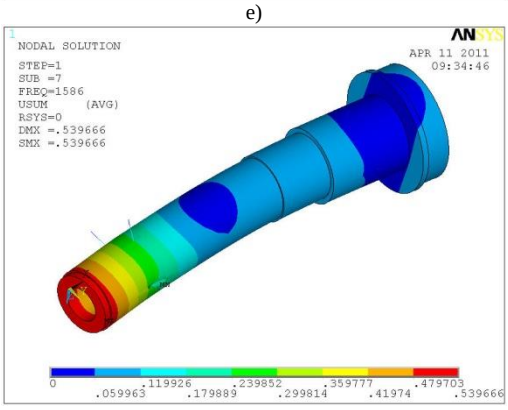
d)



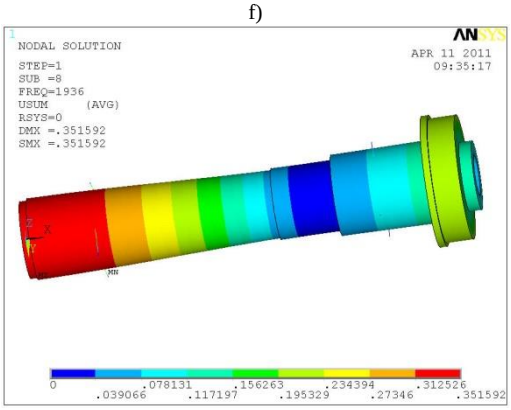
e)



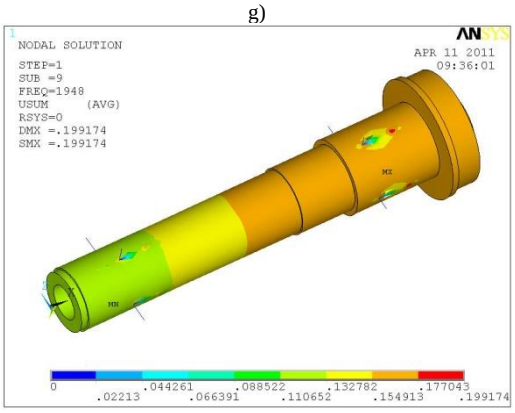
f)



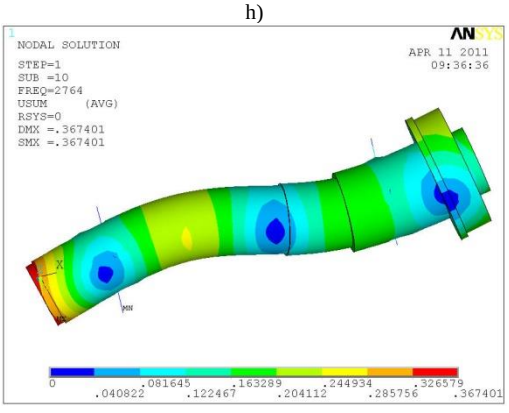
g)



h)



i)



j)

FIGURE 1 Vibration mode of spindle: a) First-order Vibration Mode, b) Second-order Vibration Mode, c) Third-order Vibration Mode, d)Four-order Vibration Mode, e) Five-order Vibration Mode, f) Six-order Vibration Mode, g) Seven-order Vibration Mode, h) Eight-order Vibration Mode, i) Nine-order Vibration Mode, j) Ten-order Vibration Mode

Through a separate analysis of all acquired vibration diagrams, combined with the natural frequencies for all orders calculated by ANSYS software, the vibration characteristics of spindle can be summarized as shown in Table 1.

TABLE 1 Natural vibration frequency and mode of spindle

Degree	Frequency(Hz)	Vibration Mode
1	0	Translational
2	865.36	First-order Z oscillation
3	865.36	First-order Y oscillation
4	1060.4	Two-order Y oscillation
5	1060.4	Two-order Z oscillation
6	1585.8	Three-order Y oscillation
7	1585.8	Three-order Z oscillation
8	1936.1	Torsion

In actual, the modal analysis of spindle can be made by use of the two-dimensional finite element analysis model as well. Although the natural frequency solved by the two-dimensional finite element analysis model is basically the same with that of three-dimensional finite element analysis model, the solved vibration mode isn't as resourceful as that of three-dimensional finite element analysis model, which is mainly caused by huge difference of nodes quantity among different unit types as well as the consideration of different treatment methods for boundary conditions. Therefore, the calculation simulation effect of three-dimensional finite element model proves to be more accurate.

2.3 ANALYSIS OF SPINDLE CRITICAL SPEED

Traditionally, critical speed of spindle mainly refers to the spindle natural frequencies at all orders. As the common spindle has the characteristic of relatively lower working speed, the influence of spindle speed on the dynamic characteristics can be ignored. Therefore, the traditionally defined critical speed is reasonable.

With the growing speed of machine tool, the traditional definition of critical speed cannot meet the requirement of high-speed spindle. In this concern, the influence of speed cannot be ignored in the dynamic analysis of high-speed spindle. The critical speed of high-speed spindle is defined as the corresponding speed at one natural frequency. In the analysis of high-speed spindle critical speed, the spindle will resonate when the rotation frequency of spindle equals its natural frequency. However, if this case should occur in actual manufacturing, it not only impairs the manufacturing quality, but also irreversibly damages the spindle, reduces its service lifetime and increases the maintenance cost.

At a specific speed, the deflection of spindle can be arbitrary value, wherein the deflection at each point can be arrayed in an elastic straight line. The specific speed is namely the critical speed as output in Equation (1).

$$w_c = l^2\sqrt{\frac{EJ}{m}} = \frac{(la)^2}{a^2}\sqrt{\frac{EJg}{A\gamma}}, \quad (1)$$

Wherein, $la = \pi, 2\pi, 3\pi$, A is the cross-sectional area of spindle segment, and γ is the material density.

Actual and ideal structure of spindle differs in the uniform section of shaft end, and the critical speed of actual structure has slightly higher critical speed as output from Equation (1). In most cases, the critical speed of spindle can be directly output through Equation (2) with the unit of speed converted to r/min.

$$n = 60f, \quad (2)$$

Wherein, n is the speed, f is the natural frequency. The critical speeds of spindle at all orders can be calculated by the natural frequency through ANSYS, but it starts from the two-order frequency because the first-order is the vibration mode of rigid body. Among all critical speeds of spindle at all orders as listed in Table 2, the low-order speed of spindle is of most referential value. However, the low-order speed value output from theoretical equation is in discrepancy with the actual value unless the design speed is within the absolute safety range; or otherwise it may easily result in irrecoverable result caused by misjudgement. In order to accurately get the critical speed of spindle, it can be calculated much closer to the actual value through ANSYS, which would avoid the misjudgement.

TABLE 2 Critical speed of spindle (Theoretical value)

Degree	Frequency(Hz)	Speed(r/min)
2	865.36	51921.6
3	865.36	51921.6
4	1060.4	63624
5	1060.4	63624
6	1585.8	95148
7	1585.8	95148
8	1936.1	116116

The speed of spindle is set as 0rad/s, 5000rad/s, 10,000rad/s and 15,000rad/s respectively for modal analysis, and Campbell Diagram [4] is eventually drawn up as shown in Fig.2.

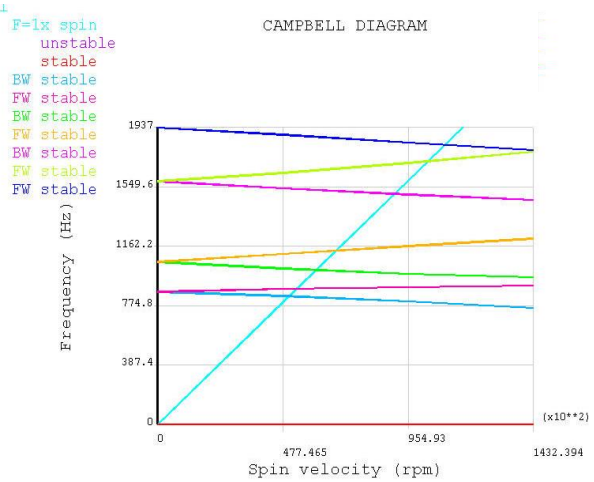


FIGURE 2 Campbell diagram for critical speed

Though Campbell Diagram we can find there are two curves diverging from each natural frequency, wherein one curve increases with the speed, while the other one decreases with the speed. These are called as forward precession frequency and backward precession frequency respectively in rotor dynamics. According to the definition of critical speed, the speed at the cross point between the 45°straight line as shown in Campbell diagram and forward precession frequency curve is the critical speed [5] at this very moment. The critical speed of spindle (Simulation value) is shown in Table 3. In comparison between the values in Table 2 and Table 3, the critical speed of spindle calculated by drawn-up Campbell diagram through ANSYS is approximately 1% higher than the critical speed by theoretical equation. The equation for critical speed of forward precession is described as Equation (3):

TABLE 3 The critical speed of spindle (Simulation value)

Degree	Frequency(Hz)	Speed(r/min)
2	865.36	52372.3
3	865.36	52372.3
4	1060.4	64707.8
5	1060.4	64707.8
6	1585.8	97252
7	1585.8	97252
8	1936.1	118458

$$\left|(-[M_1]+[J_1])\omega^2+[K_{g1}]=0\right|. \tag{3}$$

Through it we can find the critical speed would invariably increase if there is any forward precession, which is found to be in conformity with the reality; the highest design speed of HTC3250μn and HTC2550hs CNC turning centre is 6000r/min, which is lower than 2/3 of the first-order critical speed. Consequently, the spindle would not have natural vibration in actual use, which meets the running requirement.

3 Harmonic response analysis of spindle

3.1 SUMMARY OF HARMONIC RESPONSE ANALYSIS

Harmonic response analysis is a technology used to solve the structure response under the simple harmonic load when the frequency is known. The input value is the harmonic load (force and forced displacement) that its size and frequency is known, or under multiple loads at the same frequency, the output value is the harmonic displacement [3] at each degree of freedom. In actual manufacturing, cyclic exciting force will apply on the spindle when the machine tool runs. When the natural frequency of spindle equals the frequency of cyclic exciting force, resonance effect will take place. During actual manufacturing, resonance will not only reduce the manufacturing accuracy that result in rejected parts, but also impair the machine tool and cutting tool. As a result, it is necessary to make a harmonic analysis of spindle in the phase of designing the spindle.

In ANSYS, the finite model of harmonic response analysis is more or less the same with the finite model of modal analysis. Full method requires a direct integral computation, which would invariable consume a great amount of time; comparatively, the method of mode superposition directly multiplies and sums the obtained modal type through modal analysis, which has the characteristics of quickness, high efficiency and well-intended accuracy. To that regard, this paper adopts the method of mode superposition to make a harmonic response analysis.

3.2 HARMONIC RESPONSE ANALYSIS OF SPINDLE BASED ON ANSYS

Before making a harmonic response analysis of the spindle, the load changing in line with time shall be determined, or the exciting force shall be determined. The equation for typical exciting force can be expressed as follows:

$$F(t)=F_1\sin(\omega t+\phi), \tag{4}$$

wherein, F_1 stands for the amplitude of exciting force; ω stands for the range of forced vibration frequency; and ϕ stands for the phase angle. The amplitude of exciting force is actually the maximum value of external load; the range of forced vibration frequency refers to the frequency range of simple harmonic load; and the phase angle refers to the metric for load lagging or preceding the reference time [6]. In this paper, the amplitude of exciting force is turning force, which is to be put into the equation. Equation (4) in the manner of Fourier series is extended:

$$\begin{aligned}F(t)=F_0+F_1\sin(\omega t+\phi_1)\\ +F_2\sin(\omega t+\phi_2)+\cdots\end{aligned}. \tag{5}$$

If the higher order term is not considered, ϕ_1 is approximated to be "0", therefore the exciting force equation is further simplified as below:

$$F(t)=F_1\sin(\omega t). \tag{6}$$

According to general manufacturing, the vibration range is between 0 to 600Hz. In this way, the exciting force of harmonic response analysis is finally determined.

In the above part, the modal analysis is made on the spindle, which is followed by the harmonic response analysis after saving and analysing the result. In view of large time consumption in an accurate harmonic response analysis, this paper firstly analyses the entire range of vibration frequency, and gets the spindle radial response displacement curve within the range of vibration frequency by reducing the sub-steps and analysing time. Due to limited sub-steps, this curve only shows the

approximate position of change tendency and vibration points. In order to obtain the spindle radial response displacement in a more accurate way, accurate analysis could be made within one frequency range respectively. By adding more sub-steps within this frequency range, more accurate analysis result can be obtained to evaluate the response characteristics of spindle.

In time-history postprocessor of ANSYS, the response relationship between variable and frequency cannot be observed until the observable variables are defined. In reality, the definition of variable directly determines the result of spindle response analysis. Generally, dangerous points in the spindle shall be contained in those observable variables. In case some dangerous points are ignored, the evaluation of spindle response characteristics would possibly be erroneous, the dynamic rigidity of spindle would deviate, which would lead to the manufactured machine tool lower than the required precision of actual manufacturing.

In this concern, this paper intends to analyse the five dangerous points of spindle, combined with the analysis of spindle front-end, front prop, rear prop, middle point and rear-end, so that the spindle response characteristics can be obtained. In order to compare the influence of the damping of bearing on the dynamic characteristics of spindle, respective discussions under damping and absence of damping must be made. Set 0-1200Hz as the excitation frequency, through harmonic response analysis, the radial amplitude-frequency curves for spindle front-end, front prop, rear prop, middle point and rear-end can be shown in Figure 3.

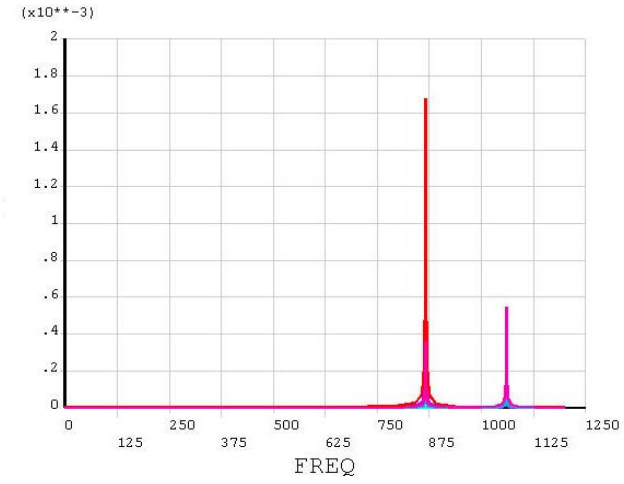


FIGURE 3 Displacement-frequency curve of spindle(0～1200Hz)

Figure 3 shows there are significant resonances at two natural frequencies: 865Hz and 1060Hz, which also shows bigger fluctuation of deformation amplitude. As a result of no damping, the vibration amplitude has the infinite tendency to increase and dynamic rigidity significantly declines, when the spindle does not meet the manufacturing requirement. In order to better reduce the influence of spindle vibration on the manufacturing precision, damping is usually introduced to reduce the vibration response amplitude of spindle. The bearing damping in this paper is set as 5000 $N \cdot S / \mu m$, and the excitation frequency is set as 0～1200Hz. Through the harmonic response analysis, the amplitude-frequency curves for spindle front-end, front prop, rear prop, and middle-point, rear-end can be shown as Figure 4. The Figure 4 shows there are significant resonances at two natural frequencies: 865Hz and 1060Hz, while in consideration of damping effect, the vibration amplitude would significantly decline. The discussion is respectively made for two resonance points in the following part.

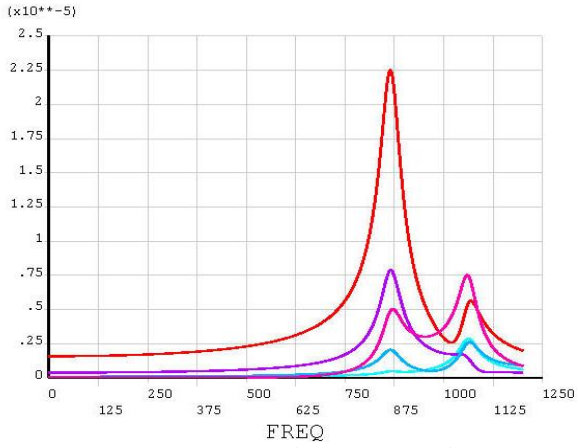


FIGURE 4 Displacement-frequency curve of spindle (0～1200Hz)

Set 400～900Hz as the range of excitation frequency, control the sub-steps as 5,000, and reanalyse the harmonic response, which is shown in Figure 5 and Figure 6.

Figure 5 and Figure 6 show that the dynamic stiffness of spindle remains almost unchanged before the excitation frequency could approximately reach the first-order natural frequency. When it closes to the natural frequency, the response displacement begins to upward while the spindle stiffness begins to downward. When it reaches the resonance point, the response displacement is the highest and the spindle dynamic stiffness is the lowest. After that, the response displacement begins to downward, while the spindle dynamic stiffness gradually upwards.

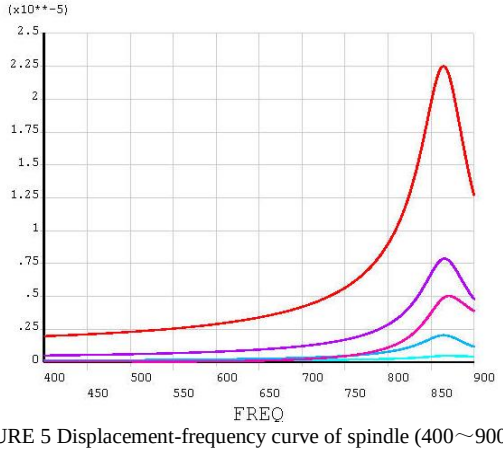


FIGURE 5 Displacement-frequency curve of spindle (400～900Hz)

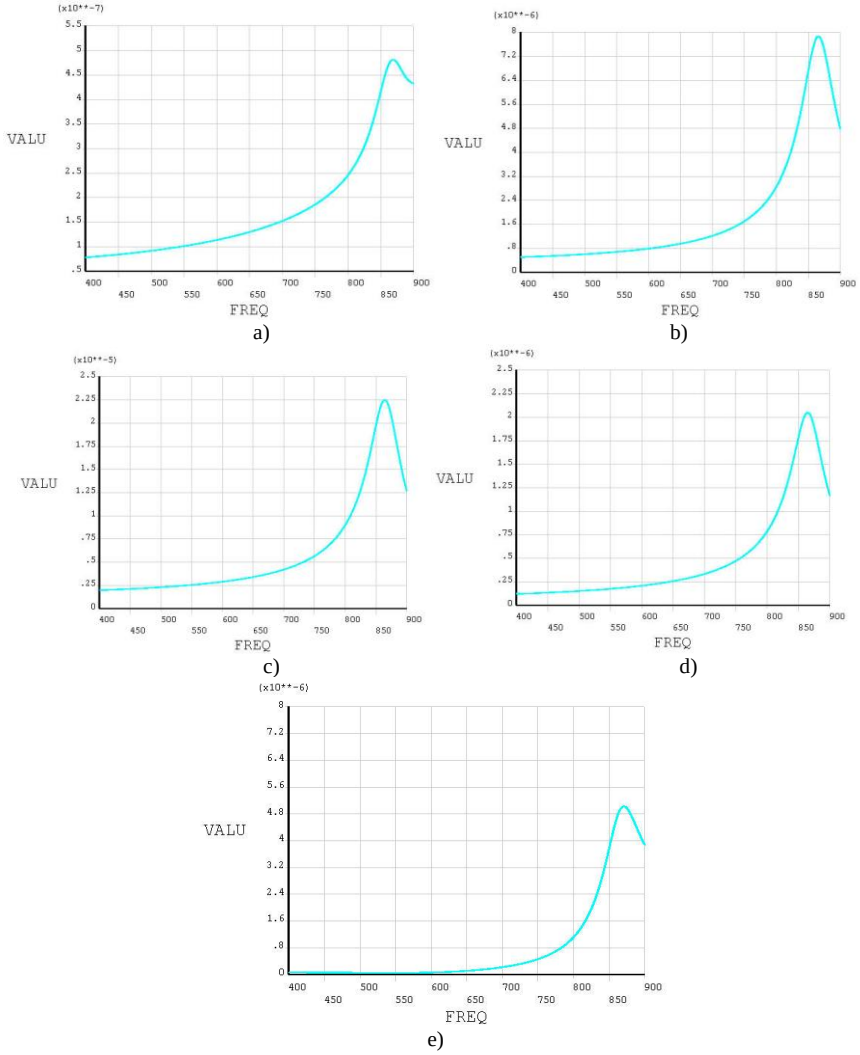


FIGURE 6 Displacement-frequency curve of spindle (400～900Hz): a) Displacement-frequency curve of front bearing, b) Displacement-frequency curve of rear bearing, c) Displacement-frequency curve of spindle front-end, d) Displacement-frequency curve of spindle rear-end, e) Displacement-frequency curve of spindle middle-point

Through the analysis of all spindle points, roughly around the first-order natural frequency, the displacement response of spindle front-end is more significant, and at 865Hz, the response displacement abruptly increases to 22.5 μm . The spindle dynamic stiffness significant downwards; after that the displacement response abruptly decreases while the spindle dynamic stiffness gradually upwards. During this process, the minimum dynamic stiffness of spindle is shown as below:

$$K = \frac{452.9}{22.5} = 20.13N / \mu m . \quad (7)$$

In view of the analysis around spindle two-order resonance points, set 900～1400 Hz as the range of excitation frequency, control sub-steps as 5,000, select five key points on the spindle to draw up the amplitude-frequency characteristic curve and make harmonic response analysis, the result is shown as Figure 7. Figure 7 shows the displacement response under the influence of spindle exciting force is lower than the first-order resonance point, when the spindle is at the two-order resonance point. However, the spindle vibration at this order remains evident, and the spindle stiffness shows the tendency of upward, downward and gradually upward after leaving the resonance point. The displacement-frequency curves for all key observing points are shown in Figure 8.

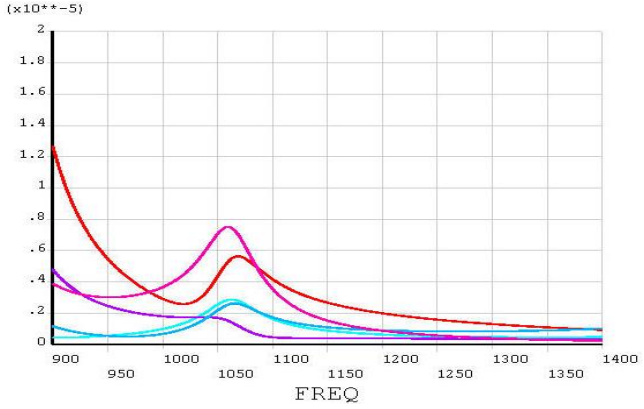


FIGURE 7 Displacement-frequency curve of spindle (900～1400Hz)

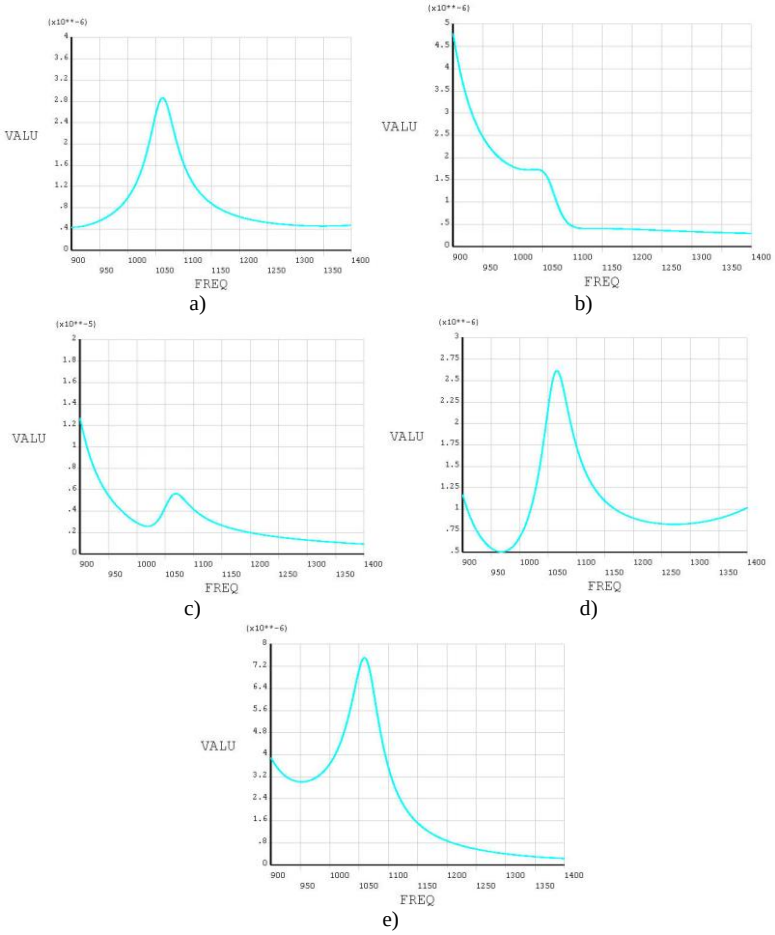


FIGURE 8 Displacement-frequency curve of spindle(900～1400Hz): a) Displacement-frequency curve of front bearing, b) Displacement-frequency curve of rear bearing, c) Displacement-frequency curve of spindle front-end, d) Displacement-frequency curve of spindle rear-end, e) Displacement-frequency curve of spindle middle-point

Through the analysis of each point on the spindle, it can be found that the displacement response of middle point is the most prominent around the second-order natural frequency. At 1060Hz, response displacement increases suddenly and the maximal displacement increases to 7.8 μm . The dynamic stiffness of spindle declines, thus it can be judged that spindle is prone to crack at the moment. Then displacement response suddenly declines and spindle dynamic stiffness gradually improves. During this process, the smallest dynamic stiffness of the spindle is as follows:

$$K = \frac{452.9}{7.8} = 58.06N / \mu m . \tag{8}$$

By comparing the amplitude of the first-order resonance point and two-order resonance point, the spindle vibration tendency matches with the modal vibration as solved in the previous part. The spindle vibration tendency does not increase strictly with the increase of natural frequency, but its amplitude usually depends on the mutual effect of characteristics and damping.

The above analysis can be used to evaluate the vibration characteristics at the time of spindle resonance. However, when applied to the actual manufacturing, as the spindle keeps off the forced resonance area at the designing phase, the analysis of the resonance point cannot completely evaluate the spindle dynamic characteristics. In order to cover this point, according to the variation range of excitation frequency at the normal state of manufacturing, the analysis result of spindle harmonic response can be shown in Figure 9. Figure 9 shows the response displacement of all spindle points increase with the excitation frequency, notably on the spindle ends and front-end bearing. At 600Hz, the maximum response amplitude is 2.9 μm , when the spindle dynamic stiffness can be expressed as:

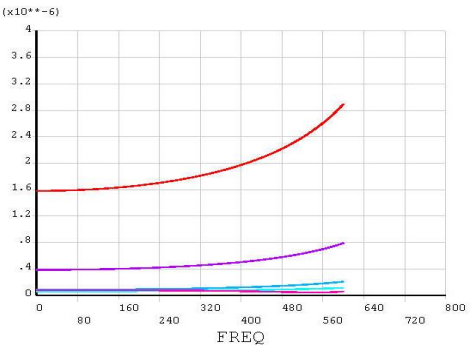


FIGURE 9 Displacement-frequency curve of spindle(0～600Hz)

$$K = \frac{452.9}{2.9} = 156.17N / \mu m . \quad (9)$$

By analysing the dynamic stiffness, it can be concluded that the spindle adopted in this paper meets the design requirement; in actual manufacturing, it can meet the precision requirement.

4 Conclusions

The performance analysis and optimization of the spindle based on ANSYS can effectively shorten the design time and improve design efficiency. Through the dynamic characteristics analysis of spindle, it can be summarized that the following method can improve the dynamic characteristics of spindle:

(1) Improving the stiffness and damping of bearings. From the vibration diagram and harmony response analysis of spindle, it can be clearly found that the deformation around the front support of spindle is more obvious. Thus, improving the stiffness and damping of the front bearing can effectively reduce the vibration deformation in the area, improve the natural frequencies of spindle and avoid the happening of resonance phenomena.

(2) Improving the structure of spindle. The configuration way of bearing, spindle span and length of overhanging end has very important influence to the dynamic characteristics of spindle. On the premise of conditions permit, the structure optimization design can significantly improve the dynamic characteristics of spindle.

Acknowledgements

Financial supports from National Science & Technology Support Program Project (ID 2012BAF12B08-04), Key Scientific and Technological Project of Liaoning Province (No. 2011216010) and the Fundamental Research Funds for the Central Universities (ID N120403002) are highly appreciated.

References

[1] Jinhua Li, Yongxian Liu, Yang Yu 2012 Static And Dynamic Characteristic Analysis of Spindle in Turning Center *Manufacturing Technology & Machine Tool* (3) 45-8
[2] Qiu Hu, Qiang Teng 2007 Dynamic Characteristics Simulation of High Speed Spindle Unit in CNC Machine Tool *Machine Tool & Hydraulics* **35**(1) 204-6
[3] Zhen Zhang, Yueqing Yu, Jianxin Yang, Guoning Si, Changhao Chen 2011 Model Analysis of The Gantry Type High-Speed Machining Center *Machinery Design & Manufacture* (2) 201-3
[4] Gagnal V, Bouzgarrou B C, Ray P, Barra C 2007 Model-Based Chatter Stability Prediction for High-Speed Spindles *Machine Tools and Manufacture* (47) 1176-86
[5] Tianbiao Yu, Xuezhi Wang, Peng Guan, Wanshan Wang 2012 Modal Analysis of Spindle System on Ultra-high Speed Grinder *Chinese Journal of Mechanical Engineering* **48**(17) 183-8

Authors

Hualong Xie

Current position, grades: Since 2010, he is an associate professor at Northeastern University.
University studies: bachelor degree in Mechanical Electronic Engineering, Master degree in Mechanical Design and Theory and PhD degree in Control Theory and Control Engineering from Northeastern University, China, in 2000, 2003 and 2006, respectively.
Scientific interests: robot, intelligent control, intelligent bionic leg and biomechanics.

Xiaopeng Mei

Current position, grades: postgraduate at School of Mechanical Engineering & Automation in Northeastern University, China.
University studies: received his bachelor degree in Mechanical Design, Manufacturing and Automation from Shen Yang University.
Scientific interests: advanced manufacturing and intelligent control.

Yongxian Liu

Current position, grades: Professor at Northeastern University.
University studies: received his Master Degree in Mechanical Design and Theory from Northeastern University, China, in 1985.
Scientific interests: product design automation, robot control, and computer simulation.

Anti-synchronization of a class of fractional-order chaotic system with uncertain parameters

Ran Ding[1, 2], Caoyuan Ma[1*], Yongyi Zhao[1], Yanfang Luo[1], Jianhua Liu[1]

[1]*College of Information and Electrical Engineering, Xuzhou, Jiangsu, 221008, China*

[2]*State Grid Jiangsu Electric Company, Nanjing, Jiangsu, 210024, China*

Received 14 August 2014, www.cmnt.lv

Abstract

In order to pull the fractional-order theory to better application, the detailed of computer numerical simulation of the Adams-Bashforth-Moulton Algorithm is proposed in this paper. Anti-synchronization of a class of fractional-order chaotic system with uncertain parameters is realized on this basis and the stability theorem of the system is presented at the same time. And thus it indicates that this method can be adapted to chaotic system with certain parameters and a class of chaotic system with not equal fractional-order. And corresponding implementation conditions is given as well. Besides, it is pointed out that the method, which unites the synchronization and anti-synchronization is also suitable for synchronization issues of the system. Finally, take classic Lorenz system for instance, track time domain and error map about drive system and response system of anti-synchronization are given. The results prove the effectiveness of the control method in the realization of anti-synchronization of a class of fractional-order chaotic system with uncertain parameters and the feasibility of fractional order computer numerical simulation

Keywords: fractional-order, chaotic system, anti-synchronization, chaos control

1 Introduction

Fractional calculus is known as a more common form of integer order, and it also has its own unique features - memory function, meanwhile, a growing number of scientific workers have been attracted to fight for it gradually [1-6]. As for fractional definitions and mathematical description, it has gained initial achievements, but how to make a more accurate numerical simulation analysis through computer is still in its infancy. It is can be predicted that computer numerical simulation of fractional order system will greatly promote the engineering applications of fractional order [7-9], meanwhile, it will certainly open up another important research area of computer applications, which will help us describe the objective world better.

Chaos is a movement pattern of nonlinear system, whose feature is a special, unstable and class of random. The system in a chaotic movement is usually called chaotic system, which was first found by the U.S. meteorologist Lorzenz in his proposed meteorological equations. With further research, the discovery and understanding of chaotic feature began to shift to the control and use of it. In 1990, physicist Ott. Grebogi and Yorke at the university of Maryland successfully controlled chaos by parameter perturbation method (i.e. OGY method), which is an iconic achievement. Closely followed, Pecora and Carroll first raised a synchronization scheme of two similar chaotic systems with different initial conditions, which is called PC method. Since then, Control of Chaotic Synchronization has aroused widespread concern. So far, various control methods of Chaotic Synchronization have been proposed by domestic and foreign scholars, such as adaptive control method [10, 11], Backstopping control method [12], fuzzy control [13, 14], sliding mode control [15, 16], etc. With the deepening of the research, the concept of synchronization has also been expanded, which includes anti-synchronization, generalized synchronization, projective synchronization, phase synchronization, trailing synchronization, etc. [17-19]. Among them, the anti-synchronization is a very interesting concept, which means a state that the drive system and response system achieve equal magnitude but opposite in sign.

In real industrial practice, the noise is ubiquitous, and many system parameters of the system cannot be accurately measured, and even unknown. In order to better realize the engineering practical value of computer numerical simulation, it is very meaningful to consider the uncertainty of noise and parameter, and it has preliminary obtained some related research results [20-22].

In summary, anti-synchronization of a class of fractional-order chaotic system with uncertain parameters is studied in this paper, and a stability theorem of the system is presented. Besides, this method is also suitable to chaotic system with certain parameters、drive and response systems with unequal Fractional Order and synchronization of a class of fractional-order chaotic

* ***Corresponding author*** e-mail: mcycumt@139.com

system. Finally, take classic Lorenz system as an example, the detailed iterative formula of computer numerical simulation of the Adams-Bashforth-Moulton Algorithm is proposed, and the results of computer numerical simulation are given.

2 Description of fractional system

Theorem 1 [23] a general fractional order linear system can be described as

$$D^q x = Ax, \tag{1}$$

where $q=[q_1,q_2,\ldots,q_i,\ldots,q_n]$ and $(0<q_i\le 1)$. If and only if all eigenvalues λ_i of matrix A is satisfied with $|\arg(\lambda_i)|>q\pi/2$, the system (1) is absolutely stable.

In this paper, a class of fractional order chaotic systems with uncertain parameters is presented as:

$$D^\alpha x=(A+\Delta A)x+f(x). \tag{2}$$

Its response system is

$$D^\beta y=(A+\Delta A)y+f(y)+U(x,y), \tag{3}$$

where $x,y\in R^n$ are all n-dimensional state vectors; f is a continuous vector function; $U(x,y)$ is the designed cont roller; α and β are the fractional orders; A is the coeffic ient matrix of the fractional order system; ΔA indicates u ncertain parameters, which satisfies $\|\Delta A\|\le\delta<M$. By th e way, M is a constant.

Our goal is to design a controller $U(x,y)$, which could make the system (2) and system (3) to achieve an equal and opposite anti-synchronization.

3 Design of adaptive sliding mode controller

In order to obtain the designed form of the controller, we first suppose is the error vector. Thus, its error system is

$$\begin{aligned} D^\beta e &= (A+\Delta A)e+f(y)-f(x)+U(x,y)\\ &= (A+\Delta A)e+F(x,y)+U(x,y)\end{aligned}, \tag{4}$$

where $F(x,y)=f(y)-f(x)$.

Here, we chose a switch sliding surface as $S(t)=D^{\beta-1}e(t)-\int_0^t(A+K)e(\tau)d\tau$, where K is a designed parameter matrices. Usually, the two conditions $S(t)=0$ and $\dot S(t)=0$ are must satisfied simultaneously.

Because $\dot S(t)=0$, we can get

$$\dot S(t)=D^\beta e(t)-(A+K)e(t)=0. \tag{5}$$

Thus, taking Eq. (4) into (5), we can get $\dot S(t)=(\Delta A-K)e(t)+F(x,y)+U(x,y)=0$.

To meet the conditions of sliding mode, we set $DS(t)=-p*sign(S)-r*abs(S)^{m/n}*sign(S)$, in which $sign(S)=\begin{cases}+1, S>0\\ 0, S=0\\ -1, S<0\end{cases}$, $p>0$, $r>0$ are both Controller gain.

On the sliding surface, because $Ds(t)=\dot s(t)=0$ is satisfied, the controller is

$$U(x,y)=(K-\Delta A)e(t)-F(x,y)-(rS+psignS). \tag{6}$$

Theorem 2: When the controller (6) is applied to the system (3), the system (2) and system (3)can achieve the anti-synchronization. In other words, the error system is zero, and it reaches absolute stability.

Proof: When the system is running on the sliding surface, namely $S(t)=0$, the error system can be simplified as $D^\beta e=(A+K)e$.

According to Theorem 1, as long as $(A+K)$ meets the condition $|\arg(\lambda_i)|>q\pi/2$, the error system is stable absolutely.

Therefore, the theorem 2 has been proved.

Lemma 1: If the parameters of the fractional order system is determined, i.e $\Delta A=0$, the controller is still valid, and also can control the drive and response system to achieve anti-synchronization.

Lemma 2: If the order α of drive system and β of response system is not equal, then it needs to introduce a compensation controller $U_I(x,y)=D^\beta(\chi x)-f(x)-(A+\Delta A)\chi x$. So, the controller (6) is still valid.

Lemma 3: If the error system is configured as $e=y-x$, this means synchronization of the drive system and the response, and the controller is still valid.

4 Numerical simulations

We consider the Lorenz system as the drive system $\begin{cases}D^{\alpha_1}x_1=a(y_1-x_1)\\ D^{\alpha_2}y_1=bx_1-x_1z_1-y_1\\ D^{\alpha_3}z_1=x_1y_1+cz_1\end{cases}$ and response system $\begin{cases}D^{\beta_1}x_2=a(y_2-x_2)\\ D^{\beta_2}y_2=bx_2-x_2z_2-y_2\\ D^{\beta_3}z_2=x_2y_2+cz_2\end{cases}$, where the parameters (a, b, c) = (10, 28, 8/3), $\alpha=\beta=0.99$, and the initial value (x_1, y_1, z_1) = (1, 0, 9), (x_2, y_2, z_2) = (1, 1, 1).

We set the parameters are p=0.2, r=6, m=3and n=2;

Coefficient matrix of the system is $A=\begin{pmatrix}-10 & 10 & 0\\ 28 & -1 & 0\\ 0 & 0 & -8/3\end{pmatrix}$;

parameter matrix $K=\begin{pmatrix}-20 & 20 & 0\\ -28 & 0 & 0\\ 0 & 0 & -1/3\end{pmatrix}$. Thus, A + K are $A+K=\begin{pmatrix}-40 & 40 & 0\\ -10 & -1 & 0\\ 0 & 0 & -3\end{pmatrix}$, and its eigenvalues are $(\lambda_1,\lambda_2,\lambda_3)=(-20.5+4.44i,-20.5-4.44i,-3)$.

According to Theorem 1, it satisfies $|\arg(\lambda_i)|>\beta_i\pi/2=0.99\pi/2$, so the error system is absolutely stable, which means the drive system and response system achieve to synchronization. Therefore, the expression of the controller is

$$\begin{cases} u &= (-30e_1+30e_2)-(5s_1+0.2sign(s_1))\\ u &= -38e_1-(5s_2+0.2sign(s_2))\\ u &= -\tfrac{1}{3}e_3-(5s_3+0.2sign(s_3)) \end{cases}.$$

The fractional order nonlinear system equation can be solved by using Adams-Bashforth-Moulton algorithm. So, the iterative formula of solving fractional Lorenz system is

$$\begin{cases} x_{n+1}=x_0+\dfrac{h^\alpha}{\Gamma(\alpha+2)}\left\{10[y^p_{n+1}-x^p_{n+1}]+\sum_{j=0}^{n}a_{1,j,n+1}10(y_j-x_j)\right\}\\ y_{n+1}=y_0+\dfrac{h^\beta}{\Gamma(\beta+2)}\left\{-x^p_{n+1}z^p_{n+1}+(24-4c)x^p_{n+1}+cy^p_{n+1}\right\}\\ \qquad +\sum_{j=0}^{n}a_{2,j,n+1}\left(-x_jz_j+(24-4c)x_j+cy_j\right)\\ z_{n+1}=z_0+\dfrac{h^\gamma}{\Gamma(\gamma+2)}\left\{[x^p_{n+1}y^p_{n+1}-\dfrac{8z^p_{n+1}}{3}]+\sum_{j=0}^{n}a_{3,j,n+1}\left(x_jy_j-\dfrac{8z_j}{3}\right)\right\} \end{cases},$$

in which

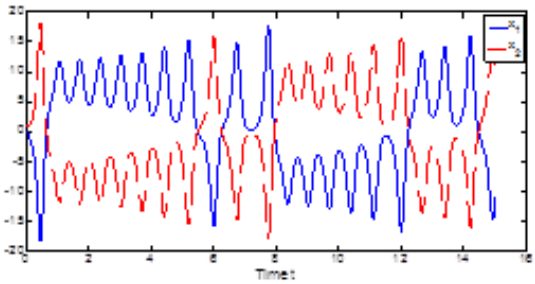
$$\begin{cases} x^p_{n+1}=x_0+\dfrac{1}{\Gamma(\alpha)}\sum_{j=0}^{n}b_{1,j,n+1}10(y_j-x_j)\\ y^p_{n+1}=y_0+\dfrac{1}{\Gamma(\beta)}\sum_{j=0}^{n}b_{2,j,n+1}(-x_jz_j+(24-4c)x_j+cy_j)\\ z^p_{n+1}=z_0+\dfrac{1}{\Gamma(\gamma)}\sum_{j=0}^{n}b_{3,j,n+1}\left(x_jy_j-\dfrac{8z_j}{3}\right) \end{cases}$$

and

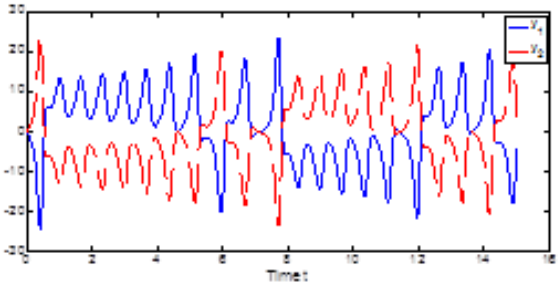
$$\begin{cases} b_{1,j,n+1}=\dfrac{h^\alpha}{\alpha}((n-j+1)^\alpha-(n-j)^\alpha, 0\le j\le n\\ b_{2,j,n+1}=\dfrac{h^\beta}{\beta}((n-j+1)^\beta-(n-j)^\beta, 0\le j\le n\\ b_{3,j,n+1}=\dfrac{h^\gamma}{\gamma}((n-j+1)^\gamma-(n-j)^\gamma, 0\le j\le n \end{cases},$$

$$\begin{cases} a_{1,j,n+1}=\begin{cases} n^\alpha-(n-\alpha)(n+1)^\alpha & j=0\\ (n-j+2)^{\alpha+1}+(n-j)^{\alpha+1}-2(n-j+1) & 0\le j\le n\end{cases}\\ a_{2,j,n+1}=\begin{cases} n^\beta-(n-\beta)(n+1)^\beta & j=0\\ (n-j+2)^{\beta+1}+(n-j)^{\beta+1}-2(n-j+1) & 0\le j\le n\end{cases}\\ a_{3,j,n+1}=\begin{cases} n^\gamma-(n-\gamma)(n+1)^\gamma & j=0\\ (n-j+2)^{\gamma+1}+(n-j)^{\gamma+1}-2(n-j+1) & 0\le j\le n\end{cases} \end{cases}.$$

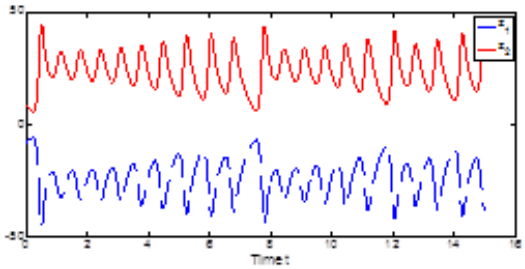
Programming numerical analysis with Matlab, we can get time-domain diagram as shown in Fig. 1 and error map as shown in Fig. 2, which demonstrates drive system and response system of Lorenz system achieve anti-synchronization. From the Fig. 1 and Fig. 2, we know the system quickly achieve to anti-synchronization, and the corresponding anti-synchronization error is 0.



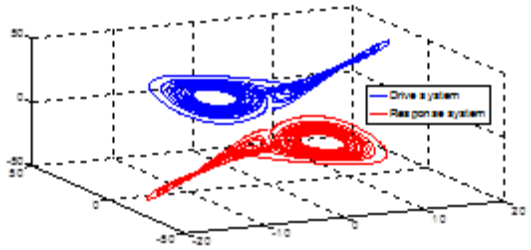
(a) x1 and x2



(b) Y1 and y2

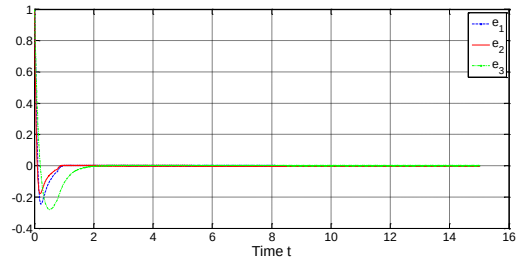


(c) Z1 and z

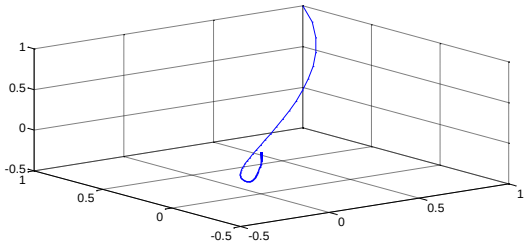


(d)Three -dimensional plot

FIGURE 1 Time-domain of Lorenz drive system and response system achieving to anti-synchronization



(a) e1, e2 and e3



(b)Three -dimensional plot

FIGURE 2 Error map of Lorenz drive system and response system achieving to anti-synchronization

5 Conclusions

In this paper, anti-synchronization of a class of fractional-order chaotic system with uncertain parameters is realized based on Sliding Mode Control Theory. Moreover, taken Lorenz system as an example, a detailed iterative formula of computer numerical simulation based on Adams-Bashforth-Moulton Algorithm is proposed, and the simulation results are presented. We can draw the following conclusions:

(1) This method has good robustness, and can well control a class of chaotic system with uncertain parameter to achieve anti-synchronization;

(2) From the three Lemmas, we can also get that this method is also suitable to chaotic system with certain parameters, unequal fractional orders;

(3) The feasibility of fractional order numerical simulation is also been proved.

Acknowledgments

This project is supported by the Natural Science Fund of China “The research of several key technologies of flexible zero residual flow Petersen coil (No: 51107143)" and Natural Science Fund of Jiangsu Province “The system stability analysis and robust fault tolerant control about burning gas turbine (No: BK20130187)”.

References

[1] Wang Yaqing, Zhou Shangbo 2011 Research on digital image encryption algorithm based on fractional Fourier transform *Application Research of Computers* **28**(7) 2738-41

[2] Mainardi F, Gorenflo R 2000 On mittag-leffler-type functions in fractional evolution processes *Journal of Computational and Applied Mathematics* **118**(1-2) 283-99

[3] Diethelm K, Ford N J 2002 Analysis of fractional differential equations *Journal of Mathematical Analysis and Applications* **265**(2) 229-48

[4] Podlubny I 1999 Fractional-order systems and PI-lambda-D-mu-controllers *IEEE Transanctions on Automatic Control* **44**(1) 208-14

[5] Li C F, Luo X N, Zhou Yong 2010 Existence of positive solutions of the boundary value problem for nonlinear fractional differential equations *Computers and Mathematics with Applications* **59**(3) 1363-75

[6] Chen Diyi, Liu Chengfu, Wu Cong, Liu Yongjian, Ma Xiaoyi 2012 A new fractional-order chaotic system and its synchronization with circuit simulation *Circuits, Systems & Signal Processing* **31** 1599-613

[7] Xu Qiang, Bao Bocheng, Hu Wen, Yang Xiaoyu 2010 Numerical analysis and circuit simulation of fractional-order chaotic system *Application Research of Computers* **27**(12) 4612-4

[8] Ma Tiedong, Jiang Weibo, Fu Jie 2012 Impulsive synchronization of fractional order hyperchaotic systems based on comparison system *Acta Physica Sinica* **61**(9) 090503

[9] Li Yan, Chen YangQuan, Podlubny I 2010 Stability of fractional-order nonlinear dynamic systems: Lyapunov direct method and generalized Mittag-Leffler stability *Computers and Mathematics with Applications* **59**(5) 1810-21

[10] Wang Zhen, Sun Wei 2012 Synchronization of fractional chaotic systems and secure communication *Application Research of Computers* **29**(6) 2221-3

[11] Chen Diyi, Chen Haitao, Ma Xiaoyi, Longyan. Hyperchaos system with only one nonlinear term and comparative study of its chaotic control *Journal of Computer Applications* (8) 2045-8

[12] Chen Di-Yi, Shi Lin, Chen Hai-Tao, Ma Xiao-Yi 2012 Analysis and control of a hyperchaotic system with only one nonlinear term *Nonlinear Dynamics* **67**(3) 1745-52

[13] Lin Tsung-Chih, Kuo Chia-Hao 2011 H-infinity synchronization of uncertain fractional order chaotic systems: Adaptive fuzzy approach *ISA Transactions* **50**(4) 548-56

[14] Lin Tsung-Chih, Kuo Chia-Hao, Balas V E 2011 Uncertain fractional order chaotic systems tracking design via adaptive hybrid fuzzy sliding mode control *International Journal of Computers Communications and Control* **6**(3) 418-27

[15] Chen Di-Yi, Liu Yu-Xiao, Ma Xiao-Yi et al 2011 No-chattering sliding mode control in a class of fractional-order chaotic systems*Chinese Physics B* **20**(12) 120506

[16] Chen Diyi, Zhang Runfan, Sprott J C, Chen Haitao, Ma Xiaoyi 2012 Synchronization between integer-order chaotic systems and a class of fractional-order chaotic systems via sliding mode control *Chaos* **22** 023130

[17] Zhang Run-Fan, Chen Di-Yi, Yang Jian-Guo, Juan Wang 2012 No-chattering anti-synchronization for a class of multi-dimensional chaotic systems based on sliding mode with noise *Physica Scripta* **85** 065006

[18] Wu Xiangjun, Li Shanzhi 2012 Dynamics analysis and hybrid function projective synchronization of a new chaotic system *Nonlinear Dynamics* **69**(4) 1979-94

[19] Chai Yuan, Chen Li-Qun 2012 Projective lag synchronization of spatiotemporal chaos via active sliding mode control *Communications in Nonlinear Science and Numerical Simulation* **17**(8) 3390-8

[20] Dormido Sebastian, Pisoni Enrico, Visioli Antonio 2012 Interactive tools for designing fractional-order PID controllers *International*

Journal of Innovative Computing Information and Control **8**(7A) 4579-90
[21] Hu Shaoxiang, Liao Zhiwu, Chen Wufan 2012 Sinogram Restoration for Low-Dosed X-Ray Computed Tomography Using Fractional-Order Perona-Malik Diffusion *Mathematical Problems in Engineering* 391050
[22] Janev Marko, Pilipovic Stevan, Atanackovic Teodor et al. 2011 Fully fractional anisotropic diffusion for image denoising *Mathematical and Computer Modelling* **54**(1-2) 729-41

Authors

Ran Ding, born in April, 1980, Puyang, Henan, China

Current position, grades: PhD student; He received his MS degrees from China University of Mining and Technology, Xuzhou, in electrical engineering; He is a senior engineer.
University studies: China university of mining and technology, in the school of information and electrical engineering
Scientific interest: Power System and Automation
Experience: He is work in State Grid Jiangsu Electic Company currently.

Caoyuan Ma, born on September 17, 1978, Chengde Hebei China

Current position, grades: joined the China University of Mining and Technology faculty in 2001 where he is currently an associate Professor; He received his MS degrees and PhD from China University of Mining and Technology, Xuzhou, all in electrical engineering;
Scientific interest: electrical safety and intelligent control;
Publications: over 10 papers in the journal "Journal of Vibration and Control", "Mining Science and Technology (China)", etc.

Yongyi Zhao, born on April 27, 1989, Fuyang Anhui China

Current position, grades: Post-graduate student of the College of information and electrical engineering, China university of mining and technology
University studies: China university of mining and technology, in the school of information and electrical engineering
Scientific interest: Electrical safety and intelligent control, Power system and automation

Yanfang Luo, born on November 19, 1988, Shangqiu Henan China

Current position, grades: Post-graduate student of the College of information and electrical engineering, China university of mining and technology
University studies: China university of mining and technology, in the school of information and electrical engineering
Scientific interest: Electrical safety and intelligent control, Power system and automation

Chuanlin Wang, born on July 15, 1990, Lu'an Anhui China

Current position, grades: Post-graduate student of the College of information and electrical engineering, China university of mining and technology
University studies: China university of mining and technology, in the school of information and electrical engineering
Scientific interest: Electrical safety and intelligent control

Validation assessment with uncertain model inputs

Liang Zhao*, Zhanping Yang

Institute of Electronic Engineering, China Academy of Engineering Physics, 621900 MianYang, China

Received 14 July 2014, www.cmnt.lv

Abstract

This paper presents a validation assessment method to measure the discrepancies between the model predictions and experimental observations under both aleatory and epistemic uncertainty. The model inputs considered in the paper are sparse point data or interval data, which leads to uncertain parameters for the distribution of the model inputs. A likelihood based method is used to represent the stochastic model inputs and it yields a single probability distribution which integrates the aleatory and epistemic uncertainty of model inputs. This representation of model inputs provides an advantage in computation efficiency for the conventional double loop sampling requirement in uncertainty propagation is collapsed into a single loop sampling. An area based validation metric is extended to compare the probabilistic model predictions obtained from uncertainty propagation with the empirical distribution function of the experimental observations, it reflects an objective quantification of the entire discrepancies between predictions and observations. The confidence interval for the validation metric, which just depends on the amount of experimental observations and confidence level is also developed. A numerical example is used to illustrate the proposed method.

Keywords: validation assessment; likelihood based method; validation metric; confidence interval

1 Introduction

Model based computer is applied in various engineering disciplines, examples range from nuclear reactor certification to understanding of the cosmos. The uncertainty and confidence in model prediction are attracting attention increasingly. Model validation has been advocated as a necessary procedure when the model is used for risk assessment with high consequence systems [1-3]. The fundamental concept of model validation has been intensively argued by professional committees [4-8], and the consensus among researchers is that a rigorous model validation should explicitly account for various uncertainties. The uncertainties can be broadly classified into aleatory and epistemic. For instance, a precise probability distribution indicates the aleatory uncertainty, and the probability distribution with uncertain parameters implies the existence of epistemic uncertainty. This paper concerns the question of how to measure the discrepancies between the model predictions and experimental observations under both aleatory and epistemic uncertainty. The question is usually called "validation assessment" [3], it is a basic process in model validation procedure, the performance of the model can be judged from the measurement.

Several approaches have recently been suggested for validation assessment, including significance testing [9-14], Bayesian method [9, 13, 14], mean based comparison [3, 15, 16] and area based method [17]. All of these approaches have drawbacks. For example, the significance testing to validation is primarily focused on identifying the evidence against a certain hypothesis, and the Bayesian method is mostly focused on the belief that the model is correct. They are rather different from the goal of validation assessment, which is interested in the objective quantification of model accuracy. Instead of making a "accept" or "reject" statement with hypothesis, the mean-based comparison method measures distance between the mean of model predictions and the estimated mean of experimental observations. The limitation of this method is that it only takes the central tendency of predictions and observations into account, while the distribution of predictions contain amount of detail which may be represented insufficiently with a comparison of means. With the aim of measuring the discrepancies of the entire distributions between predictions and observations, the area based method uses the area between the prediction distribution and the observation distribution as a validation metric, but it does not provide a confidence level of the metric due to the lack of sufficient observations. Besides, of the aforementioned drawbacks, there is a serious limitation about theses method that they are only suitable for validation assessment under aleatory uncertainty, and cannot be used directly when model inputs are quantities with epistemic uncertainty.

Typical epistemic uncertainty regarding a model input can be a stochastic quantity with uncertain distribution parameters [18], this is usually due to sparse point data or interval data. Several methods such as evidence theory [19], second order probability method [20], fuzzy sets [21], etc. have been proposed for quantification of the epistemic uncertainty, but the results of these methods which are usually in forms of distribution bounds cannot

* ***Corresponding author*** e-mail swjtu_zhaoliang@126.com

be used in validation assessment directly. This paper extends the area-based validation assessment method to account for epistemic uncertainty arising from sparse or interval data with model inputs. A likelihood based approach [22] is used to construct a single probability distribution for model inputs with uncertain distribution parameters. With the likelihood based representation of uncertainty, the model predictions can be characterized as a single probability distribution after uncertainty propagation, it facilitates the comparison between model predictions and probabilistic experimental observations that are the essential of area based validation metric. The confidence interval of the validation metric is provided by calculating the infimum and supremum of the discrepancy between the cumulative probability distribution function (CDF) of model predictions and the possible experimental empirical distribution functions (EDF) which are bounded by the Kolmogorov - Smirnov limit theorem.

The paper is organized as follows. Section 2 describes the likelihood based method to represent the epistemic uncertainty with an unique probability distribution for model inputs. Section 3 derives the validation metric based on area method, accounting for both aleatory and epistemic uncertainty. Furthermore, this section presents a confidence interval for the validation metric, the interval is associated with the amount of experimental observations. Section 4 demonstrates the proposed method with a numerical example. Finally, Section 5 offers some concluding remarks.

2 Likelihood based representation of epistemic uncertainty

Consider model input X that has a probability density function (PDF) $f_X(x|\boldsymbol{P})$ where $\boldsymbol{P}$ denotes the distribution parameters. With the distribution type is known, the PDF is conditioned on the choice of $\boldsymbol{P}$. The likelihood function $L(\boldsymbol{P})$ is defined as the probability of observing data x given $\boldsymbol{P}$ [22]. When the information regarding X is available with independent point data x_i(i=1,2,⋯,n), the likelihood for $\boldsymbol{P}$ can be calculated as Eq.(1) with a finite precision ε:

$$\begin{aligned} L(\boldsymbol{P}) &\propto \prod_{i=1}^{n} P_i\left(X \in (x_i-\varepsilon, x_i+\varepsilon) \mid \boldsymbol{P}\right) \\ &= \prod_{i=1}^{n} \int_{x_i-\varepsilon}^{x_i+\varepsilon} f_X(x \mid \boldsymbol{P})\mathrm{d}x \\ &\propto \prod_{i=1}^{n} f_X(x_i \mid \boldsymbol{P}) \end{aligned} \quad . \tag{1}$$

Here the independence assumption in Equation (1) is modest in statistical analysis.

Since the likelihood in Equation (1) is actually integrated in an infinitely small interval instead of calculating at data points, it is straightforward to apply the definition to X in the form of interval. When the information regarding X is available with intervals [a_i, b_i] (i=1,2,⋯,m), the likelihood for $\boldsymbol{P}$ of PDF $f_X(x|\boldsymbol{P})$ can be expressed as:

$$\begin{aligned} L(\boldsymbol{P}) &\propto \prod_{i=1}^{m} P_i\left(X \in (a_i, b_i) \mid \boldsymbol{P}\right) \\ &= \prod_{i=1}^{m} \int_{a_i}^{b_i} f_X(x \mid \boldsymbol{P})\mathrm{d}x \end{aligned} \quad . \tag{2}$$

Accounting for both point values and intervals in the available data, the likelihood function of $\boldsymbol{P}$ can be represented as:

$$L(\boldsymbol{P}) \propto \left[\prod_{i=1}^{n} f_X(x_i \mid \boldsymbol{P})\right]\left[\prod_{i=1}^{m} \int_{a_i}^{b_i} f_X(x \mid \boldsymbol{P})\mathrm{d}x\right]. \tag{3}$$

It is popularly known that the $\boldsymbol{P}$ can be estimated by maximizing the likelihood function. However, the validation assessment is interested in the usage of entire likelihood function to construct the PDF of $\boldsymbol{P}$ rather than maximizing the likelihood. Thereby, consider the joint probability density of distribution parameters $\boldsymbol{P}$, denoted by $f_P(\boldsymbol{P})$, it can be calculated as Equation (4) using Bayes theorem with an uniform prior PDF $f_P'(\boldsymbol{P})=h$ (over the whole range of $\boldsymbol{P}$).

$$f_{\boldsymbol{P}}(\boldsymbol{P}) = \frac{L(\boldsymbol{P})h}{\int L(\boldsymbol{P})h\mathrm{d}\boldsymbol{P}} = \frac{L(\boldsymbol{P})}{\int L(\boldsymbol{P})\mathrm{d}\boldsymbol{P}}. \tag{4}$$

After calculating the $f_P(\boldsymbol{P})$, the PDF for X can be constructed based on principles of conditional probability and total probability, as represented in Equation (5).

$$f_X(x) = \int f_X(x \mid \boldsymbol{P}) f_{\boldsymbol{P}}(\boldsymbol{P})\mathrm{d}\boldsymbol{P}. \tag{5}$$

Let Y denotes the model predictions, the PDF of Y represented by $f_Y(y)$ can be obtained using uncertainty propagation analysis. The single PDF of X facilitates the uncertainty propagation for it requires just one level of Monte Carlo sampling to construct the $f_Y(y)$. In contrast, the conventional uncertainty quantification methods such as second order probability method [20] require a two level Monte Carlo sampling: first, draw samples of $\boldsymbol{P}$ from $f_P(p)$, each of them determines a PDF of X: $f_X(x|\boldsymbol{P})$; second, several samples from each $f_X(x|\boldsymbol{P})$ are drawn to calculate a distribution of Y, the process generates a family of $f_Y(y)$ for different $\boldsymbol{P}$ ultimately. The two level Monte Carlo sampling strategy is so computationally costly that it might not always be affordable in practical model application. Furthermore, the family of model output distributions leads to a difficulty in measuring discrepancies between model predictions and experimental observations. The $f_X(x)$ in Eq.(5) can be interpreted as the expected value of $f_X(x|\boldsymbol{P})$, which depends on the choice of $\boldsymbol{P}$. The two level of uncertainties considered in second order probability method are integrated into the single PDF $f_X(x)$, and it can be evaluated numerically. Based on the single PDF for

model input, the $f_Y(y)$ can be calculated using a straightforward Monte Carlo method. The resultant single PDF for model prediction brings an advantage in validation assessment. The following section implements the area based method [17] to measure the discrepancies between model predictions and experimental observations. With the likelihood based method explained in section 2, the area based method can account for both aleatory uncertainty and epistemic uncertainty in validation assessment.

3 Validation assessment using area based method

3.1 AREA METRIC

The aforementioned model prediction can be characterized as a CDF, represented as $F_Y(y)$. Consider the experimental observations are provided as a set of point data y_i, (i=1,2,⋯), the EDF for the data set is represented as $S(y)$, it preserves almost all statistical information in the data set The area based method proposed by Ferson [18] uses the area between the $F_Y(y)$ and the $S(y)$ as the measurement of the discrepancies. The mathematical expression for the area metric can be written as Equation (6):

$$d\left(F_Y, S\right) = \int_{-\infty}^{+\infty} \left|F_Y(y) - S(y)\right| \mathrm{d}y \,, \tag{6}$$

The $d(F_Y, S)$ in Eq.(6) is essentially a special case of the Wasserstein distance. The important merit of the area metric is that $d(F_Y,S)$ measures the discrepancies between the entire distributions from predictions and observations. Since the epistemic uncertainty and aleatory uncertainty in the model inputs have already been quantified with the single PDF $f_X(x)$, the area based method which makes use of the $f_X(x)$ can be applicable in the validation assessment with both epistemic uncertainty and aleatory uncertainty.

3.2 CONFIDENCE INTERVAL FOR THE AREA METRIC

Consider the sample uncertainty in experimental observations, the Kolmogorov-Smirnov statistics can be used to bound the EDF of observations as Equation (7):

$$\begin{aligned} \overline{S}(y) &= \min(1, \max(0, S(y) + D)) \\ \underline{S}(y) &= \min(1, \max(0, S(y) - D)) \end{aligned}, \tag{7}$$

Where $\overline{S}(y)$ and $\underline{S}(y)$ refer to the upper bound and lower bound of the experimental EDF respectively, and the D denotes the critical value for the Kolmogorov-Smirnov statistic [23], it just depends on the confidence level and the amount of experimental observations. After calculating the $\overline{S}(y)$ and $\underline{S}(y)$ with specified confidence level (1-α)×100%, the confidence interval for the area metric can be expressed as Equation (8):

$$\left[\inf_{S(y)\in K} \int_{-\infty}^{+\infty} \left|F_Y(y) - S(y)\right| \mathrm{d}y, \sup_{S(y)\in K} \int_{-\infty}^{+\infty} \left|F_Y(y) - S(y)\right| \mathrm{d}y \right] , \tag{8}$$

where inf denotes the infimum, sup denotes the supremum, and the K represents the region encompassed by $\overline{S}(y)$ and $\underline{S}(y)$, it contains all possible EDFs of the experimental observations. The task of calculating the infimum or supremum taking the amount of experimental observations into account is sometimes challenging, intelligence optimization algorithms such as genetic algorithm can be used to handle this problem. The genetic algorithm randomly generates a population of possible experimental EDF constrained by Equation (7), then evaluates the fitness of each EDF using the area metric. For the infimum calculation task, a small area between possible EDF and model prediction CDF indicates high fitness for this EDF, whereas the supremum calculation task pursues large area as high fitness. The set of parents forming the next generation can be selected based on the fitness of individual EDF, where the high fitness members have more chances of being chosen. The mutation and crossover operation are applied to the parents to create next generation of possible experimental EDF. The selection, mutation and crossover are repeated until the maximum fitness in the population meets the criterion.

The area metric provides a quantitative measurement of the discrepancies between model predictions and experimental observations. The choice of the threshold for the metric is another question in model validation, which is usually called “adequacy decision” [3]. The adequacy decision considers that whether the validation metric is “great” enough to draw a conclusion that the discrepancies are significant. This question is derived from the concept that the threshold should be separated from the validation metric [9, 16, 17]. If the area metric is presented below the chosen threshold, one can conclude that the model is adequate for the intended use. Furthermore, consider the sampling uncertainty of the experimental observations, the upper bound of the confidence interval for area metric can be used to compare with the threshold, it claims that whether the model is adequate with (1-α)×100% confidence. Usually, the threshold quite depends on the intended use of the model for different requirement of model accuracy. Note that the distribution of model predictions calculated on the basis of $f_X(x)$ is not parametrically available, and the value of $d(F_Y,S)$ is expressed in physical units rather than statistical units. It implies that the engineers or project managers may be more rational than the mathematicians who developed the validation assessment method to decide the choice of the threshold. As stated in section 1 earlier, the paper concerns only the question of how to measure the discrepancies between the model predictions and experimental observations, the choice of the threshold for the metric is out of the research scope.

4 Numerical example studies

A numerical example is presented in this section to illustrate the proposed validation assessment method. Consider a material layer slab with thickness L, volumetric heat capacity ρC and thermal conductivity k, it is exposed to a heat flux, the temperature (T) at a specified time (t) after exposure to the heat flux (q) needs to be predicted. The model for the temperature prediction can be written as:

$$T(k,\rho C,T_0,L,q;l,t) = T_0 + \frac{qL}{k}\left[\frac{kt/\rho C}{L^2} + \frac{1}{3} - \frac{l}{L} + \frac{l^2}{2L^2} - \sum_{n=1}^{6}\frac{2}{\pi^2 n^2}\exp\left(-\frac{n^2\pi^2 kt}{L^2\rho C}\right)\cos\left(n\pi\frac{l}{L}\right)\right], \quad (10)$$

where T_0 is the initial ambient temperature, it is fixed at 25°C, l is the location within the material. The controllable inputs (L, q) are assumed to be known exactly, i.e. L=0.0127m, q=1000W/m^2, while the (ρC, k) are uncertain inputs. For the sake of illustration, suppose that the probability distribution of the volumetric heat capacity ρC and thermal conductivity k are described as normal by experts but with uncertain distribution parameters. The observations of ρC are available as three point data {4.52E+05J/m^3°C, 4.10E+05J/m^3°C, 4.02E+05 J/m^3°C}, and the data about k is available in the form of intervals as [0.0601W/m°C, 0.0604W/m°C], [0.0545 W/m°C, 0.0547W/m°C]. Let the purpose of the model be to predict the temperature at the surface of the layer (l=0m) after exposure to the heat flux for 500 seconds (t=500s). Experiment in the circumstance consistent with the intended use of the model is carried out four times repeatedly, the observations are four point data as {210.6°C, 214.6°C, 186.9°C, 219.0°C}. It is required to measure the discrepancies between the model predictions and experimental observations.

Consider that model inputs ρC and k are described with epistemic uncertainty, the likelihood based approach explained in section 2 is used to represent the two model inputs with single probability distribution respectively. Note that the normal distribution describing the model inputs ρC and k is parameterized by mean and standard deviation, so the likelihood in Equation (3) is integrated for the mean and the standard deviation together. The entire admissible ranges for the mean and the standard deviation are both $(0,+\infty)$, for the sake of calculation, here we draw finite bounds for the two ranges. For instance, assume that the mean of ρC can vary in (2.5E+05, 5.5E+05). Based on the observations we believe the finite ranges are still wide enough to cover any possible value of the distribution parameters. By combining Equation (4) and (5), the distributions of the ρC and k are calculated as it is shown in Figure 1 and Figure 2.

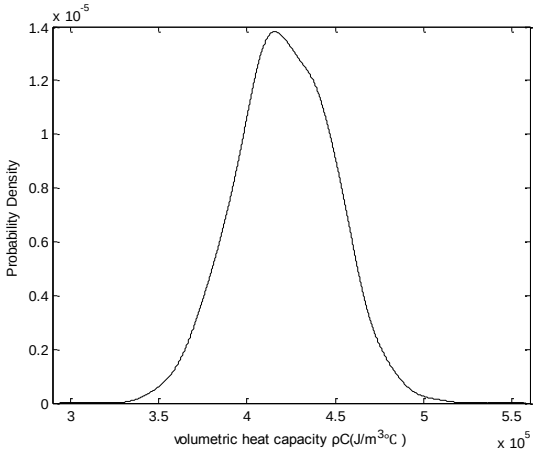


FIGURE 1 Distribution of volumetric heat capacity ρC

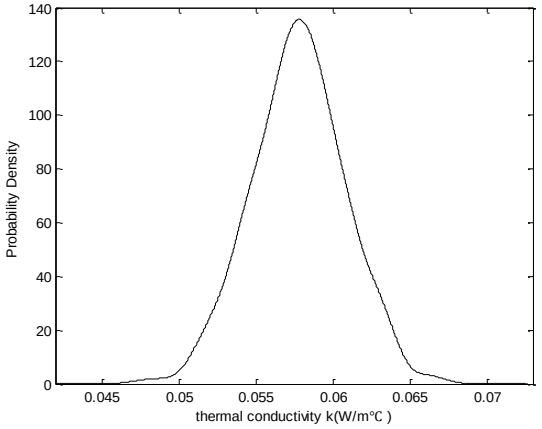


FIGURE 2 Distribution of thermal conductivity k

The distributions presented in Figure 1 and Figure 2 include both epistemic uncertainty and aleatory uncertainty in ρC and k. Note that the integration in Equation (5) is calculated numerically for $f_P(\boldsymbol{P})$ is not parametrically available, so the resultant PDFs for ρC and k are not analytical. An Monte Carlo method is used to propagate the uncertainty of ρC and k through the thermal model in Equation (9), the PDF of the material layer surface temperature at t=500s is shown in Figure 3.

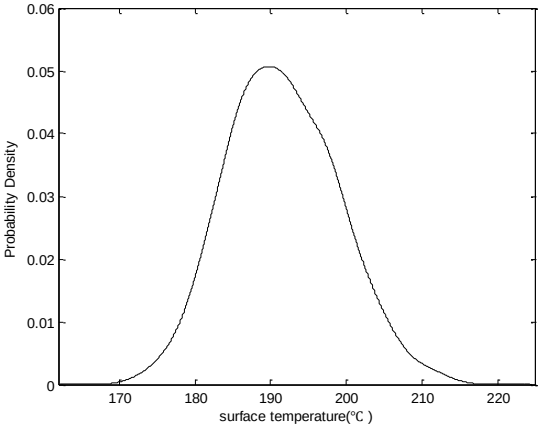


FIGURE 3 Distribution of model predictions

The area metric is used to measure the discrepancies between model predictions and experimental observations. Figure 4 illustrates this area with shaded region between smooth black curve and blue step function. The black curve is the CDF of the model predictions, and the blue step function is the EDF for experimental observations consisting of four point value.

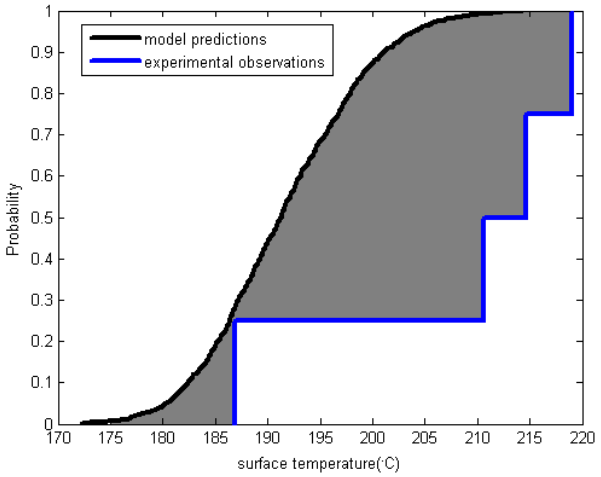


FIGURE 4 EDF of experimental observations against CDF of model predictions

Using Equation (6), the area metric is obtained to be 16.21ºC. A genetic algorithm is employed to find the infimum and supremum defined in Equation (8) to get the confidence interval for the area metric. The exact value of the confidence interval with 95% confidence level is [4.13ºC, 26.88ºC]. The result of the validation assessment provides an objective quantification of the discrepancies between model predictions and experimental observations. As stated early, the threshold for the metric depends on the specific requirement of model accuracy. Here for the sake of illustration, assume the threshold for the discrepancies is 2ºC. Since the lower bound of the metric's confidence interval is beyond the threshold, we can conclude with 95% confidence that the accuracy of the model is insufficient, it indicates that the model needs to be improved.

5 Conclusions

This paper presents a framework for validation assessment when there is both aleatory and epistemic uncertainty in model inputs. The likelihood based method is applied to representation of stochastic quantities with uncertain distribution parameters which are due to sparse point data or interval data. The method's result for an uncertain model input is a single probability distribution that facilitates the following uncertainty propagation and validation metric. It provides an obvious advantage in computation efficiency for the conventional double loop sampling strategy is collapsed into a single loop sampling. The probability distribution of model output obtained from uncertainty propagation is compared with the EDF of experimental observations using area based validation metric, it reflects an objective quantification of the entire discrepancies between predictions and observations. A confidence interval for the validation metric which just depends on the amount of experimental observations and confidence level is also developed. It is helpful for the decision making which follows the validation assessment. The numerical example demonstrates the validation assessment framework presented in this paper.

The discussion of the framework in this paper is limited to the univariate case, which implies the model output to be a single response quantity following a statistical distribution. The validation assessment also looks for a metric, which have the flexibility of measuring the discrepancies between predictions and observations in a multivariate case. There are several types of multivariate case such as multiple location for model response or various response quantities at a single location. The validation metric for the multivariate case is required to provide an overall performance measurement for the model. There are still difficulties in aggregating individual metrics accounting for confidence level and correlation among multiple quantities. Future work in this direction will extend the validation assessment to multivariate case.

References

[1] Romero V J 2006 Some issues and needs in quantification of margins and uncertainty in complex coupled systems *47th AIAA/ ASME/ ASCE/ AHS/ ASC Structures, Structural Dynamics, and Materials Conference* Newport

[2] Helton J 2009 *Conceptual and Computational Basis for the Quantification of Margins and Uncertainty* Sandia National Laboratories: Albuquerque

[3] Oberkampf W L, Trucano T G 2007 *Verification and Validation Benchmarks* Sandia National Laboratories: Albuquerque

[4] American Institute of Aeronautics and Astronautics 1998 *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations* AIAA: Reston VA

[5] The American Society of Mechanical Engineers 2008 *ASME V&V 20-2008 Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer* ASME: New York

[6] National Aeronautics and Space Administration 2008 *NASA-STD-7009 Standard for Models and Simulations* NASA: Washington DC

[7] Trucano T G, Swiler L P, Igusa T, Oberkampf W L, Pilch M 2006 Calibration, validation, and sensitivity analysis: What's what *Reliab Eng Syst Safety* **91**(10-11) 1331-57

[8] Roy C J, Oberkampf W L 2011 A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing *Comput Method Appl M* **200**(25-28) 2131 – 44

[9] Liu Y, Chen W, Arendt P and Huang H Z 2011 Toward a better understanding of model validation metrics *J Mech Design* **133**(7) 48-60

[10] Hills R G, Trucano T G 1999 *Statistical Validation of Engineering and Scientific Models: Background* Sandia National Laboratories: Albuquerque

[11] Hills R G, Trucano T G 2002 *Statistical Validation of Engineering and Scientific Models: A Maximum Likelihood Based Metric* Sandia National Laboratories: Albuquerque

[12] Buranathiti T, Cao J, Chen W 2006 Approaches for Model Validation: Methodology and Illustration on a Sheet Metal Flanging Process *J Manuf Sci E* **128**(2) 588-97

[13] Rebba R, Mahadevan S 2006 Validation of models with multivariate output *Reliab Eng Syst Safety* **91**(8) 861-71
[14] Rebba R, Mahadevan S 2008 Computational Methods for Model Reliability Assessment *Reliab Eng Syst Safety* **93**(8) 1197-207
[15] Oberkampf W L, Trucano T G, Hirsch C 2004 Verification, Validation, and Predictive Capability in Computational Engineering and Physics *Appl Mech Rev* **57**(3) 345-84
[16] Oberkampf W L, Barone M F 2006 Measures of agreement between computation and experiment: validation metrics *J Comput Phys* **217**(1) 5-36
[17] Ferson S, Oberkampf W L, Ginzburg L 2008 Model validation and predictive capability for the thermal challenge problem *Comput Method Appl M* **197**(29-32) 2408-30
[18] Baudrit C, Dubois D 2006 Practical representations of incomplete probabilistic knowledge *Computational Statistics and Data Analysis* 86-108
[19] Agarwal H, Renaud J E, Preston E L and Padmanabhan D 2004 Uncertainty quantification using evidence theory in multidisciplinary design optimization *Reliab Eng Syst Safety* **85**(1-3) 281-94
[20] Swiler L P, Giunta A A 2007 *Aleatory and epistemic uncertainty quantification for engineering applications* Sandia National Laboratories: Albuquerque
[21] Rao S, Annamdas K 2009 An evidence based fuzzy approach for the safety analysis of uncertain systems *50th AIAA/ ASME/ ASCE/ AHS/ ASC Structures, Structural Dynamics, and Materials Conference* California
[22] Sankararaman S, Mahadevan S 2011 Likelihood based representation of epistemic uncertainty due to sparse point data and/or interval data *Reliab Eng Syst Safety* **96**(7) 814-24
[23] Miller L H 1956 Table of percentage points of Kolmogorov statistics *Journal of the American Statistical Association* **51** 111-21

Authors

Liang Zhao, born in 1983, MeiShan, China

Current position, grades: Ph.D candidate
University studies: Computer science in Southwest Jiaotong University
Scientific interest: Quantification of uncertainty and model validation.
Publications: 5
Experience: 2011-2014: Ph.D in system science, China Academy of Engineering Physics; 2009-2011: System engineer in JEZETEK; 2006-2009: Master in computer science, Southwest Jiaotong University

Zhanping Yang, born in 1966, ChongQing, China

Current position, grades: Researcher, Chief Engineer of Institute of Electronic Engineering
University studies: System science in Beijing Institute of Technology
Scientific interest: Modelling and assessment of complex system
Publications: 27
Experience: 2001-2014: Electronic system design in Institute of Electronic Engineering; 1998-2001: Ph.D in system science, Beijing Institute of Technology

Fuzzy set and rough set based evaluation algorithm of web customers

Yun Peng*

Computer and Information Engineering College, Jiangxi Normal University, Nanchang 330022, China

Received 1 July 2014, www.cmnt.lv

Abstract

A fuzzy algorithm of web customers evaluation based on rough set is presented. Key attributes can be gotten through rough set. The evaluation from the data objects based on key attributes can reduce the data size and algorithm complexity. After Clustering analysis of customers, then the evaluation analysis will process to the clustering data. There are a lot of uncertain data in customer clusters, so the traditional method of classification and evaluation to the incomplete data is very difficult. Superposition evaluation algorithm based on fuzzy set can improve the reliability and accuracy of web customer evaluation. Evaluation of the web customer also can improve efficiency, service quality and profitability of web businesses.

Keywords: Rough Set, Membership Function, Cluster Analysis, Fuzzy Set

1 Introduction

Following the rapid development of electronic commerce, which brought a huge number of web customer groups, and how to classify and evaluate the customers is very important. According to the customer's transactions, characteristics of the customer information, customers can be classified and evaluated and given the levels. Web services can be improve efficiency in the future based on the actual conduct of targeted customer, and increase customer attention and interest level, thereby enhancing the profitability of web. Because of the ambiguity of customer behaviour, a comprehensive evaluation is designed based on fuzzy superposition algorithm in this paper [1].

2 The proposed algorithm

Customer's attribute can be web transactions based on behaviour characteristics, also it can be static characteristics of the customer, and the attribute can be looked as the basis for cluster analysis. In the classification process, the keywords are divided into different classes for different levels of customer types. Membership function will convert all the data values less than or equal to 1, so the next step of the fuzzy clustering can be carried out [2].

Fuzzy set A in the domain U, with a membership function to describe it, namely: $U \rightarrow [0,1]$, for any $u \in U$, there is $u \rightarrow \mu_{\tilde{A}}(u)$, $\mu_{\tilde{A}}(u) \in [0,1]$, $\mu_{\tilde{A}}(u)$ is the membership of element u to set A, which represents the degree of u belonging to A. The designing of fuzzy superposition evaluation algorithm is shown as follows [3].

(1) Let U be the customer domain, U_i is the No.i customer, $i \in 1,2,3 \ldots n$, A_j is the No.j attribute of U, $j \in 1,2,3 \ldots m$, S_{ij} is the value of U_i and A_j, $P(K_j)$ is the weighting coefficients of A_j. The membership function of keyword property value is shown in (1).

$$\mu_{\tilde{A}}(S_{ij}) = F(S_{ij})/(S_{1j} + S_{2j} \ldots S_{nj}) * P(K_j). \quad (1)$$

After pre-treatment customers, we can design an appropriate fuzzy clustering algorithm of customer data.

(2) Set up the fuzzy similarity relation R of U. The order of R matrix is $|U|$, m is the number of attributes. Using Euclidean distance formula shown in (2), we can calculate the matrix elements of r_{ij} in R [4].

$$r_{ij} = \begin{cases} 1 & i = j \\ \sqrt{\frac{1}{m}\sum_{k=1}^{m}(sik - sjk)^2} & i \neq j \end{cases}. \quad (2)$$

(3) The graph $G = (V, E)$ can be obtained by R, and the maximum spanning tree as $T = (V, TE)$ from G can be calculated using Prim algorithm.

(4) According to the practical problems to set an appropriate $\lambda \in [0,1]$, $T(e)$ is the weight of edge e, if $T(e) < \lambda$, edge e will be removed, the connected component is the classification based on λ.

(5) The set of attributes reduction can be derived from discernibility matrix and discernibility function [5]. Discernibility matrix is shown in (3).

$$M(B) = \{m(i, j) | n \times n, 1 \leq i, j \leq n\}, \quad (3)$$

* ***Corresponding author*** e-mail 332952797@qq.com

where $m(i, j) = \{a \in A \mid a(i) \neq a(j) \text{ and } d(i) \neq d(j)\}$, $n = |U|$.

Discernibility function is shown in (4). Σ is "∨ ", Π is "∧ ".

$$\Delta = \prod_{(i,j)\in U\times U} \sum m(i,j) \quad . \tag{4}$$

(6) Attribute reduction and nuclear can be derived from the minimal disjunctive of distinction function, which can deduce an attributes set of critical evaluation [6].

(7) Set up evaluation set $V = \{v_1, v_2, ...v_m\}$, the weight distribution of U_i is: $\underset{\sim}{A_i} = \{a_{i1}, a_{i2}, \ldots, a_{in_i}\}$. The evaluation factors are the fuzzy mapping from U to $F(V)$ shown in (5).

$$\begin{aligned} &\underset{\sim}{f}: U \to F(V), \forall u \in U \\ &u_i \mapsto \underset{\sim}{f}(u_i) = \frac{r_{i1}}{v_1} + \frac{r_{i2}}{v_2} + \cdots + \frac{r_{im}}{v_m}, \end{aligned} \tag{5}$$

where $0 \le r_{ij} \le 1$, $1 \le i \le n$, $1 \le j \le m$

$\underset{\sim}{R_i}$ is the single factor evaluation matrix of U_i, so the first-class comprehensive evaluation is as follows: $\underset{\sim}{R_i} = \underset{\sim}{A_i} \circ \underset{\sim}{R_i} = (b_{i1}, b_{i2}, ..., b_{im})$, $(i = 1,2, ...s)$.

(8) As an element for each U_i, using $\underset{\sim}{B_i}$ as its single factor assessment, the evaluation matrix can be derived.

(9) According to some property of a higher level, a subset S can be divided into more advanced sub-set, then return step (7) and (8). Finally, we can constitute a multi-class fuzzy evaluation.

3 Example analysis

According to the type of attribute value and the membership function, we can calculate the membership degree of each attribute, and the membership value is as the initial value for classification. Original customer data is shown in Table 1, where C is the customer, CA is the transaction attribute [7].

TABLE 1 Original customer data

C	CA_1	CA_2	CA_3	CA_4	CA_5
C_1	322	524	401	112	123
C_2	121	510	502	142	280
C_3	126	228	224	89	118
C_4	210	217	565	321	217
C_5	260	162	865	212	322
C_6	215	159	625	112	265
C_7	129	195	587	219	285
C_8	139	112	495	120	279
C_9	562	255	750	250	105
C_{10}	409	596	362	201	419

(1) Through the mapping of membership function, data can be initialized shown in Table 2, the values are changed to the values less than or equal to 1, and the values reflect the dependence of the attribute.

TABLE 2 Initialization of customer data

C	CA_1	CA_2	CA_3	CA_4	CA_5
C_1	0.134	0.140	0.401	0.500	0.390
C_2	0.343	0.144	0.300	0.257	0.144
C_3	0.257	0.451	0.500	0.457	0.310
C_4	0.544	0.457	0.400	0.357	0.229
C_5	0.457	0.467	0.067	0.487	0.300
C_6	0.450	0.057	0.500	0.257	0.320
C_7	0.257	0.437	0.520	0.457	0.400
C_8	0.542	0.400	0.184	0.400	0.457
C_9	0.434	0.300	0.540	0.460	0.387
C_{10}	0.544	0.500	0.330	0.213	0.420

(2) Fuzzy similarity relation matrix R shown in Table 3 can be calculated using Euclidean.

TABLE 3 Fuzzy similarity relation matrix R

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
C_1									
C_2	0.032								
C_3	0.213	0.245							
C_4	0.273	0.216	0.130						
C_5	0.139	0.130	0.270	0.230					
C_6	0.247	0.237	0.126	0.140	0.270				
C_7	0.232	0.242	0.089	0.093	0.255	0.089			
C_8	0.052	0.079	0.223	0.234	0.133	0.257	0.251		
C_9	0.186	0.191	0.073	0.099	0.260	0.126	0.145	0.211	
C_{10}	0.245	0.235	0.090	0.093	0.258	0.066	0.042	0.138	0.147

(3) The graph $G = (V, E)$ can be derived by R, and the maximum spanning tree $T = (V, TE)$ from G can be calculated using Prim algorithm, where $|V| = 10$, $|TE| = 9$, the maximum spanning tree is shown in Figure 1.

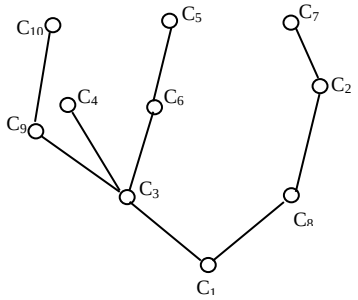


FIGURE 1 The maximum spanning tree

The three clusters above is in accordance with the feature of customers through the data analysis. Sometimes the λ is a variable value, we can draw the dynamic graph by the max-tree. The dynamic classification graph is as Figure 2.

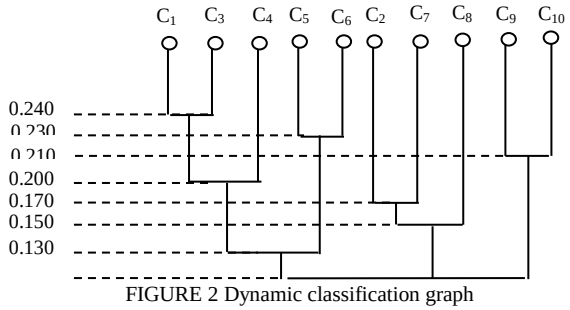


FIGURE 2 Dynamic classification graph

Set $\lambda = 0.130$, customer classification can be obtained as follows:

C1 = {C_1, C_3, C_4, C_5, C_6 }, C2 = {C_2, C_7, C_8}, C3 = {C_9,C_{10}}.

(4) The distinction matrix is shown in Table 4.

TABLE 4 Distinction matrix

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
C_1									
C_2	2,5								
C_3	3,4	4,5							
C_4	3	2,4,5	4						
C_5	1,2,3	1,2,3,4	1,3,4	1,2,3,4					
C_6	2,5	1,2,4,5	1,2,4	1,2,4,5	3,4				
C_7	2,4	2,3,5	1,2	1,2	3,4,5	4,5			
C_8	1,2	2,3.4,5	2,4	2,4	1,2,3,4,5	1,2,5	1,2,4		
C_9	2,3,4	1,2,3,5	1,2	1,2	3,4,5	4,5	2	3.4,5	
C_{10}	1,3,4,5	2,3,4,5	2,3,5	1,4,5	1,3	3,4	2,4,5	2,5	1,3,5

According to Table 4, we can get distinction function as follows:

Δ=($CA_2 \vee CA_5$)$\wedge$ ($CA_3 \vee CA_4$)$\wedge$ ($CA_4 \vee CA_5$)$\wedge CA_3 \wedge$ ($CA_2 \vee CA_4 \vee CA_5$)$\wedge CA_4 \wedge$ ($CA_1 \vee CA_2 \vee CA_3$)$\wedge$ ($CA_1 \vee CA_2 \vee CA_3 \vee CA_4$)$\wedge$ ($CA_1 \vee CA_2 \vee CA_3 \vee CA_4$)$\wedge$ ($CA_1 \vee CA_2 \vee CA_4 \vee CA_5$)$\wedge$ ($CA_1 \vee CA_2 \vee CA_4$)$\wedge$ ($CA_2 \vee CA_4$)$\wedge$ ($CA_2 \vee CA_3 \vee CA_5$)$\wedge$ ($CA_1 \vee CA_2$)$\wedge$ ($CA_3 \vee CA_4 \vee CA_5$)$\wedge$ ($CA_2 \vee CA_3 \vee CA_4 \vee CA_5$)$\wedge$ ($CA_1 \vee CA_2 \vee CA_3 \vee CA_4 \vee CA_5$)$\wedge$ ($CA_1 \vee CA_2 \vee CA_5$)$\wedge$ ($CA_2 \vee CA_3 \vee CA_4$)$\wedge$ ($CA_1 \vee CA_2 \vee CA_3 \vee CA_5$)$\wedge CA_2 \wedge$ ($CA_1 \vee CA_3 \vee CA_4 \vee CA_5$)$\wedge$ ($CA_1 \vee CA_4 \vee CA_5$)$\wedge$ ($CA_1 \vee CA_3$)$\wedge$ ($CA_2 \vee CA_5$)$\wedge$ ($CA_1 \vee CA_3 \vee CA_5$)

$=CA_1 \wedge CA_4 \wedge CA_5$

After reduction we can get the key attributes as {CA_1, CA_4, CA_5} and the second layer attribute as {CA_2, CA_3}. A two-tiered evaluation system for fuzzy evaluation can carried out then. To analyse the evaluation system conveniently, we simplify the system accordingly, but does not affect the algorithm analysis [5].

(5) Customer evaluation grades are divided into four grade (excellent, good, middle, bad). The weight of first-class is: $\underset{\sim}{W}$ ={0.30,0.40,0.20,0.10}.

The weight of second-class is:

$\underset{\sim}{W}_1$ ={0.35,0.45,0.20};

$\underset{\sim}{W}_2$ ={0.30,0.25,0.10,0.20,0.15};

$\underset{\sim}{W}_3$ ={0.45,0.30,0.25};

$\underset{\sim}{W}_4$ ={0.32,0.30,0.20,0.18}.

The weight of third-class is:

$\underset{\sim}{W}_{11}$={0.50,0.30,0.20};

$\underset{\sim}{W}_{12}$={0.30,0.20,0.15,0.25,0.10}.

(6) The fuzzy evaluation matrix is shown in (6) and fuzzy evaluation result of second-class is shown in (7).

$$\underset{\sim}{R}_{11} = \begin{bmatrix} 0.25 & 0.45 & 0.20 & 0.10 \\ 0.30 & 0.20 & 0.35 & 0.15 \\ 0.28 & 0.22 & 0.39 & 0.11 \end{bmatrix} \quad (6)$$

$$\underset{\sim}{B}_{11} = \underset{\sim}{W}_{11} \circ \underset{\sim}{R}_{11} = \underset{\sim}{B}_{11}$$

$$\begin{bmatrix} (0.50 \wedge 0.25) \vee (0.30 \wedge 0.30) \vee (0.20 \wedge 0.28) \\ (0.50 \wedge 0.45) \vee (0.30 \wedge 0.20) \vee (0.20 \wedge 0.22) \\ (0.50 \wedge 0.20) \vee (0.30 \wedge 0.35) \vee (0.20 \wedge 0.39) \\ (0.50 \wedge 0.10) \vee (0.30 \wedge 0.15) \vee (0.20 \wedge 0.11) \end{bmatrix} = (0.30,0.45,0.30,0.15),$$

the normalized result is obtained: = (0.25,0.40,0.25,0.10).

Using the same algorithm can be obtained:

$\underset{\sim}{B}_{12}$ = (0.30,0.40,0.20,0.10);

$\underset{\sim}{B}_{13}$= (0.25,0.35,0.20,0.20).

$$\underset{\sim}{B}_1 = \underset{\sim}{W}_1 \circ \underset{\sim}{R}_1 = \begin{bmatrix} (0.35 \wedge 0.25) \vee (0.45 \wedge 0.30) \vee (0.20 \wedge 0.25) \\ (0.35 \wedge 0.40) \vee (0.45 \wedge 0.40) \vee (0.20 \wedge 0.35) \\ (0.35 \wedge 0.25) \vee (0.45 \wedge 0.20) \vee (0.20 \wedge 0.20) \\ (0.35 \wedge 0.10) \vee (0.45 \wedge 0.10) \vee (0.20 \wedge 0.20) \end{bmatrix} = , \quad (7)$$

$$(0.30, 0.40, 0.25, 0.20)$$

The normalized result is obtained:

$\underset{\sim}{B}_1$ = (0.25,0.35,0.25,0.15).

Using the same algorithm can be obtained:

$\underset{\sim}{B}_2$ = (0.25,0.35,0.20,0.20);

$\underset{\sim}{B}_3$ = (0.30,0.45,0.10,0.15);

$\underset{\sim}{B}_4$ = (0.25,0.30,0.20,0.25).

(7) According to the second-class evaluation results matrix, first-class level evaluation shown in (8) can be derived.

$$\underset{\sim}{B} = \underset{\sim}{W} \circ \underset{\sim}{R} = \begin{bmatrix} (0.30 \wedge 0.25) \vee (0.40 \wedge 0.25) \vee (0.20 \wedge 0.30) \vee (0.10 \wedge 0.25) \\ (0.30 \wedge 0.35) \vee (0.40 \wedge 0.35) \vee (0.20 \wedge 0.45) \vee (0.10 \wedge 0.30) \\ (0.30 \wedge 0.25) \vee (0.40 \wedge 0.20) \vee (0.20 \wedge 0.10) \vee (0.10 \wedge 0.20) \\ (0.30 \wedge 0.15) \vee (0.40 \wedge 0.20) \vee (0.20 \wedge 0.15) \vee (0.10 \wedge 0.25) \end{bmatrix} = . \quad (8)$$

$$(0.25, 0.35, 0.25, 0.20)$$

The normalized result: $\underset{\sim}{B}$ = (0.24,0.34,0.24,0.18).

Based on the maximization of fuzzy set membership, the evaluation results of customer class C1 can be rated as “good”.

4 Conclusions

The evaluation of web customers adopts multi-class structure and takes the static and dynamic attributes into account. To delete the redundant attributes and improve the computation efficiency, attributes reduction is necessary. The evaluation result is an important reference for commercial distribution of web businesses. The weight factors and the fuzzy evaluation algorithm can be adjusted according to the actual customer data. In designation of superposition algorithm to evaluate web customers, compliance with requirements of enterprise, true reflection of the needs and behaviour of customers and adaptive algorithm for data processing are all to be considered. Tested by actual data analysis, cluster

analysis can reduce the size of customer data and data noise, and the key class evaluation analysis can improve the evaluation efficiency of web customers.

Acknowledgments

This work is supported by Social Science Planning Project of Jiangxi Province under grant No. 13TQ08, and Science Research Project of Jiangxi Normal University under grant No.4546.

References

[1] Kopylov A V, Moskovtsev A F 2013 The model of assessment of strategic resources of industrial enterprise on the basis of fuzzy set theory *World Applied Sciences Journal* **23**(1) 37-45
[2] Deng-Neng Chen, Jeng B, Wei-Po Lee, Cheng-Hung Chuang 2008 An agent-based model for consumer-to-business electronic commerce *Expert Systems with Applications* **34**(1) 469-81
[3] Yan Zou, Zhi Xiao 2008 Data analysis approaches of soft sets under incomplete information *Knowledge-Based Systems* **21**(8) 941-5
[4] Cho J, Lee J 2013 Development of a new technology product evaluation model for assessing commercialization opportunities using Delphi method and fuzzy AHP approach *Expert Systems with Applications* **40**(13) 5314-30
[5] Jia X, Tang Z, Liao W, et al. 2014 On an optimization representation of decision-theoretic rough set model *International Journal of Approximate Reasoning* **55(**1) 156-66
[6] Li B, Tang P, Chow T W S 2013 Quantization of rough set based attribute reduction *Journal of Software Engineering and Applications* **5**(12) 117-23
[7] Eid M I 2011 Determinants of web customer satisfaction, trust, and loyalty in Saudi Arabia *Journal of Electronic Commerce Research* **12**(1) 78-93

Authors

Yun Peng, born on October 27, 1972, China

Current position, grades: a Lecturer from 2002 to 2007 and currently an Associate Professor in the Computer and Information Engineering College at Jiangxi Normal University
University studies: B.E. degree in electronic engineering from Nanchang University, Nanchang, P.R.China, in 1994 and the M.E. degree in computer software and theory from Jiangxi Normal University
Experience: He is a member of CCF and ACM. His research interests include artificial intelligence, data mining and software engineering.

Coordinated scheduling on single serial batching machine with transportation considerations

Cunchang Gu[1, 2*], Xiaoyan Xu[2]

[1]*School of Management, Qufu Normal University, Rizhao 276826, China*

[2]*College of Science, Henan Universities of Technology, Zhengzhou 450001, China*

Received 15 May 2014, www.cmnt.lv

Abstract

The coordination of production scheduling and transportation has recently received a lot of attention in logistics and supply chain management. We study a coordinated scheduling problem, in which each job is transported to a single serial batching machine for further processing, each batch to be processed occurs a processing cost, and the objective is minimizing the sum of the makespan and the total processing cost. Under the condition of the jobs' processing times are equal, if the job assignment to the vehicles is predetermined, we provide a polynomial time dynamic programming algorithm, for the general problem, we prove it is NP-hard. When the returning time of vehicle is zero, we present the approximation algorithm and prove that the worst case ratio of the algorithm is not greater than $2-\frac{1}{m}$, and the bound is tight.

Keywords: supply chain scheduling, dynamic programming algorithm, complexity, worst case analysis

1 Introduction

The classical scheduling problems usually assume that there is an infinite number of facilities for processing jobs, and the transportation between the warehouse and the facility can be done instantaneously. As a result, the job delivery and the machine scheduling are separately considered without effective coordination between the two. Coordination between job delivery and machine scheduling becomes more practical.

The coordination of production scheduling and transportation has recently received a lot of attention in logistics and supply chain management. Semi-finished jobs are transported from a holding area to a manufacturing facility for further processing by transporters in many manufacturing systems. Another motivation arises in many industries where the coordination of production and transportation can help to save energy and reduce fuel consumption. This is particularly true in the iron and steel industry.

In this paper, motivated by applications in the iron and steel industry, we study a coordinated scheduling problem of transportation and production. The jobs located at a holding area need to be transported by m vehicles to a serial batching machine for further processing. Each vehicle can transported one job at a time, and the serial batching machine can process several jobs at a time. Each batch of jobs to be processed on the serial batching machine occurs a processing cost; assume that batch processing cost is proportional to the batch number. The problem is to find a joint schedule of transportation and production such that the objective is to minimize the sum of makespan and total processing cost.

In the last decade, the coordination of transportation and scheduling has become one of the most important topics in production and operations management research. Chungyee L, Zhilong C [1] considers two types of transportation. The first type is intermediate transportation in a flow shop where jobs are transported from one machine to another for further processing. The second type is the transportation necessary to deliver finished jobs to the customer.

The batching machine scheduling is an important research topic. Recent reviews of batch scheduling research are provided by Potts C N and Kovalyov M Y [2] and Brucker P et al. [3]. The scheduling problems on the batching machine can be divided into two categories of parallel batch and serial batch according to batch processing pattern. In parallel batching scheduling problems, many jobs can be processed on the machines at any time; the processing time of batch is equal to the maximum of all jobs' processing times of this batch. Yuzhong Z, et al. [4] consider single parallel machine scheduling problem which the jobs have the sizes and only two different arrival time, present the approximation algorithm that the worst case ratio is not greater than $\frac{33}{14}$ when the processing times and the sizes are agreeable.

In serial batching scheduling problems, only one job can be processed on the machines at any time, and the

* ***Corresponding author*** e-mail gucunchang@163.com

jobs are processed one by one simultaneously, the processing time of batch is equal to the setup time plus the total processing times of the jobs in this batch. Coffman E G et al. [5] study minimizing the total completion times scheduling problem on single serial batch machine. Liji S et al. [6] consider the serial batch scheduling problem embedded in a job shop environment to minimize makespan, propose a tabu search algorithm which consists of various neighbourhood functions, multiple tabu lists and a sophisticated diversification structure. Glass C A et al. [7] study a problem of scheduling and batching on two machines in a flow-shop and open-shop environment, the aim is to make batching and sequencing decisions, which specify a partition of the jobs into batches on each machine, and a processing order of the batches on each machine, respectively, so that the makespan is minimized. Webster S and Baker K R [8] consider scheduling groups of jobs on a single machine on three basic models known as family scheduling with item availability, family scheduling with batch availability, and batch processing. Baptiste P [9] analyses the single parallel batching and serial batching scheduling problems that all jobs are arrived at time 0, present the dynamic programming algorithms respectively. But, the attention is relatively less for the coordination of transportation and serial batching scheduling at present.

In this paper, we investigate the problem of coordinated scheduling on single serial batching machine with transportation. At time 0, the n jobs $J_1, J_2, \ldots, J_n$ located at a holding area are transported by the m vehicles to a serial batching machine for further processing, the transportation times are $t_1, t_2, \ldots, t_n$, respectively. In the transportation stage, each vehicle can deliver only one job at a time, the transportation time of one job transported by different vehicles is identical, and the empty moving time of each vehicle from the batching machine back to the holding area is t. In the production stage, the processing time of the job J_j is p_j, $j = 1, 2 \ldots, n$, the serial batching machine can process several jobs simultaneously as a batch, the maximum number of jobs that can be processed simultaneously in the serial batching machine is called capacity c of that batching machine, the completion time C_j of any job J_j in each batch is the fixed set-up time s plus the total processing times of the jobs in this batch. Once processing of a batch is initiated, it cannot be interrupted and other jobs cannot be introduced into the machine until processing is completed. Each batch to be processed on the batching machine occurs a processing cost, b is the number of batches to be processed on the batching machine, $\alpha(b)$ is the linear non-decreasing function of b, and denote the processing cost function. The objective is minimizing the sum of the makespan and the total processing cost $C_{\max} + \alpha(b)$. This problem is denoted as $D \to 1 \mid m \geq 1, c \geq 1, s - batch \mid C_{\max} + \alpha(b)$.

The paper is organized as follows. In Section 2, we give some preliminaries and a useful lemma. Section 3 studies a special case, and provides a polynomial time dynamic programming algorithm. In Section 4, we study the complexity of the problem; propose the TSPT-DP algorithm to solve the problem. Finally, some conclusions are made in Section 5.

2 Preliminaries

Let π and π^* be a feasible solution and an optimal solution, respectively. We will use the following notations and definitions frequently in the reminder of this paper.

$b^{\pi}(b^*)$: the total number of batches in $\pi(\pi^*)$.

$B_l^{\pi}(B_l^*)$: the l-th processing batch in $\pi(\pi^*)$, $l = 1, 2, \ldots, b^{\pi}(b^*)$.

$n_l^{\pi}(n_l^*)$: the number of jobs processed up to the l-th batch, clearly, $n_{k^{\pi}}^{\pi} = n_{k^*}^* = n$.

$\mid B_l^{\pi} \mid (\mid B_l^* \mid)$: the number of jobs processed in $B_l^{\pi}(B_l^*)$, clearly, $\mid B_l^{\pi} \mid = n_l^{\pi} - n_{l-1}^{\pi} (\mid B_l^* \mid = n_l^* - n_{l-1}^*)$.

$r_j^{\pi}(r_j^*)$: the arrival time of the job J_j, i.e., the time when J_j arrives at the serial batching machine.

$r_{[j]}^{\pi}(r_{[j]}^*)$: the arrival time of the j-th arrived job.

$S_l^{\pi}(S_l^*)$: the starting time of the batch $B_l^{\pi}(B_l^*)$.

$C_j^{\pi}(C_j^*)$: the completion time of the job J_j.

$C^{\pi}(l)(C^*(l))$: the completion time of the batch $B_l^{\pi}(B_l^*)$, clearly, $C_j^{\pi} = C^{\pi}(l)(C_j^* = C^*(l))$, for $J_j \in B_l^{\pi}(B_l^*)$.

The notations can be abbreviated to $b, n_l, r_j, r_{[j]}, B_l$, $\mid B_l \mid$, et al. Let $n_0 = 0$, then, $C(i) = S_i + s + \sum_{j \in B_i} p_j$, $C_{\max} = C(b) = S_b + s + \sum_{j \in B_i} p_j$. In situation where the dimension on the measurements $C_{\max}$ and $\alpha(b)$ is difficult to unify, we may adjust cost function $\alpha(b)$ to uniform dimension with the maximum completion time. For ease of presentation, denote our problem as TSBSM.

Although the problem considered in this paper is different from that in Hall N G and Potts C N [10], some properties can be found in the same way.

Lemma 1. For problem $D \to 1 \mid m \geq 1, c \geq 1, s - batch, p_j = p \mid C_{\max} + \alpha(b)$, there exists an optimal solution π^* in which:

(i) There is no idle time between the jobs transported on each vehicle in the transportation part.

(ii) All jobs assigned to the same vehicle are scheduled in the non-decreasing order of transportation times.

(iii) The starting time of each batch on the serial batching machine is made either an arrival time of some job on the machine or immediately at a time when the machine becomes available.

(iv) earlier arrived jobs are processed no later than those arrived later.

Usually, the quality of an approximation algorithm (denoted by H) is measured by the worst case ratio of the algorithm, which is defined as the smallest number ρ such that $F^H \le \rho F^*$ for all instances, where F^H and F^* denote the objectives of solution produced by H and the optimal algorithm, respectively.

3 A polynomial time algorithm for a special case

In this section, we consider a special case where the job assignment to the vehicles is predetermined. It is evident that the problem reduces to an optimal serial batching problem in this case. This special case characterizes the practical situation where the arrival times of the jobs are known. Now, we can provide a dynamic programming algorithm to solve the optimal serial batching problem in a polynomial time as follow.

Re-index all jobs $J_j, j = 1,2,\ldots,n$, in accordance with the job arrival time on the serial batching machine, i.e., $r_1 \le r_2 \le \cdots \le r_n$, the processing time p_j of any job J_j on the serial batching machine is p. It suffices to consider one job sequence and apply it to the processing of jobs on the serial machine. So the starting time of each batch on the machine need to be decided, and this can be done by dynamic programming.

By Lemma 1(iii), the starting time S_l of the batch B_l on the serial batching machine is either the completion time $C(l-1)$ of the batch B_{l-1} or the arrival time r_{n_l} of the last arrived job J_{n_l} in this batch. In the first case, $S_l = C(l-1) = S_{l-1} + s + |B_{l-1}|\, p$, let B_j denote the earliest processing batch which the jobs are processed consecutively until the batch B_l, the starting time of the batch B_j must be the arrival time r_{n_j} of the last arrived job J_{n_j} in this batch, so,

$$\begin{aligned} S_l &= r_{n_j} + (l-j)s + (n_l - n_{j-1})p, \\ &= r_{n_j} + (l-j)s + (|B_j| + \cdots + |B_{l-1}|)p. \end{aligned} \quad (1)$$

Denote $n_l - n_{j-1} = q_l$, satisfy $q_l + j \le n$. In the second case, $S_l = r_{n_l}$.

Hence, the possible starting time of the bath B_l on the serial batching machine can be $r_j, r_j + s + xp, \ldots, r_j + zs + yp$, where $\lceil n/c \rceil \le z \le l-1 \le n,\ x \le y \le n-j$, and $S_0 = 0$.

Define $f(k, j, S_l)$ as the minimal makespan to schedule the first k jobs $J_1, J_2, \ldots, J_k$, provided that the current last batch B_l contains jobs $J_j, J_{j+1}, \ldots, J_k$, and starts to be processed at time S_l , where $k - j + 1 \le c,\ S_l \ge r_k$. If we know the available time $f(j-1, i, S_{l-1})$ of the batching machine before process jobs $J_j, J_{j+1}, \ldots, J_k$, then the starting time of the batch B_l is actually fixed, i.e., $S_l = \max\{f(j-1, i, S_{l-1}), r_k\}$.

$f(k, j, S_l)$ satisfies the following three properties:

(i) $0 < k - j + 1 \le c$;

(ii) $S_l = r_j, r_j + s + xp, \ldots, r_j + zs + yp$, and $S_l - S_{l-1} \ge s + p, l = 2,3,\ldots,b$, where $x \le y \le n$, $\lceil n/c \rceil \le z \le l-1$;

(iii) $\lceil k/c \rceil \le l \le k$.

Otherwise, $f(k, j, S_l) = \infty$.

Dynamic Programming Algorithm *DP*

Initial condition:

$f(0,0,0) = 0$;

Recursive relation:

$f(k, j, S_l) = \min\{\max\{f(j-1, i, S_{l-1}), r_k\} + (s + (k-j+1)p) \mid$ all possible states $(i, S_{l-1})\}$,where $j-1, i, S_{l-1}$ satisfy the three condition described above.

Optimal solution:

$F(n) = \min\{f(n, j, S_b) + \alpha(b) \mid$ all possible states $(j, S_b)\}$.

By recording all the necessary information in the above process, an optimal schedule can be calculated. It is not difficult to see that the time complexity of the algorithm *DP* is $O(cn^4)$. The following theorem can be obtained.

Theorem 1. Algorithm *DP* finds an optimal schedule for problem $D \to 1 \mid m \ge 1, c \ge 1, s - batch, p_j = p \mid C_{\max} + \alpha(b)$ in $O(cn^4)$ time when the job assignment to the vehicles is predetermined.

We now demonstrate the above solution method with a numerical example.

Example. Job set $J = \{J_1, J_2, J_3, J_4\}$, $m = 2, t = 1, c = 3$, $t_1 = 1, t_2 = 4, t_3 = 2, t_4 = 1, s = 1, p = 2, \alpha(b) = 5b$, assume that J_1 and J_2 are transported by one vehicle, J_3 and J_4 are transported by another vehicle. From the above method, we know $r_1 = 1, r_2 = 2, r_3 = 4, r_4 = 6$.

We have the following results by the dynamic programming algorithm *DP* :

$f(1,1,S_1) = \max\{f(0,0,0), r_1\} + (s+p) = 4$;

$f(2,1,S_1) = \max\{f(0,0,0), r_2\} + (s+2p) = 7$;

$f(2,2,S_2) = \max\{f(1,1,S_1), r_2\} + (s+p) = 7$;

$f(3,1,S_1) = \max\{f(0,0,0), r_3\} + (s+3p) = 11$;

$f(3,2,S_2) = \max\{f(1,1,S_1), r_3\} + (s+2p) = 9$;

$f(3,3,S_2)=\max\{f(2,1,S_1),r_3\}+(s+p)=10$;
$f(3,3,S_3)=\max\{f(2,2,S_2),r_3\}+(s+p)=10$;
$f(4,2,S_2)=\max\{f(1,1,S_1),r_4\}+(s+3p)=13$;
$f(4,3,S_2)=\max\{f(2,1,S_1),r_4\}+(s+2p)=12$;
$f(4,4,S_2)=\max\{f(3,1,S_1),r_4\}+(s+p)=14$;
$f(4,3,S_3)=\max\{f(2,2,S_2),r_4\}+(s+2p)=12$;
$f(4,4,S_3)=\min\{\max\{f(3,2,S_2),r_4\}+(s+p)=12,\ \max\{f(3,3,S_2),r_4\}+(s+p)=13\}=12$;
$f(4,4,S_4)=\max\{f(3,3,S_3),r_4\}+(s+p)=13$;
$F(4)=\min\{f(4,j,S_b)+\alpha(b)\mid \text{all possible states } (j,S_b)\}=f(4,3,S_2)+\alpha(2)=22$.

The optimal schedule on the serial batching machine is finally found as $\pi^*=\{\{J_1,J_3\},\{J_4,J_2\}\}$ with the optimal objective value of 22.

4 Complexity and approximation algorithm

In this section, we firstly show that for TSBSM is NP-hard by reducing the 2-Partition problem to the decision version of this problem when the job assignment to the vehicles is not predetermined, secondly propose an approximation algorithm for the problem TSBSM.

4.1 COMPLEXITY

2-Partition problem: Given $h+1$ positive integers a_j, $j=1,2,\ldots,h$, and a, such that $\sum_{i=1}^{h}a_j=2a$. The question asks if the set $H=\{1,2,\ldots,h\}$ can be divided into two disjoint subsets G_1 and G_2, such that $G_1+G_2=H$, $\sum_{j\in G_1}a_j=\sum_{j\in G_2}a_j=a$.

Theorem 2. The problem $D\to 1\mid m\ge 1,c\ge 1,s-batch, p_j=p\mid C_{\max}+\alpha(b)$ is NP-hard even if $m=2$.

Proof. To any instance of the 2-Partition problem, we construct an instance of TSBSM as follow. There are $n=2h$ jobs split into two job sets: the J'-jobs (partition jobs) denoted by $J_j',j=1,2,\ldots,h$, the J''-jobs (auxiliary jobs) denoted by $J_j'',j=1,2,\ldots,h$. Their transportation times and other parameters are given as follows:

Transportation times: $t_{J_j'}=2a_j,t_{J_j''}=0,j=1,2,\ldots,h$;

Returning time: $t=0$;

Set-up time of the serial batch: $s=a$;

Processing time: $p_j=\frac{a}{h},j=1,2,\ldots,2h$;

Processing cost: $\alpha(b)=3ba$;

Machine capacity: $c=h$;

Threshold value: $y=10a$.

We are going to show that for the constructed scheduling problem instance, a schedule π with $C_{\max}+\alpha(b)\le y$ exists if and only if the 2-Partition problem has a solution.

If there is a solution to the 2-Partition problem instance, we show that there is a schedule π with $C_{\max}+\alpha(b)\le y$ for the scheduling instance. Suppose that the 2-Partition problem instance has a solution G_1 and G_2. Now construct the following schedule π:

Since the transportation times of J''-jobs and the returning time of the vehicles are equal to 0, J''-jobs are transported to the serial batching machine, and two vehicles are available at time 0, the machine can first process these jobs. Due to $c=h$, the J''-jobs are processed as the first batch at time 0, and the completion time $C(1)$ of this batch is equal to $S_1+s+\sum_{j\in J''}p_j=0+a+h\frac{a}{h}=2a$. Next, vehicle 1 transports the jobs of G_1 one by one, and vehicle 2 transports the jobs of G_2. Let T_i denotes the total running time of vehicle i, for $i=1,2$, we can see that $T_1=\sum_{j\in G_1}t_{J_j'}=2a,T_2=\sum_{j\in G_2}t_{J_j'}=2a$. The J'-jobs are processed as the second batch at time $2a$, and the completion time $C(2)$ of this batch is equal to $S_2+s+\sum_{j\in J'}p_j=2a+a+h\frac{a}{h}=4a$. It is easy to check that $C_{\max}+\alpha(b)\le y$. (See Figure 1)

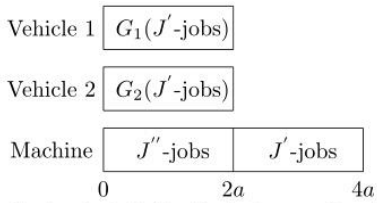


Figure 1. A schedule for the instance of problem TSBSM

If there exists a schedule π with $C_{\max}+\alpha(b)\le y$ to the instance of TSBSM, we show that the 2-Partition problem has a solution.

First, the following properties hold in schedule π:

(i) Schedule π exactly contains two batches, i.e., each batch contains exactly h jobs.

(ii) All J''-jobs are processed as the first batch at time 0.

(iii) All J'-jobs are processed as the second batch at time $2a$.

Next, we prove three properties as follows:

(i) Suppose that there are b batches in schedule π. Due to $c=h,\alpha(b)=3ba$, we have $2\le b\le 3$. If $b=3$, the sum of the makespan and the processing cost

$C_{\max}+\alpha(b)$ is more than $s+\sum_{j\in J''\cup J'}p_j+9a=12a$, which is a contradiction.

Hence, schedule π exactly contains two batches and each batch contains h jobs.

(ii) Denote S_1 and S_2 as the starting time of the first batch and the second batch on the serial batching machine, respectively. Note that $S_1+s+h\frac{a}{h}=S_1+2a\le S_2$. Since the processing time of each batch on the machine is $2a$, we have $C_{\max}+\alpha(b)=S_2+2a+6a\le 10a$. Hence, we obtain that $S_1=0$ and $S_2=2a$. At time 0, there are only J''-jobs available. Thus, all J''-jobs are processed as the first batch at time 0.

(iii) From (i) and (ii), we know that all J'-jobs are processed as the second batch at time $2a$.

Let G_1 and G_2 be a partition of J'-jobs. We assume that vehicle 1 transports the jobs of G_1 one by one, and vehicle 2 transports the jobs of G_2 one by one. Let T_i denotes the total running time of vehicle i, for $i=1,2$. Based on the above discussion, the starting time of the second batch satisfies $S_2=2a\ge\max\{T_1,T_2,2a\}$, where $T_1=\sum_{j\in G_1}2a_j\le 2a, T_2=\sum_{j\in G_2}2a_j\le 2a$. Due to $T_1+T_2=\sum_{j\in H}2a_j=4a$, we have $\sum_{j\in G_1}a_j=\sum_{j\in G_2}a_j=a$. Then it is easy to see that G_1 and G_2 form a solution to the 2-Partition problem instance. Theorem 2 follows.

4.2 APPROXIMATION ALGORITHM

The proof of theorem 2 indicates the problem $D\to 1\,|\,m\ge 1,c\ge 1,s-batch,p_j=p\,|\,C_{\max}+\alpha(b)$ is NP-hard even if $t=0$. When the returning time of vehicle is 0, we present the following approximation algorithm and analysis its worst case ratio.

TSPT-DP Algorithm

Step 1: In the transportation part, assign jobs to vehicles by the SPT rule of the transportation times, i.e., always assign the shortest transportation time job to the available vehicle ;

Step 2: Renumber the jobs in the non-decreasing arrival time order, i.e., $r_1\le r_2\le\cdots\le r_n$;

Step 3: In the processing part, use the dynamic programming algorithm *DP* for batching and processing jobs.

Let π be the solution generated by the TSPT-DP algorithm for the problem $D\to 1\,|\,m\ge 1,t=0,c\ge 1,$ $s-batch,p_j=p\,|\,C_{\max}+\alpha(b)$. We denote by π' the solution generated by assigning jobs to vehicles by the SPT rule of the transportation times, and processing the same number of jobs in each batch coincide with that of the optimal solution π^* as early as possible (maybe wait).

Therefore, π' transports jobs the same as π and uses the same batching policy as π^*. Consequently, we have $r_j^{\pi'}=r_j^{\pi}, j=1,2,\ldots,n$, and $n_i^{\pi'}=n_i^*, 1\le i\le b^*=b^{\pi'}$, $\alpha(b^*)=\alpha(b^{\pi'})$. By Theorem 1,

$$C_{\max}^{\pi}+\alpha(b^{\pi})\le C_{\max}^{\pi'}+\alpha(b^{\pi'}). \tag{2}$$

Lemma 2. $r_j^{\pi'}\le(2-\frac{1}{m})r_{[j]}^*, 1\le j\le n$.

Proof. Since $r_{[j]}^*$ is the arrival time of the j-th arrived job in π^*, it follows $r_{[1]}^*\le r_{[2]}^*\le\cdots\le r_{[n]}^*$, and $t_1,t_2,\ldots,t_j$ are the j minimum transportation times. Thus, we have

$$r_{[j]}^*\ge\max\{t_{[1]},t_{[2]},\cdots,t_{[j]},\frac{1}{m}\sum_{l=1}^{j}t_{[l]}\}\ge\max\{t_j,\frac{1}{m}\sum_{l=1}^{j}t_l\}. \tag{3}$$

On the other hand, as π' uses the SPT rule of the transportation times to assigning jobs to the vehicles,

$$r_j^{\pi'}\le(1-\frac{1}{m})t_j+\frac{1}{m}\sum_{l=1}^{j}t_l\le(2-\frac{1}{m})r_{[j]}^*, 1\le j\le n. \tag{4}$$

Lemma 3. $C^{\pi'}(b^{\pi'})\le(2-\frac{1}{m})C^*(b^*)$.

Proof. Since π' and π^* have the same number of jobs in each batch, we denote $n_i^{\pi'}=n_i^*=n_i$, $|B_i^{\pi'}|=|B_i^*|=|B_i|=n_i-n_{i-1}, 1\le i\le b^*=b^{\pi'}$. $r_{n_i}^{\pi'}$ and $r_{[n_i]}^*$ are denoted as the arrival time of the last arrived job $J_{n_i}^{\pi'}$ and $J_{[n_i]}^*$ in the i-th batch in π' and π^*, respectively.

By Lemma 2 and (4), we have

$$\begin{aligned}C^{\pi'}(1)&=r_{n_1}^{\pi'}+s+|B_1|\,p\le(2-\frac{1}{m})r_{[n_1]}^*+s+|B_1|\,p\\&\le(2-\frac{1}{m})(r_{[n_1]}^*+s+|B_1|\,p)=(2-\frac{1}{m})C^*(1).\end{aligned} \tag{5}$$

By Lemma 1(iii) and (5),

$$\begin{aligned}C^{\pi'}(2)&=\max\{r_{n_2}^{\pi'},C^{\pi'}(1)\}+s+|B_2|\,p\\&\le(2-\frac{1}{m})\max\{r_{[n_2]}^*,C^*(1)\}+s+|B_2|\,p\\&\le(2-\frac{1}{m})(\max\{r_{[n_2]}^*,C^*(1)\}+s+|B_2|\,p)\\&=(2-\frac{1}{m})C^*(2).\end{aligned} \tag{6}$$

Repeat the above process, we get the result

$$\begin{aligned}C^{\pi'}(b^{\pi'})&=\max\{r_{n_{b^{\pi'}}}^{\pi'},C^{\pi'}(b^{\pi'}-1)\}+s+|B_{b^{\pi'}}|\,p\\&\le(2-\frac{1}{m})\max\{r_{[n_{b^*}]}^*,C^*(b^*-1)\}+s+|B_{b^*}|\,p\end{aligned}$$

$$\le (2-\frac{1}{m})C^*(b^*).$$

Theorem 3. For the problem $D \to 1 \mid m \ge 1, t = 0, c \ge 1, \quad s-batch, p_j = p \mid C_{\max} + \alpha(b)$, the worst case ratio of TSPT-DP algorithm is not greater than $2-\frac{1}{m}$, and the bound is tight.

Proof. By (2) and Lemma 3, we have

$$\begin{aligned} C^{\pi}_{\max} + \alpha(b^{\pi}) &\le C^{\pi'}_{\max} + \alpha(b^{\pi'}) = C^{\pi'}(b^{\pi'}) + \alpha(b^{\pi'}) \\ &\le (2-\frac{1}{m})C^*(b^*) + \alpha(b^*) \\ &\le (2-\frac{1}{m})(C^*_{\max} + \alpha(b^*)). \end{aligned} \tag{7}$$

We can conclude that the worst case ratio of TSPT-DP algorithm is at most $2-\frac{1}{m}$. To show the tightness, let us consider the following instance: $n = m^2 - m + 1$, $t = 0$, $t_1 = t_2 = \cdots = t_{m^2-m} = 1, t_{m^2-m+1} = m$, $s = p = 0$, $c = 1$, and $\alpha(b) = 0$. Clearly, both TSPT-DP and the optimal algorithm must process jobs in $m^2 - m + 1$ batches. Consequently, we have $C^{\pi}_{\max} = 2m - 1$, $C^*_{\max} = m$, and $\alpha(b^{\pi}) = 0$, $\alpha(b^*) = 0$. Thus, the desired result follows.

5 Conclusion

In this paper, we studied the problem of coordinated scheduling on a serial batching machine with transportation. Under the condition of the job processing times are equal, if the job assignment to the vehicles is predetermined, we provide a polynomial time dynamic programming algorithm, for the general problem, we prove it is NP-hard. When the returning time of vehicle is 0, we present the approximation algorithm TSPT-DP and prove that the worst case ratio of the algorithm is not greater than $2-\frac{1}{m}$, and the bound is tight.

Several possible extension to this research can be considered, such as, minimizing maximum job tardiness, developing effective heuristics to solve the general problem, et al.

Acknowledgments

This research is partly supported by the projects as follows:
(i) Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20123705110003);
(ii) National Natural Science Foundation of China (Grant No. 11071142, 11201259, 11201121, 1271338);
(iii) Natural Science Foundation of Education Department of Henan Province (Grant No. 2011B110007, 2011B110008);
(iv) Natural Science Foundation of Shandong Province (Grant No. ZR2010AM034).

References

[1] Chungyee L, Zhilong C 2001 *J. Sched.* **4** 3-24
[2] Potts C N, Kovalyov M Y 2000 *Eur. J. Oper. Res.* **120** 228-49
[3] Brucker P, Gladky A, Hoogeveen H, Kovalyov M Y, Potts C N, Tautenhahn T, Van de Velde S L 1998 *J. Sched.* **1** 31-54
[4] Yuzhong Z, Qingguo B, Jianteng X 2006 *Oper. Res. Trans.* **10** 99-105
[5] Coffman E G, Yannakakis M, Magazine J, Santos C 1990 *Ann. Oper. Res.* **26** 135-47
[6] Liji S, Buscher U 2012 *Eur. J. Oper. Res.* **221** 14-26
[7] Glass C A, Potts C N, Strusevich V A 2001 *INFORMS J. Comput.* **13** 120-37
[8] Webster S, Baker K R 1995 *Oper. Res.* **43** 692-703
[9] Baptiste P 2000 *Math. Methods Oper. Res.* **52** 355-67
[10] Hall N G, Potts C N 2005 *Ann. Oper. Res.* **135** 41-64

Authors

Cunchang Gu, born in 1980

Current position and grades: a Ph. D. candidate at Qufu Normal University, and an assistant professor at Henan University of Technology in China.
University studies: BSc and MSc degrees from Department of Mathematics, Zhengzhou University, Zhengzhou City, China, in 2002 and 2009, respectively. He is currently working towards Ph.D. degree at School of Management, Qufu Normal University, Rizhao City, China.
Research interests: scheduling, combinatorial optimization, and supply chain management.

Xiaoyan Xu, born in 1979

Current position and grades: a lecturer at Henan University of Technology in China.
University studies: BSc and MSc degrees from Department of Mathematics, Henan University, Kaifeng, City, China, in 2002 and 2010, respectively. She is currently working as a teacher at a College of Science, Henan University of Technology, Zhengzhou City, China.
Research interests: combinatorial optimization, linear programming and statistics theory.

Chaos control of unified chaotic system base on tridiagonal matrix stability theory and adaptive hybrid synchronization

Yaolin Huang*, Shuangli Wang, Chengjun Xie, Zhuo Jiang, Jinyan Zhao

College of Computer Science and Technology, Beihua University, Jilin, 132021, China

Received 13 June 2014, www.cmnt.lv

Abstract

This paper presents a new chaos control method to control the unified chaotic system to zero. In order to control a unified chaotic system to zero, the first step is to design a stable system, which stable to zero base on tridiagonal matrix stability theory, the stable system as the master system. The second step is to make the unified chaotic system with controlled as slave system, the third step to make the master system and the slave system synchronization. Different system state apply different adaptive synchronize method to realize synchronization. The adaptive control law and parameter update law are obtained base on Lyapunov stability theory. Numerical simulations are presented to demonstrate the effectiveness of theoretical analysis.

Keywords: chaos control, adaptive hybrid synchronization, tridiagonal structure matrix stability theory

1 Introduction

In 1990, Ott, Grebogi and Yorke presented the OGY method on chaos control [1]. After their pioneering work, chaos control has gained more interest in nonlinear problems and there have been a lot of progress in this field [2-5]. Generally, there are two control ways: feedback control and nonfeedback control. Feedback methods [6-10] are used to stabilize the unstable periodic orbit of chaotic systems by feedback their states. Nonfeedback methods [11, 12] are adopted to suppress chaotic behaviour by apply periodic perturbations to some parameters or variables. With the development of chaos control technology, all kinds of methods are developed.

This paper proposes a new chaos control method to control the unified chaotic system to zero, the new method base on tridiagonal structure matrix stability and hybrid synchronization.

This paper is organized as follows. In the next section, we analyse the tridiagonal structure matrix stability theory and system with tridiagonal structure matrix. In section 3, we introduce the hybrid synchronization theory. In section 4, we make use of tridiagonal matrix stability theory and adaptive hybrid synchronization to control the unified chaotic system to zero. In section 5, some numerical simulations are done to test the effectiveness of theoretical analysis. Finally, some conclusions are drawn in section 6.

2 The theory

2.1 THE TRIDIAGONAL STRUCTURE MATRIX STABILITY THEORY

Lemma 1 [13, 14] If the nonlinear system has the following forms of tridiagonal structure:

$$\dot{X} = \begin{bmatrix} -k_1 & f_1(x) & & \\ -f_1(x) & -k_2 & \ddots & \\ & \ddots & \ddots & f_{n-1}(x) \\ & & -f_{n-1}(x) & -k_n \end{bmatrix} X \text{ , where}$$

$X(t) = (x_1, x_2, \cdots, x_n)^T \in R^n$ is state vectors, $f_j(x_i)(j = 1, 2, \ldots, n-1, i = 1, 2, 3, \ldots, n)$ are function about x_i , $k_i \in R^+(i = 1, 2, \ldots, n)$, the state vectors x of nonlinear system will be globally asymptotically stabilized to origin.

Proof: A positive definite function is as follows, $V = \frac{1}{2} X^T X$.

The derivative of V , $\dot{V} = -\sum_{i=1}^{n} K_i x_i^2 < 0$.

Due to V is the Lyapunov function and x will be globally asymptotically stabilized to origin [14].

2.2 SYSTEM WITH TRIDIAGONAL STRUCTURE MATRIX

Due to permanent magnet synchronous motor (PMSM) system has tridiagonal structure, we introduce PMSM system.

* ***Corresponding author*** e-mail hyl200199@163.com

Model of PMSM can be expressed as follows [15],

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix}=\begin{bmatrix}-1 & x_3 & 0\\ -x_3 & -1 & \gamma\\ 0 & \sigma & -\sigma\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}. \quad (1)$$

The parameters γ and σ ($\sigma>0$) can decide the system (1) is chaos or not [16]. Fig. 1 show the chaotic character of system (1) with $\gamma=20$ and $\sigma=3$, initial values $x_1(0)=0.1, x_2(0)=0.2, x_3(0)=0.4$. From Figuge 1 it can be concluded that system (1) is in chaotic state.

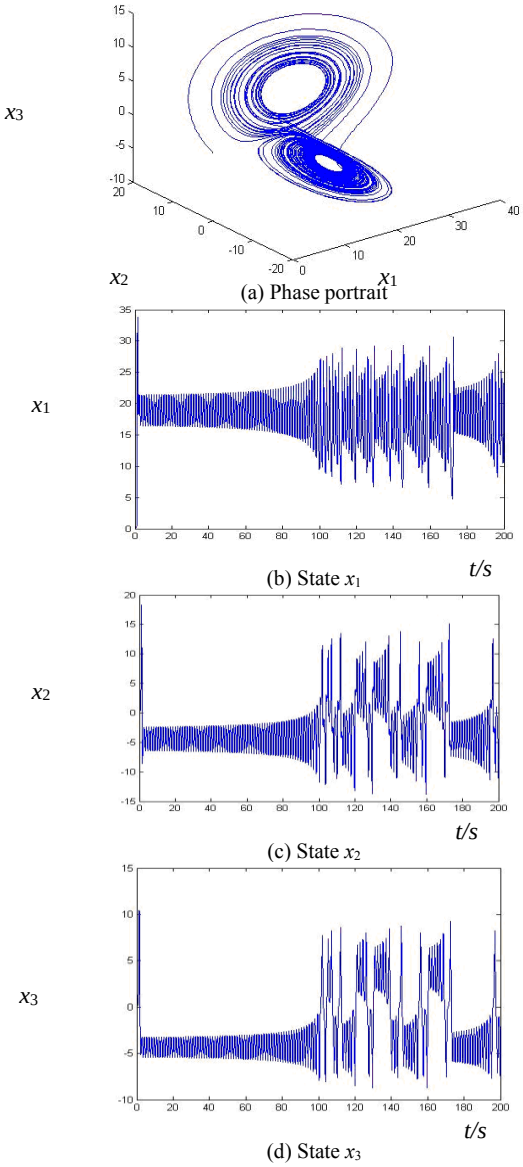


FIGURE 1 the chaotic character of system (1) when $\gamma=20$ and $\sigma=3$

System (1) has tridiagonal structure matrix form when parameters $\gamma=$-σ, which can be expressed as follows,

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix}=\begin{bmatrix}-1 & x_3 & 0\\ -x_3 & -1 & -\sigma\\ 0 & \sigma & -\sigma\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}. \quad (2)$$

Figure 2 shows states changes of system (2) when parameters $\sigma=5$, initial values $x_1(0)=0.1, x_2(0)=0.2, x_3(0)=0.4$. From Figure 2, it can be concluded that system (2) is stabilized to zero.

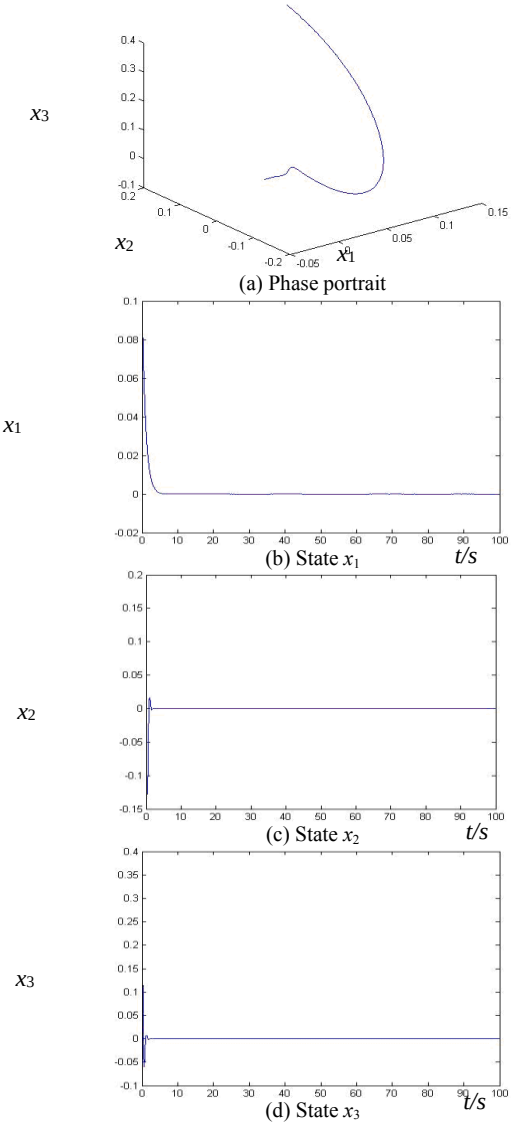


FIGURE 2 Trajectories of the system (2) states when $\sigma=5$

3 Hybrid Synchronization

Consider the following three-dimensional master system (3) and slave system (4),

$$\dot{x}=f(t,x), \quad (3)$$

$$\dot{y}=g(t,y)+u(t,x,y), \quad (4)$$

where $x(t)=(x_1,x_2,x_3)^T\in R^3$ and $y(t)=(y_1,y_2,y_3)^T\in R^3$ are master and slave state vectors respectively, $f:R^3\to R^3$ and $g:R^3\to R^3$ are continuous nonlinear vector functions and $u(t,x,y)=(u_1,u_2,u_3)^T\in R^3$ is control input for synchronization between master system (3) and slave system (4).

Definition 1. It is said that hybrid synchronization occurs between master system (3) and slave system (4) if satisfy the following conditions:

Define $s(r)$ is a function, where r is state vector. Satisfy

$$\lim_{t\to\infty}\|e_1(t)=y_1(t)-s(x_1(t))\|=0.$$

Define λ is constant and satisfy

$$\lim_{t\to\infty}\|e_2(t)=y_2(t)-\lambda x_2(t)\|=0.$$

Define $m(t)$ is a function and satisfy

$$\lim_{t\to\infty}\|e_3(t)=y_3(t)-m(t)x_3(t)\|=0.$$

4 Method

The unified chaotic system [17] is expressed as follows,

$$\begin{aligned}\dot{x}_1&=(25\theta+10)(x_2-x_1)\\ \dot{x}_2&=(28-35\theta)x_1+(29\theta-1)x_2-x_1x_3\\ \dot{x}_3&=x_1x_2-\frac{8+\theta}{3}x_3\end{aligned}\quad(5)$$

where $\theta\in[0,1]$. System (5) is chaotic for $\theta\in[0,1]$.

When $\theta\in[0,0.8)$, system (5) reduces to the general Lorenz system; when $\theta=0.8$, it becomes the general Lü system; and $\theta\in(0.8,1]$, system (5) is the general Chen system.

We assume that the system (2) with $\sigma=5$ as master system, system (5) with controlled as the slave system (6) given by,

$$\begin{aligned}\dot{y}_1&=(25\theta+10)(y_2-y_1)+u_1\\ \dot{y}_2&=(28-35\theta)y_1+(29\theta-1)y_2-y_1y_3+u_2\\ \dot{y}_3&=y_1y_2-\frac{8+\theta}{3}y_3+u_3\end{aligned}\quad(6)$$

where u_1,u_2,u_3 are the control input.

By the definition of hybrid synchronization, the error is given by

$$\begin{aligned}e_1&=y_1-s_1(x_1)\\ e_2&=y_2-\lambda x_2\\ e_3&=y_3-m(t)x_3\end{aligned}\quad(7)$$

The error dynamics system is then given as

$$\begin{aligned}\dot{e}_1&=\dot{y}_1-\dot{s}_1(x_1)\dot{x}_1\\ \dot{e}_2&=\dot{y}_2-\lambda\dot{x}_2\\ \dot{e}_3&=\dot{y}_3-\dot{m}(t)x_3-m(t)\dot{x}_3\end{aligned}\quad(8)$$

By substituting Eqs. (2) And (6) in the above equation, we obtain

$$\begin{aligned}\dot{e}_1&=(25\theta+10)(y_2-y_1)-\dot{s}_1(x_1)(-x_1+x_3x_2)+u_1\\ \dot{e}_2&=(28-35\theta)y_1+(29\theta-1)y_2-y_1y_3-\lambda(-x_2-x_3x_1-5x_3)+u_2\\ \dot{e}_3&=y_1y_2-\frac{8+\theta}{3}y_3-\dot{m}(t)x_3-m(t)[5(x_2-x_3)]+u_3\end{aligned}\quad(9)$$

Here, our goal is to achieve hybrid synchronization between two systems with different initial conditions. For this end, the following control laws are designed:

$$\begin{aligned}u_1&=-(25\hat{\theta}+10)(y_2-y_1)+\dot{s}_1(x_1)(-x_1+x_2x_3)-k_1e_1\\ u_2&=-(28-35\hat{\theta})y_1-(29\hat{\theta}-1)y_2+y_1y_3+\lambda(-x_2-x_3x_1-5x_3)-k_2e_2\\ u_3&=-y_1y_2+\frac{8+\hat{\theta}}{3}y_3+\dot{m}(t)x_3+5m(t)(x_2-x_3)-k_3e_3\end{aligned}\quad(10)$$

where $\tilde{\theta}=\theta-\hat{\theta}$, we define $\hat{\theta}$ as estimate of θ of system (6), $\tilde{\theta}$ is estimate error, $k_i>0(i=1,2,3)$.

Substituting Eqs. (10) In Eqs. (9), we obtain

$$\begin{aligned}\dot{e}_1&=25\tilde{\theta}(y_2-y_1)-k_1e_1\\ \dot{e}_2&=-35\tilde{\theta}y_1+29\tilde{\theta}y_2-k_2e_2\\ \dot{e}_3&=-\frac{\tilde{\theta}}{3}y_3-k_3e_3\end{aligned}\quad(11)$$

The parameter update rule for $\hat{\theta}$ is chosen as follows

$$\dot{\hat{\theta}}=\frac{1}{3}e_3y_3-25e_1(y_2-y_1)+e_2(35y_1-29y_2).\quad(12)$$

Theorem 1. The system (6) can be stabilized to zero by hybrid synchronization between master system (2) and slave system (6) when choose the control laws (10) and parameter update rule (12).

Proof. Choose the following Lyapunov function

$$V=\frac{1}{2}(e_1^2+e_2^2+e_3^2+\tilde{\theta}^2),\quad(13)$$

where $\tilde{\theta}=\theta-\hat{\theta}$, the time derivative of Eq. (13) along the trajectory of error system is

$$\dot{V}=\dot{e}_1e_1+\dot{e}_2e_2+\dot{e}_3e_3+\tilde{\theta}\dot{\hat{\theta}},\quad(14)$$

and substitute Eqs. (11) and Eq. (12) in Eq. (14), $\dot{V} = -k_1 e_1^2 - k_2 e_2^2 - k_3 e_3^2$.

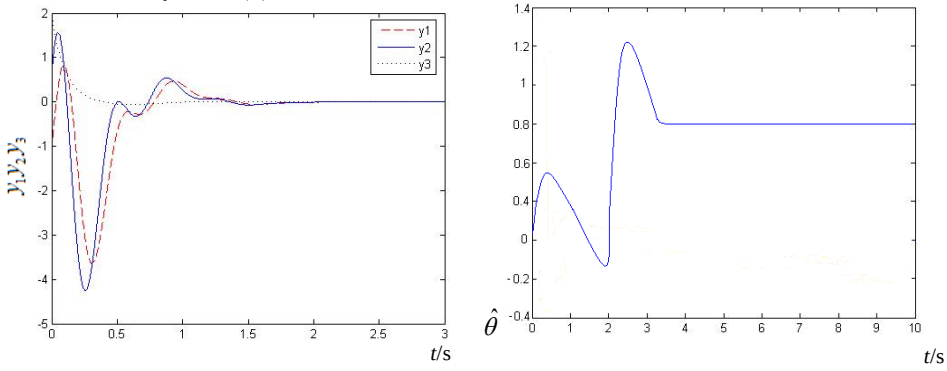
Since $\dot{V} < 0$, the master system (2) and the slave system (6) achieve synchronization. The slave system (6) can be stabilized to zero.

5 Numerical simulations

Numerical simulations are presented to demonstrate the effectiveness of the proposed synchronization controller. Fourth-order Runge-Kutta method is used to solve master system (2) and slave system (6) with time step size 0.0001, the parameter of master system (2) are chosen to be $\sigma = 5$ so that the master system (2) exhibits stable behaviour. The initial conditions of the master system (2) as follows, $x_1(0) = 0.1, x_2(0) = 0.2, x_3(0) = 0.3$, and those of the slave system (6) are $y_1(0) = -1, y_2(0) = 0.6, y_3(0) = 2$.

The estimate parameter of slave system (6) is chosen as $\hat{\theta} = 0.1$ real parameter is $\theta = 0.8$. Moreover, $s_1(r) = \sin r, s_1(x_1) = \sin x_1$, $m(t) = \sin t + 1$, $\lambda = 30$. The control gains are chosen as $k_1 = 0.05$, $k_2 = k_3 = 10$.

The simulations results are illustrated in Fig. 3. From Fig. 3 (a) it can be concluded that system (6) can be stabilized to zero. From Fig. 3 (b) it can be concluded that and estimate parameter value research 0.8.



(a) Trajectories of the states y_1, y_2, y_3 **(b) Parameter identification**

FIGURE 3 System (6) control process

6 Conclusions

In this paper, we propose a chaos control method to control the unified chaotic system with uncertain parameter to zero. The first step is to design a stable system, which stabilizes to zero bases on tridiagonal matrix stability theory, the stable system as master system. The second step is to make the unified chaotic system with controlled as the slave system. The third step is to make the slave system and the master system synchronize. The variables of system apply adaptive function synchronization, adaptive projective synchronization and the adaptive generalized projective synchronization to realize the synchronization of master system and the slave system. Thus, the slave system can be stabilized to zero.

References

[1] Ott E, Grebogi C, Yorke J A 1990 Controlling chaos *Physical Review Letters* **64** 1196-9
[2] Arena A, Lacarbonara W 2012 Nonlinear parametric modeling of suspension bridges under aeroelastic forces: torsional divergence and flutter, nonlinear dynamics **70** 2487-510
[3] Hussain I, Shah T, Gondal M A 2012 A novel approach for designing substitution-boxes based on nonlinear chaotic algorithm, nonlinear dynamics **70** 1791-4
[4] Guan X P, Fan Z P, Chen C L, Hua C C 2002 *Chaotic control and its application on secure communication* Beijing: National Defence Industry Press 168-225
[5] Wang X Y 2003 *Chaos in the complex nonlinearity system* Beijing: Electronics Industry Press 28-32
[6] Jang M J, Chen C L, Chen C K 2002 Sliding mode control of hyperchaos in Rössler systems *Chaos, Solitons & Fractals* **14** 1465-76
[7] Wang X, Chen G, Yu X 2000 Anticontrol of chaos in continuous-time systems via time-delay feedback *Chaos* **10** 771-9
[8] Zheng Y A 2006 Controlling chaos using Takagi-Sugeno fuzzy model and adaptive adjustment *Chinese Physics* **15** 2549-52
[9] Gong L H 2005 Study of chaos control based on adaptive pulse perturbation *Acta Physica Sinaca* **54** 3502-7
[10] Wang X Y, Wu X J 2006 Chaos control of a modified coupled dynamos system *Acta Physica Sinaca* **55** 5083-93
[11] Wang L Z, Zhao W L 2005 Suppression of chaotic motion in a class of piecewise-smooth systems by using sine periodic force *Acta Physica Sinaca* **54** 4038-43
[12] Chen L, Wang D S 2007 Nonfeedback control of chen's chaotic system *Acta Physica Sinaca* **56** 91-4
[13] Liu B, Zhang Z 2007 Stability of nonlinear systems with tridiagonal structure and its applications *Acta Atomatica Sinica* **33** 442-5
[14] Liu B, Zhou Y M, Jiang M, Zhang Z K 2009 Synchronizing chaotic systems using control based on tridiagonal structure *Chaos Solitons & Fractals* **39** 2274-81
[15] Elmas C, Ustun O 2008 A hybrid controller for the speed control of a permanent magnet synchronous motor drive *Control Engineering Practice* **16** 260-70
[16] Wang X Y, Wang M J 2008 A hyperchaos generated from Lorenz system *Physica A* **387** 3751-8
[17] Lü J H, Chen R G, Cheng D Z 2002 Bridge the gap between the Lorenz system and the Chen system *Int J Bifur Chaos* **12** 2917
[18] Perchaotic Lü 2006 Attractor via state feedback control *Physica A* **364** 103-10
[19] Mohammad H, Mahsa D Impulsive synchronization of Chen

Authors

Huang Yaolin, born in July, 1979, Jilin, Jilin, China

Current position, grades: lecturer in the school of computer science and technology Beihua University, he and as school of computer science and technology college secretary, he was involved in the project of Jilin Provincial Department of science and technology, this project and won the three prize of scientific and technological progress in Jilin province.

Wang Shuangli, born in December, 1972, Jilin, Jilin, China

Current position, grades: associate professor at the school of computer science and technology, was a master's degree in engineering.
Publications: 2 papers in EI retrieval, the provincial papers 2, complete the student innovation project of national level 2. Is mainly engaged in object oriented programming, VC++ programming technology, network technology and Web technology in teaching and research activities, neural network and chaos theory, related project development and maintenance work practice.
Experience: In 2012 obtained the Jilin province science and technology progress prize 3 item, independently of the city Department of project 1, involved in the preparation of teaching materials 1

Xie Chengjun, born in September, 1963, Jilin, Jilin, China

Current position, grades: professor in the school of computer science and technology Beihua University, Ph.D. graduate, received a doctor degree in engineering. School of computer science and technology Beihua University Dean, top-notch innovative talents in Jilin Province, Jilin Province Department of education focus on training of young teachers, Jiangcheng elite talent, director of research Chinese computer education in higher normal school, director of the Jilin computer society, the expert review committee of project of national Natural Science Fund, hall of Jilin province science and technology project evaluation experts, master of computer application technology tutor, Beihua University computer professional academic leaders, academic committee of Beihua University, Journal of Beihua University (NATURAL SCIENCE EDITION) editorial board.

Jiang Zhuo, born in March, 1978, Jilin, Jilin, China

Current position, grades: lecturer in the school of computer science and technology, to obtain a master's degree in engineering, science and technology, she is also the director of computer
Scientific interest: mainly teaches data structure, software engineering, object oriented 8 courses analysis and design, guiding students curriculum design, graduation design. Presided over the completion of the "software engineering" quasi bilingual demonstration course project, participated in the "excellent course network teaching model research and practice", "local college computer professional basic course platform is arranged on the reform and practice of teaching research project"
Experience: participated in "software engineering" one textbook. Participate in program design, data structure, software engineering best construction. She also won the two prize of the Beihua University fourth session of young teachers and multimedia teaching contest, teaching excellence award. Beihua University outstanding teachers, outstanding communist title.

Zhao Jinyan, born in October, 1963, Jilin, Jilin, China

Current position, grades: professor at the school of computer science and technology, was a master's degree in engineering. She is also vice president Institute of computer science and technology.
Publications: 12 scientific research, editor of 2 books, published 9 papers
Experience: Once obtained the middle-aged backbone teacher, "outstanding communist", "Jilin city's plastic model", "three eight" red banner pacesetter in Jilin province and the educational system ethics advanced individual title. Access to the provincial teaching achievement 2 (first prize, three prize 1 items), the completion of the provincial research projects 3; awarded at provincial level or above education technology achievement award 5 items (including the Ministry of Education 1, Provincial Department of education, 4)

The impact on collaborate level of cluster

Lirong Qiu*

School of Information Engineering, Minzu University of China, Beijing, China

Received 1 March 2014, www.cmnt.lv

Abstract

Due to the impact of the characters of nodes in complex network on collaboration level, we put forward a new iterated game model based on conformist mechanism. In this model, nodes can update tactic not only according to their payoffs but also to their species, which they belong to. The new model can assure that nodes in the same species adopt the same tactic. Simulation results show that the collaboration level of the networks that adopt conformist mechanism is higher than the networks that adopt normal mechanism. In the other words, the collaboration level is in inverse proportion to the species number. On the other hand, we find that the average payoffs increased with the penalty gene instead of increasing alternately. So the new model can promote the collaboration level and the average payoffs of the nodes in network at the same time.

Keywords: Conformist Mechanism, Iterated Game, Collaborate Level, Complex Network

1 Introduction

Traditional evolutionary game theory considers the main nature, which prefers to select selfish and stronger individuals. However, there are lots of cooperation behaviours in nature, which contradicts the natural section rule. Game theory is an efficient tool in the field of studying biological, economic, and social relations of complex network etc. Especially after the notion of Prisoner's Dilemma Game (PD game) proposed by Neumann and Morgenstern, various related research works are applied into many fields. The prisoner's dilemma is a canonical example of a game analysed in game theory that shows why two purely "rational" individuals might not cooperate, even if it appears that it is in their best interests to do so. It was originally framed by Merrill Flood and Melvin Dresher working at RAND in 1950. Albert W. Tucker formalized the game with prison sentence rewards and named it "prisoner's dilemma" (Poundstone, 1992).

The basic idea of PD game is as follows: two thieves are arrested and imprisoned. Each thief is in solitary confinement with no means of speaking or exchanging messages with the other. Here is how it goes:

(1) If A and B both defect from the other, each of them serves 2 years in prison.

(2) If A defects from B but B remains silent, A will be set free and B will serve 3 years in prison (and vice versa).

(3) If A and B both remain silence (which means that they cooperate), both of them will only serve 1 year in prison (on the lesser charge).

It is implied that the prisoners will have no opportunity to reward or punish their partner other than the prison sentences they get, and that their decision will not affect their reputation in future. Because defect from a partner offers a greater reward than cooperating with them, all purely rational self-interested prisoners would defect from the other, and so the only possible outcome for two purely rational prisoners is for them to defect from each other. The interesting part of this result is that pursuing individual reward logically leads both of the prisoners to betray, when they would get a better reward if they both cooperated. In reality, humans show a systematic bias towards cooperative behaviour in the similar games, much more than predicted by simple models of "rational" self-interested action. A model based on a different kind of rationality, where people forecast how the game would be played if they formed coalitions and they maximize their forecasts, has shown to make better predictions of the rate of cooperation in this and similar games given the payoffs of the game.

Generally, in PD game, each party (denoted as P1 and P2) has two choices: Cooperation (C for short) and Defection (D for short). Utility is defined as follows:

The utility is T for P1 when P1 adopts D and P2 adopts C;

The utility is S for P1 when P1 adopts C and P2 adopts D;

Both P1 and P2 get R if they adopt C;

Both P1 and P2 get P if they adopt D.

We often use a matrix A to describe the relationships between P1 and P2.

$$A=\begin{bmatrix} R & S \\ T & P \end{bmatrix},\ T>R>P>S,\ 2R>T+S.$$

For one-short PD game, where the game is played only once, both parties would rather to adopt D since D is dominating strategy for them. However, the optimal result for both parties is (C, C) since R is bigger than P.

* ***Corresponding author*** e-mail: qiu_lirong@126.com

Therefore, the problem is how to encourage both parties to cooperation in PD game. One solution for this problem is iterated PD game, where parties interact for several rounds. The reason is that in iterated PD game, parties can adopt certain retaliatory measures to punish those who adopt D. Thus, parties who adopt D may not defect in the following rounds PD games. Therefore, both parties would like to cooperate in iterated PD games.

On the other hands, cooperation often appears in complex network. Santosand Pacheco [1] fink that the scale-free network is beneficial to cooperation emergence and maintenance. Furthermore, they also discuss snow-drift game model in scale-free network [2] and the results show that mixed network is more conductive of cooperation emergence. Vainstein and Arenzon [3] find that some potential disruption in sparse grid can strengthen the density of cooperation. Riolo [4] studied a game model, where does not exists any reciprocity.

In this model, parties decide whether to cooperate according to the similarity of their opponents. Although this model may boost cooperation, it has a strong assumption. That is each party in this game must interact with those who are similar to him. This game is degenerated to PD game once this assumption is released. Therefore, cooperation may not appear in the model without reciprocity mechanism.

This paper proposes a conformity mechanism, which considers the impact of conformity to the results of game theory. Here parties should not only consider his own utility but also majority parties in the same set. Firstly, this paper describes the basic idea of conformity scheme. Secondly, discuss the simulation results of the scheme. The results show that this scheme can greatly boost cooperation among parties in network. Furthermore, conformity can avoid the average utility rise alternately such that it has a proportional relationship with punishment factor.

2 The conformity scheme

In order to boost cooperation among parties, scientists propose various methods and schemes. A simplest strategy is Tit-for-Tat (TFT for short).

Tit for tat is a highly effective strategy in game theory for the iterated PD game. The strategy was first introduced by Anatol Rapoportin Robert Axelrod's two tournaments [1], held around 1980. Notably, it was (on both occasions) not only the simplest strategy but also the most successful in direct competition. An agent using this strategy will first cooperate, and then subsequently replicate an opponent's previous action. If the opponent previously was cooperative, the agent is cooperative. If not, the agent is not. This is similar to super-rationality and reciprocal altruism in biology. The success of the tit-for-tat strategy is astonishing, which is largely cooperative despite that its name emphasizes an adversarial nature. Arrayed against strategies are produced by various teams it won in two competitions. After the first competition, new strategies formulated specifically to combat tit-for-tat failed due to their negative interactions with each other; a successful strategy other than tit-for-tat would have had to be formulated with both tit-for-tat and itself in mind.

The results may give insight into how groups of animals (and particularly human societies) live in largely (or entirely) cooperative societies, rather than the individual way, which is in "red in tooth and claw" way that might be expected from individuals engaged in a Hobbesian state of nature. This, and particularly its application to human society and politics, is the subject of Robert Axelrod's book “The Evolution of Cooperation”.

Moreover, the tit-for-tat strategy has been of beneficial used to social psychologists and sociologists in studying effective techniques to reduce conflict. Research has indicated when individuals who have been in competition for a period of time no longer trust one another, the most effective competition reverser is the use of the tit-for-tat strategy. Individuals commonly engaged in behavioural assimilation, a process in which they tend to match their own behaviours to those displayed by cooperating or competing group members. Therefore, if the tit-for-tat strategy begins with cooperation, then cooperation ensues. On the other hand, if the other party competes, then the tit-for-tat strategy will lead the alternate party to compete as well. Ultimately, each action by the other member is countered with a matching response, competition with competition and cooperation with cooperation.

In the case of conflict resolution, the tit-for-tat strategy is effective for several reasons: the technique is recognized as clear, nice, provocable, and forgiving. Firstly, it is a clear and recognizable strategy. Those using it quickly recognize its contingencies and adjust their behaviour accordingly. Moreover, it is considered to be nice as it begins with cooperation and only defects in following competitive move. The strategy is also provocable because it provides immediate retaliation for those who compete. Finally, it is forgiving as it immediately produces cooperation should the competitor make a cooperative move.

Individuals who employ the tit-for-tat strategy are generally considered to be tough but fair—a disposition that is often respected in the business/organization world. Those who always cooperate with a competitor are often viewed as weak, while those who consistently compete are perceived as unfair. In any case, the implications of the tit-for-tat strategy have been of relevance to conflict research, resolution and many aspects of applied social science.

In addition, if parties in game theory can observe other parties’ strategy and assign values on reputation, then this indirect reciprocity can explain why cooperation emergence. However, it is a controversial topic which scheme leads to this cooperation. Recently, some research works show that some different indirect reciprocity can effectively result in cooperation emergency.

The models without reciprocity can also lead to cooperation emergence such as Riolo model. Although this model has some shortcomings it has caught wide interests since it makes parties to cooperate on the basis of similarity. If interaction is not random, then the group of co-operators can exist in the settings without cheaters. That is cooperation mechanism based on similarity can lead to a high level of cooperation.

Previous models only consider his own utility while they do not consider the impact when other parties update their strategy. In fact, parties' strategy are easy affected by other parties' strategy. This is what we called conformity. More specifically, parties would like to adopt the strategies which are adopted by most parties in network although this strategy may be not optimal.

In interpersonal network, nodes denote parties in games, lines between nodes denote the relationships of two parties. Previous research work often neglects some properties of nodes. In this paper, we set up a network considering those properties. For example, a node has commercial benefits and then we can divide nodes into several groups according to their different characters. The conformity idea is similar to Riolo model. The difference lies in that Riolo model release the limits for parties, where they cannot only interact with parties in the same group but also parties in other groups. Furthermore, the utility matrix is changed with strategy. In iterated PD games, we adopt dynamic matrix. The game model is set up as follows.

(1) Choose Zachary [6] network as research object. Zachary's karate club: social network of friendships between 34 members of a karate club at a US university in the 1970s. Please cite W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research 33, 452-473 (1977).

(2) Set up utility matrix. We adopt the matrix as Nowak and May A=$\begin{bmatrix} 1 & 0 \\ b & 0 \end{bmatrix}$, where 1<b<2.

(3) Randomly initialize strategies for each party.

(4) Node I play games with his neighbour j. If i adopt strategy C then we do not update utility matrix.

Otherwise, the game proceed into next round. Before entering the next round, let b=b-Q, where Q is punish factor. The utility of each neighbour lj_i of party i is $S(lj_i)$. After party i interact with all neighbours, choose maximize utility $\max(S(lj_i))$ among $lj_{\max i}$. Finally update the strategy of party i as his neighbours' strategy $lj_{\max i}$.

(5) Entering the next round, divide different groups according to their strategies. That is, parties belong to the same group if they adopt same strategies. Here in order to contrast the efficiency of new schemes, we adopt a reference experiment. In the reference experiment nodes adopts the strategy of those who have highest utility.

(6) Repeat step (4), (5) till the game end.

In the new model, nodes in network first play game with his neighbours and adopt strategies of those who have the highest one. If the game proceeds into the next round, nodes will adopt the strategy, which is adopted by most parties in the same group. This agrees with conformity and is presented in Figure 1.

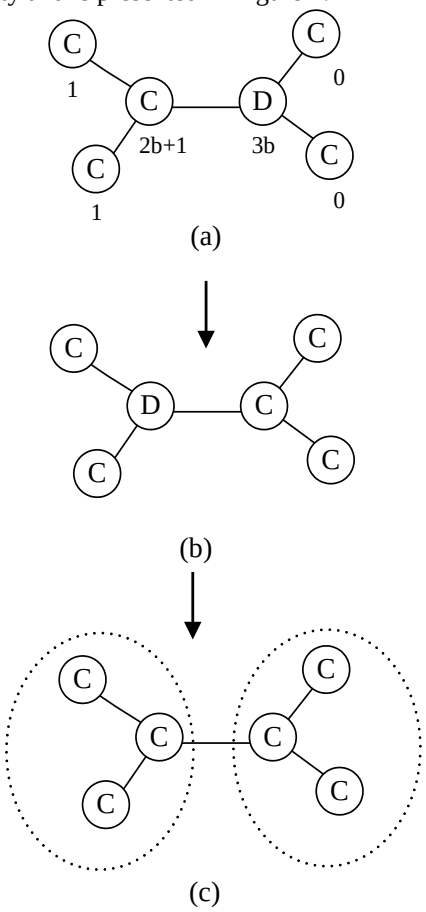
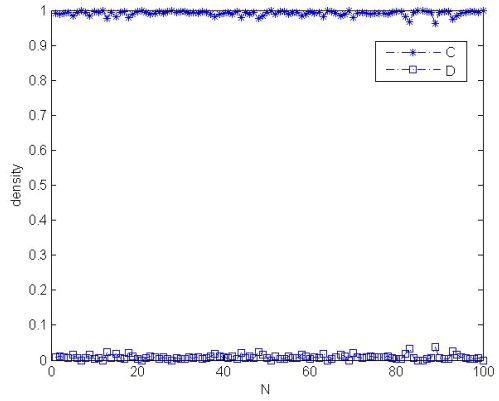


FIGURE 1 The proceed of conformity

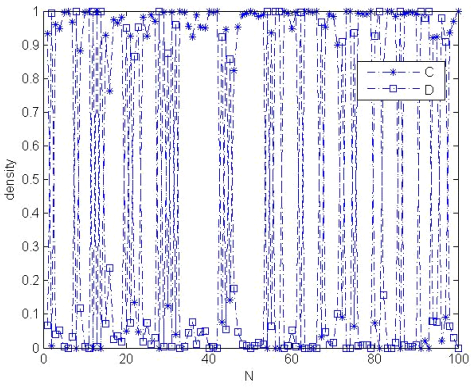
Since initial strategy is randomly set, the initial utility of the node is shown in Figure 1(a). Then after one round of game, node A adopt the strategy of node B. Since node B has the highest utility among all neighbours of node A, node A adopt the strategy of node B in next round. For the same reason, node B adopts the strategy of node A. This proceed is shown in 1(b). However, node A should adopt the strategy of those most parties adopt in the same group according to the conformity scheme as shown in 1(c), where dotted line mean two groups. We can see that the cooperation level in the network is improved due to conformity.

3 Simulation results

When node is in the same group, this model reduces to the model in [4]. When parties update their utility after one round, then this model reduce to the model in [7]. When the network is divided into three groups, the impact of this model and models in [7] is shown in Figure 2.



(a)

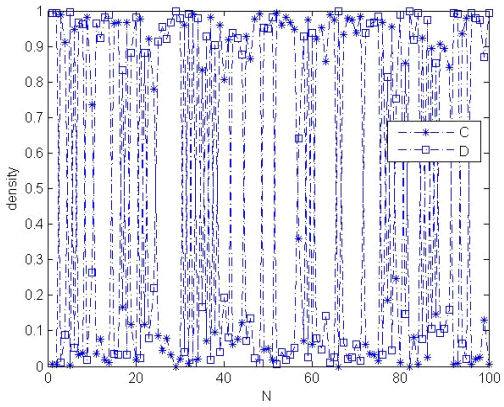


(b)

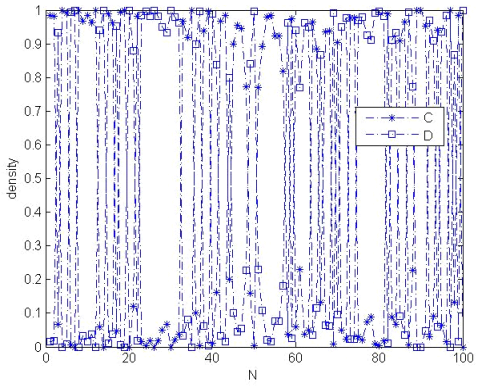
FIGURE 2 Network strategy density (when there are 3 groups)

Figure 2(a) denotes the network strategy density when the network is divided into 3 groups and the rounds number is 100. Figure 2(b) denotes that each node update their utility according to his own strategy. As shown in Figure 2, the cooperation and defect strategy rise alternatively in Tomochi and Kono and there are no obvious bounds. And cooperation level is obviously higher than defect level in our model. This illustrate that conformity greatly improve the cooperation level in the network. Iterated games can improve cooperation level. Therefore, most parties in the network are like to adopt cooperation. Furthermore, nodes in the whole group choose their strategy according to conformity scheme. So we can regard that individual party may give up his own benefit in order to conform to the benefit of the whole group.

Note that conformity scheme may lose his efficiency when the number of nodes is equal to the whole network. That is, each node belongs to one group. The reason is that each node will only consider the strategy of him. There is no group to conform. In order to study the impact of group scale to cooperation level, we discuss the cooperation and defect density when there are 10 groups in one network. The results are shown in Figure 3. Just as before, we also adopt Tomochi and Kono as a reference.



(a)



(b)

FIGURE 3 Network strategy density (when there are 10 groups)

Compared with 3(a) and Figure 2(a), we find that cooperation and defect rise alternatively in both figures and there are also no obvious bound in these two figures. This results show that the impact of conformity scheme when the group number is large is less obvious than that when the group number is small. Compared 3(a) and Figure 3(b), we find that the former has a higher cooperation levels, which shows that conformity scheme is dominating.

Furthermore, we also find that conformity can also improve the average utility in the whole network. The average utility alternatively increases in [7]. However, when we adopt conformity scheme, there is a linear relation between average utility and punishment factor Q. Figure 4 denotes the average utility when the group number is 3 and 10. The results show that the average utility becomes small when the number of groups becomes larger.

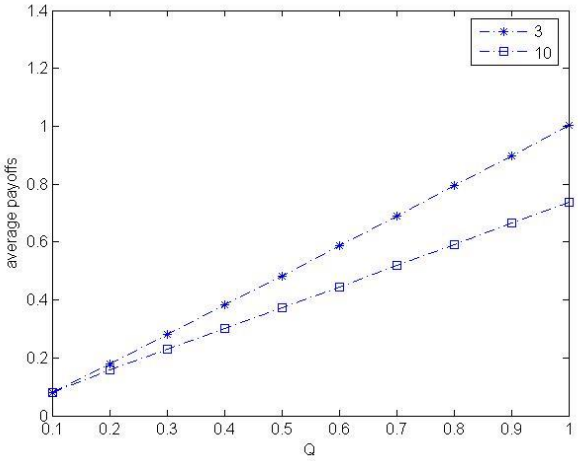


FIGURE 4 Network average utility

4 Conclusions

With the development of complex network, the structure of complex network becomes more and more popular especially for the cooperation strategy. Game theory on complex networks will be more and more interesting and become to be the centre of the evolution theory. The emergence of group cooperative behaviour and stability maintenance are confused phenomenon for scientists and the structure evolution plays an important role in complex system.

This paper proposes a conformity scheme to boost cooperation among parties. The basic idea is to use strategies in iterated games such that each node adopts the strategy of those who have the highest utility in each round. That is, if the strategy conforms to the majority strategies in the group, then parties remain to adopt this strategy. Otherwise, parties adopt the strategy, which is adopted by most nodes in the same group. The simulation results show that this conformity scheme can greatly improve the cooperation level in the network.

Acknowledgements

Our work is supported by the National nature science foundation of China (No. 61103161), the Program for New Century Excellent Talents in University (NCET-12-0579) and the "985" special funds in School of information engineering, Minzu university of China.

References

[1] Pacheco J M, Traulsen A, Nowak M A 2006 Coevolution of strategy and structure in complex networks with dynamical linking *Physical Review Letters* **97**(25) 103-18
[2] Van Segbroeck, Sven, et al. 2011 Selection pressure transforms the nature of social dilemmas in adaptive networks *New journal of physics* **13**(1) 137-50
[3] Nowak M A 2006 Five rules for the evolution of cooperation *Science* **314**(5805) 1560-3
[4] Riolo R L, Cohen M D, Axelrod R 2001 Evolution of cooperation without reciprocity *Nature* **414**(6862) 441-3
[5] Nowak M A 2006 Five rules for the evolution of cooperation *Science* **5805** 1560-63
[6] Zachary W W 1977 An Information Flow Model for Conflict and Fission in Small Groups *Journal of Anthropological Research* **33** 452-73
[7] Tomochi Masaki, Mitsuo Kono 2002 Spatial prisoner's dilemma games with dynamic payoff matrices *Physical Review E* **65**(2) 112-20

Authors

Lirong Qiu

Current position, grades: associate professor of Information Engineering Department, Minzu University of China.
University studies: Ph.D. degree in Computer Science from Chinese Academy of Science (2007).
Scientific interests: natural language processing, artificial intelligence and distributed systems.

A gait recognition system based on BP neural network and plantar pressure

Tengda Shi, Kaifeng Su, Linqian He, Lei Yan*

[1]*School of Technology, Beijing Forestry University, Beijing, China, 100083*

Received 6 August 2014, www.cmnt.lv

Abstract

In order to get a faster, more effective and stable control over lower extremity exoskeleton of power assist robot, precise examination on gait information is necessary, thus it is so important to design and establish a gait recognition system with accurate detection. In this paper, a wireless in-shoe wearable plantar pressure acquisition system based on ATmega16 and 8 FSR sensors will be applied to data acquisition for the gaits which consist of standing, walking, jumping and going upstairs. And four volunteers (2 males and 2 females) will be invited in this research to collect the pressure information. The NNT of MATLAB will be applied to establish an 8-12-4 BP neural net model. The input factors come from the eight sensors of plantar pressure system, the output is gait category. Proved by a great deal of experiments, the gait recognition method proposed in this research is quite feasible.

Keywords: gait recognition, BP neural networks, plantar pressure, lower extremity exoskeleton of power assist robot

1 Introduction

1.1 RESEARCH STATUS AND APPLICATION PROSPECTS OF GAIT RECOGNITION SYSTEM

As a developing research field about the technology in recognition of biological characteristics, gait recognition [1] has become the focus of computer fans. How to obtain effective gait information and the accuracy of gait recognition have become the key points and difficulty in recent study.

Bobick [2] has determined human's characteristics of step and extensibility of legs with the aid of structured analysis method in Georgia Tech. The outcome was concluded by some experiments, which involved the problems of shadow and sunlight, because the experiments were finished indoors and outdoors. Kale [3] used the human's side profiles in binarized images as the features of images. Then for each person, they used c-means algorithm to choose a sample in a single gait cycle. In the process of recognition, the FED vector between each frame in a gait sequence and the sample was calculated and recognized by HMM. Lee [4] used seven ovals to represent different body parts in the binarized images of human's side profile. Using four characteristic values including barycenter to represent the height of the body image's barycentre. Meanwhile all 29 characteristic values were used to represent the image of a person's body side. The gait recognition was completed via the method of template matching. In Little and Boyd's [5] method, the gait recognition was achieved with the help of gait sequence images' optical frequency and phase. While Phillips [6] utilized different time-space domain template to accomplish the gait recognition. For Cunado [7], a pendulum module involved thigh and shank was established so that gait characteristics could be detected from the frequency component of the pendulum angel signal.

Gait recognition is a kind of technology that is used for detecting within a certain distance in detecting human's instinct biological characteristics like fingerprint identification and face recognition, some problems related to close contact in physics always exist. In this way, gait recognition can be applied well and has a bright future.

1.2 THE BASIC PRINCIPLES AND CHARACTERISTICS OF ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN), also called Neural Networks(NN), is a kind of artificial bionic model based on the theory which about synaptic connections in the brain. It is made up of a great deal of neurons [8].

Neuron, the most basic unit in the neural network, plays a quite important role in the network. Different networks are made of different neurons which connects with each other in different ways, which means the structures of neural networks are also various. So they have various functions, too.

ANN has the following outstanding characteristics:

1) it can make the input sample approximate complex non-linear function relation well;

2) it has good fault-tolerant ability and robustness;

3) NN uses the paralleled distributed methodology to manipulate the input sample's information, which makes scientists can finish big-scale arithmetical operation faster.

* ***Corresponding author*** e-mail: mark_yanlei@bjfu.edu.cn

1.3. NEURAL NETWORK MODEL

BP neural network, also called Error Back Propagation Neural Network, is a kind of forward multi-layered neural network. In BP network, signals propagate forwards, while errors propagate backwards. BP network often includes input layer, hidden layer and output layer. Theoretically, it can increase the accuracy and expression ability of the network by adding hidden layers.

According to Kolmogorov Theory, a three-layer BP network can form any mapping from m-dimensions to n-dimensions, which means that just one hidden layer is enough. Such a NN can approximate any continuous function correctly [9] (Figure 1).

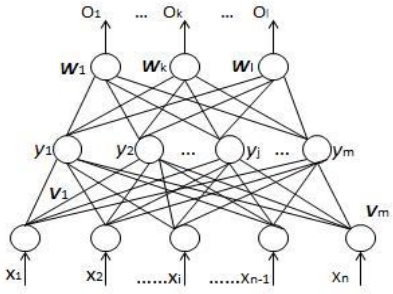


FIGURE 1 A single hidden layer of BP network model

2 BP neural network established for gait recognition

2.1 IN-SHOE WEARABLE PLANTAR PRESSURE ACQUISITION SYSTEM

Due to several outstanding characteristics which include portability, flexibility and great convenience, plantar pressure acquisition system has been one of the most important application techniques in gait recognition and has been extended to other relevant fields, such as medical testing, sport science and robotics research.

In-shoe plantar pressure acquisition system, which is able to measure plantar forces and detect gait-phases of human at the same time, has become an attractive alternative for ground mounted force platform. Researchers in this field can make more products by using it [10].

Many researchers have proposed many different numbers of force sensors to recognize the gait phase. For instance, a research group used four force sensors to identify the gait phases with a classification algorithm and has obtained good results [11], Faivre employed eight sensors in the in-shoe plantar pressure system according to the literature [12], and flexible force sensor(Tekscan Inc., USA) and FSR sensor(Interlink Electronics, USA) has been commonly used.

As we illustrated in the Figure 2, the heel, metatarsals and hallux are the main regions to sustain people's weight [13]. Thus, we designed to place four force sensors configured at the Heel, Meta 2nd, Meta 1st and Hallux for each foot to obtain plantar pressure and detect gait phase. These sensors were packaged in two pairs of insole and the insole was adhered to a pair of shoes, as shown in Figures 3 and 4.

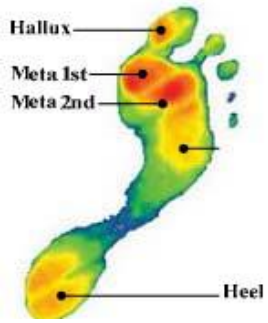


FIGURE 2 Plantar pressure distribution [14]

FIGURE3 The package of sensors in the shoes

FIGURE4 The position of sensors in insole

2.2 DATA ACQUISITION AND PROCESSING

In order to get more universal and accurate data of plantar pressure, we invited four volunteers: two females and two males to conduct the data acquisition of this experiment. The brief information of these four volunteers is shown in Table 1.

TABLE 1 General information of volunteers

Gender	Age	Height	Weight	Feet size
Female	23	167cm	55kg	39cm
Female	22	160cm	48kg	37cm
Male	20	174cm	60ke	41cm
Male	22	167cm	65kg	40cm

A network can have a plurality of output variables, in this study, there are four output variables, because of the input of numeric variables, the output is also corresponding to numerical quantity. Actually, network's output is only 0 to 1 or -1 to 1, so the desired output must

need to be normalized.

Normalization is a method of data processing by means of the network input and output limits in the [0,1] or [-1,1]. Normalization process is performed in order to avoid absolute error components that the large value is large and the small one is small, the network will just adjust the weight value of the total error, as a result, the output component that has a small share of the total error will have larger relative error. In this paper, the output of all the types of gait is normalized, thus [1,0,0,0] was chosen to represent 'standing'; [0,1,0,0] represented 'walking'; [0,0,1,0] for 'jumping' and [0,0,0,1] for 'going upstairs'.

2.3 ESTABLISH GAIT RECOGNITION SYSTEM BASED ON BP NEURAL NETWORK MODEL

The purpose of this research is to achieve information of human gait which gathered based on plantar pressure system and then use MATLAB software to simulate the BP algorithm to find the suitable BP network model to distinguish the type of gait. A total of eight input characteristic value and four output type value. After a large number of experiments, according to preparation of the train-dataset, test-dataset and design of the BP network' structure, the following three-tier network model constructed by MATLAB R2012a's BP neural network toolbox [15], a good gait recognition system was obtained in this research:

1) eight input-layer neurons;

2) output-layer nodes in the hidden layer was set to twelve according to the number of training samples and the dimension of the input and output layer nodes;

3) the dimension of the output layer equals the types of gait, four; the construction principle of the output layer is: different types of output in the corresponding position is 1 and the rest position is 0, for example, [1,0,0,0] represented 'standing', [0,1,0,0] represented 'walking', [0,0,1,1] for 'jumping' and [0,0,0,1] for 'going upstairs'.

4) logsig is selected as the network error transfer function from the input layer to the hidden layer and logsig from hidden layer to the output layer.

5) we chose Powell-Beale connection gradient BP training function 'trainlm' as the learning function;

6) The learning numbers of the network is set to 3000, the learning rate is set to 0.0005 and training objectives set to 0.004. We used an 800x8 dataset which contained 200 data from 'standing', 200 data from 'walking', 200 data from 'jumping' and 200 data from 'going upstairs' to train the network. A 400x8 data that comprised 100 data from 'standing', 100 data from 'walking', 100 data from 'jumping' and 100 data from 'going upstairs', with the accuracy of the test results and the square error value we can judge whether the types of gait can be classified.

3 The training and simulation results of BP neural network model

After the BP neural network's hidden layers, the number of neurons, the input and output layers are all determined, it is virtually certain that the network model can be simulated by programming in MATLAB R2012a(as shown in Figure 5)

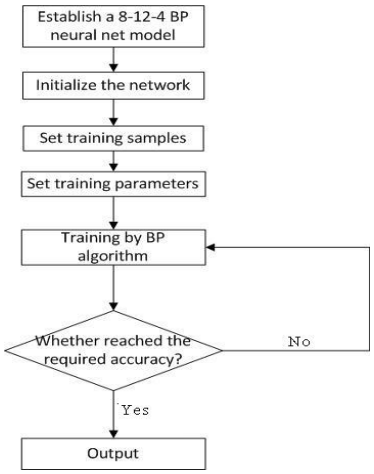


FIGURE 5 System architectures of based on BP Neural Networks of the gait recognition system

The final classification model based on BP neural network was done by the algorithm designed by this research through neural network toolbox of MATLAB. And result of final gait recognition was shown in Figure 6.

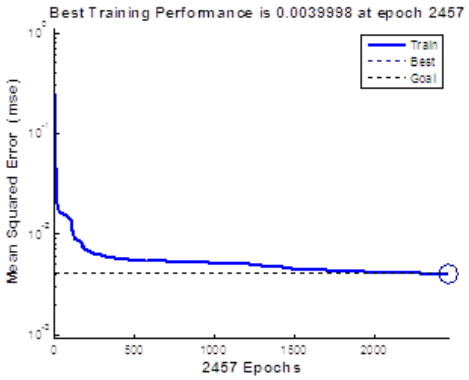


FIGURE 6 The results of training MSE

From the experimental result of the final gait recognition shown in Table 2, it can be concluded that we obtain a better gait recognition system which is based on BP neural network and plantar pressure. The accuracy of the gait recognition is more than 90% except for 'going upstairs' because of its complication among other three gaits.

TABLE 2 The result of final gait recognition

Gaits	True number of gaits	Classified number of gaits	Accuracy of the gait recognition
Standing	100	99	99%
Walking	100	90	90%
Jumping	100	91	91%
Going upstairs	100	85	85%

4 Discussions and conclusions

From the results of the final experiment, we obtain a high recognition rata as shown in the Table2. Consequently, using in-shoe wearable plantar pressure acquisition system to collect gait information is a good method for recognizing the types of gait with the function of classification of BP neural network. It can provide much more accurate gait information of human for the lower extremity exoskeleton of power assist robot, making the robot provide better services to help the human body.

Although we get a high recognition rate from the results of the final experiment, this is just a preliminary study, there are still many issues that need further study, such as finding a better algorithm to improve the recognition accuracy of 'going up-stairs', studying more complicate gait and so on. Thereby, there are lots of works to do to improve the accuracy rate of gain recognition.

Acknowledgment

This work was financially funded by National Training Programs of Innovation and Entrepreneurship for Undergraduates (Grand No.: G201410022041). And we would like to acknowledge that the four volunteers are studying in Beijing Forestry University.

References

[1] Boulgouris N V, Hatzinakos D, Plataniotis K N 2005 *IEEE signal processing magazine* **22**(6) 78-90
[2] Lee L, Grimson W E L 2002 Gait analysis for recognition and classification *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition 2002* 148-155
[3] Kale A, Rajagopalan A N, Cuntoor N, Kruger V 2002 Gait-based Recognition of Humans Using Continuous HMMs *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition 2002* 336-41
[4] Wang L, Tan T, Ning H, Hu W 2003 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(12) 1505-18
[5] Bobick A F, Johnson A Y 2001 Gait recognition using static, activity-specific parameters *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2001 CVPR 2001* **1** I-423-I-430
[6] PJ Phillips, S Sarkar, I Robledo, P Grother, K Bowyer 2002 Baseline Results for the Challenge Problem of Human ID Using Gait Analysis *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition 2002* 130-5
[7] Cunado D, Nixon M S, Carter J N 2003 Automatic Extraction and Description of Human Gait Models for Recognition Purposes *Computer Vision and Image Understanding* **90**(1) 1-41
[8] Basheer I A, Hajmeer M 2000 Artificial neural networks: fundamentals, computing, design and application *Journal of microbiological methods* **43**(1) 3-31
[9] Finlay J, Beale R 1993 Neural Networks and Pattern Recognition in Human-computer Interaction *ACM SIGCHI Bulletin* **25**(2) 25-35
[10] Lemaire E D, Biswas A, Kofinan J 2006 Plantar Pressure Parameters for Dynamic Gait Stability Analysis *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2006 EMBS'06* 4465-8
[11] Okaza Y, Yoshida N, Oguchi K 2006 Using Sole Pressure Signals to Analyze Walking Posture *International Conference on Biomedical and Pharmaceutical Engineering 2006 ICBPE 2006* 197-200
[12] Faivre A, Dahan M, Parratte B, Monnier G 2004 Instrumented Shoes for Pathological Gait Assessment *Mechanics Research Communications* **31**(5) 627-32
[13] Cavanagh P R, Rodgers M M 1987 Pressure distribution under symptom-free feet during barefoot standing *Foot & Ankle International* **7**(5) 262-78
[14] Nathan D G 1998 Clinical research: perceptions, reality, and proposed solutions *Jama* **280**(16) 1427-1431.
[15] Chua L O, Yang L. 1998 *IEEE Transactions on Circuits and Systems* **35**(10) 1273-90

Authors

Tengda Shi, born in June, 1993, Beijing Forestry University, Beijing, P.R. China

Current position, grades: a senior student (BSc) in Electrical Engineering at Beijing Forestry University, China.
Scientific interests: image processing, pattern recognition, computer simulation.
Publications: more than 3 papers.
Experience: teaching experience of 2 years, 3 scientific research projects.

Kaifeng Su, born in October, 1992, Beijing Forestry University, Beijing, P.R. China

Current position, grades: A senior student (BSc) in Electrical Engineering at Beijing Forestry University, China.
Scientific interests: pattern recognition.
Publications: more than 1 paper.
Experience: teaching experience of 2 years, 3 scientific research projects.

Linqian He, born in April, 1995, Beijing Forestry University, Beijing, P.R. China

Current position, grades: A sophomore student (Bsc) in Electrical Engineering at Beijing Forestry University, China.
Scientific interests: pattern recognition.
Experience: teaching experience of 1 years, 3 scientific research projects.

Lei Yan, born in June, 1979, Beijing Forestry University, Beijing, P.R. China

Current position, grades: Associate Professor of School of Technology, Beijing Forestry University, China.
University studies: M.Sc. at Jilin University in China. PhD at Kyungpook National University in Korea.
Scientific interests: development of automatic control systems, image processing, pattern recognition.
Publications: more than 40 papers.

Automatic verification of embedded system based on EFSM

Jinjiang Liu*, Jingjing Liang

School of Computer and Information Technology, Nanyang Normal University, Nanyang 473061, Henan Province, China

Received 6 July 2014, www.cmnt.lv

Abstract

To ensure the correctness of embedded system, automation of test case generation is necessary in industrial. This paper present a technique for specifying coverage criteria and a method for generating test suites for embedded systems whose behaviours is depend on its interactive environment. The embedded system under test can be described as extended finite state machines (EFSM) and the coverage criteria can be specified as monitor automata with parameters, which monitor and accept traces that cover a given test criterion of an EFSM. The flexibility of the technique is demonstrated by specifying a number of well-known coverage criteria based on control- and data-flow information using observer automata with parameters. We also develop a method for generating test cases from coverage criteria specified as observers. It is based on transforming a given observer automata into a bitvector analysis problem that can be efficiently implemented as an extension to an existing state-space exploration such as, e.g. SPIN or Uppaal.

Keywords: EFSM (Extended Finite State Machine), embedded system, test case generation, model-based testing

1 Introduction

A Model based test case generation has in recent years been developed as a prominent technique in testing of reactive software systems. A model serves both the purpose of specifying how the system should respond to inputs from its environment, and of guiding the selection of test cases, e.g., using suitable coverage criteria. Typical notations for such models are state machines in some form, often extended with data variables. Test cases can be selected as individual "executions" of the model, checking that the outputs from the system under test (SUT) conform to those specified by the mode.

In this paper, we present a technique for specifying coverage criteria in a simple and flexible manner, and a method for generating test cases according to such coverage criteria. The technique fits well as an extension of a state space exploration tool, such as, e.g., SPIN [2] or Uppaal [4], which performs enumerative or symbolic state-space exploration. It can also be used to generate monitors that measure the coverage of a specific test suite by monitoring the test execution.

Most related work on test case generation from models of reactive systems employs some rather specific selection of coverage criteria. Explicitly given test purposes have been considered, both enumerative [5-7] and symbolic [9]. Test purposes in these works can in some sense be regarded as coverage observers, but are not used to specify more generic coverage criteria and do not make us of parameterization, as in our work.

Some approaches present more flexible techniques for specifying a variety of coverage criteria. Hong et al [11-14] describe how flow-based coverage criteria can be expressed in temporal logic. A particular coverage item is expressed in CTL, and a model checker generates a trace, which covers the coverage item. In our approach, we use monitors instead of temporal logic, which avoids some of the limitations of temporal logic [15]. Our technique using monitors whit parameters can let one pass of a state-space exploration tool generate a test suite that covers a large set of coverage items, whereas the above approaches invoke a run of a model checker for each coverage item.

The remainder of the paper is structured as follows. We present EFSM in the next section, and monitors in Section 3. In Section 4-5, we show how our definitions of coverage can be used for test case generation, and report on a partial implementation of the technique. Section 6 concludes the paper

2 Extended finite state machine

In this section, it is assumed that a System Under Test (SUT) interacts with its environment through events. Whenever the SUT receives an input event, it responds by performing some local computation and emitting an output event. To a given SUT, we associate a set A of event types, each with a fixed arity. An event is a term of form *a(d1, . . . , dk)* where a is an event type of arity *k* and *d1, . . . , dk* are the parameters of the event. The set *A* of event types is partitioned into *input event types* and *output event types*. A trace is a finite sequence

$$a_1(\bar{d}_1)/b_1(\bar{d}_1') \quad a_2(\bar{d}_2)/b_2(\bar{d}_2') \ \ldots \ a_n(\bar{d}_n)/b_n(\bar{d}_n')$$

of *input/output event* pairs. Intuitively, the trace represents a behaviour where the SUT, starting from its initial state, receives the input event $a_1(\bar{d}_1)$ and responds with the

*Corresponding author's e-mail: nytcc@sina.com

output event $b_1(\bar{d}'_1)$. Thereafter, it receives the input event $a_2(\bar{d}_2)$ and so on. An input sequence is a finite sequence of input events.

Assume a set A_I of input event types, and a set A_O of output event types. An Extended Finite State Machine (EFSM) over (A_I, A_O) is a tuple $< L, L_0, \bar{v}, E >$ where:

1) L is a finite set of locations (control states).
2) $l_0 \in L$ is the initial location.
3) $\bar{v}$ is a finite set of state variables.
4) E is a finite set of edges, each of which is of form:

$$e : l \xrightarrow{a(\bar{w}), g \to \bar{u} := \overline{\exp r} / b(\overline{\exp r'})} l',$$

where e is the name of the edge, l is the source location, and l_0 is the target location. $a(\bar{w})$ is an input event type, and $\bar{w}$ is a tuple of formal parameters of event a, g is a guard, $u := \overline{\exp r}$ is an assignment of new values to a subset $\bar{u} \subseteq \bar{v}$ of the state variables, and $b(\overline{\exp r'})$ is an expression which evaluates to an output event.

Intuitively, an edge of the above form denotes that whenever the EFSM is in location l and receives an event of form $a(\bar{w})$, then, provided that the guard g is satisfied, it can perform a computation step in which it updates its state variables by $u := \overline{\exp r}$, emits the output event $b(\overline{\exp r'})$ and moves to location l'. We require the EFSM to be deterministic, i.e., that for any two edges with the same source location l and parameterized input event $a(\bar{w})$, the corresponding guards are inconsistent.

A system state is a tuple $< l, \sigma >$ where l is a location, and σ is a mapping from $\bar{v}$ to values. We can extend σ to a partial mapping from expressions over $\bar{v}$ in the standard way. The initial system state is the tuple $< l_0, \sigma_0 >$, where l_0 is the initial location, and σ_0 gives a default value to each state variable. A computation step is of the form $< l, \sigma > \xrightarrow{a(\bar{d}')/b(\bar{d}')} < l', \sigma' >$ consist of system state $< l, \sigma >$ and $< l', \sigma' >$, an input event $a(\bar{d}')$ and an output event $b(\bar{d}')$.Such that there is an edge of the (above) form $l \xrightarrow{a(\bar{w}), g \to \bar{u} := \overline{\exp r} / b(\overline{\exp r'})} l'$, for which $\sigma(g[\bar{d} / \bar{w}])$ is true, $\sigma' = \sigma[u \to \sigma(\overline{\exp r}[\bar{d} / \bar{w}])]$ and $\bar{d}' = \sigma(\overline{\exp r'}[\bar{d} / \bar{w}])$. A run of the EFSM over a trace $a_1(\bar{d}_1) / b_1(\bar{d}'_1) \ \ldots \ a_n(\bar{d}_n) / b_n(\bar{d}'_n)$ is a sequence of computation steps:

$$< l_0, \sigma_0 > \xrightarrow{a_1(\bar{d}_1)/b_1(\bar{d}'_1)} < l_1, \sigma_1 > \xrightarrow{a_2(\bar{d}_2)/b_2(\bar{d}'_2)} \cdots \xrightarrow{a_n(\bar{d}_n)/b_n(\bar{d}'_n)} < l_n, \sigma_n >,$$

labeled by the input-output event pairs of the trace.

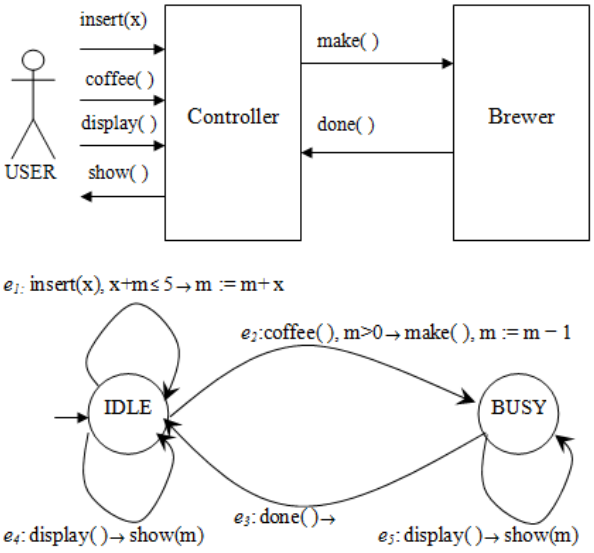


FIGURE 1 EFSM Model of the controller of a simple coffee machine

Example 1: an EFSM specifying for the simple coffee controller: Figure 1 demonstrates an EFSM specifying the behaviour of the controller of a simple coffee machine which interacts with a user and a brewer unit is shown. The controller has L = {*IDLE, BUSY*}, l_0 = *IDLE*, $\bar{v}$ = {m}, A_I = {*insert, coffee, display, done*}, A_O = {*show, make*}, and E = {e_1, e_2, e_3, e_4, e_5}. The parameter x and the variable m take values that are integers in the range [0 ... 5].

3 Monitors

In this section, we present how to use observers to specify coverage criteria for test generation or test monitoring.

As a very simple example, the coverage item "visit location l of the EFSM" can be represented by an monitor with one initial state, and one accepting location, named loc(l), which is entered when the EFSM enters location l. The coverage criterion "visit all locations of the EFSM" can be represented by a parameterized observer with one initial state, and one parameterized accepting location, named loc(L), where L is a parameter that ranges over locations in the EFSM. For each value l of L, the location loc(l) is entered when the EFSM enters location l.

Formally, an *monitor* is a tuple (Q, q_0,Q_f,B) where:

1) Q is a finite set of observer locations
2) q_0 is the initial observer location.
3) $O_f \in Q$ is a set of accepting locations, whose names are the corresponding coverage items.
4) B is a set of edges, each of form $q \xrightarrow{b} q'$, where b is a predicate that can depend on the input event received by the SUT, the mapping from state variables of EFSM to their values after performing the current computation step, and the edge in the EFSM that is executed in response to the current input event.

3.1 OBSERVER PREDICATE

In the following we introduce a more specific syntax for the predicates b occurring on observer edges. The predicates will use a set of predefined *match variables* that are given values at the occurrence of:

1) an event $a(\bar{d})$,

2) an edge $e: l \xrightarrow{a(\bar{w}), g \to \bar{u} := \overline{\exp r}} l'$ of the EFSM, traversed in response to $a(\bar{d})$,

3) the computation step $< l, \sigma > \xrightarrow{a(\bar{d})} < l', \sigma' >$ generated in response to $a(\bar{d})$.

For a traversed EFSM edge we use the following match variables (with associated meaning):

1) *event type* is the event type a of the occurring event;
2) *event-pars* is the list $\bar{d}$ of parameters of the event;
3) *edge* is the name *e*;
4) *target_loc* is the target location *l'*;
5) *guard* is the guard expression *g*;
6) *assignments* is the set *u* := $\overline{expr}$ of assignments;
7) *target_val* is the function from EFSM state variables to values.

To be able to express more interesting properties we also introduce a set of operations that can be used together with the match variables:

3.2 FUNCTION DEFINITION OF PREDICATE

With the match variables and operations above we define new functions that can be used as tests in the observer. In this paper, we shall make use of:

1) *def (v)* which is true *iff* the variable *v* is defined by the transition in the EFSM. This can be expressed as: $v \in map(lhs, assignments)$.

2) *use (v)* which is true *iff* the variable *v* is used (in a guard or assignment) by the transition in the EFSM. This can be expressed as: $v \in vars(map(rhs, assignments)) \vee v \in vars(guard)$.

3) *da* (v_1, v_2) which is true *iff* the variable v_1 is on the right hand side and variable v_2 is on the left hand side of the same assignment in the EFSM specification. The function can intuitively be understood to be true if v_1 directly affects v_2. This can be expressed as: $map\ (affect\ (v_1, v_2), assignments) \neq \varnothing$.

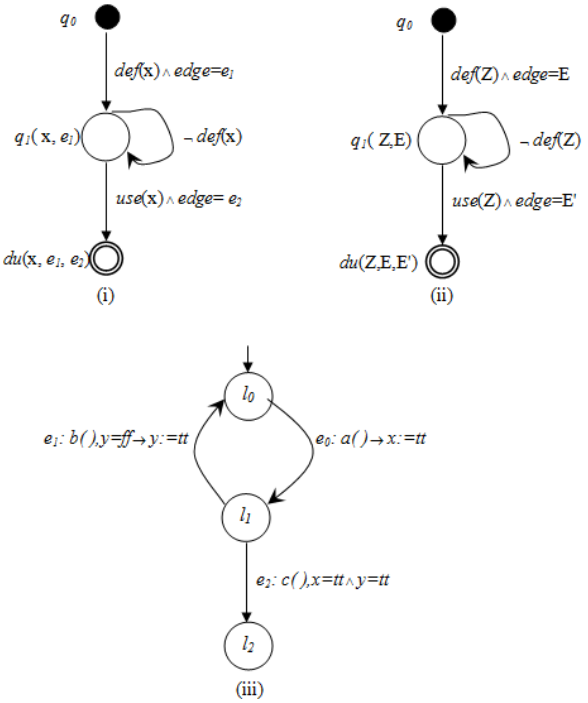


FIGURE 2 Examples of (i) a non-parameterized monitor (ii) a parameterized monitor, and (iii) a simple EFSM

Example 2: example for non-parameterized and parameterized monitor: apparently, the monitor in Figure 2(i) is non-parameterized which specifies definition-use pair coverage for a specific variable m, and specific edges e_1 and e_2. Figure 2(ii) shows a corresponding (parameterized) monitor that specifies definition-use pair coverage for any EFSM variable *Z*, and EFSM edges *E* and *E*'. This is done by parameterizing the location q1 with any variable and any edge, and the accepting location du with any variable and any two edges. The edges are parameterized in a similar way. For example, there is one observer edge from location $q_1(z, e)$ to location *du* (*z*, *e*, *e*') for each EFSM variable *z*, and each pair *e*, *e*' of EFSM edges.

4 Monitoring test generation by superposing a monitor on an EFSM

In test case generation or when monitoring test execution of a SUT, an observer observes the events of the SUT, and the computation steps of the EFSM. Reached accepting locations correspond to covered coverage items. We formally define the execution of an observer in terms of a composition between an EFSM and an observer, which has the form of a superposition of the observer onto the EFSM. Each state of this superposition consists of a state of the EFSM, together with a set of currently occupied observer locations.

Say that a predicate *b* on a monitor edge is satisfied by a computation step $< l, \sigma > \xrightarrow{a(\bar{d})} < l', \sigma' >$ of an EFSM, denoted $< l, \sigma > \xrightarrow{a(\bar{d})} < l', \sigma' > \models b$ if *b* holds for the

event $a(\bar{d})$, the computation step $<l,\sigma> \xrightarrow{a(\bar{d})} <l',\sigma'>$ and the edge $e: l \xrightarrow{a(\bar{w}),g \to \bar{u}:=\overline{\exp r}} l'$ from which the computation step is derived.

Formally, the superposition of an monitor (Q, q_0, Q_f,B) onto an EFSM < L,L_0, $\bar{v}$,E >is defined as follows.

1) *States* are of the form $<<l,\sigma>||Q>$, where $<l,\sigma>$ is a state of the EFSM and Q is a set of locations of the monitor.

2) The *initial state* is the tuple $<<l_0,\sigma_0>||\{q_0\}>$, where $<l_0,\sigma_0>$ is the initial state of the EFSM and q_0 is the initial location of the monitor.

3) A *computation step* is a triple $<<l,\sigma>||Q> \xrightarrow{a(\bar{d})} <<l',\sigma'>||Q'>$ such that $<l,\sigma> \xrightarrow{a(\bar{d})} <l',\sigma'>$ and

$$Q' = \{q' \mid q \xrightarrow{b} q' \text{and } q \in Q \text{ and } <l,\sigma> \xrightarrow{a(\bar{d})} <l',\sigma'> \models b\}.$$

4) A state $<<l,\sigma>||Q>$ of the superposition *covers the coverage item* represented by the location $q_f \in Q_f$ if $q_f \in Q$.

$$<<l_0,\{x=tt,y=tt\}>|| \{q_0,q_1(x,e_0),q_1(y,e_1)\}> \xrightarrow{a()} <<l_1,\{x=tt,y=tt\}>||\{q_0,q_1(x,e_0),q_1(y,e_1)\}>$$

5 Experimental results for three crucial embedded systems

5.1 ALGORITHM

At test case generation, we use the superposition of an observer onto an EFSM, and views the test case generation problem as a search exploration problem. To cover a coverage item q_f is then the problem of finding a trace $tr = <<l_0,\sigma_0>||\{q_0\}> \xrightarrow{a(\bar{d}) \cdots a'(\bar{d}')} <<l,\sigma>||Q>$, such that $q_f \in Q$.

An abstract algorithm to compute test case is shown as below:

1) $Pass := \varnothing$, Max:=0, $w_{max} := w_0$
2) $Wait$:={ $<<s_0||\{q_0\}>,w_0>$ }
3) **while** Wait $\neq \varnothing$ **do**
4) select $<<s||Q>,w>$ from $Wait$
5) **if** $|q_f \cap Q| > Max$ **then**
6) $w_{max} := w$, Max:= $|q_f \cap Q|$
7) **if** for all $<s||Q'>$ in $Pass$: $Q \subsetneq Q'$ **then**
8) add $<s||Q>$ to $Pass$
9) for all $<s''||Q''>$
10) such that $<s||Q> \xrightarrow{a} <s''||Q''>$:
11) add $<<s''||Q''>,wa>$ to $Wait$
12) **return** w_{max} and Max

To improve the presentation, we use s to denote a system of the form $<l,\sigma>$ and s_0 to denote the initial system state $<l_0,\sigma_0>$ and a to denote an input action $a(\bar{d})$. The algorithm computes the maximum number of coverage items that can be visited (*Max*), and returns a trace with maximum coverage (w_{max}). The two main data structures *Wait* and *Pass* are used to keep track of the states waiting to be explored, and the states already explored, respectively.

Initially, the set of already explored states is empty and the only state waiting to be explored is the extended state $<<s_0||\{q_0\}>,w_0>$, where w_0 is the empty trace (in Line2). The algorithm then repeatedly examines extended states from Wait (in Line 3). If a state <s||Q> found in *Wait* is included in a state <s||Q> in *Pass*, then obviously <s||Q> does not need to be further examined(in Line 7-8). If not, all successor states reachable from <s||Q> in one computation step are put on Wait, with their traces extended with the input action of the computation step from which they are generated (in Line 9-11). The state <s||Q> is saved in *Pass*. The algorithm terminates when *Wait* is empty.

The variables w_{max} and *Max* are initially set to the empty trace and 0, respectively (in Line 1). They are updated whenever an extended state is found in *Wait* which covers a higher number of coverage items than the current value of *Max* (in Line 5-6). Throughout the execution of the algorithm, the value of *Max* is the maximum number of coverage items that have been covered by a single trace, and w_{max} is one such trace. When the algorithm terminates (in Line 12), the two values *Max* and w_{max} are returned.

5.2 BITVECTOR IMPLEMENTATION

In order to efficiently represent and manipulate the set Q of observer locations we shall use bitvector analysis [15]. Let the set Q be represented by a bitvector where each bit represents an observer location q'. Then each bit is updated by the following function:

$$f_{q'}(q') = \bigvee_{<b,q> \in in(q')} q \wedge b ,$$

where $in(q') = \{<b,q> \mid q \xrightarrow{b} q' \in B\}$ is the set of pairs of predicates b and source locations q of the edges ingoing to the location q'. That is, given a state of the superposition $<<l,q>||Q>$ and an EFSM transition $<l,\sigma> \xrightarrow{a(\bar{d})} <l',\sigma'>$ the bit representing q' is set to 1 if there is an monitor edge $q \xrightarrow{b} q' \in B$, such that $q \in Q$ and $<l,\sigma> \xrightarrow{a(\bar{d})} <l',\sigma'> \models b$. Otherwise the bit representing q' is set to 0. It should be obvious that this

corresponds precisely to the semantics of an monitor superposed onto an EFSM, described in Section 4.2

Example 4: Interpreting monitor states set Q into bitvector: when the monitor in Figure 2(ii) is superposed onto the EFSM in Figure 2(iii), we have: $E = E' = E = \{e_0, e_1, e_2\}$ and $Z = \bar{v} = \{x, y\}$. Thus, we have that:

$$Q = \{q_0\} \cup \{q_1(z, e_a) \mid z \in \bar{v} \wedge e_a \in E\} \cup \{du(z, e_a, e_b) \mid z \in \bar{v} \wedge e_a \in E \wedge e_b \in E\}.$$

Any enumeration of the set can be used as index in the bitvector. As the observer has three locations with parameters we get three types of bitvector functions:

$$f_{q_0}(q_0) = q_0 \wedge tt, \quad (1)$$

$$f_{q_1(v_i, e_j)}(q_1(v_i, e_j)) = (q_0 \wedge def(v_i) \wedge edge = e_j) \vee (q_1(v_i, e_j) \wedge \neg def(v_i)), \quad (2)$$

$$f_{du(v_i, e_j, e_k)}(du(v_i, e_j, e_k)) = (q_1(v_i, e_j) \wedge use(v_i) \wedge edge = e_k) \vee (du(v_i, e_j, e_k) \wedge tt). \quad (3)$$

There is one function of Equation (1), six of Equation (2), and 18 of Equation (3). Note that Equation (1) is always true and that Equation (3) will remain true once it becomes true, due to implicit self-loops in these locations.

6 Conclusion

This paper has presented a technique for testing the remote environment control systems. Our technique have shown to be a flexible tool in model checking and run-time monitoring, and by this paper we have shown that they are a versatile tool for specifying coverage criteria for test case generation and test monitoring.

In particular, the parameterization mechanism, as used in this paper, allows a succinct specification of several standard generic coverage criteria. In this way, test case generation can be transformed into a reachability problem, which can be general used in verification of environmental control systems.

References

[1] Hu C, Zhu L 2010 The analysis and the evaluation of complicated network software *LNCS* **13**(10) 1-5

[2] Yunfeng Wang, Hongde Xia, Raomei Yan 2008 The analysis of the social network and the study of the application cases of NetDraw *Modern education technology* **18**(4) 85-89

[3] Pothen A, Simon H, Liou K P 1990 Petitioning sparse matrices with eigenvectors of graphs S*IAM Journal on Matrix Analysis and Applications* **11** 430-6

[4] Grivan M, Newman M E J 2001 Community structure in social and biological networks *Proc Natl Acad Sci* **99**(12) 7821-6

[5] Newman M E J, Grivan M 2004 Finding and evaluating community structure in networks *Physical Review E* **39**(10) 69-84

[6] Toyoda M, Kitsuregawa M 2003 Extracting evolution of web communities from a series of web archives *Proceedings of the fourteenth ACM conference on Hypenext and hypermedi*a **101** 78-87

[7] Palla G, Derényi I, Vicsek T 2007 The Critical Point of *k-groups* Percolation in the Erdös-Rényi Graph *Journal of Statistical Physics* **128**(1) 219-27

[8] Palla G, Vicsek T, Barabási A-L 2007 Community dynamics in social networks *Noise and Stochastics in Complex Systems and Finance* **6601**(3) 273-87

[9] Xu C, Zhang Y, Yang D 2011 Ontology based Image Semantics Recognition using Description Logics *IJACT: International Journal of Advancements in Computing Technology* **3**(10) 1-8

[10] Ju C, Wei J 2012 Research on Multi-interest Profile Based on Resource Clustering *JCIT: Journal of Convergence Information Technology* **7**(21) 582-90

[11] Gargantini A, Heitmeyer C 1999 Using Model Checking to Generate Tests From Requirements Specifications *In Software Engineering – ESEC/FSE'99: 7th European Software Engineering Conference held jointly with the 7th ACM SIGSOFT Symposium on the Foundations of Software Engineering* **1687**(1) 146-62

[12] Kim J-M, Porter A 2009 A history-based test prioritization technique for regression testing in resource constrained environments *In ICSE '09: Proceedings of the 31th International Conference on Software Engineering* **139**(3) 119-29

[13] Rothermel G, Untch R H, Chu C, Harrold M J 2009 Test case prioritization: An empirical study *Proceedings of the IEEE International Conference on Software Maintenance* **168**(11) 179-83

[14] Srikanth H, Williams L 2005 On the economics of requirements-based test case prioritization *Proceedings of the 7th international workshop on Economics-driven software engineering research* **153**(1) 1-3

Authors

Jinjiang Liu, born in October, 1974, Nanyang County, Henan Province, P.R. China

Current position, grades: associate professor in School of Computer and Information Technology of Nanyang Normal University, China.
University studies: MSc in Computer Applications at Wuhan University of Science & Technology in China.
Scientific interests: computer modeling, data mining.
Publications: more than 6 papers.
Experience: teaching experience of 15 years, 6 scientific research projects.

Jingjing Liang, born in October, 1981, Nanyang County, Henan Province, P.R. China

Current position, grades: instructor at the School of Computer & Information Technology of Nanyang Normal University, China.
University studies: BSc in University of Electronic Science and Technology of China.
Scientific interests: software engineering, formal modelling.
Publications: more than 5 papers.
Experience: teaching experience of 10 years.

Damage characteristic of rock sample with circular defect based on the distinct element and moment tensor methods

Baowen Hu[1, 2], Changhong Li[2*]

[1]*School of Civil and Environmental Engineering, University of Science and Technology Beijing, Beijing, China*

[2]*State Key Laboratory of High-Efficient Mining and Safety of Metal Mines, Ministry of Education, Beijing, China*

Received 2 August 2014, www.cmnt.lv

Abstract

Based on the distinct element and moment tensor theory, the damage process and mechanism of rock sample with circular hole were researched. The crack evolution patterns corresponding to the laboratory test showed that there were mainly four stages. Firstly, the microcracks were randomly distributed in the rock sample. Secondly, the larger primary cracks are formed around top of circular hole with the increasing of microcracks. Then, more microcracks were localized near the boundary of circular hole, which formed the competition of several possible sets of fractures. Finally, the rupture zone was formed along one of crack zones. The size effect of circular hole showed that there would take on different rupture forms with increasing of diameter. Meanwhile, for better probing into damage mechanism, the acoustic emission (AE) algorithm based on moment tensor theory was implanted into the whole loading process. The AE magnitudes of all parts of rock sample were shown in AE contour maps, and these contour maps showed that the formations of rupture zone were contributed by different stress or energy levels.

Keywords: circular hole, damage mechanism, acoustic emission, distinct element, moment tensor

1 Introduction

There has been a substantial effort by engineers to probe into the damage and fracture propagation around underground excavations. It is generally thought that, for a circular opening in a brittle rock subjected to different stress fields, there mainly exist three types of fracture: spalling fracture caused by compressive stress field, primary fracture caused by tensile stress field, and remote fracture caused by mixture effect of compressive and tensile stress field [1-3]. For further understanding the damage process, various research methods were employed and developed. One approach often used is to build theoretical models and find analytical solutions. Sharan [4] finished the analytical solution for stress and displacement around a circular opening in a generalized Hoek-Brown rock. Zhang [5] obtained the stress and deformation solution of circular opening in strain-softening rock mass. In addition, the effect of seepage force was considered in the theoretical models [6, 7]. Although the analytical methods can facilitate us to understand the mechanical mechanism of deformation and failure characteristics from perspective of continuum mechanism, after all, some of results are lack of evidences from field and laboratory experiments [8], and cannot easily probe into the entire fracture process from meso-scale. Therefore, the physical test and numerical simulation methods are adopted by many researchers in recent years. However, some of simulation works focused on the loading type failure of Berea sandstone and didn't consider the size effect of circular hole [9, 10]. As we know, the lithology and size effect of circular hole would influence on the rupture characteristic. With respect to the simulation method, the distinct element method would be more practical in modelling the entire fracture process, which involves the initiation, propagation, and coalescence of micro-cracks through to the formation of a full-scale macro-crack in the rock materials [11], compared other methods like fast lagrangian analysis of continua or finite element method [10, 12]. Because mesh discretization is not necessarily considered in this way, namely the mesh sensitivity problem is not present.

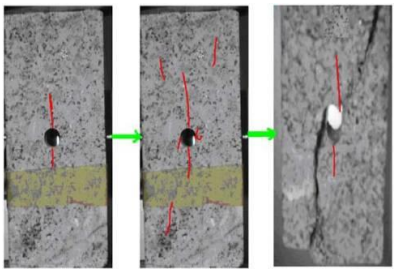
For these reasons, take the previous laboratory experiment done by Lv [13] as reference, the aim of our work is to further reveal progressive failure leading to collapse around a circular opening in granite specimen by virtue of simulation of particle flow code (PFC2D). The opening size effect is considered. Meanwhile, for better understanding the damage process and mechanism, the acoustic emission (AE) algorithm based on the moment tensor theory is used in the simulation.

2 Laboratory experimental model

The rock samples used in the test were granite materials which were obtained from Dagangshan Hydropower Station in Sichuan Province of China. The uniaxial compression strength is 195MPa, the Young's modulus is 47.6GPa, the Poisson ratio is 0.23, and the tensile strength is 6.2MPa. To simulate the underground opening,

[*]***Corresponding author*** e-mail: lch@ustb.edu.cn

rectangular rock plates containing a 10mm diameter hole in centre were tested in servo-controlled hydraulic test machine (type: RMT-150C). The size of the rock plate was designed to 60 mm × 20 mm × 120 mm (length, width, height). Then the rectangular rock plates with a circular opening were loaded to failure under uniaxial compression test. Partial experimental results finished by Lv [13] are listed above. Figures 1a, 1b, 1c shows the failure process of rock plate, which represents the initiation of primary cracks, the expanding of remote cracks and compression-shear cracks, and the generation of rupture zone due to the coalescence of cracks respectively. Figure 2 shows the variation of compressive stress and AE events with axial strain.



a b c

FIGURE 1 Failure process of rock samples

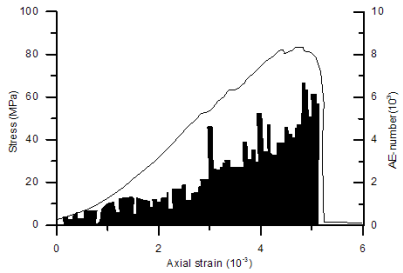


FIGURE 2 Compressive stress and AE event number versus axial strain

3 Establishment of numerical model and loading test

PFC2D is a distinct element computer program designed to simulate the mechanical behaviour of bonded or unbounded granular materials. Therefore the rock sample can be treated as assembly of circular particles that can interact through normal and shear springs. Although PFC2D can simulate a particulate media, any circular element in this program does not necessarily model a particle in the real material as two-dimensional nature of the program limits particles to disks or cylinders [9]. So we used this software as an attempt to mimic the basic mechanical features of actual material.

Before establish the numerical rock sample, numerical calibration test were necessary [14]. The micro-mechanical parameters were needed to be adjusted repeatedly, and finalized until the macroscopic properties such as Young's modulus and the uniaxial compressive strength calculated are basically consistent with physical macroscopic parameters.

The micro-parameters required to be adjusted are as follows: ρ is ball density, $R_{\min}$ is minimum ball radius, R_ratio is ball size ratio, $\overline{\lambda}$ is parallel-bond radius multiplier, E_c is ball-ball contact modulus, $\overline{E_c}$ is parallel-bond modulus, k_n / k_s is ball stiffness ratio, $\overline{k}_n / \overline{k}_s$ is parallel bond stiffness ratio, μ is ball friction coefficient, $\overline{\sigma}_c$ is parallel bond normal strength, and $\overline{\tau}_c$ is parallel-bond shear strength. The values of micro-parameters are listed in Table 1.

TABLE 1 Micro-mechanical parameters of numerical samples

$\rho/(kg.m-3)$	$R_{\min}$ / (mm)	R_ratio	μ	
2500	0.3	2.0	0.5	
$\overline{\sigma}_{mean}$ / (MPa)	$\overline{\tau}_{mean}$ / (MPa)	$\overline{\sigma}_{sdev}$ / (MPa)	$\overline{\tau}_{sdev}$ / (MPa)	
100	100	36	36	
k_n / k_s	$\overline{k}_n / \overline{k}_s$	E_c / (GPa)	$\overline{E_c}$ / (GPa)	$\overline{\lambda}$
2.5	2.5	37.6	37.6	1.0

There are mainly four steps for the creation of corresponding rock samples according to the micromechanics parameters that have been determined by calibration tests. Firstly, a rectangular specimen consisting of arbitrarily placed particles confined by four frictionless walls is generated by radius expansion method. Secondly, radii of all particles are changed uniformly to achieve a specified isotropic stress so as to reduce the magnitude of locked-in stresses that will develop after subsequent bond-installation and specimen-unloading steps. In this paper the isotropic stress is set to 0.1MPa. Thirdly, the floating particles that have less than three contacts are eliminated. Fourthly, the parallel bonds are installed throughout the assembly between all particles that are in near proximity to finalize the specimen.

As the generation of rock sample was completed, then circular holes with diameter from 10mm to 50mm were excavated respectively since the size effect of circular hole on failure characteristic of rock sample would be researched. In addition, the AE algorithm was implanted into the whole failure process for recording the AE event and magnitude.

4 Acoustic emission algorithm based on moment tensor theory

There are essentially two ways that have been adopted to extract seismic or AE information from bond breakages in PFC [15-17]. Initially, the magnitude of AE events from the kinetic energy of the particles in each crack was calculated, and moment tensor was also calculated from the force at the contact as breakage of bond occurs. However, this method often yields magnitudes that are too large. Besides, it fails to account for shear events where the entire force at the contact is not lost. For solving the problem, the moment tensor of AE events was recalculated by observing the force changes at contacts around the source particles, and moment magnitudes were then

calculated form the moment tensor matrix. Therefore the second computing method was used in this paper to further probe into the damage mechanisms of rock samples with circular hole.

The Equation for calculating a moment tensor for an AE event is as following:

$$M_{ij} = \sum_{S} \Delta F_i R_j , \tag{1}$$

where ΔF_i is *i*-th component of the change in contact force, and R_j is *j*-th component of the distance between the contact point and event centroid. The centroid of the event is assumed to be the geometric centre of multiple-crack events since each AE event could be composed of more than one bond breakage. The sum is performed over source boundary which is a surface that completely encloses the displacement discontinuity caused by the 'fault'. As we know the moment tensor is a function of time because it will be calculated at each time step over the duration of the event. The duration of AE event is determined by assuming a fracture propagates at half the shear wave velocity of the rock. Therefore, for reducing memory consumption, the final moment tensor that is independent of time is determined to be the moment tensor at the time of maximum scalar moment. The formulation of the scalar moment is as following:

$$M_0 = \left(\frac{\sum_{j=1}^{3} m_j^2}{2} \right)^{1/2} , \tag{2}$$

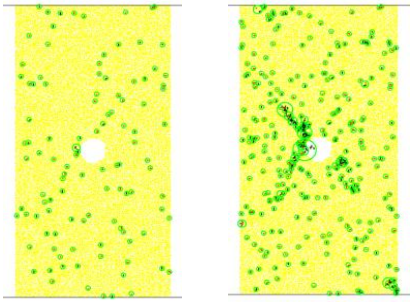
where m_j are the eigenvalues of the moment tensor matrix. By virtue of the scalar moment, the moment magnitude of the event can be calculated by this following Equation:

$$M_w = \frac{2}{3} \log M_0 - 6 \tag{3}$$

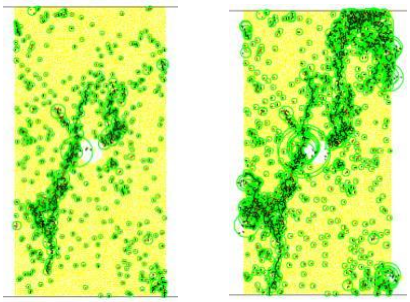
5 Numerical experiment results and discussion

5.1 CRACK EVOLUTION AND AE CHARACTERISTIC

Figure 3 shows the numerically simulated failure process around a circular opening with diameter 10 mm. The black and red spots represent the bond breakage as the result of its normal and shear strength being exceed respectively. Each green circle represents an AE event, and its radius represents the range of AE event, i.e. one AE event may contain many bond breakages. At the beginning loading stage (Figure 3a), there are some independent microcracks (bond breakages) distributed randomly in the rock sample, which lead to the production of AE events. In this period, most of AE events consist of a series of single microcrack event, only few are made up of more than one microcrack events. In the second stage (Figure 3b, axial strain is about 1e-3), it can be observed that more microcracks are generated in the rock sample, and the larger primary cracks are formed around top of circular hole. As the axial strain is getting to around 1.5e-3 (Figure 4), more microcracks localize near the lateral boundaries of the circular opening and begin to form rupture zones on both sides of specimen (Figure 3c). Meanwhile, the remote cracks are produced and expanded. The number of AE events starts to increase, and the radius of some green circles have become larger, which means the larger scale cracks form, and magnitudes of partial AE events also become increasingly. Finally, as the axial strain gets to around 2e-3, the rock sample shows a macroscopic discontinuity along the rupture zone (Figure 3d), and the number of AE events gets to the maximum nearby the peak of loading stress (Figure 4).



(a) Random damage (b) Local damage



(c) Damage expanding (d) Rupture zone

FIGURE 3 Simulated failure process of rock sample

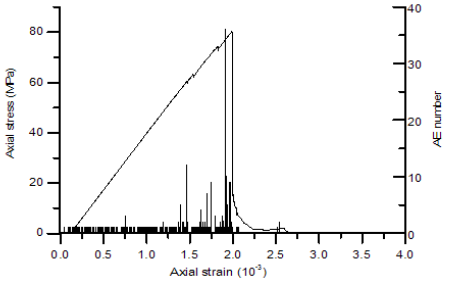


FIGURE 4 Simulated compressive stress and AE event number versus axial strain

The Figure 5 demonstrates the contour map of moment magnitude of AE event, which is corresponding to the Figure 3d. As we can see, the magnitudes of main rupture zone represented by red colour are between -4.9 and -4.7 which means such kind of magnitude can fully satisfy the

generation of rupture zone developed by the coalescence of the microcracks. Outsides of main rupture zone, the magnitudes are mainly between -5.5 and -5, which means these smaller scale damages fail to connect each other to produce the larger scale cracks. In short, such kind of simulation based on the distinct element and moment tensor theory can better reveal and demonstrate the damage process and mechanism of rock sample. The results of magnitude calculated by scalar moment match very well with simulated damage distribution and actual lab test.

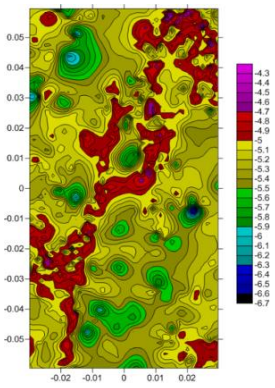


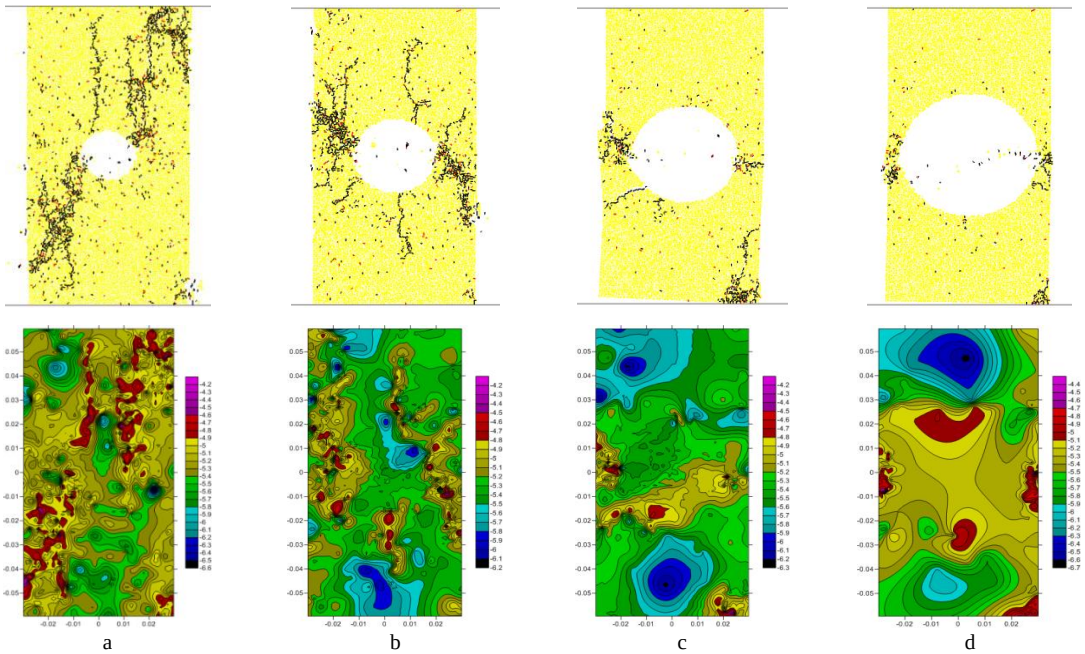
FIGURE 5 Contour map of magnitude of AE event

Although the simulated curve of compressive stress versus the axial strain shows a slightly stiffer response, the whole variation trend and peak value are basically in agreement with the actual laboratory test. As for number of AE events versus axial strain, the variation trend also has similarity with actual lab test despite of concrete number of AE event, because we only care about basic mechanical features and laws.

5.2 THE SIZE EFFECT OF CIRCULAR HOLE

The influences of different sizes of circular hole on the final rupture zones of the rock sample are considered in this section. As the diameters are between 10 mm and 30 mm (Figure 3d, Figures 6a-6b), there are longitudinal crack zones generated nearby the top or bottom of the circular hole, especially for the diameter 30 mm, because the increment of size of circular hole reduces the tensile strength on the periphery of hole, and the values of tensile stress on the top and bottom of hole are also larger than other direction. However, more microcracks localize near the lateral boundaries of the circular hole and form notches, and finally the failure plane extends from the notches in the inclined direction; As the diameters are greater than 40mm, the rock sample shows the symmetrical failure along the rupture zone of which the direction is sub-horizontal.

It is interesting to note that the maps of contour of magnitude of AE event, there still take on red colour in the core area of rupture zones. However, periphery of red colour is mainly encompassed by the yellow colour, part of which is still in the rupture zone, and of which the magnitudes are lower than red colour. Such kind of phenomenon is more obvious as the diameter of circular hole is greater than 20 mm. There are mainly two reasons we think lead to this kind of phenomenon.



a b c d

FIGURE 6 Crack patterns of rock sample and contour maps of magnitude of AE event and under different diameters of circular hole including 20 mm, 30 mm, 40 mm, 50 mm

It is interesting to note that the maps of contour of magnitude of AE event, there still take on red colour in the core area of rupture zones. However, periphery of red colour is mainly encompassed by the yellow colour, part of which is still in the rupture zone, and of which the magnitudes are lower than red colour. Such kind of phenomenon is more obvious as the diameter of circular hole is greater than 20 mm. There are mainly two reasons we think lead to this kind of phenomenon. On the one hand, the different stress or energy levels can come together to promote propagation of the rupture zone. Some core ruptures are mainly finished by the higher magnitude of stress or energy then the lower stress or energy finishes extension and coalescence of the rupture zone on the basis of core rupture zone. On the other hand, the increasing of size of circular hole equivalently increases the damage degree of rock sample and decreases the compressive strength as we can see from the Figure 7, which make the relatively lower stress or energy easier to generate the cracks.

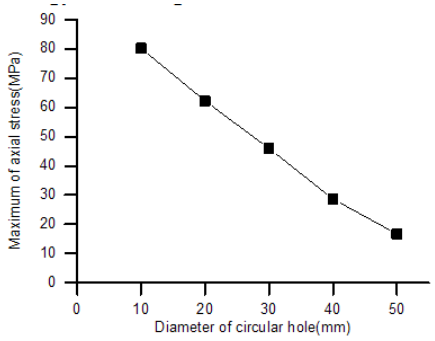


FIGURE 7 the maximum of axial stress versus the diameter of circular hole

6 Conclusions

In this study, the failure process of granite sample with circular hole was simulated by the distinct element method. The numerically simulated results reproduced the development of cracks around a circular opening in rock, which were in very agreement with the laboratory test and also can reveal microscopic damage characteristic. The microcrack pattern showed that competition of several possible sets of fractures, eventually the cracks on both sides of circular hole prevailed and continued to propagate until the final rupture zone was formed. The simulated number of AE events would get peak as failure occurred, which was consistent with actual test.

The AE algorithm based on the moment tensor theory was used in the simulation, which can better reveal the damage mechanism for the loading test. Especially, the AE magnitude contour map can tell us intuitively the damage degree in different parts of rock sample.

The size effect of circular hole was considered in this paper. It showed that the crack development and distribution were influenced by the size of diameter, because the size of circular hole determines the prefabricated damage degree of the rock sample. In addition the contour map of AE magnitude showed that the formation of rupture zone were contributed by different stress or energy levels, and it would be more obvious with increasing of size of circular hole.

Acknowledgments

The authors are grateful for the financial support from the National Basic Research Program of China (973 Program) (2010CB731500). The authors also thank the professor Hazzard J.F. for providing the help about AE algorithm

References

[1] Martin C D 1997 The effect of cohesion loss and stress path on brittle rock strength *Canadian Geotechnical Journal* **34**(5) 159–68
[2] Napier J A L, Hildyard M W 1992 Simulation of fracture growth around openings in highly stressed brittle rock *Journal of the South African Institute of Mining and Metallurgy* **92**(6) 149-58
[3] Hoke E, Brown E T 1980 Underground excavation in rock *Metallurgical Industry London*
[4] Sharan S K 2008 Analytical solutions for stresses and displacements around a circular opening in a generalized Hoek-Brown rock *International Journal of Rock Mechanics and Mining Sciences* **45**(1) 78-85
[5] Zhang Q, Jiang B S, Wang S l, Ge X R 2012 Elasto-plastic analysis of a circular opening in strain-softening rock mass *International Journal of Rock Mechanics and Mining Sciences* **50**(2) 38-46
[6] Zareifard M R, Fahimifar 2014 A Effect of seepage forces on circular openings excavated in Hoek-Brown rock mass based on a generalised effective stress principle *European Journal of Environmental and Civil Engineering* **18**(5) 584-600
[7] Yang X L, Huang F 2010 Influences of strain softening and seepage on elastic and plastic solutions of circular openings in nonlinear rock masses *Journal of Central South University of Technology* **17**(3) 621-7
[8] Wang S, Hagan P, Cheng Y 2013 Experimental research on the instability characteristics of double-layer rock plates based on MTS-AE system *Applied Mathematics & Information Sciences* **7**(1L) 339-45
[9] Fakhimi A, Carvalho F, Ishida T 2002 Simulation of failure around a circular opening in rock *International Journal of Rock Mechanics and Mining Sciences* **39**(4) 507-15
[10] Wang S Y, Sloan S W, Sheng D C 2012 Numerical analysis of the failure process around a circular opening in rock *Computers and Geotechnics* **39**(1) 8-16
[11] Potyondy D, Cundall P, 2001 The PFC model for rock: predicting rock-mass damage at the underground research laboratory *Itasca Consulting Group*, Inc. Report no. 06819-REP-01200-10061-R00
[12] Li D, Li X, Li C C 2010 Experimental and numerical studies of mechanical response of plate-shape granite samples containing prefabricated holes under uniaxial compression *Chinese Journal of Rock Mechanics and Engineering* **30**(6) 1198-206 (*in Chinese*)
[13] Lv S, Chen W, Jia S, Tan X J 2009 Experimental study on brittle rock failure *Chinese Journal of Rock Mechanics and Engineering* **28** (SUPPL1) 2772-7
[14] Wang S R, Hagan P, Hu B W, Gamage K, Cheng Y, Xu D 2014 Rock-arch instability characteristics of the sandstone plate under different loading conditions *Advances in Materials Science and Engineering* Article ID 950870 1-9
[15] Hazzard J F, Young R P 2000 Simulating acoustic emissions in bonded-particle models of rock *International Journal of Rock Mechanics and Mining Sciences* **37**(5) 867–72
[16] Hazzard J F, Young R P 2002 Moment tensors and micromechanical models *Tectonophysics* **356**(1) 181–97 (*in Chinese*)
[17] Hazzard J F, Young R P 2004 Dynamic modelling of induced seismicity *International Journal of Rock Mechanics and Mining Sciences* **41**(8) 1365-76

Authors

Baowen Hu, born in May, 1984, Beijing, China

Current position, grades: the doctoral student of School of Civil and Environmental Engineering, University of Science and Technology Beijing, China.
University studies: M.Sc in Ecology from Chinese Academy of Sciences.
Scientific interest: mathematical model and numerical simulation of rock mechanics.
Publications: 10 papers.
Experience: 3 scientific research projects.

Changhong Li, born in October, 1962, Beijing, China

Current position, grades: the professor of School of Civil and Environmental Engineering, University of Science and Technology Beijing, China.
University studies: PhD in in Mining Engineering from University of Science and Technology Beijing.
Scientific interest: rock mechanics
Publications: 40 papers.

Network intrusion clustering based on fuzzy C-Means and modified Kohonen neural network

Hongwei Ye[1*], Lianjiao Zhang[1], Xiaozhang Liu[2]

[1]*School of Electron and Information Engineering, Heyuan Polytechnic, Dong Huan Road, Heyuan, China*

[2]*School of Computer Science, Dongguan University of Technology, No.1, Daxue Rd Songshan Lake, Dongguan, China*

Received 6 July 2014, www.cmnt.lv

Abstract

Kohonen neural network recognizes and clarifies substantive network data, but with a long running time and a slow convergence process. To solve this problem, a network intrusion clustering method is presented in this paper. Specifically, the training data is pretreated using Fuzzy C-Means (FCM). Then some selected data will be trained with using Kohonen neural network. Meanwhile, to speed up the convergence process of Kohonen neural network and to form a better optimized network topology, a neighbourhood function is established for the competing neuron. Each neuron has neighbourhood topology collections. The data simulation results demonstrate the efficiency and effectiveness of the proposed algorithm.

Keywords: Kohonen neural network; neuron, FCM, network intrusion clustering

1 Introduction

Network intrusion is an unauthorized access to the completeness, privacy and availability of a computer or network. Intrusion detection is an intrusion detection system that attempts to discover unauthorized malicious activity by analysing information collected by the computer or network. Kohonen neural network (KNN) is applied to the detection of network intrusion. The improvement of KNN can be achieved by a set suitable threshold to the competition layer in the neural network to avoid the occurrence of necrotic neuron [1]. Also fuzzy bias degree is introduced in some documents, which aims to avoid dead neurons [2]. To speed up the convergence process of KNN, energy function is introduced [3]. An initial value is set by a simple splitting algorithm to improve KNN. The more data it has, the longer time KNN training will take. To speed up the process, each kind of training is FCM clustered. Data is chosen according to the degree of membership for KNN training. Meanwhile, KNN will be improved in two aspects. One is neighbourhood function established among neurons, and the other is topology identification of neural neighbours, in which the training data, after being correctly clarified, reach suitable topology and then the iteration breaks.

2 FCM

FCM is a cluster algorithm based on objection function. It is developed from Hard C- Means (HCM). It is a algorithm that determines the likelihood of a data by using the membership degree. If X is a vector of a limited data set in eigenspace R^N. According this algorithm, X is divided into several fuzzy cluster, each centre vector forms a set $V=\{v_1, v_2, \cdots v_c\}$, a matrix of $N\times C$ dimension $U=(u_{ij})$, $u_{ij}\in[0,1]$, a membership matrix of each sample, and here $i=1,2,...,N;\ j=1,2,...,C$ makes the fuzzy objective function minimum. Fuzzy clustering objective function is presented in Equation (1) and Equation (2).

$$J=\sum_{i=1}^{N}\sum_{j=1}^{C}u_{ij}\left\|x_i-v_j\right\|^2, \tag{1}$$

$$u_{ij}=\begin{cases}\left[\sum_{k=1}^{c}\dfrac{\left\|x_i-v_j\right\|^{\frac{2}{m-1}}}{\left\|x_i-v_k\right\|^{\frac{2}{m-1}}}\right]^{-1} & \left\|x_i-v_k\right\|\neq 0\\ 1 & \left\|x_i-v_k\right\|=0\&\&k=j\\ 0 & \left\|x_i-v_k\right\|=0\&\&k\neq j\end{cases}. \tag{2}$$

$$v_j=\frac{\sum_{i=1}^{n}u_{ij}^m x_i}{\sum_{i=1}^{n}u_{ij}^m}$$

Here, u_{ij} is the degree of membership for the x_i in cluster j, m is the fuzzy weight index, v_j is the centre vector for cluster j.

FCM algorithm schema:

Step 1: fix discriminant function C, and fuzzy weight index m.

Step 2: initial centre vector for cluster v.

[*]***Corresponding author*** e-mail: aboyhw@163.com

Step 3: compute the membership matrix *u*.

Step 4: compute each cluster centre *v*.

Step 5: compute the fuzzy cluster objective value to determine if it meets the given termination conditions. If it meets, the iteration ends. If not, jump to step 3.

3 Modified Kohonen neural network

The model was first described as an artificial neural network by the Finnish professor Teuvo Kohonen in 1981, and is sometimes called a Kohonen map or network. It is sometimes called SOFM [5]. It automatically gets the centre of each cluster by competitive learning training coefficient, so it is widely used in fields like pattern identification and patter control. Kohonen believes that neurons in different part do their own jobs [6]. Different parts of the network will respond accordingly to certain input patterns. Kohonen network can learn both the distribution characters and the topology structure of input vector of training data [7, 8]. Typical structure of KNN is shown in Figure 1.

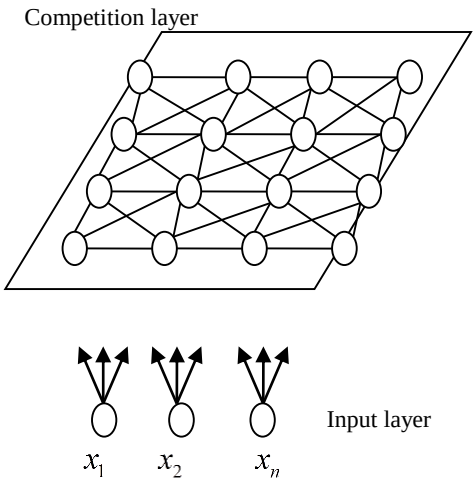


FIGURE 1 The structure of KNN

The training algorithm for the Kohonen network as follows:

(1) Initialise the weight vectors, $w_j(0)$,where j is an index denoting the neuron number and runs form 1,2,…,*K*, where *K* is the total number of neurons int the output grid. The number in () afterwards is used to denote the time-step. The weight vectors can be initialised by setting each component of each weight vector to a small random number.

(2) Draw a connection vector, $X=(x_1,x_2,\cdots x_m)^T$, from the training data without replacement.

(3) Find the winning neuron, j^* ,This is the neuron whose weight vector is closed to *X* in a Euclidean sense, calculated as follows:

$$d_j=\|X-W_j\|=\sqrt{\sum_{i=1}^{m}(x_i(t)-w_{ij}(t))^2}\ .$$

(4) Adjust the weight vectors of all neurons, as follows:

$$\triangle w_{ij}=w_{ij}(t+1)-w_{ij}(t)=\eta(t)(x_i(t)-w_{ij}(t))$$

where $\eta(t)$ is the learning rate at epoch number *t*, usually $\eta(t)=\frac{1}{t}$ or $\eta(t)=0.2\left(1-\frac{t}{T}\right)$, *T* is the total number of training.

(5) Repeat from step 2 until all training examples have been presented. This constitutes on epoch.

Repeat form step 2 until the desired number of epochs is reached.

According to the biology principal of neurons, neurons will be in a sorted order after self-organized training. Neurons of the same class will be next to each other, while those different will be in a distant [9-12], as Figure 2 shows. For example, a competition layer with 4 rows and 4 columns has 16 neurons. Among them, number 1 is the winning neuron, and then number 2 and number 5 neurons is more possible to win. But number 16 neuron is not likely to win. So, the first a probability matrix needs to be established for each neuron before network training.

$$B=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix},$$

where, b_{ij} is the transit probability between the *i*-th neuron and the *j*-th neuron. $i=1,\ldots,n$, $j=1,\ldots,n$. Here *n* is the total amount of the neurons in the competition layer. Then the neighbourhood function $G(t)$, follows a Gaussian probability distribution as defined by Equation (3), where $j(t)$ denotes the neuron in the neighbourhood of the winning neuron, and then adjust the weight vectors of all neurons as given in Equation (4).

$$G(t)=\exp\left(-\frac{\left(j(t)-j^*\right)^2}{2}\right), \tag{3}$$

$$Vw_{ij}=w_{ij}(t+1)-w_{ij}(t)=g(t)\eta(t)(x_i(t)-w_{ij}(t))\ . \tag{4}$$

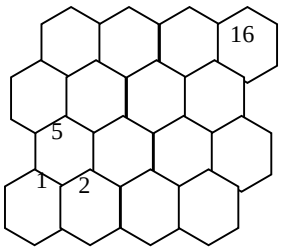


FIGURE 2 Topology of neurons in competition layer with 4 ×4

Identification of neighbour topology. Build a neighbour set U for each neuron. For example, the

neighbour set of number 1 neuron is {2,5}, and number 5 is {1,2,6,9,10}. In the process of KNN training, we set minimum learning times. If the training times exceed the set one, KNN begins to self-classification and identification with the aim to see if the neurons are correctly classified. For example, in a competition Kohonen neural network with 6 rows and 6 columns, there are 5 training kinds. The winning neuron in each kind forms a set, namely {6,12}, {7,13}, {16,21,22}, {25,31}, {30,36}. Take the intersection of each two sets. If all of them are disjoint, then it shows that KNN can classify correctly. At the same time, conduct neighbourhood topology identification.

4 Simulation results

4.1 GET DATA

Intrusion Detection has become a pertinent part of network security. An Intrusion Detection System can monitor the events happening in a system, and identify whether they are attacks or legitimate accessed. There are nine 'basic features' in Intrusion Detection System, such as duration of the connection; total bytes sent to destination host; total bytes sent to source host; service type ,such as FTP, HTTP, Telnet; number of wrong fragments. In our experiment, we get five attack categories as shown in Table1.

TABLE 1 Category attack data use by our experiment

Category	Description
1	HTTP
2	FTP
3	Mail
4	SQL
5	Remote shell

4.2 FILTER DATA

The attack dataset is $N_{4500\times39}$, we use the 4000 data as the training data, the lest 500 as the test data. Each given 5 classified 4000×38 training data is FCM clustered. A membership matrix of each data will be formed. Arrange the matrixes in the order of membership degree. Choose the first 20 data as representatives. The algorithm of FCM clustered is described as Algorithm 1.

Algorithm 1:Filter data using FCM

Input: $N_{4500\times39}$

Output: $\text{TrainData}_{100\times38}$

```
for i=1:category
{
DataIndex=find(T1==i);
data=inputn(DataIndex,:);
[center,U,obj_fcn] = fcm(data,1);
[su,index]=sort(U,'descend');
FilterData{i,:}=inputn(DataIndex(index(1:20)),:);
}
TrainData=[FilterData{1,1};FilterData{2,1};
FilterData{3,1};FilterData{4,1};FilterData{5,1}];
```

4.3 TRAINING KNN

We get a training data of 100×38. The dimension of the data is 38 from 5 different network invasion patterns. So there are 38 nodal points in the input layer. The nodal point in the competition lay represents the potential kinds of the input numbers. Usually the nodal point in the in the competition is far more than the actual data kinds. Here the author assumes the nodal point is 36 and the neuron topology arrangement is as shown in Figure 3.

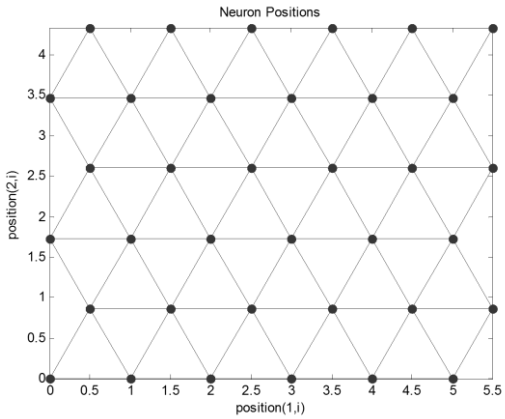


FIGURE 3 Topology arrangements of neurons in the competition layer

The data will be trained by the modified KNN (MKNN). The Process of simulation experiments as shown in Figure 4. The winning neurons's topology arrangement is as shown in Figure 5.

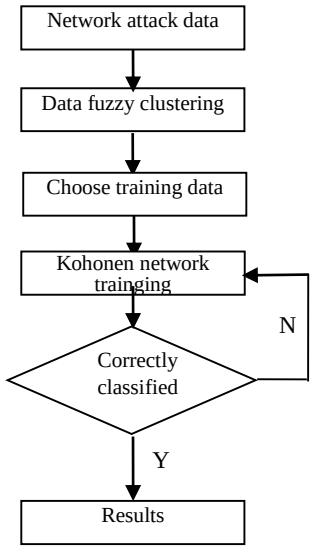


FIGURE 4 Process of simulation experiment

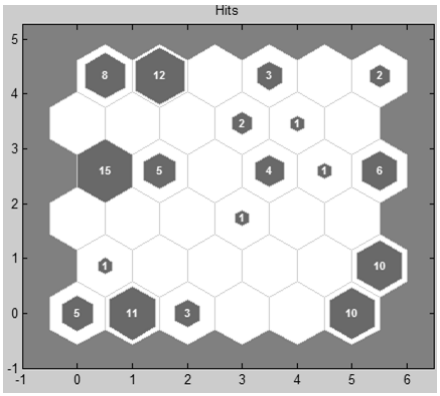


FIGURE 5 The winning neuron topology

The List of winning neurons as shown in Table 2.The computer adopted in the experiment is Intel Core i3-2370M 2.4G with 6G RAM and a runtime environment of Matlab R2012b. It takes the traditional KNN 30 minutes and 42 seconds to deal with the training data, while only less 5 minutes using the method mentioned in this essay and the topology of the winning neurons are better than those of the traditional. The traditional KNN Cluster and the MKNN Cluster is shown in Figure 6 and Figure7.

TABLE 2 List of winning neurons

kind	number
The first kind	6,12
The second kind	31,32
The third kind	19,20
The fourth kind	1,2,3,7
The fifth kind	16,22,23,24,28,29,34,36

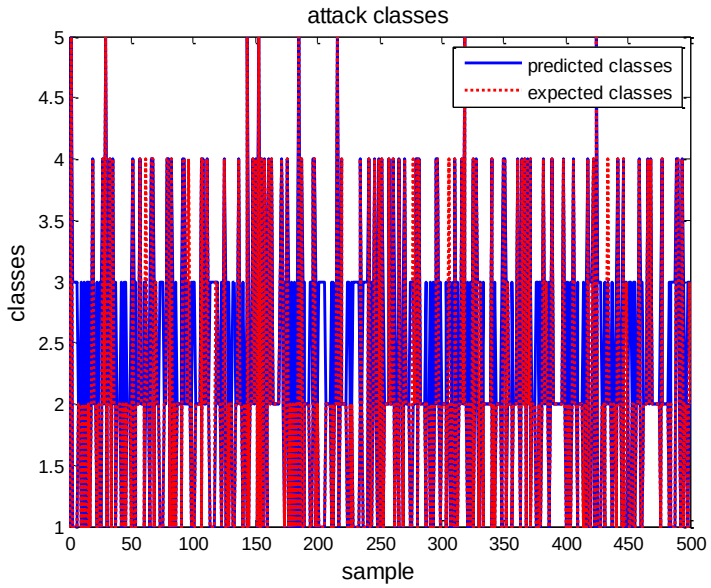


FIGURE 6 The traditional KNN Cluster

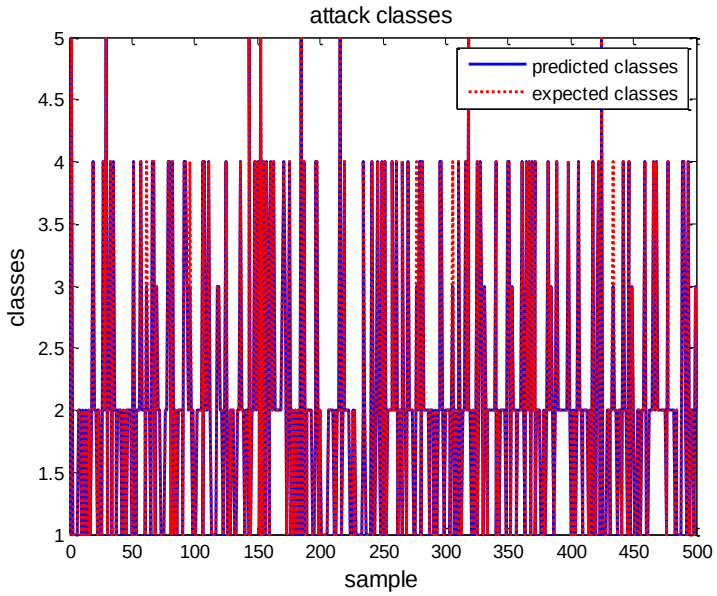


FIGURE 7 The MKNN Cluster

9 Conclusions

Based on the high clarification efficiency of KNN, this essay presents a method of FCM clustering in case of large amount of network invasion data. Typical data is chosen to be trained, which greatly cut down the training time of network classification. Also the introduction of neural neighbourhood function and neural neighbourhood topology identification improves the neural topology distribution and classifies the data more accurately. The result of the simultaneous experiment shows the efficiency of the algorithm in network invasion clustering.

References

[1] Fang H 2013 Track correlation algorithm based on modified Kohonen neural network *Journal of Computer Applications* **33** (5) 1476-80

[2] Xu M–j 2009 Improved clustering algorithm based on fuzzy Kohonen clustering network *Computer Simulation* **26** (4) 228-32

[3] Lei H 2005 Identification of porn images based on improved kohonen neural network and BP Network *Computer Engineering* **31**(10) 164-7

[4] Liu Y 2011 The research of fuzzy clustering algorithm of data mining based on shuffled frog leaping algorithm *LanZhou:LanZhou University of Technology*

[5] Yang C 2010 The research of kohonen neural network in telecom frauds forecast *ShangHai:EastChina Normal University*

[6] Duan L-z, Zhu M, Wang L-m 2009 A web user access pattern mining algorithm based on the dual kohonen neural Network *Computer Engineering & Science* **31**(9) 95-8

[7] De Almeida C W D, De Souza R M C R 2013 Fuzzy Kohonen clustering networks for interval data *Neurocomputing* 65-75

[8] Bianchi D, Calogero R, Tirozzi B 2007 Kohonen neural networks and genetic classification *Mathematical and Computer Modelling* **45**(1) 34-60

[9] Obimbo C, Zhou H, Wilson R 2011 Multiple SOFMs working cooperatively in a vote-based ranking system for network intrusion detection *Procedia Computer Science* **6** 219-24

[10] Kayacik H G, Zincir-Heywood A N 2007 A hierarchical SOM-based intrusion detection system *Engineering Applications of Artificial Intelligence* **20** 439-51

[11] Powers ST, He J 2008 A hybrid artificial immune system and Self Organising Map for network intrusion detection *Information Sciences* **178** 3024-42

[12] Ballabio D, Consonni V, Todeschini R 2008 The Kohonen and CP-ANN toolbox: A collection of MATLAB modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks *Chemometrics and Intelligent Laboratory Systems* **98** 115-22

Authors

Hongwei Ye, born in March, 1979, Heyuan City, Guangdong Province, P.R. China

Current position, grades: the lecture of School of Electronic and Information, Heyuan Polytechnic, China.
University studies: MSc in Computer Software And Theory from Sun Yat-sen University in China.
Scientific interest: computer software, and neural network.
Publications: 11 papers.
Experience: teaching experience of 11 years, 2 scientific research projects.

Lianjiao Zhang, born in October, 1980, Heyuan City, Guangdong Province, P.R. China

Current position, grades: the lecture of School of Electronic and Information, Heyuan Polytechnic, China.
University studies: MSc from Huazhong University of Science and Technology in China.
Scientific interest: information engineering.
Publications: 6 papers.
Experience: teaching experience of 11 years, 3 scientific research projects.

Xiaozhang Liu, born in November, 1978, Dongguan City, Guangdong Province, P.R. China

Current position, grades: the associate professor of School of Computer Science, Dongguan University of Technology, China.
University studies: PhD in Computational Science from Sun Yat-sen University in China.
Scientific interest: pattern recognition, and wireless sensor networks.
Publications: 20 papers.
Experience: teaching experience of 10 years, 2 scientific research projects.

Numerical analysis of sheet metal U-Bending process

Jiansheng Xia*, Shasha Dou

Yancheng Institute of Technology, Yancheng, 224051, China

Nanjing University of Aeronautics & Astronautics, nanjin, 211106, China

Received 16 July 2014, www.cmnt.lv

Abstract

The sheet metal U-bending forming is a complex process, based on the assumption of Prandtl-Reuss flow rule and Von Mises yield criterion, the incremental elastoplastic large deformation finite element model was established based on the Updated Lagrangian Formulation (ULF). The elastoplastic conversions of boundary and deformation are reduced with rmin rule. The friction phenomenon of slippage and viscosity at the boundary interface is revised with increment of revision Coulomb rule. The increment rules are led into the whole stiffness matrix, and derived out the stiffness equation. The studies show that the influence on U-Bending deformation of sheet metal is influenced by die structure and parameter. The results show there is a good consistency between the finite element simulation and experimental result.

Keywords: elastoplastic, FEM simulation, sheet metal, U-Bending, warpage

1 Introduction

Sheet metal forming is a common material processing method, which can be divided into several kinds, such us: stretching, bending, drawing and flanging, etc. Bending is the important forming that pushes sheet mental into a certain shape with different angle and curvature. Bending sheets is commonly used in the area of sheet metal forming, especially in the automotive metal, industrial products and household appliances. The common types of bending are V-Bending, U-Bending and L-Bending [1].

There are some problems in U-Bending, such us thinning, thickening or warping, and unable to obtain the desired shape [2, 3].

In this paper, based on the axisymmetric large deformation elastic-plastic theory to analyse the U-bending of sheet metal, the relationships between the U-Bending quality and parameters are analysed such us: punch die radius, clearance, friction coefficient and plate thickness etc. Finally, the theoretical analysis is verified with the experimental results.

Kim and Thomson [4] found that spring-back and the sidewall curl were proportional to die radius, die clearance, work hardening rate and yield strength after considering the friction factor in U- bending

Lee and Yang [5] analysed U-draw bending process parameters with the finite element method, such us: punch velocity, penalty function, damping ratio, and evaluated the important factor which influencing the spring-back with the taguchi method, and found the element size and arc radius were the significant factor.

Liu [6] analysed U-draw bending spring-back under the variable and fixed plate force with the NUMISHEET93 method, and found the variable power can obtain the better forming quality.

Huang and Chen [2] developed a set of incremental elastic-plastic large deformation finite element program to simulate the V bending spring-back and saddle phenomenon and verified experiment.

Sousa [3] optimized design parameters of V and U bending with the finite element method and Genetic algorithm (GA), those parameters including: punch and die radius, punch displacement and plate force.

2 Experimental Section

2.1 FUNDAMENTAL THEORY

2.1.1 Virtual work principle

It descripts the elastic-plastic deformation with the updated Lagrangian formulation ULF [10], the Virtual work principle formulation can be shown as follows:

$$\int_{V^E}\left(\ddot{\sigma}_{ij}-e\sigma_{ik}\dot{\varepsilon}_{kj}\right)\delta\dot{\varepsilon}_{ij}dV+\int_{V^E}\sigma_{jk}L_{ik}\delta L_{ij}dV=\int_{S_f}\dot{f}\,\delta v_i dS\,, \quad (1)$$

where, $\ddot{\sigma}_{ij}$ is the Cauchy stress tensor, $\dot{\varepsilon}_{kj}$ is the rate of stress tensor, $\dot{\varepsilon}_{ij}$ is the strain tensor, σ_{jk} is the rate of strain tensor, $\delta\dot{\varepsilon}_{ij}$ is the virtual strain tensor of the point, δL_{ij} is the virtual velocity gradient tensor of the point, δv_i is the velocity component, $\dot{f}$ is surface force component, L_{ij} is velocity gradient tensor, V is unit volume, S is unit surface area.

***Corresponding author** e-mail: xiajiansheng@163.com

2.1.2 Constitutive relation

In preparing the theory of elasto-plasticity, we have made certain assumptions:

1) The material is homogeneous and isotropic;
2) There is no strain before manufacturing;
3) Temperature effect do not consider when manufacturing;
4) It obeys the laws of the Hooke's Law in elastic stage;
5) It obeys the von Mises yield rule and Prandtl-Reuss plastic flow rule;
6) It contains Isotropic strain hardening in constitutive equation;
7) There are elastic strain stage and plastic strain stage in material strain rate;
8) Punch, die and holder are steel structure;
9) The Bauschinger effect does not consider in reverse unloading.

After assuming above, the constitutive relation can be written as follows:

$$\overset{\circ}{\sigma}_{ij} = C^{ep}_{ijmn}\dot{\varepsilon}_{mn}, \tag{2}$$

$$C^{ep}_{ijmn} = C^{e}_{ijmn} - \frac{C^{e}_{ijkl}C^{e}_{uv}\dfrac{\partial f}{\partial \sigma_{kl}}\dfrac{\partial f}{\partial \sigma_{uv}}}{C^{e}_{kluv}\dfrac{\partial f}{\partial \sigma_{kl}}\dfrac{\partial f}{\partial \sigma_{uv}} + H'\dfrac{\sigma_{uv}\dfrac{\partial f}{\partial \sigma_{uv}}}{\bar{\sigma}}}, \tag{3}$$

where $\overset{\circ}{\sigma}_{ij}$ is Jaumann differential of σ_{ij}, C^{ep}_{ijmn} is the elastic-plastic module, C^{e}_{ijmn} is Elastic module, f is the initial yield function, H' is the strain hardening rate, $\bar{\sigma}$ is Von Mises yield function, the Matrix form of C^{ep}_{ijmn} can be expressed as below:

$$\left[C^{ep}\right] = \left[C^{e}\right] - \frac{1}{S}\begin{bmatrix} S_1^2 & S_1S_2 & S_1S_3 & S_1S_4 & S_1S_5 & S_1S_6 \\ & S_2^2 & S_2S_3 & S_2S_4 & S_2S_5 & S_2S_6 \\ & & S_3^2 & S_3S_4 & S_3S_5 & S_3S_6 \\ & & & S_4^2 & S_4S_5 & S_4S_5 \\ & & & & S_5^2 & S_5S_6 \\ symm & & & & & S_6^2 \end{bmatrix}, \tag{4}$$

where,

$$S = \frac{4}{9}\bar{\sigma}^2H' + S_1\sigma'_{xx} + S_2\sigma'_{yy} + S_3\sigma'_{zz} + 2S_4\sigma'_{yz} + 2S_5\sigma'_{zx} + 2S_6\sigma'_{xy}, \tag{5}$$

$$S_1 = 2G\sigma'_{xx},\ S_2 = 2G\sigma'_{yy}\ S_3 = 2G\sigma'_{zz}, \tag{6}$$

$$S_4 = 2G\sigma'_{yz},\ S_5 = 2G\sigma'_{zx},\ S_6 = 2G\sigma'_{xy}, \tag{7}$$

where σ'_{ij} is deviator of σ_{ij}, G is the friction flow potential, $G = \sigma_1^2 + \sigma_2^2$, $\left[C^{e}\right]$ is the equation in minimum strain, which can be expressed as:

$$\left[C^{e}\right] = \frac{E}{1+\nu}\begin{bmatrix} \frac{1-\nu}{1-2\nu} & \frac{1-\nu}{1-2\nu} & \frac{1-\nu}{1-2\nu} & 0 & 0 & 0 \\ & \frac{1-\nu}{1-2\nu} & \frac{1-\nu}{1-2\nu} & 0 & 0 & 0 \\ & & \frac{1-\nu}{1-2\nu} & 0 & 0 & 0 \\ & & & \frac{1}{2} & 0 & 0 \\ & & & & \frac{1}{2} & 0 \\ symm & & & & & \frac{1}{2} \end{bmatrix}, \tag{8}$$

where E is modulus of elasticity, v is Poisson's ratio. If the material is homogeneous and isotropic, the elastoplastic rate equation can be written:

$$\overset{\circ}{\sigma}_{ij} = \frac{E}{1+\nu}\left[\delta_{ik}\delta_{jl} + \frac{\nu}{1-2\nu}\delta_{ij}\delta_{kl} - \frac{3\alpha\left(\dfrac{E}{1+\nu}\right)\sigma'_{ij}\sigma'_{kl}}{2\bar{\sigma}^2\left(\dfrac{2}{3}H' + \dfrac{E}{1+\nu}\right)}\right]\dot{\varepsilon}_{kl}. \tag{9}$$

If α=1, it is a plastic stage; when α =0, it is a elastic stage or unloading stage.

Equivalent stress and equivalent plastic strain relations can express by *n*-power law equation:

$$\dot{\sigma} = C\left(\varepsilon_0 + \dot{\varepsilon}_p\right)^n, \tag{10}$$

where C is material constant, n is strain hardening index; $\dot{\sigma}$ is the equivalent stress, $\dot{\varepsilon}_p$ is the equivalent plastic strain, ε_0 is the initial strain.

2.1.3 The finite element equation

Finite element analysis is the method that the structure is divided into many small units called discrete entity. Based on Large deformation stress and stress rate relation, the finite deformation of update Lagrangian formulation, material constitution relationship, the velocity distribution of each unit is show below:

$$\{v\} = \leq [N]\{\dot{d}\}, \tag{11}$$

$$\{\dot{\varepsilon}\} = \leq [B]\{\dot{d}\}, \tag{12}$$

$$\{L\} = \leq [M]\{\dot{d}\}, \tag{13}$$

where $[N]$ is shape function, $\{\dot{d}\}$ is nodal velocity, $[B]$ is strain rate-velocity matrix, $[M]$ is velocity gradient-velocity matrix

The principle of virtual work equation and the constitutive equation based on update Lagrangian are linear equation. The formula can be written by the form of incremental representation.

After finite element discretization, the large deformation rigid general equation is written as below:

$$[K]\{\Delta u\}=\{\Delta F\}, \tag{14}$$

where:

$$[K]=\sum_{\langle E\rangle}\int_{V^e}[B]^T\left([C^{ep}]-[Q]\right)[B]dV+\sum_{\langle E\rangle}\int_{V^e}[E]^T[Z][E]dV, \tag{15}$$

$$\{\Delta F\}=\sum_{\langle E\rangle}\int_{S^\varepsilon}[N]^T\{\dot{f}\}dS\Delta t, \tag{16}$$

where $[K]$ is the overall elastoplastic stiffness matrix; $\{\Delta F\}$ is the nodal displacement increment; $\{\Delta u\}$ is the nodal forces incremental; [Q] and [Z] are stress correction matrix.

2.1.4 Friction processing

There is friction in sheet forming process, so we need to pay attention to materials and tools of the interface conditions. When the material moves along the tool surface curve of the slide, the contact force can be expressed as:

$$F=F_l l+F_n n, \tag{17}$$

where F_l is radial force and F_n is normal force, and differential equation of F can be expressed as:

$$\dot{F}=\dot{F}_l l+F_l\dot{l}+\dot{F}_n n+F_n\dot{n}, \tag{18}$$

where differentials of $\dot{l}$ and $\dot{n}$ are expressed as:

$$\dot{l}=-\Delta u_l^{rel}/R, \tag{19}$$

$$\dot{n}=\Delta u_l^{rel}/R, \tag{20}$$

where R is tool radius, Δu_l^{rel} is the local relative velocity between the tool and node, and the nodes relative speed can be expressed as:

$$\Delta u_l^{rel}=\Delta u_l-\dot{u}_{tool}\sin\theta, \tag{21}$$

where Δu_l is the contact tangent displacement increment of nodes, $\dot{u}_{tool}$ is the displacement increment of tooling, θ is the rotation angle.

The increment equation of $\dot{F}$ is expressed as follow:

$$\dot{F}=\left(\dot{F}_l-F_n\Delta u_l/R+F_n\dot{u}_{tool}\sin\theta/R\right)\cdot l+\left(\dot{F}_n-F_l\Delta u_l/R-F_l\dot{u}_{tool}\sin\theta/R\right)\cdot n, \tag{22}$$

Rigid matrix governing equation of the contact nodes is expressed below:

$$\begin{bmatrix}K & \cdots & \cdots\\ \cdots & K_{11}+F_n/R & K_{12}\\ \cdots & K_{21}+F_n/R & K_{22}\end{bmatrix}\cdot\begin{Bmatrix}\cdots\\ \Delta u_l\\ \Delta u_n\end{Bmatrix}=\begin{Bmatrix}\cdots\\ \dot{F}_l+F_n\dot{u}_{tool}\sin\theta/R\\ \dot{F}_l-F_l\dot{u}_{tool}\sin\theta/R\end{Bmatrix}. \tag{23}$$

2.1.5 Incremental steps of R_{min} method

Based on the elastic plastic finite element method with large deformation method, the Yamada r_{min} method [9] is used to judge the function. Each incremental step value is equal to incremental displacement of initial deformation increment of the tooling. Adopted the method of updated Lagrangian formulation, calculated each increment of displacement, strain, stress, load, spring-back value after forming the final shape of sheet mental in unloading condition, the value of load incremental in each step is controlled by $r_{\min}$ method, which is shown as below:

$$r_{\min}=MIN(r_1,r_2,r_3,r_4,r_5), \tag{24}$$

where, r_1 is the maximum allowable strain increment, r_2 is the maximum allowable rotation increment, r_3 is the minimum value in all elastic elements, r_4 is the contact position between free node and tooling, r_5 is the discontent position between free node and tooling.

2.2 NUMERICAL ANALYSIS

2.2.1 *Numerical analysis flow*

Based on the finite deformation theory, ULF equation and $r_{\min}$ method, a set of effective analysis of sheet metal forming process is established. Firstly, a 3D part and mold is designed with the NX software, and then is meshed with NASTRAN software. Secondly, the meshed model is drawn into the data file and did finite element analysis. The simulation flow chart is shown in Figure 1.

Based on the theory upwards, the bending process of U plate type is studied, including relationship between the punch load and displacements, stress and strain, thickness, spring-back and warpage. Some simulation experiments were carried out, of which parameters are friction coefficient (μ), punch radius (r_p), die radius (R_d). The results were verified by the experiment, and optimized and served a reference for U bending designer.

The whole experimental structure is composed of punch, die and blank holder. The experimental model picture was shown as Figure 2 below.

The initial relation of part and die is shown in Figure 2a, also, the punch down a certain stroke is shown Figure

2b. It takes two coordinates to solve the problem, which are fixed coordinates (X, Y, Z) and local coordinates (ξ, η, ζ). based on the right-hand rule, it uses the fixed coordinates (X, Y, Z) when nodes do not contact with the tool, and uses the local coordinates (ξ, η, ζ) when nodes contact with the tool. L-axis is the tangential direction of contact line between the part and tools when n-axis is the normal direction.

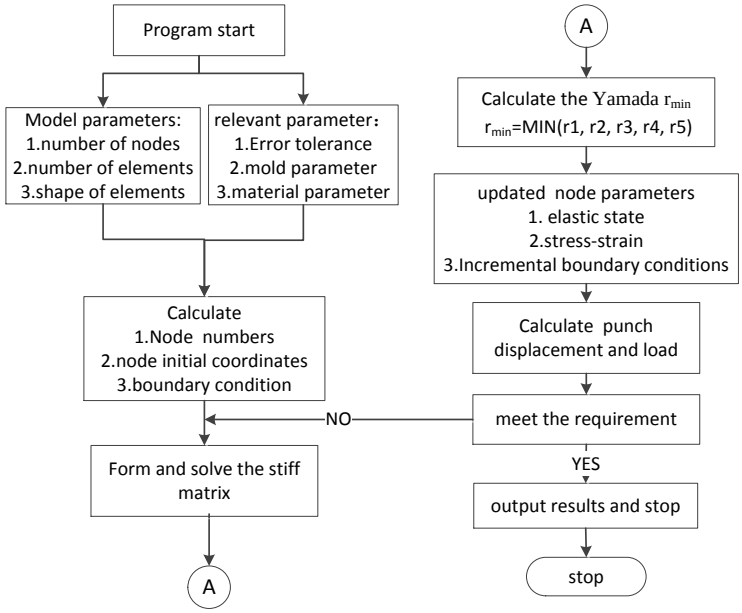
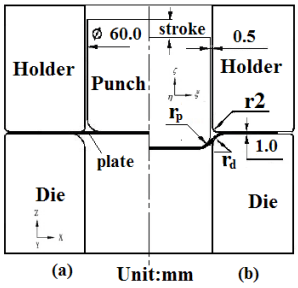


FIGURE 1 Numerical simulation of flow chart



a) before deformation, b) after deformation
FIGURE 2 Sheet metal and die size chart

The contact condition of each node of plates will change which based on deformation in sheet metal forming. When the displacement increment is zero, the boundary conditions of increment displacement of the next node will changes to free node boundary conditions. After checking the contact condition between the part and tool, generalized Rmin method is used. if sheet contacts the mold, the boundary condition will be changed to the contact condition.

Because the structure of sheet model is symmetrical, we take the 1/4 model to analysis.

It uses the quadrilateral segmentation of degenerated shell element in sheet metal meshing, when the die meshing uses the triangle segmentation.

Material parameters for numerical simulation are shown in Table 1 below, the shape of part is rectangular, Length L=120mm, Width W=50mm.

TABLE 1 Material property parameters.

Thickness	T=1mm
Length	L=120mm
Width	W=50mm
Yield stress	σ_y=131MPa
Young's modulus	E=210000MPa
Poisson's ratio	ν=0.3
Friction factor	μ=0.14

3 Results and discussion

3.1 STRESS DISTRIBUTION ANALYSIS

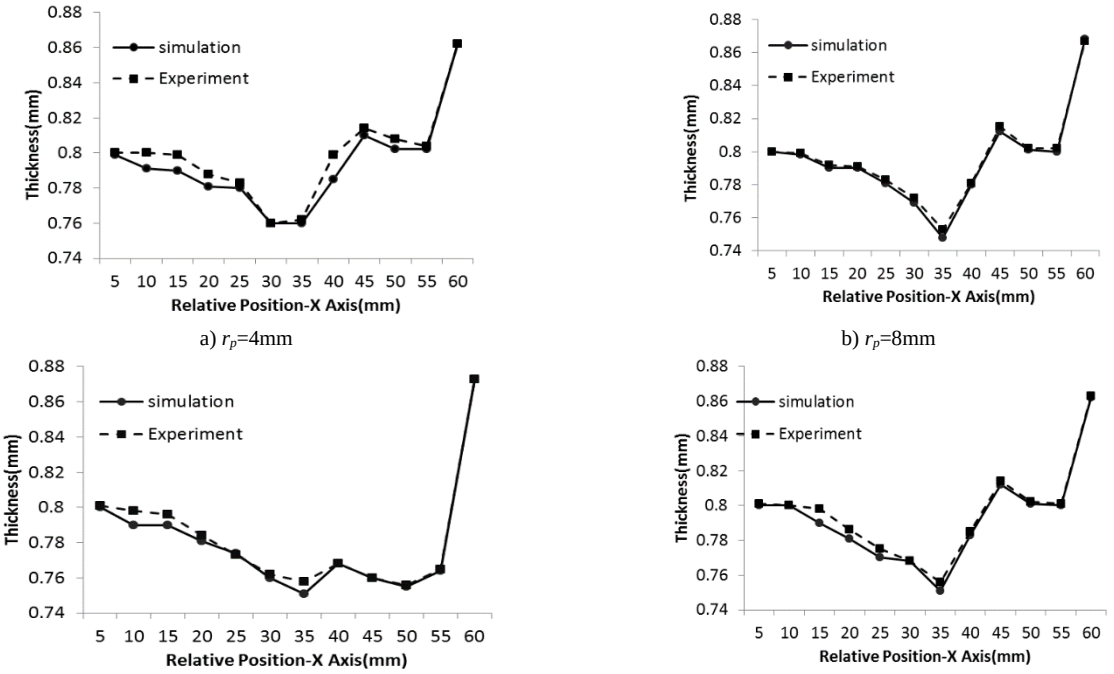
As can be seen from the result of simulation in Fig. 3, the initial stress mainly concentrates in fillet of die and punch. When the stroke is 12mm, the mainly stress concentrates in the area of punch fillet, some stress slightly and warpage occurs in bottom. The stress gradually shifted to the side wall plate with the increasing of stroke. When the stroke is 35mm, the sheet detaches the holder, the warping stress eliminates gradually, but the wall tress also exists.

3.2 THICKNESS ANALYSIS

After pressing, there is different stress in U-bending process of the sheet mental. Thickness would be chicken and thinning. In order to study the changes regulation between the thickness and fillet radius of die and punch, four simulation experiments are arranged, such us: r_p=4mm, r_p=8mm, R_d=4mm, R_d=8mm. The thickness of sheet in different position of the X-Axis are measured and recorded in Table 2 below.

TABLE 2 The thickness of sheet in different Relative position-X Axis under different parameters

Parameters	Simulation	Experiment	Simulation	Experiment	Simulation	Experiment	Simulation	Experiment
Relative position – X-Axis(mm)	r_p=4mm		r_p=8mm		R_d=4mm		R_d=8mm	
5	0.799	0.800	0.800	0.800	0.800	0.801	0.800	0.801
10	0.791	0.800	0.798	0.799	0.790	0.798	0.800	0.800
15	0.790	0.799	0.790	0.792	0.790	0.796	0.790	0.798
20	0.781	0788	0.790	0.791	0.781	0.784	0.781	0.786
25	0.780	0.783	0.781	0.783	0.774	0.773	0.770	0.775
30	0.760	0.760	0.769	0.772	0.760	0.762	0.768	0.768
35	0.760	0.762	0.748	0.753	0.751	0.758	0.751	0.756
40	0.785	0.799	0.780	0.781	0.768	0.768	0.783	0.785
45	0.810	0.814	0.812	0.815	0.760	0.760	0.812	0.814
50	0.802	0.808	0.801	0.802	0.755	0.756	0.801	0.802
55	0.802	0.804	0.800	0.802	0.764	0.765	0.800	0.801
60	0.862	0.862	0.868	0.867	0.873	0.873	0.862	0.863



a) r_p=4mm b) r_p=8mm

c) R_d=4mm d) R_d=8mm

FIGURE 4 Thickness distribution chart under different parameters

As can be seen from the Figure 4, Thickness distribution of sheet can be divided into three regions. The relative positions are from the 0mm to 20mm; the sheet contacts to the punch closely; the thickness is no obvious change. The second area is from the 20mm to 40mm, the sheet is under the axial stress in the area of die fillet, and thin gradually. The third area is from the 40mm to 60mm, the sheet is squeezed into die accompany with punch movement; the sheet flow is hindered, and the thickness increase.

3.3 WARPAGE AND SPRING-BACK ANALYSIS

The reason of warpage in U-Bending is that the sheet which affected by die constraint effect moves along the die side wall, when the punch moves down and sheet slips into the die. The outside edge of sheet is continuous, and reverse force is generated in curve bend, and forces the sheet to deform and warp at bottom of punch.

Measurement method of warpage and springback is shown as Figure 5.

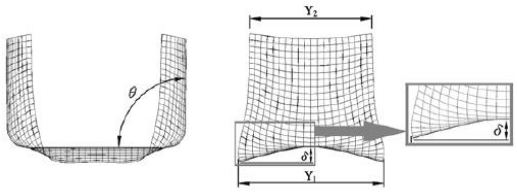


FIGURE 5 Measurement location of warpage and the Springback angle

As can be seen from the Figure 6, the relationship is shown between the relative position and warpage with the different punch stroke. When the stroke is 12.0mm, the warpage is 2.13mm. When the stroke is 16.0mm, the warpage is 3.62mm. When the stroke is 40.0mm, the

warpage is 4.72mm. When unload; the warpage is 4.71mm. The warpage mainly form when the stroke is from 0mm to12.0mm, and it changes slowly outside that scope.

In order to study the change law between warpage and fillet radius of different punch and die, improve the quality of sheet, the simulation experiments, which follow the arrangement below are carried out. Firstly, the punch fillet radius is fixed 6mm, die fillet radius are change from 2mm, 4mm, 6mm, 8mm to 10mm. Secondly, the die fillet radius is fixed 6mm, punch fillet radius are change from 2mm, 4mm, 6mm, 8mm to 10mm. the simulation and experiment results are recorded in Tab3.

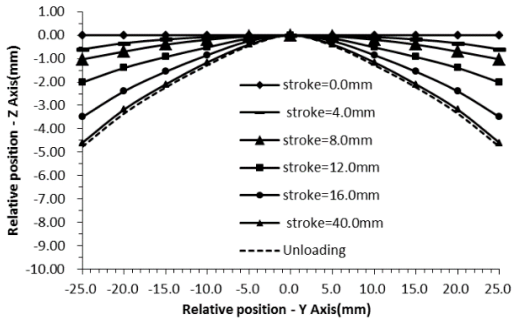


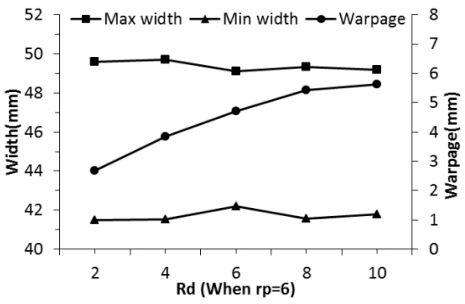
FIGURE.6 The relationship between the relative position and warpage with the different punch stroke

TABLE 3 Analysis results of warpage and angle in different parameters

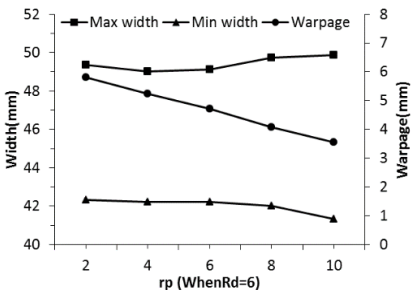
Parameters			Warpage (mm)	Angle (°)	Max width (mm)	Min width (mm)
r_p	R_d					
6.0	2.0	Sim.	2.67	90.1°	49.6	41.48
		Exp.	2.68	90.1°	49.56	41.53
	4.0	Sim.	3.85	90.1°	49.72	41.52
		Exp.	3.82	90.5°	49.38	41.63
	6.0	Sim.	4.72	90.1°	49.12	42.21
		Exp.	4.87	91.3°	48.92	42.82
	8.0	Sim.	5.42	90.5°	49.32	41.57
		Exp.	4.96	91.1°	49.03	41.75
	10.0	Sim.	5.63	90.1°	49.17	41.78
		Exp.	5.69	90.8°	49.34	42.03
2.0	6.0	Sim.	5.81	90.2°	49.36	42.32
		Exp.	6.52	91.5°	49.42	42.81
4.0		Sim.	5.24	92.3°	49.03	42.21
		Exp.	5.81	91.5°	48.92	42.81
6.0		Sim.	4.72	90.1°	49.12	42.21
		Exp.	4.87	91.3°	48.92	42.82
8.0		Sim.	4.08	90.1°	49.72	42.03
		Exp.	3.85	90.1°	49.65	42.62
10.0		Sim.	3.54	90.6°	49.87	41.32
		Exp.	3.50	90.2°	49.82	41.43

Note: sim.: simulation value, exp.: experiment value, r_p: Punch radius, R_d: Punch radius.
Sheet size: L=120.0mm, W=50.0mm.

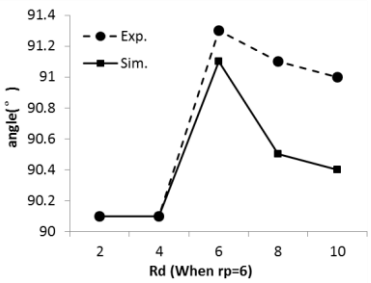
As can be seen from the Figure 7, the experimental and simulation results are basically similar. When the punch radius is unchanged, die radius is smaller and warpage is smaller. When the die radius is unchanged, punch radius is smaller but warpage is greater. The best warpage is appear when punch radius r_p=6mm, die radius R_d=2mm, the Min value is 2.67, the changes of Spring-back and width are relatively small.



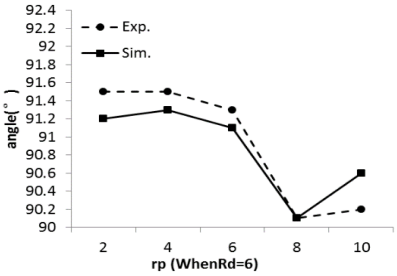
a)



b)



c)



d)

FIGURE 7 Warpage and spring back angle distribution chart under different parameters

4 Conclusion

Based on the numerical analysis and experimental results, combined with finite element method with the incremental elasto-plastic theory, analysed warpage phenomenon in U-bending process of sheet metal, the following conclusions are obtained as follows.

1) The maximum stress is mainly concentrated in the fillet of punch and die. With the stroke's moving, the stress shifts to the side wall and the distribution is quite uniform.

2) From simulation results it is shown the thinnest of sheet occurs at the fillet of punch, and if rupture, the position will be here. In addition, the end piece of long-axis occurs the thickening.

3) From simulation results it is shown the warpage occurs at tangent of punch and die fillet. Maximum warping appears when the stroke is from 0mm to 12.0mm.When the punch radius is constant, the die radius is smaller, the warpage is smaller too, Figure 7a. When the die radius is constant, the punch radius is greater, but the warpage is smaller, Figure 7b.

For additional experiments analysis, the min warpage can be predicted that when punch radius is 10.0mm, the die radius is 2.0mm, and the warpage can reduce to the minimum 2.52mm. These results are applied to engineering.

Acknowledgment

This work was financially supported by the Key Laboratory for Advanced Technology in Environmental Protection of Jiangsu Province (AE201039) (AE201038). Natural science foundation of Jiangsu Province (BK2012250). The Natural Science Foundation for General Universities of Jiangsu Province (China). (12KJD520010). The composite processing machine of small and medium-sized aircraft fuselage part (2014ZX04001071).

References

[1] Xia J, Dou S 2013 Experimental Research on the Flanging Height Adjustment Based on Environmental Way *International Journal of Applied Environmental Sciences* **8**(20) 2470-89
[2] Huang Y M, Chen T C 2005 Influence of Blank Profile on the V-Die Bending Camber Process of Sheet Metal *International Journal of Advanced Manufacturing Technology* (25) 668-77
[3] Sousa L C, Castro C F, António C A C 2006 Optimal Design of V and U Bending Processes Using Genetic Algorithms *Journal of Material Processing Technology* **172** 35-41
[4] Kim J K, Thomson P F 1989 Springback and Side-Wall Curl of Galvanized and Galvalume Steel Sheet *Journal of Mechanical Working Technology* **19** 223-238
[5] Lee S W, Yang D Y 1998 An Assessment of Numerical Parameters Influencing Springback in Explicit Finite Element Analysis of Sheet Metal Forming Process *Journal of Material Processing Technology* **20** 80-1
[6] Liu G, Lin Z, Xu W, Bao Y 2002 Variable Blankholder Force in U-Shaped Part Forming for Elimination Springback Error *Journal of Material Processing Technology* **120** 259-64
[7] Samuel M 2000 Experimental and numerical prediction of springback and side wall curl in u-bending of anisotropic sheet metals *Journal of Materials Processing Technology* **3**(105) 382-93

Authors

Jiansheng Xia, born in September, 1980, Yancheng City, Jiangsu Province, P.R. China

Current position, grades: lecturer at Yancheng Institute Of Technology, China.
University studies: BSc in Mechanical engineering and automation at Northeast Agricultural University in China. MSc at the Northeast Agricultural University.
Scientific interests: mold design, CAD/CAE/CAM technologies.
Publications: more than 101 papers.
Experience: teaching experience of 8 years, 2 scientific research projects.

Shasha Dou, born in December, 1982, Yancheng City, Jiangsu Province, P.R. China

Current position, grades: lecher of Yancheng Institute Of Technology, China.
University studies: BSc In Mechanical engineering and automation at Shandong University of Technology in China. MSc at Shandong University of Technology.
Scientific interests: plastic mold design, CAD/CAE/CAM technologies.
Publications: more than 101 papers.
Experience: teaching experience of 8 years, scientific research projects.

Frequency analysis of the area error in the triangular partition of a discrete global grid

Ming Tao*

College of Civil Engineering and Architecture, China Three Gorges University, No.8,Daxue Street, Xiling, District, Yichang, Hubei, China

Received 5 November 2014, www.cmnt.lv

Abstract

The discrete global grid is a method of hierarchical space expression with global scale, multi-resolution, and multi-scale transform features. It is a basic theory related to global geographic information systems and spatial grid computation. The main challenge in the practical application of discrete global grids is the generation of errors. Thus, four typical kinds of global discrete grid models were investigated in this study of partition errors. Area was considered an indicator of partition errors. Furthermore, this study computed the area errors in levels 3 to 10 partitions of each kind of global discrete grid model. The frequency of each indicator was also analysed. Analysis results show that the Synder models have the smallest area deformation.

Keywords: discrete global grid, triangular partition, frequency analysis

1 Introduction

The discrete global grid is a method of hierarchical space expression with global scale, multi-resolution and multi-scale transform features [1]. In this research field, however, errors are a significant problem, especially the deformation of discrete global grid model partitions. This deformation influences the geometrical homogenization of the discrete global grid model [2]. The basic partition deformation problems occur in terms of area, distance, and direction. In this study, the frequency analysis method is employed to investigate the area error in triangular partitions. The partition deformation change rule is analysed by computing the area error and the corresponding frequency characteristics.

2 Study objects

Four kinds of discrete global grid models were selected as the study objects. They represented typical discrete global grid categories.

The QTM model has an octahedral inscribed geometry [3]. Each of the chords between the adjacent nodes of the octahedral is divided into two equal segments. The midpoint of each chord is projected onto the surface of the sphere and becomes a new node. This new node forms a new geometry along with the original nodes in a process called partition, which can be repeated. Each partition is a project process from plane to sphere.

SQT has an icosahedral inscribed geometry [4]. The surface of the sphere is divided into 20 partitions of equal area. Each partition is recursively divided into four spherical triangles of similar area.

The Synder model also has an icosahedral inscribed geometry [5]. The projection transformation equation of this model was developed by Synder in 1992, and its partition method is similar to that of the Fuller model.

STQIE model [6] uses a projection transformation equation based on ERLPR projection. The projection plane is divided into triangles of equal area. The vertices of each triangle are then reversely projected onto the surface of the sphere to form many spherical triangles. These spherical triangles in turn comprise the STQIE grid.

3 Computation method of partition area error

The ideal discrete global grid model is an evenly partitioned unit sphere. Moreover, nodes are uniformly distributed on its surface. The partitions of the ideal model are equal in terms of area and are not afflicted with geometric deformation. Therefore, the ideal model does not display systematic errors. The area values of the ideal model can be regarded as true values. Once determined, the difference between true and calculated values of each discrete global grid model can then be used to analyse the errors.

The area calculation formula of ideal partition is as follows:

$$S_n = 4 \times \frac{\pi}{2^n}, \tag{1}$$

where S_n is the area of a partition in level n and n is the depth of the partition level.

The actual partition area is calculated using the following spherical triangle area formula:

* ***Corresponding author*** e-mail: mingtaomail@163.com

$$S = E \times \pi \times R^2 \times 180^0, \quad (2)$$

where E is spherical excess, and R is spherical radius.

A, *B*, and *C* are set as the three angles of a spherical triangle. Thus, the spherical excess Equation is as follows:

$$E = A + B + C - 180^0. \quad (3)$$

The area error Equation is as follows:

$$S_{\triangle} = E - S_n. \quad (4)$$

In discrete global grids with inscribed geometries, the vertices of the inscribed geometry distribute evenly on the surface of the sphere prior to partition. Thus, the partitions formed by these vertices are equal in terms of area and do not exhibit any errors. These partitions are called level 1 partitions. Level 2 partitions display errors because of the partition algorithms and the systematic error in the model. According to the results of previous research [2], deformation changes in the partitions stabilized after level 10. Thus, the aforementioned grid models are partitioned from levels 3 to 10 in this study to analyse the error change rule.

The area errors in the partitions from levels 3 to 10 are analysed through frequency analysis. This analysis evenly divides the interval between the maximum and the minimum statistical values. Furthermore, the frequency analysis of partition error can reflect whether it is influenced by model systematic error. It can also reflect the influence of subdivision methods on partition deformation.

4 Frequency analysis of partition area error

Figures 1 to 32 present the frequency analyses of the STQIE, SQT, Synder, and QTM models. Each model is divided into partitions ranging from levels 3 to 10.

Figures 1 to 8 depict a single peak shape in the representation of STQIE partition error frequency. The error intervals are both positive and negative. Furthermore, the entire interval is 9.775E-8. The error rate of the negative interval is 53.26%, whereas that of the positive interval is 46.74%.

Figures 9 to 16 also display a single peak shape, but the interval is positive. The entire interval is 3.595E-7, and the error interval of SQT is 3.67 times that of STQIE. Moreover, the distribution of area error in the latter is more concentrated than that in the former. In addition, STQIE has both positive and negative intervals. By contrast, SQT has only a negative interval. Therefore, the area deformation in the STQIE partition is smaller than that in the SQT partition.

The interval of area error frequency in the Synder model is symmetric, and the centre is located at point 0 as shown in Figures 17 to 24. Error distribution is increasingly concentrated in line with the increase in level depth. Furthermore, the partition area is little deformed by systematic error.

Figures 25 to 32 indicate that the frequency representation of the QTM model is irregular in shape. Partition area deformation is greater in this model than in the three other models under the influence of systematic error. Thus, the QTM partition method generates severe area error.

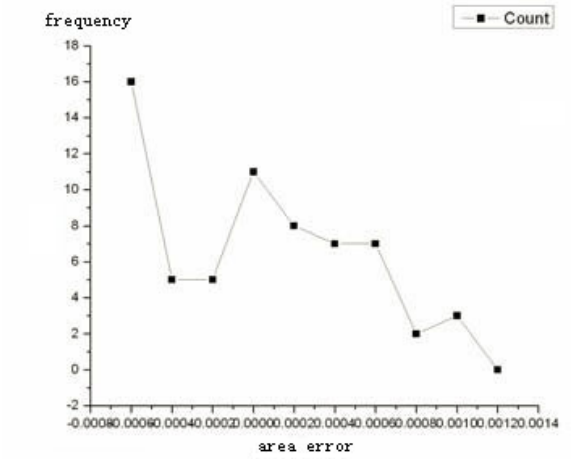


FIGURE 1 Frequency of the STQIE partition error in level 3

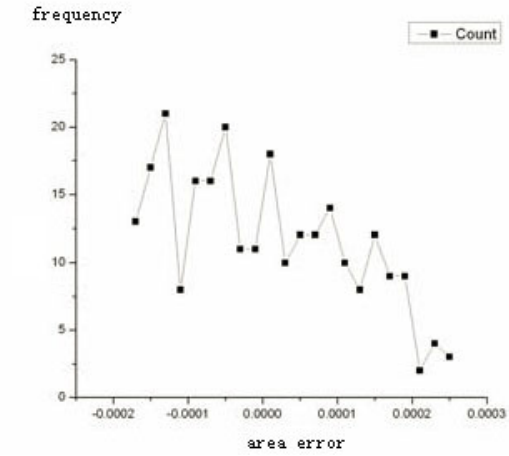


FIGURE 2 Frequency of the STQIE partition error in level 4

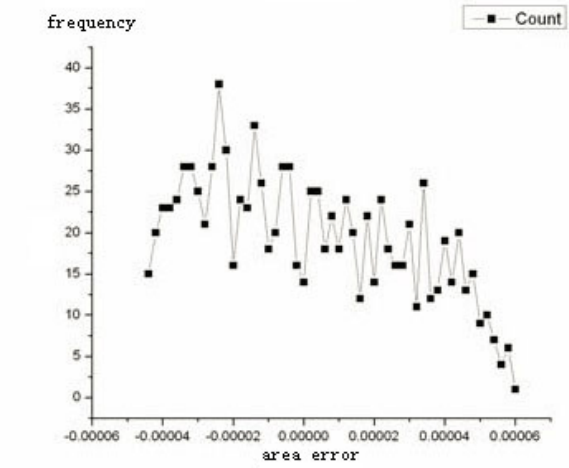


FIGURE 3 Frequency of the STQIE partition error in level 5

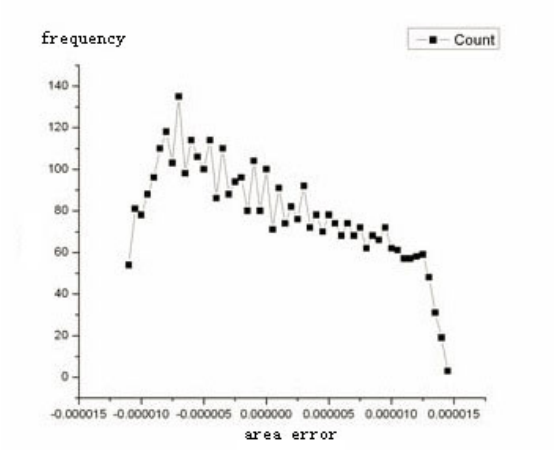


FIGURE 4 Frequency of the STQIE partition error in level 6

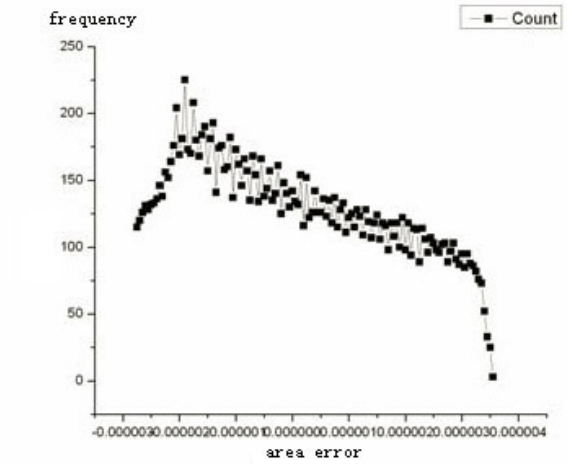


FIGURE 5 Frequency of the STQIE partition error in level 7

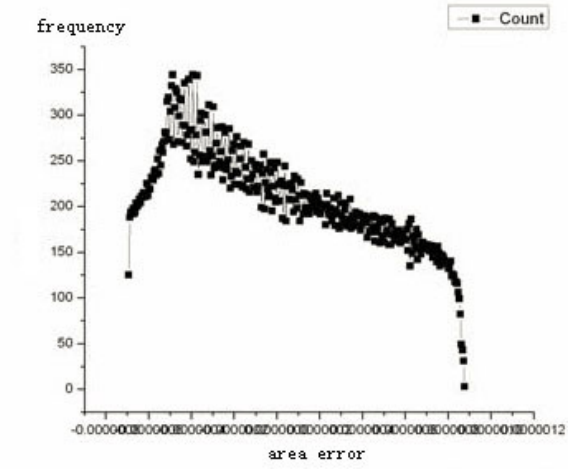


FIGURE 6 Frequency of the STQIE partition error in level 8

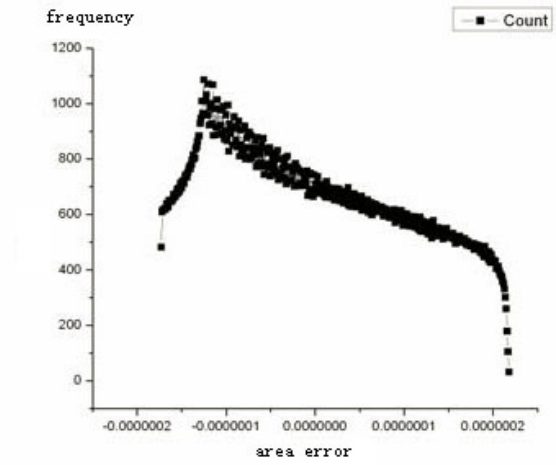


FIGURE 7 Frequency of the STQIE partition error in level 9

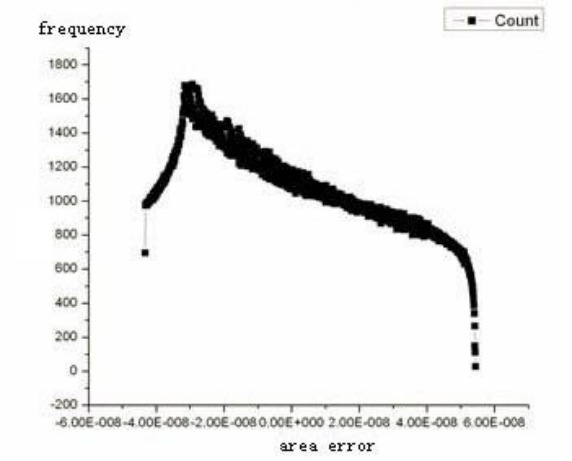


FIGURE 8 Frequency of the STQIE partition error in level 10

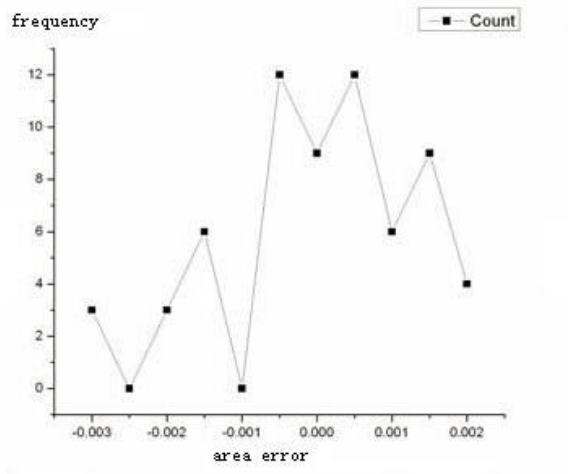


FIGURE 9 Frequency of the SQT partition error in level 3

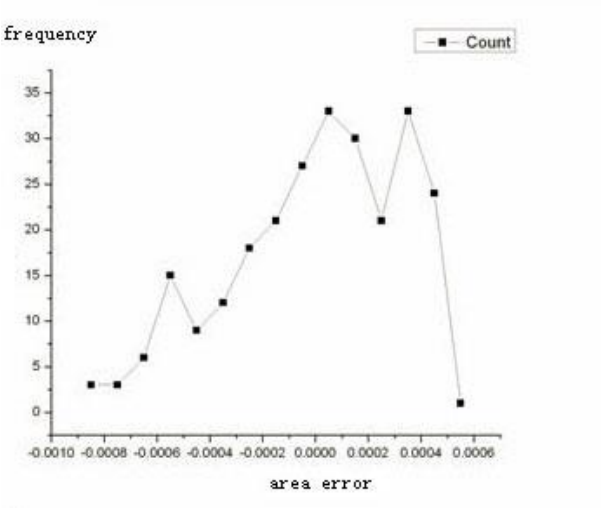


FIGURE 10 Frequency of the SQT partition error in level 4

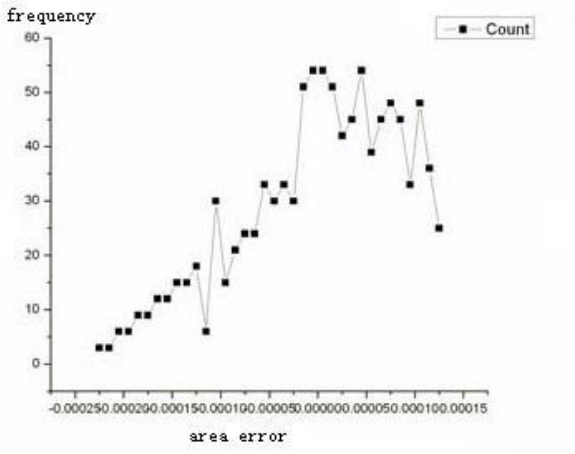


FIGURE 11 Frequency of the SQT partition error in level 5

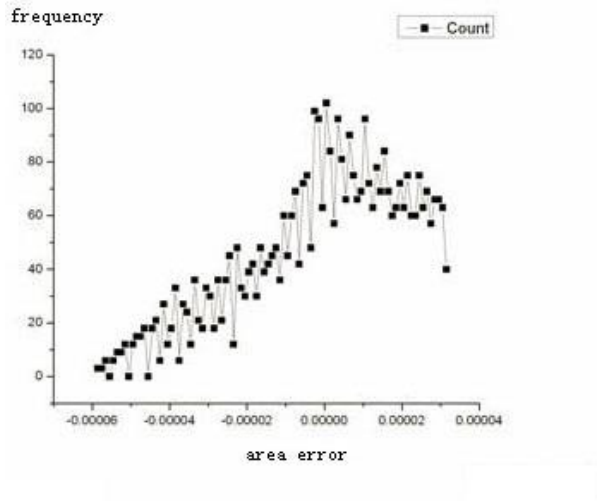


FIGURE 12 Frequency of the SQT partition error in level 6

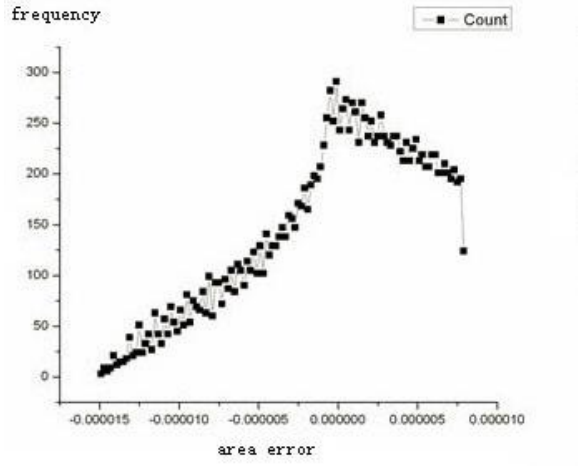


FIGURE 13 Frequency of the SQT partition error in level 7

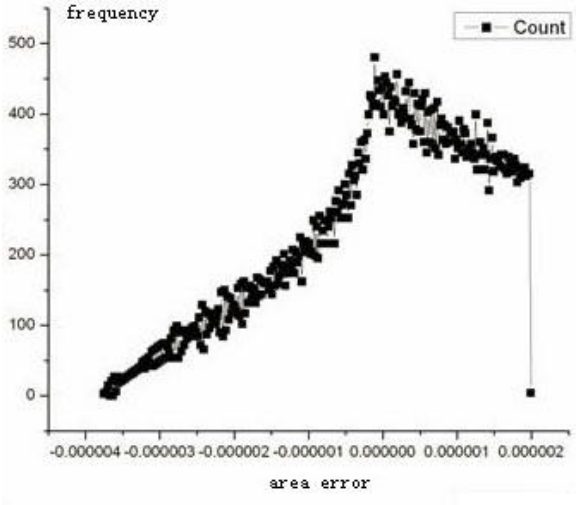


FIGURE 14 Frequency of the SQT partition error in level 8

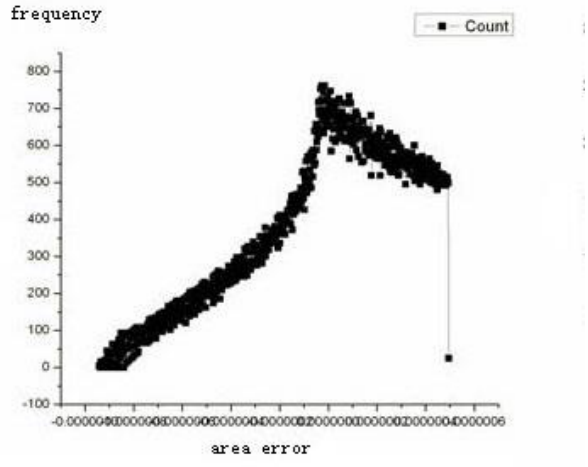


FIGURE 15 Frequency of the SQT partition error in level 9

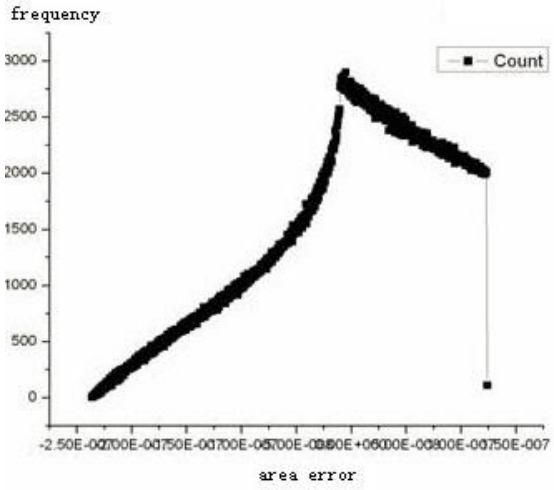


FIGURE 16 Frequency of the SQT partition error in level 10

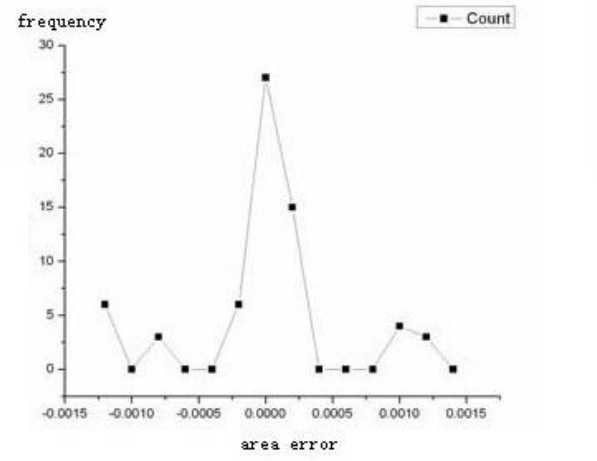


FIGURE 17 Frequency of the Synder model partition error in level 3

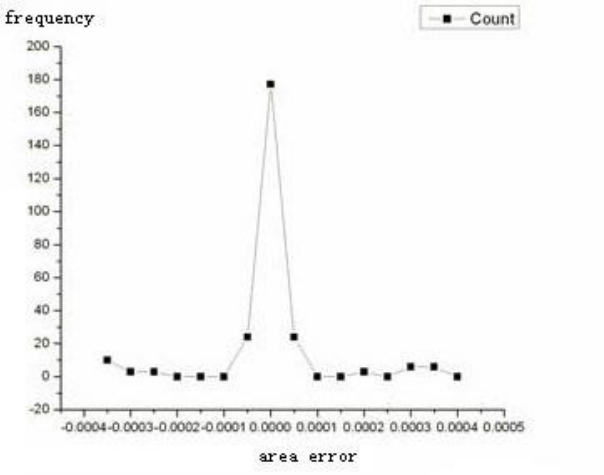


FIGURE 18 Frequency of the Synder model partition error in level 4

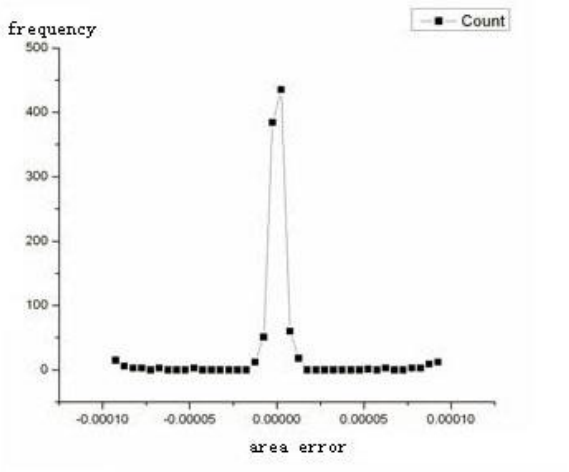


FIGURE 19 Frequency of the Synder model partition error in level 5

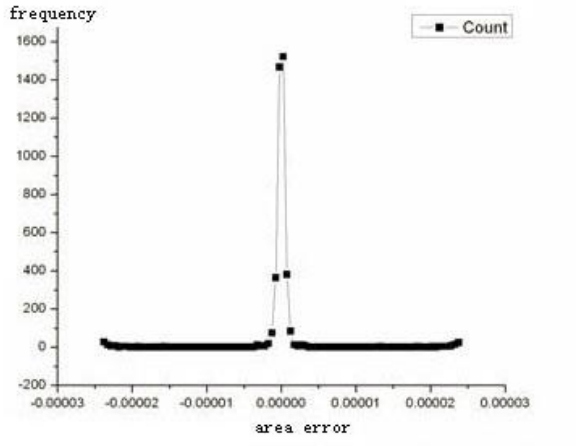


FIGURE 20 Frequency of the Synder model partition error in level 6

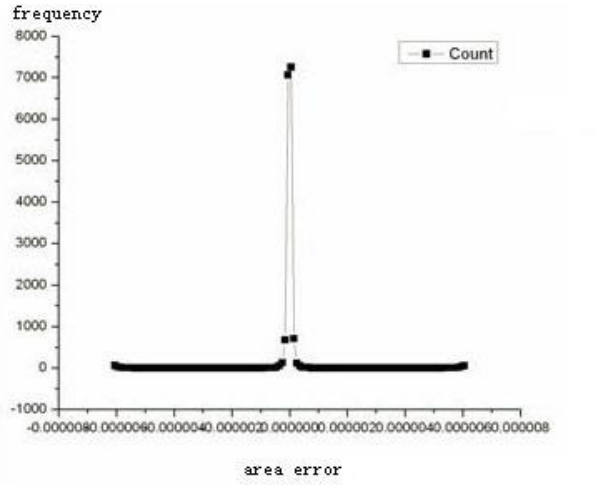


FIGURE 21 Frequency of the Synder model partition error in level 7

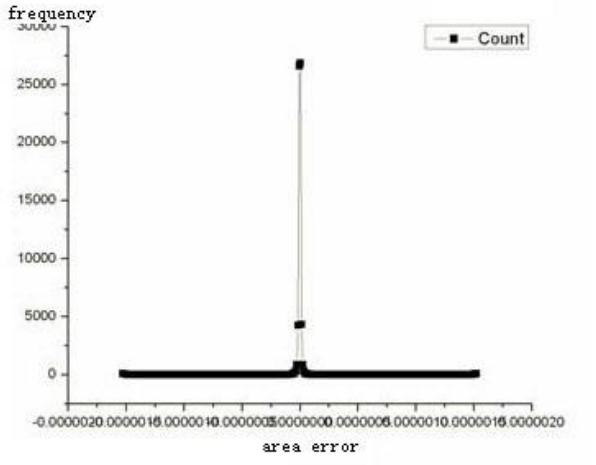


FIGURE 22 Frequency of the Synder model partition error in level 8

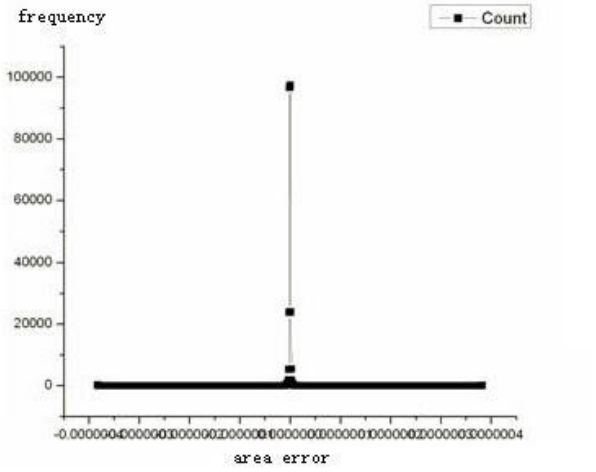


FIGURE 23 Frequency of the Synder model partition error in level 9

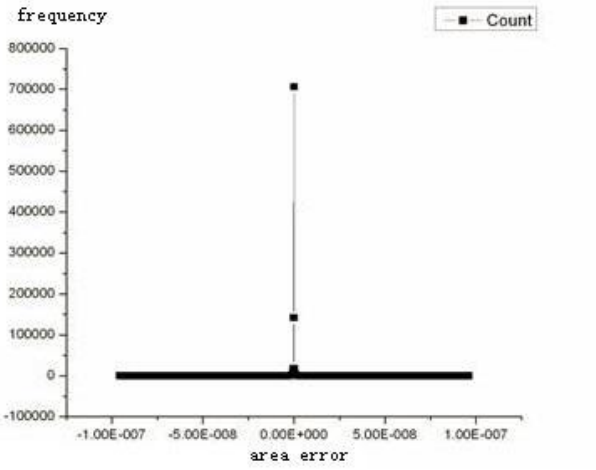


FIGURE 24 Frequency of the Synder model partition error in level 10

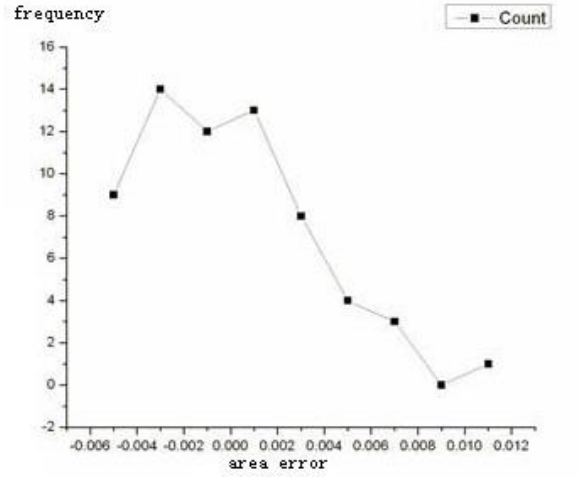


FIGURE 25 Frequency of the QTM partition error in level 3

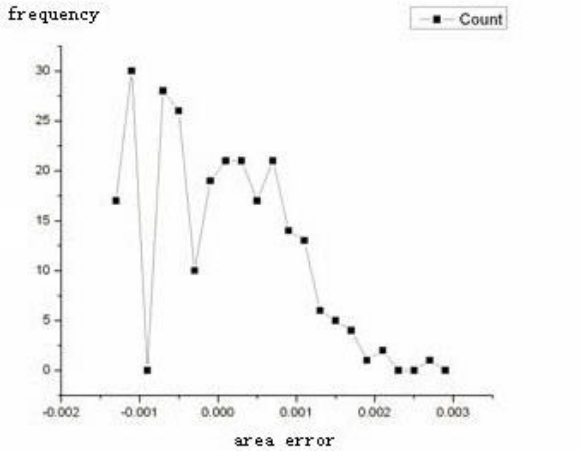


FIGURE 26 Frequency of the QTM partition error in level 4

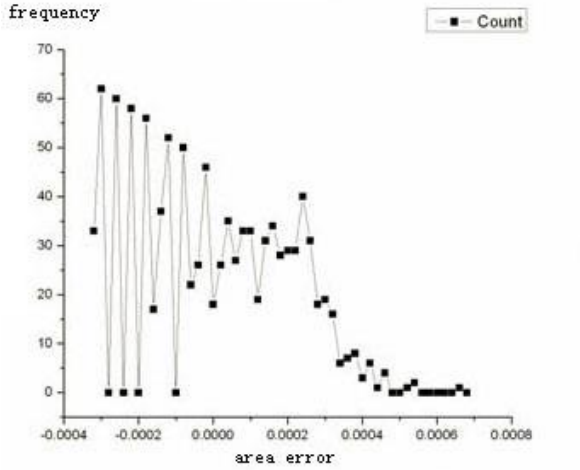


FIGURE 27 Frequency of the QTM partition error in level 5

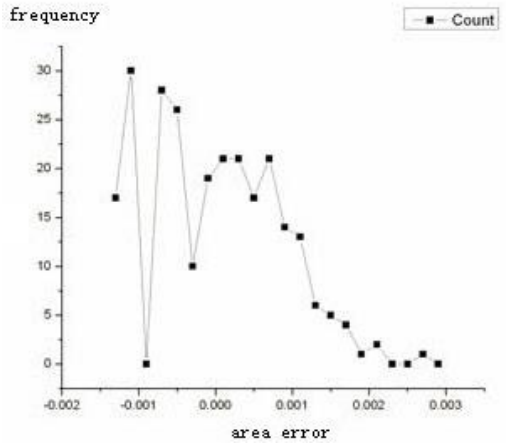


FIGURE 28 Frequency of the QTM partition error in level 6

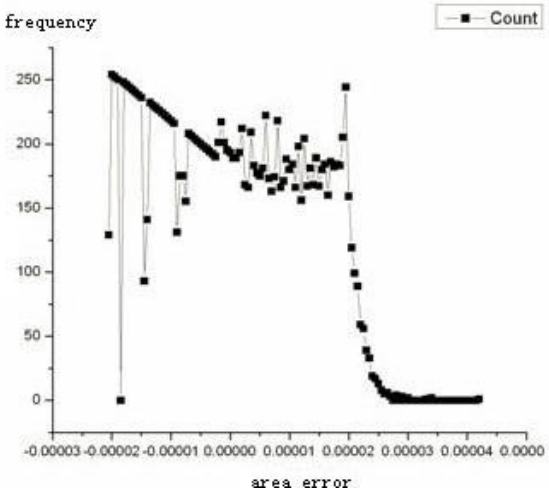


FIGURE 29 Frequency of the QTM partition error in level 7

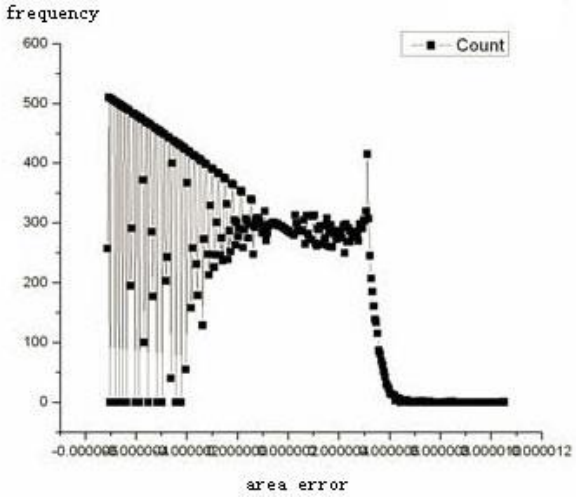


FIGURE 30 Frequency of the QTM partition error in level 8

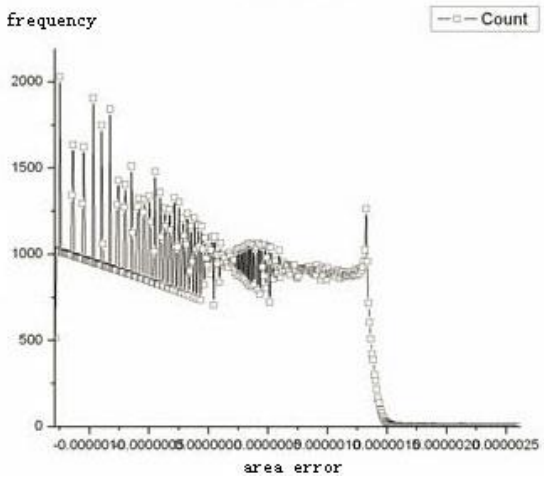


FIGURE 31 Frequency of the QTM partition error in level 9

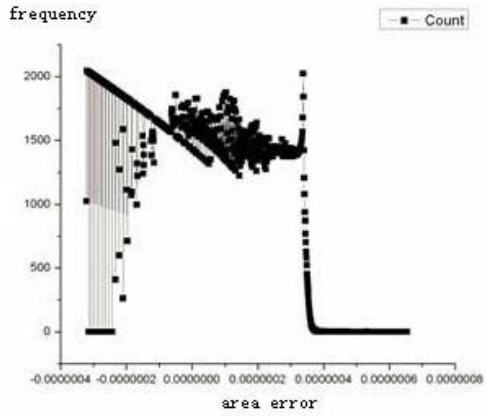


FIGURE 32 Frequency of the QTM partition error in level 10

5 Conclusions

A frequency analysis can clearly reflect the characteristics of area error in different grid model types. In particular, the distribution of these errors in the Synder model is symmetrical, and they are mainly induced by the accidental error generated during the computational process. This phenomenon is consistent with the characteristics of equal area projection in this model.

As a result of this projection, the Synder model displays the smallest area error compared with the other three models as per frequency analysis. By contrast, the QTM model exhibits the largest area error among the four models because of its partition method and octahedral inscribed geometry. With respect to area error, the four models are thus ranked in ascending order as follows: Synder, STQIE, SQT, and QTM.

Acknowledgments

This research was funded by the China State Natural Sciences Foundation (No. 40801161) and the China Three Gorges University Sciences Foundation of (No. KJ2014B004).

References

[1] Sahr.K, White D, Kimerling A J 2003 Geodesic discrete global grid systems *Cartography and Geographic Information Science* **30**(2) 121-34
[2] Tao M, Zhuang D, Wen Y, Qiu D, Wang Z 2007 The study on the geometrical homogenization of discrete global grid model *Chinese high technology letters* **17**(8) 40-3 (*in Chinese*)
[3] Dutton G 1999 A Hierarchical Coordinate System for Geoprocessing and Cartography *Springer-Verlag Berlin*
[4] Fekete G 1990 Rendering and managing spherical data with sphere quadtrees *Proceedings of Visualization'90* 176-86
[5] Snyder J 1992 An equal-area map projection for polyhedral globes *Cartographica* **29**(1) 10-21
[6] Yuan W, Ma A, Guan X 2005 A new projection for spherical triangle:equal angle ratio projection (EARP) *Acta Geodaetica et Cartographica Sinica* **34**(1) 78-84

Author

Ming Tao, born in September, 1978, Yi Chang City, Hubei Province, P.R. China

Current position, grades: the lecturer of College of Civil Engineering and Architecture, China Three Gorges University, China.
University studies: M.E. from Central South University in China, D.SC. from Institute of Geographic Sciences and Natural Resources Research, CAS.
Scientific interest: GIS and RS.
Publications: 10 papers.
Experience: teaching experience of 7 years, 8 scientific research projects.

Research on the application of clustering algorithm based on minimum spanning tree

Chen Ye*

College of Science and Technology, Ningbo University, Ningbo City, Zhejiang Province, China, 315212

Received 28 October 2014, www.cmnt.lv

Abstract

Cluster analysis is one of the most important technologies in data mining. Minimum spanning tree (MST) is an advanced algorithm in cluster analysis. Studying MST has important practical significances. Firstly, this paper analysed partitioning, hierarchical, density and grid clustering algorithms based on MST thoroughly. Secondly, implementation principles and shortcomings of these four algorithms were discussed. Finally, practical applications of clustering algorithm based on MST were introduced, aiming to solve some practical problems.

Keywords: data mining, clustering algorithm, MST model

1 Introduction

Nowadays, mature database technology has been developed and data application has been promoted to a new high. People are facing with big data every day. To utilize such big data effectively, researchers developed knowledge discovery in database (KDD) and improve it continuously. Various data mining technologies for different algorithms have been developed [1].

Cluster analysis is an important technology of data mining. However, clustering technology is still developing. Research on cluster analysis not only has theoretical significance, but also can widen its practical applications. Figure 1 shows the whole process of cluster analysis.

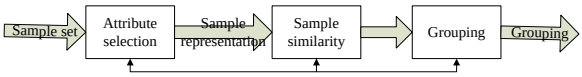


FIGURE 1 Process of cluster analysis

As things of one kind come together, researchers divide data objects into different groups according to their attributes. This is known as cluster analysis (Figure 2)

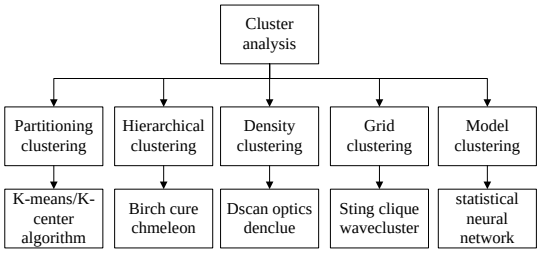


FIGURE 2 Cluster analysis

To overcome shortcomings of classical clustering analysis algorithms, a clustering algorithm based on MST was developed. It is an advanced cluster analysis and can eliminate blindness of clustering analysis significantly. With the continuous development and improvement of MST, four clustering algorithms based on MST have been developed: partitioning clustering algorithm based on MST, hierarchical clustering algorithm based on MST, density clustering algorithm based on MST and grid clustering algorithm based on MST [2] (Figure 3).

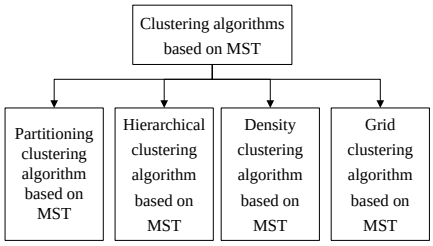


FIGURE 3 Classification of clustering algorithms based on MST

2 Classification of clustering algorithms based on MST

2.1 PARTITIONING CLUSTERING ALGORITHN BASED ON MST

Principle: in the partitioning clustering algorithm based on MST, features of dataset *D* are analysed and its complete undirected graph with weight (*G*) is constructed firstly. Secondly, researchers will divide the MST (*T*) into *k* subtrees according to the characteristics of *G*. So an initial cluster valued *k* is gained. When dividing the *T*, there's a common practice: longest *k*-1 sides in *T* will be found and deleted, so that the distance sum of sides in all subtrees will be the smallest and *k* initial clusters are acquired. This is based on the principle that different points can be grouped into different clusters according to different side lengths. In *D*, distance between sides of different clusters is longer than distance between points within the same cluster.

* ***Corresponding author*** e-mail: yechen@nbu.edu.cn

Therefore, we will get a globally optimal solution. Finally, *k* clusters can be generated by k-means clustering and adjusting affiliation of data objects in different clusters.

Steps of partitioning clustering algorithm based on MST are:

Step 1: Users construct the complete undirected graph with weight (*G*) of the clustering dataset (*D*) according to its characteristics.

Step 2: Users build the MST (*T*) of *G* by using appropriate algorithm.

Step 3: Users set parameters and divide *T* into *k* subtrees according to a specific criterion. *k* initial clusters and the initial cluster centre are generated.

Step 4: Users adjust the cluster centre continuously until get the desired one.

To sum up, partitioning clustering algorithm based on MST has significantly higher accuracy than other classical ones. This is contributed by its scientific and practical generation of initial cluster. Instead of random choose, it generates *k* initial clusters according to the principle of MST. Although the partitioning clustering algorithm based on MST has higher scientific and practical value than traditional ones, it still has some shortcomings for its intrinsic properties:

1) Low generation efficiency of *T*. The time complexity of *T* generation is the function of n^2: $0(n^2)$.

2) Low flexibility. When dividing the *T*, we have to delete the longest *k* sides. This is impractical for many datasets with different geometric distributions. In Step 3, the actual geometric distribution of dataset shall maintain close to the initial cluster in order to get good clustering effect. However, we have to set some parameters and use different division rules. Therefore, improvement is needed to make improved MST satisfy actual cluster distribution [3].

2.2 HIERARCHICAL CLUTERING ALGORITHM BASED ON MST

Analysis: Users build MSTs of the clustering dataset. Later, they cluster data objects using different hierarchical processes according to some characteristics of these MSTs.

Principle: Hierarchical clustering algorithm is an important algorithm. It is composed of agglomeration and splitting. Hierarchical clustering that decomposes from bottom to up is called as agglomeration. In agglomeration, each object is viewed as an independent class. Then, features of these objects will be analysed and similar classes will be combined. It ends until classes are combined into one class. Hierarchical clustering that decomposes from upper to bottom is known as splitting. It concentrates all data objects together as a class. Later, these data objects will be iterated continuously to split the class into smaller classes. It ends until each small class contains only one data object. Both agglomeration and splitting enable users to set different end conditions according to desired classes. Distance between classes is an important reference index to agglomeration and splitting. Common measurements of distance between classes are listed:

Minimum distance:

$$d_{\min}(c_i - c_j) = \min_{p \in c_i, p' \in c_j} \left| p - p' \right|. \quad (1)$$

Maximum distance:

$$d_{\max}(c_i - c_j) = \max_{p \in c_i, p' \in c_j} \left| p - p' \right|. \quad (2)$$

Mean distance:

$$d_{mean}(c_i - c_j) = \min_{p \in c_i, p' \in c_j} \left| m_i - m_j \right|. \quad (3)$$

Average distance:

$$d_{avg}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{p \in C_i} \sum_{p' \in C_j} \left| p - p' \right|. \quad (4)$$

In hierarchical clustering algorithm based on MST, users build MSTs according to different features of the clustering dataset, which then will be put in proper orders. Subsequently, proper combination objective function is selected to judge whether two adjacent sides shall be combined. To ensure the accuracy, all sides are judged. Finally, several sides are combined gradually until get the appropriate clustering result.

Although it has been improved a lot compared to classical hierarchical clustering algorithms, it still has some shortcomings:

1) Low efficiency. Since the time-consuming Prim or Kruskal algorithm is the main algorithm used to build MST, the time complexity of hierarchical clustering algorithm based on MST is the function of n^2: $0(n^2)$. It is almost inapplicable to clustering of big dataset.

2) Great fluctuation of clustering results. It often uses appropriate objective function for side combination, which requires users to set many experimental parameters. As a result, different users with different experiences will come to significantly different clustering results [4].

2.3 DENSITY CLUSTERING ALGORITHM BASED ON MST

In many division approaches, users often groups objects according to their distances. This often ends with a ball cluster. To get different shaped clusters, researchers developed a clustering algorithm based on density. In density clustering algorithm, users preset a threshold. If density of adjacent areas is higher than the threshold, data points will be clustered continuously. In other words, given a class, there must be data points higher or equal to a numerical value in the analysing area. Density clustering algorithm can process "noise" data and get clusters of different shapes. However, it has low efficiency. Its time complexity is generally a function of n^2: $0(n^2)$. Moreover, it performs unsatisfying to dataset with uneven density [5].

Since parameter setting in DBSCAN algorithm is very

complicated, Ankerst proposed the OPTICS (Ordering Points to Identify the Clustering Structure) algorithm based on the ordering of classes. It shows good performances to dataset with uneven density. In China, Zhao Yanchang et al. developed an isodense clustering algorithm. In most clustering algorithms based on density, parameter setting affects clustering structure directly and causes violent fluctuation of clustering results.

Analysis: Firstly, users build the MST of given dataset. Secondly, users divide the MST into many subtrees in view of its characteristics. Thirdly, these subtrees will be clustered according to densities of these subtrees, getting corresponding clustering structure.

Principle: Firstly, users analyse feature of given dataset and build the MST. Secondly, they find out the longest $k-1$ sides in the MST and delete them to get an initial cluster valued *k*. The MST after division can be expressed by $T[k]$. Viewed from the space perspective of dataset, the dataset is divided into *k* local areas. Data objects with similar density are divided into the same area. Thirdly, an appropriate function is selected to calculate the function value of each initial cluster. The core object (*d*) of $T[k]$ is determined. All objects within the direct density of d will be clustered. Now, the density clustering result based on MST is acquired.

Such improved density clustering algorithm is significantly superior to the classical ones. However, it still has some shortcomings because of its intrinsic properties.

1) Low efficiency. Since the time-consuming Prim or Kruskal algorithm is the main algorithm used to build MST, the time complexity of density clustering algorithm based on MST is the function of n^2: $0(n^2)$. Its application to clustering of big dataset is restricted.

2) Great fluctuation of clustering results. Users have to set parameters of $T[k]$ when judging objects within the direct density of d. These parameters often have no fixed reference standard and shall be set according to users' experiences. As a result, different users with different experiences will come to significantly different clustering results.

2.4 GRID CLUSTERING ALGORITHM BASED ON MST

Grid clustering algorithm is to develop a network structure by quantize an object space into limited units. It is advantageous for quick operation, but disadvantageous for low clustering accuracy.

To further accelerate the operation of grid clustering algorithm, Wang et al. put forward grid-based multiresolution clustering algorithm. It is quick in operation and has higher data processing efficiency. Moreover, it can make real-time data processing upon data adding and updating in the dataset and produce new clustering result. Though it has evident advantages, it is inferior in clustering accuracy.

Schikuta E. proposed two improved grid clustering algorithms which achieves outstanding performances in clustering of big dataset. These grid clustering algorithms are superior for no consideration to data input sequence, low requirement on data distribution and flexible data dimension and size, but inferior for low clustering accuracy [6].

Analysis: Firstly, users divide the dataset space into many basic grids and then distribute data objects to different grids. Based on users' density threshold, density of grids will be calculated and served as the basis for clustering. Secondly, users can build the MST of dense grids through appropriate algorithm and get *k* clusters by deleting longest *k*-1sides.

Although grid clustering algorithm based on MST has obvious advantages than other classical ones, it is inferior for fluctuating clustering results. As the basis of grid clustering algorithm based on MST, dense grid often needs various parameters to build the MST. Since these parameters often have no fixed reference standard, different users will set different parameters, thus getting different dense grid structures. As a result, they will build different MSTs and finally achieve different clustering results [7].

3 Application of clustering algorithms

Cluster analysis plays an important role in our daily lives. Clustering algorithms, an important mean of cluster analysis, are widely used. For example, they are often used in pattern recognition, research of market prospect, image processing, document classification, etc. [8].

In market analysis, cluster analysis is useful when marketers want to implement different marketing strategies to different customer groups. It can divide customers into groups according to purchase mode and make corresponding marketing strategies to win high market acceptance of their products. Cluster analysis is also highly appreciated in urban planning. It helps analysers and designers to divide and design the region into different types of residential area [9]. In seismic study, researchers can make a cluster analysis on geological faults and divide existing seismic centre into different clusters, so that analysers can get a comprehensive understanding on the distribution of seismic belt. In biological field, cluster analysis on genes of animals and plants enables scientists to classify genes with similar functions and discover deeper information. In real estate sales, salesmen can make different marketing strategies according to the cluster analysis of different commercial residential building [10].

The application of cluster analysis based on MST in genetic noise reduction is introduced. Figure 4 is the model of collected genetic data points, in which there are six noises. Noises in the model were processed using a clustering algorithm based on MST under different *k* and *q*. Results are listed in Table 1. When $k=3$ and $q=2$ or 3, the cluster analysis based on MST achieved the best noise reduction effect.

To verify reasonability of parameter settings, a proof test was conducted. DNA data were acquired using probing tools and 7 elements of the probe data interface $(a_0, a_1, a_2, a_3, a_4, a_5, a_6)$ were collected. Subsequently, the rotation matrix R Equation (5) and the displacement matrix Equation (6) were used to process data. Genetic data points simulated according to Equation (7) are drawn onto a coordinate system. Finally, noises of data points were reduced under the optimal k and q. The test results find good accordance with the cluster analysis.

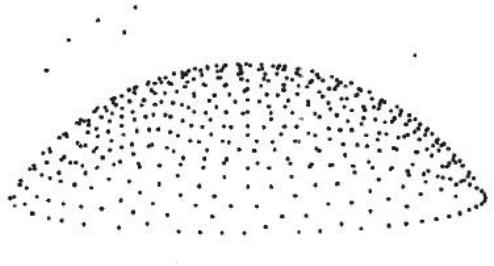


FIGURE 4 Model of genetic data points

$$R = \begin{vmatrix} a_0^2 + a_1^2 - a_2^2 - a_3^2 & 2(a_1a_2 - a_0a_3) & 2(a_1a_3 + a_0a_3) \\ 2(a_1a_2 + a_0a_3) & a_0^2 + a_2^2 - a_1^2 - a_3^2 & 2(a_2a_3 - a_0a_1) \\ 2(a_1a_3 - a_0a_2) & 2(a_2a_3 + a_0a_1) & a_0^2 + a_3^2 - a_1^2 - a_2^2 \end{vmatrix}, \quad (5)$$

$$T = (a_4 a_5 a_6)^T, \quad (6)$$

$$X = R_1^{-1}(R_2 X_2 + T_2 - T_1). \quad (7)$$

TABLE 1 Noise reduction results under different k and q

q	k=1		k=2		k=3		k=4		k=5	
	Noises reduced	**Errors**	**Noises reduced**	**Errors**	**Noises reduced**	**Errors**	**Noises reduced**	**Errors**	**Noises reduced**	**Errors**
1	3	9	3	8	6	2	6	4	6	6
2	3	10	4	8	6	0	6	2	6	7
3	2	6	4	5	6	0	6	5	6	7
4	2	7	5	5	6	0	6	7	6	9
5	3	8	6	4	6	2	6	7	6	10

4 Conclusions

This paper analyses four clustering algorithms based on MST: partitioning clustering algorithm based on MST, hierarchical clustering algorithm based on MST, density clustering algorithm based on MST and grid clustering algorithm based on MST. Principles, shortcomings and applications of these four clustering algorithms are introduced.

Acknowledgements

A Project Supported by Scientific Research Fund of Zhejiang Provincial Education Department (Y201119684).

References

[1] Zhou Y, Grygorash O, Hain T F 2011 Clustering with minimum spanning trees *International Journal on Artificial Intelligence Tools* **20**(1) 139-77
[2] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81
[3] Torkestani J A, Meybodi M R 2011 Learning automata-based algorithms for solving stochastic minimum spanning tree problem *Applied Soft Computing* **11**(6) 4064-77
[4] Yildirim A A, Özdoğan C 2011 Parallel WaveCluster: A liner scaling parallel clustering algorithms implementation with application to very large datasets *Journal of Parallel and Distributed Computing* **71**(1) 955-62
[5] Zhang C, Huang L, Zhao Z 2013. Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34
[6] Reddy D, Jana P K 2012 Initialization for K-means Clustering using Voronoi Diagram *Procedia Technology* **4** 395-400
[7] Zhang R, Kabadi S N, Punnen A P 2011 The minimum spanning tree problem with conflict constraints and its variations *Discrete Optimization* **8**(2) 191-205
[8] Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103
[9] Tang D 2010 *Research on clustering analysis and its application* PhD thesis of University of Electronic Science and Technology 54-8
[10] Wang X, Liu Q, Lu C 2009 Minimum spanning tree clustering algorithm *Journal of Chinese Computer Systems* **30**(5) 577-822 *(in Chinese)*

Author

Chen Ye, born in February, 1974, Ningbo, Zhejiang Province, P.R. China

Current position, grades: lecturer of the College of Science and Technology, Ningbo University, Zhejiang Province, China.
Scientific interests: statistical pattern recognition and stochastic processes.
Publications: more than 10 papers.
Experience: Teaching experience of 14 years, 4 research projects.

Routing method of quantum genetic algorithm

Jing Yang*

Baotou Vocational & Technical College, Baotou, Inner Mongolia, 014035, China

Received 1 May 2014, www.cmnt.lv

Abstract

Quantum genetic algorithm was applied in the work to solve multiple-QoS routing problems of bandwidth-delay constraint, thus meeting current requirements for multimedia messaging. Mathematical model and constraint condition of QoS routing were established, analysing algorithms principles and steps that how quantum genetic algorithm optimized QoS routing. By comparing average fitness and maximum polymerization fitness under different running times, quantum genetic algorithm can effectively find optimal solutions to multiple-QoS routing problem.

Keywords: genetic algorithm, quantum genetic algorithm, mathematical model, route choice, fitness

1 Introduction

Quantum genetic algorithm is a probability evolutionary algorithm with combination of quantum computing and genetic algorithm. State vector expression of quantum was introduced to genetic code, using quantum logic gates for chromosomal evolution, achieving better results than genetic algorithm. Therefore, there are important theoretical and practical values in solving practical problems with quantum genetic algorithm. QoS routing involves multi-objective optimization problems, thus the selection of appropriate algorithm has critical influence on the control of network bandwidth, delay and cost [1-3].

2 Quantum genetic algorithm

2.1 QUANTUM BIT ENCODING

A two-state quantum system, acted as an information storage unit in quantum computer, was called quantum bit. Quantum bit or qubit is the smallest information unit in quantum computing. A quantum bit has three states, namely $|0>$ state, $|1>$ state, and superposition state between $|0>$ and $|1>$. Therefore, the state of any quantum bits can be described as [4]:

$$|\varphi>=\alpha|0>\beta|1>,$$

where α, β are called probability amplitude of corresponding state of quantum bits, satisfying the normalization condition:

$$|\alpha|^2+|\beta|^2=1.$$

2.2 QUANTUM GATE RENEWAL

As the executing agency of evolution operations, quantum gates are selected according to specific issues. Quantum Rotating Gate was chosen in the work, with the following adjusted operation:

$$U(\theta_i)=\begin{bmatrix}\cos(\theta_i) & -\sin(\theta_i)\\ \sin(\theta_i) & \cos(\theta_i)\end{bmatrix}. \quad (1)$$

Renewal process is as follows:

$$\begin{bmatrix}\alpha_i'\\ \beta_i'\end{bmatrix}=U(\theta_i)\begin{bmatrix}\alpha_i\\ \beta_i\end{bmatrix}=\begin{bmatrix}\cos(\theta_i) & -\sin(\theta_i)\\ \sin(\theta_i) & \cos(\theta_i)\end{bmatrix}\begin{bmatrix}\alpha_i\\ \beta_i\end{bmatrix}. \quad (2)$$

From Equation (2) we can obtain:

$$\begin{cases}\alpha_i'=\alpha_i\cos(\theta_i)-\beta_i\sin(\theta_i)\\ \beta_i'=\alpha_i\sin(\theta_i)+\beta_i\cos(\theta_i)\end{cases}. \quad (3)$$

Therefore,

$$|\alpha_i'|^2+|\beta_i'|^2=[\alpha_i\cos(\theta_i)-\beta_i\sin(\theta_i)]^2+[\alpha_i\sin(\theta_i)+\beta_i\cos(\theta_i)]^2=|\alpha_i|^2+|\beta_i|^2=1. \quad (4)$$

Then it can be transformed as:

$$|\alpha_i'|^2+|\beta_i'|^2=1. \quad (5)$$

3 Qos routing problem

Currently, Internet network mostly adopts *best – effort* routing protocol with the shortest-path routing strategy, thus the requirements cannot be satisfied for receiving and transmitting multimedia information. From network users' perspective, QoS (Quality of Service) routing algorithm should firstly meet users' QoS request, namely finding a

***Corresponding author** e-mail: yangjing8825@126.com

transmission path corresponding to various conditions from source node to destination node. From facilitators' perspective, the use of network resources can be optimized through QoS routing algorithm [5, 6].

Users' Qos is decided by various parameters of network. QoS defined by RFC2216 refers to packet transmission characteristics described through parameters of bandwidth, delay, delay jitter and packet loss rate. Only by meeting users' requirements of QoS parameters can network reach the required services quality of network.

Typically, the topology and link state information of network can be abstracted as a weighted graph $G(V,E)$ in Figure 1 [7]:

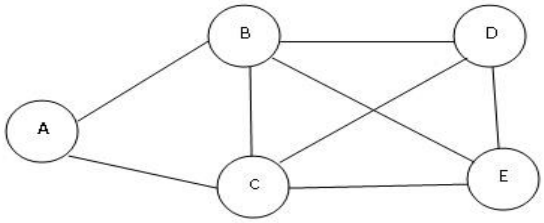


FIGURE 1 Network topology diagram of QoS routing

Vertices in Figure 1 represented network nodes; edges referred to network links; $V=(V_1,V_2,...,V_n)$ was the set of all switching nodes of router; $E=(E_1,E_2,...,E_n)$ indicated the link set connecting routers. QoS routing selection was to find the route satisfying requirements of network QoS from Figure 1.

4 QoS routing based on quantum genetic algorithm

4.1 MATHEMATICAL MODEL OF QOS ROUTING

A mathematical model of QoS routing should be established for better analysis and solutions of problems. A,B,C,D,E in Figure 1 represented the routes, while ligatures between two routes were pathways of routes. Each link was expressed with cost, width and delay, respectively.

Based on above analysis, issues of cost, bandwidth and delay are key factors affecting QoS routing. Not all factors should be considered in design of specific routing algorithm, because too complex algorithm will affect actual application of algorithm. Therefore, appropriate algorithms should be designed for different practical needs, dealing with different constraints [8-10].

For any links, (a,b) and $D(a,b)$ expressed the delay of link generation; $B(a,b)$ represented available bandwidth of the link; $C(a,b)$ was the cost of communication process.

Assuming the delay of path between u,v is:

$$delay(u,v)=\sum_{(a,b)\in P(a,b)} D(a,b)\,. \tag{6}$$

Available bandwidth of $P(u,v)$ is:

$$width(u,v)=\min_{(a,b)\in P(u,v)} B(a,b)\,. \tag{7}$$

The cost of $P(u,v)$ in communication process is:

$$\cos t(u,v)=\sum_{(a,b)\in P(a,b)} C(a,b)\,. \tag{8}$$

The mathematical model of QoS routing can be implied from Equations (6), (7) and (8), as well as constraint condition of QoS routing. The mathematical model is as follows:

Objective function:

$$\cos t(T)=\min(\sum_{(a,b)\in Er} C(a,b))\,. \tag{9}$$

Constraint condition:

$$\begin{aligned}&\forall v\in M,\ \sum_{(a,b)\in P_T(u,v)} D(a,b)\le D_{\max}\\ &\forall v\in M, Width(P_T(s,v))\ge W_{\min}\end{aligned}\,. \tag{10}$$

4.2 ALGORITHM FLOW OF QOS ROUTING BASED ON QUANTUM GENETIC ALGORITHM

Based on principles of quantum genetic algorithm and the mathematical model of QoS routing, algorithm flow of the work is as follows:

1) Initialize population $Q(t_0)$, and randomly generate n chromosomes encoded with quantum bits;

2) Measure each individual of the initial population $Q(t_0)$ to obtain corresponding definite solutions;

3) Evaluate the fitness of each definite solution;

Fitness function:

$$\cos t(T)=\min(\sum_{(a,b)\in Er} C(a,b))\,. \tag{11}$$

In the Equation, $C(a,b)$ represented the required cost of passing links in communication process. The smaller of fitness function value, the lower of communication cost.

4) Record optimal individual and corresponding fitness;

x_i was the i-th place of current chromosome; $best_i$ the i-th place of optimal chromosome; $f(x)$ fitness function; $s(\alpha_i,\beta_i)$ rotation angle direction; $\Delta\theta_i$ rotation angle size.

The calculated fitness $f(x)$ was compared with $f(best_i)$ - the fitness of current optimal individual in its population $f(best_i)$. If $f(x)>f(best_i)$, then the corresponding quantum bits should be adjusted, making probability amplitude (α_i,β_i) evolve toward the direction conductive to emergence of x_i. If $f(x)>f(best_i)$, then

probability amplitude (α_i, β_i) should evolve toward the direction conductive to emergence of $f(best_i)$.

5) Determine whether the calculation process can be completed: exit if the end conditions were met, or continue calculation;

6) Measure each individual of population $Q(t)$, and obtain corresponding definite solutions;

7) Evaluate the fitness of each definite solution;

8) Adjust individuals with quantum revolving gat $U(t)$, and obtain new population $Q(t+1)$;

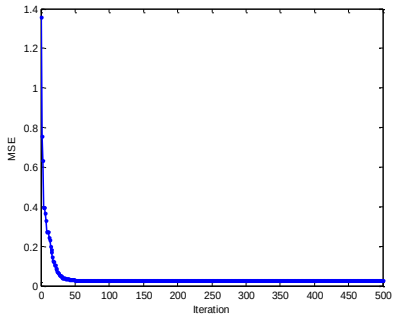
9) Record optimal individual and corresponding fitness;

10) Return to Step 5) with iterations $t = t+1$.

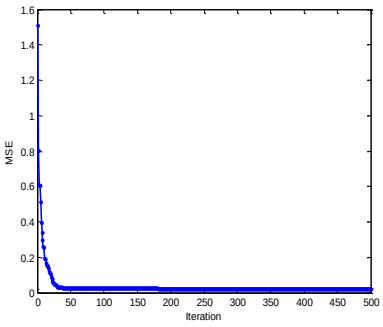
5 Simulation analysis

In order to verify the validity and effectiveness of the algorithm, corresponding parameters should be set as follows. Maximum genetic algebra: 500; population size: 40; maximum delay constraints: $D_m = 10$; minimum bandwidth constraints: $W_m = 100$; crossover probability: $p_c = 0.3$; mutation probability $p_m = 0.15$; passed path: $num\,\mathrm{var} = 8$.

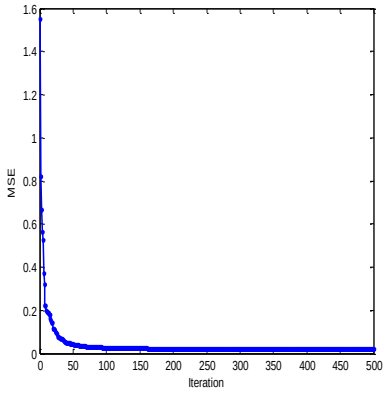
Convergent graph of different running times was shown in Figure 2:



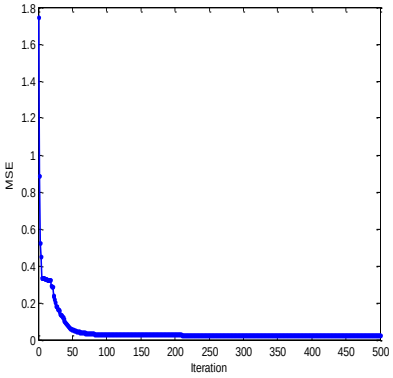
a) 1 running time



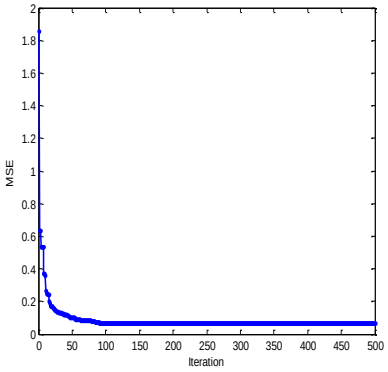
b) 5 running times



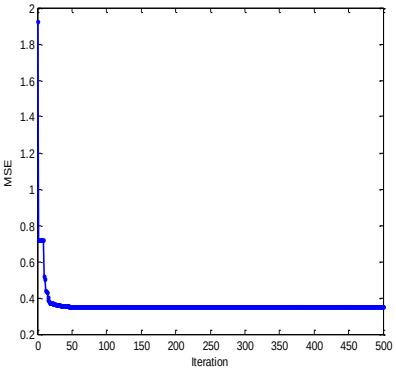
a) 10 running times



b) 20 running times



a) 25 running times



b) 50 running times

FIGURE 2 Convergent graph of different running time

Simulation results were shown in Table 1:

TABLE 1 Optimization results of fitness function

Running times	Average fitness	Maximum polymerization fitness
1	11.2616	0.1848
5	10.7399	0.1872
10	10.5234	0.1895
20	10.1735	0.1935
25	9.6875	0.1957
50	9.3012	0.1984

Table 1 showed that average fitness value presented a decreasing trend with increase of running times. The required cost of individuals continued decreasing, while maximum polymerization fitness value showed an increasing trend with increase of running times. This indicated that individuals' fitness more and more adapted to required conditions of specific issues. Meanwhile, better results could be obtained through running times of quantum genetic algorithm.

6 Conclusions

In the work, quantum genetic algorithm—integrating principles of quantum computing with genetic algorithm—was applied to QoS routing. Objective function and constraint condition were determined by constructing a mathematical model of QoS routing. Then, simulation solving was conducted in accordance with the principles and steps of the algorithm in the work. Simulation results showed that individual required cost presented a decline trend with increase of running times; individual fitness showed a rising trend with the increased of running times. Therefore, better results can be obtained with multiple running times of quantum genetic algorithm.

References

[1] Laska J, Kirolos S, Duarte M,et al 2007 Theory and implementation of an analog to information converter using random demodulation *Proceedings of the IEEE Inlsymp.On Circuits and Systems (ISCAS)[C] Piscataway Institute of Electrical and Electronics Engineers Inc 2007 1959-1962.*

[2] Tropp J, Gilbert A 2007 *IEEE Trans on Information Theory* **53**(12) 4655-66

[3] Donoho D, Tsaig Y 2006 Fast solution of ell-1-norm minimization problems when the solution may be sparse *Stanford University Department of Statistics Technical Report* 18

[4] Figueiredo M A T, Nowak R D, Wright S J 2007 *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing* **1**(4) 586-98

[5] Egiazarian K, Foi A, Katkovnik V 2007 Compressed sensing image reconstruction via recursive spatially adaptive filtering *(Preprint, 2007)*

[6] Duarte M, Davenport M, Takhar D, Laska J, Sun T, Kelly K, Baraniuk R 2008 *IEEE Signal Processing Magazine* **25**(2) 83-91

[7] Wakin M, Laska J, Duarte M, Baron D, Sarvotham S, Takhar D, Kelly K, Baraniuk R 2006 An architecture for compressive imaging *Int Conf on Image Processing (ICIP) Atlanta Georgia*

[8] Wakin M, Laska J, Duarte M, Baron D, Sarvotham S, Takhar D, Kelly K, Baraniuk R 2006 Compressive imaging for video representation and coding *Proc Picture Coding Symposium (PCS) Beijing China*

[9] Takhar D, Laska J, Wakin M, Duarte M, Baron D, Sarvotham S, Kelly K, Baraniuk R 2006 A new compressive imaging camera architecture using optical-domain compression *Computational Imaging IV at SPIE Electronic Imaging San Jose California*

[10] Haupt J, Nowak R 2006 Compressive sampling vs conventional imaging *Int Conf on Image Processing (ICIP) Atlanta Georgia*

Author

Jing Yang, born in March, 1974, Shuozhou, Shanxi Province, China

Current position, grades: an associate professor in Baotou Vocational & Technical College, China.
University studies: computer science and technology
Scientific interest: computer software and theory, algorithm.
Publications: 22 papers.

A novel method for K-Means clustering algorithm

Jinguo Zhao*

School of Computer and Information Science, Hunan Institute of Technology, Hunan, 421002, China

Received 1 March 2014, www.cmnt.lv

Abstract

This paper investigated K-means algorithm, a well-known clustering algorithm. K-means clustering algorithms have some shortfalls and defects, and one defect is reviewed in this study. One of the disadvantages of K-means clustering algorithms is that they can produce clusters that do not always include all the correct components. It is due to the presence of the error rate during the clustering process. The purpose of this research was to decrease error rates in the K-means clustering algorithm and to reduce iteration of running this algorithm. A novel method is proposed to calculate the distance between cluster members and cluster centre. To evaluate the algorithm proposed in this study, seven well-known data sets consisting of Balance, Blood, Breast, Glass, Iris, Pima and Wine data sets were used. This investigation revealed that the performance of K-means algorithms was increased and resulted in valid clusters and that it reduced error rates, run time and iteration.

Keywords: K-means, clustering algorithm, error rate, iteration, reduction, stable

Introduction

Clustering is an important technique used in many fields such as knowledge discovery and information retrieval. It helps researchers find related information more quickly [33]. As a result, researchers are kept up to date with new findings in their fields. Clustering is the process of grouping or dividing a set of objects into subsets (called clusters) so that the objects that are similar to one another are placed within the same cluster and dissimilar objects are placed in other clusters [26]. In other words, an object is similar to at least one other object in the same cluster and dissimilar to objects in other clusters in terms of predefined distance or similarity measure [31]. Currently, clustering as a tool for classification, pattern analysis, information extraction and decision making, has attracted the tendency of numerous investigators. Numerous techniques and approaches have been introduced in the literature. Each of these methods includes a certain measure, and has its own disadvantages and advantages. In general, there is no comprehensive technique and measure for optimal clustering of any kind of data [6].

In this study, a new understanding of the clustering algorithm was expressed. The most prominent, the most commonly used and the most popular clustering algorithm is the K-means algorithm, and it is used in this study. Among clustering algorithms, the K-means clustering algorithm can be used in many fields, including image and audio data compression, pre-process system modelling with radial basis function networks and task decomposition of heterogeneous neural network structure. One problem of clustering algorithms is that the clustering results are not always stable. In repeating the clustering algorithm several times, correct answers may be found in some trials but in others it may not find the correct answers due to instability. The clustering algorithm should be constant and stable, which is reviewed in this survey. This problem and gap as mentioned in the fourth part are related to the summary of a section of a Jain article [23].

Cormack (1971) first proposed that clusters should be internally integrative and externally segregated, suggesting a certain degree of uniformity within clusters and heterogeneity between clusters. So, many investigators tried to operationalize this description by minimizing within-group disparity [11, 14, 15, 45]. Following these efforts at maximizing within-group uniformity, Sebestyen (1962) and MacQueen (1967) separately developed the K-means technique as a strategy that tries to discover optimal partitions. Based on this significant advancement, K-means has become very popular, earning a place in a variety of textbooks on multivariate techniques [28, 32, 46], cluster analysis [17], pattern recognition [12], statistical learning [19, 43]. There are many surveys in K-means clustering algorithm field, yet this algorithm has still not been completely improved. In this paper, we reduced the error rate of the clustering algorithm and increased the stability of this algorithm.

K-means clustering algorithm has a number of disadvantages and problems, and one problem was reviewed in this study. This paper is organized as follows. Section 2 and 3 review the literature about clustering algorithms and K-means clustering algorithm. Section 3 describes the proposed method and research methodology used in this study. Section 5 explains the experiment conducted as a part of this study in the K-means clustering algorithm and improved K-means clustering algorithm, and the results are evaluated in Section 6. Finally, conclusions are drawn and discussed in Section 7.

***Corresponding author** e-mail: jinguo2014@126.com

2 Related works

In this section, the brief literature of the clustering algorithms is examined in which different researchers have previously expressed and improved these algorithms. Forgy's technique [16] randomly allocates each point to one of the K clusters homogeneously. The centres are then given with the centroids of these primary clusters. This technique has not basis of theoretical as, for example, random clusters have not homogeneity of internal [2]. Jancey's technique [25] allocates to each centre a combinatorial point randomly generated within the space of data. However, as the data set fills the space, a number of these centres may be too distant from any of the points [2], which might lead to the formation of unfilled clusters [13].

MacQueen (1967) suggested two different techniques. The first technique is the default choice in the Quick Cluster method of IBM SPSS Statistics [38], which obtains the first K points in X as the centres. An obvious disadvantage of this technique is its sensitivity into data ordering. The second technique selects the centres randomly from the data points. The foundation behind this technique is that random choice is likely to result in the selection of points from dense regions, points are suitable applicants to be centres. Ball and Hall's technique [5] obtains the centre of X, as the first centre. It then crosses the points in optional order and obtains a point as a centre if it is at least T units apart from the formerly selected centres until K centres are taken. The aim of the distance threshold T is to make sure that the seed points are well parted. The Simple Cluster Seeking technique [47] is the same as Ball and Hall's technique with the distinction that the first point in X is obtained as the first centre. This technique is applied in the FASTCLUS method of SAS [13, 22].

Maximin technique [30] selects the first centre c1 randomly and the i-th (i{2,3,...,K}) centre ci is selected to be the point that has the most minimum distance to the formerly chosen centres, that is c1,c2,...,ci-1. This technique was originally expanded as an approximation to the K-centre clustering problem. It should be referred that, motivated with a vector quantization request, Katsavounidis et al.'s variant [30] obtains the point with the greatest Euclidean standard as the first centre.

Al-Daoud's density technique [1] first regularly partitions the data space into M decomposed hyper-cubes. It then randomly chooses K Nm/N points as of hypercube m (m {1,2,…,M}) to take a total of K centres where Nm is the points number in hypercube m. Bradley and Fayyad's technique [7] begins by randomly partitioning the data set into J subsets. These subsets are clustered by k-means initialized through MacQueen's second technique producing J sets of intermediate centres, each with K points. These centre sets are united into a superset that is then clustered through k-means J times, each time initialized by a diverse centre set. Members of the centre set that give the least SSE are then taken as the final centres. Pizzuti [40] advanced Al-Daoud's density-based technique using a solution grid method. This technique begins through 2D hypercube and iteratively divides these as the number of points they accept increases. The k-means++ technique [3] interpolates between maximin technique and MacQueen's second technique. It selects the first centre randomly and the i-th (i{2,3,...,K}) centre is selected to be x, where md(x) denotes the distance of minimum from a point x to the previously chosen centres.

The PCA-Part technique [44] applies a divisive hierarchical system based on PCA (Principal Component Analysis). In this method, starting from a first cluster that contains the all data set, the technique iteratively chooses the cluster with the greatest SSE and divides it into two sub-clusters by a hyper-plane that it passes with the centre of cluster and is orthogonal to the way of the basic eigenvector of the covariance matrix. This method is repeated until K clusters are taken. The centres are then given through the centres of these clusters. Lu et al.'s technique [35] applies a two phase pyramidal method. The attributes of each point are first encoded as integers. These points of integer are considered to be at stage 0 of the pyramid. In the phase of bottom-up, starting from stage 0, adjacent data points at stage k (k{0,1,...}) are averaged to take weighted points at stage k+1 until at least 20 K points are taken. Onoda technique [39] first computes K Independent Components (ICs) [21] of X and then selects the i-th (i{1,2,...,K}) centre as the point that has the least cosine distance [13].

3 K-means clustering algorithm

The aim of data clustering, also known as cluster analysis, is to discover the normal grouping of a set of points, objects or patterns. The Merriam-Webster dictionary defines cluster analysis as ‘‘a statistical classification method for discovering whether the individuals of a population fall into different groups by making quantitative comparisons of multiple characteristics.” The goal is to develop a clustering algorithm that will find the normal groupings in the data of unlabelled objects [23]. Cluster analysis or clustering is a method of assigning a set of data objects into clusters where all the objects in a cluster are considered to be similar based on common features. Clustering is an unsupervised learning-based technique for statistical data analysis used in many fields including data mining, pattern recognition, image analysis, and bioinformatics [8]. Selecting clusters of optimally is an NP-hard problem [48]. Clustering algorithms include many algorithms, and K-means algorithm is the most popular. -means algorithm is a rather simple but well-known algorithm for grouping objects [29]. This algorithm is so well known and has widely applied that researchers consider it the equivalent of clustering algorithms.

The term "K-Means" was first used by James MacQueen in 1967, though the idea originates with Hugo Steinhaus in 1956. A standard algorithm was first proposed by Stuart Lloyd in 1957 as a technique for pulse-

code modulation, though it was not published until 1982. The classical K-means clustering algorithm aims to detect a set C of K clusters Cj with cluster mean cj to reduce the sum of squared errors. Number of clustering C is a very important parameter [20].

The K-means algorithm is a greedy algorithm, which can only converge to a local minimum, even though recent study has exposed the enormous possibility that K-means could converge to the overall optimum when clusters are well detached [27,37]. K-means begins with a primary partition with K clusters and allocates patterns to clusters so as to decrease the squared error.

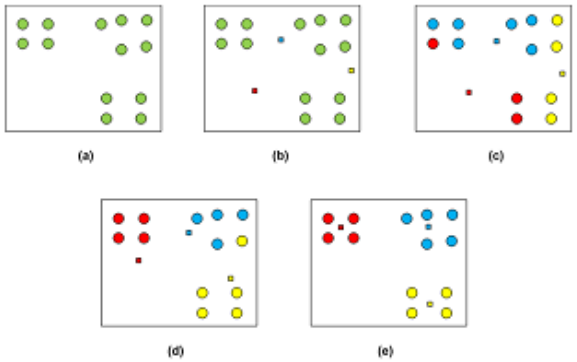


FIGURE 1 K-means clustering algorithm for 3 clusters

The Figure 1 expresses an illustration of the standard K-means algorithm on a dataset of two-dimensional with three clusters. Figure 1, sets out a design for a K-means clustering algorithm. (a) Two-dimensional input data with three clusters; (b) three seed points chosen as cluster centres and initial assignment of the data points to clusters; (c) & (d) intermediate iterations updating cluster label and the centres; (e) final clustering obtained by K-means clustering algorithm at convergence [23]. Clustering algorithms have many applications, but there are problems with this algorithm. One problem is that the clustering algorithm and K-mean algorithm are not always constant. The clustering algorithm may generate correct answers several times in some trials but in some other trials it may not find the correct answers due to instability. In order for the clustering algorithm to stabilize, it must reduce the number of errors and the number of iteration steps in the algorithm. Therefore, the proposed method tries to reduce the error rate and iteration in the K-means algorithm.

4 The proposed method

In the method used in this study, numbers were taken from outside of the cluster centre. Data sets were added and the K-means clustering algorithm was implemented. In this study is calculated the distance in the clustering algorithm to determine the best method. When the data distance was calculated correctly, cluster errors in the algorithm were reduced. The principal goal of the research methodology used in this study was error reduction in the K-means algorithm. In this section, the initialization of the proposed method is first checked. Then, the improved K-means clustering algorithm is expressed. Last, the problem formulation and proportion is expressed.

4.1 INITIALIZING THE CENTRE

In this study, initial value is randomly selected, after the data set was applied as cluster centres are selected randomly in the initial stage. All members of the dataset attributes must be an integer. A set of datasets was generated using MATLAB for testing the effect of various parameters and size of the problem on the time taken through the algorithm. By selecting the required number of cluster centres randomly in the domain [1, number of rows], which chosen randomly is a normal distribution. First, the dataset is applied to MATLAB. If the dataset format is more usable in MATLAB, it must be converted to the format used. Text format for the datasets have been used in this study. Second, the number of rows in the dataset is determined and then the number of clusters is selected as random numbers from 1 to the number of rows. For example, if the number of clusters is three and number of rows in the dataset is 150, three random numbers from 1 to 150 will be selected. So, selected attributes of these rows are initial cluster centres.

4.2 PROPOSED ALGORITHM

In the proposed algorithm (Reduction of Error Rates in K-Means algorithm or RER-K-Means algorithm), equation is used to calculate the distance between the members of dataset and cluster centres. Another difference between RER-K-means algorithm and K-means algorithm is that comparing and finding the minimum distance used a better method, which it is described more in the next section. Actually, equation is a compatibility function that would calculate and minimize the intra cluster distance. The equation has K clusters of N data vectors classified according to the distance from each cluster centre; it is located at one of the clusters. In this equation, the total aggregate of Euclidean distance of all the data vectors from cluster centres that they own is calculated and added to each other.

Therefore, by determining the optimal, centres can easily be clustered and the answer is that one that is best clustered. Using the equation, the number of clustering errors is reduced and it is close to being stable. It follows that the main objective of this equation is to ensure that minimum distances between the centres of the clusters are optimized, till, K-means clustering algorithm is to be improved. This study calculated the distance between the centre of the cluster and the cluster members using one of the best ways to calculate distance, which is the Euclidean function in MATLAB. Also, calculations of the distance between centres of the cluster and the cluster members are eliminated as additional unnecessary operations have a negative impact on the calculation. Accordingly, the proposed algorithm will be clustered; cluster members will

be assigned to data sets, reducing the error rate and stabilizing clustering algorithms.

4.3 PROBLEM FORMULATION

The RER-K-means clustering algorithm is further described in this section. The programming code was written using MATLAB software. A coding program was used to reduce the complexity of the algorithm, and the best method for clustering data was calculated in the K-mean algorithm. In the following algorithm, the RER-K-means clustering algorithm that was implemented in MATLAB is shown. In algorithm 2, the RER-K-means clustering algorithm is described, which the Euclidean method was used to calculate the distance between clusters. The RER-K-means clustering algorithm (Reduction of Error Rates in K-means clustering algorithm) has eleven stages, which are described below.

Step 1: At first, the target dataset is applied to the MATLAB software. The dataset must have the clustering conditions.

Step 2: In this step, the number of rows of the dataset is found followed by selecting desired numbers of rows randomly as cluster centres. The selected attributes of the random rows are assumed to be initial cluster centres.

Step 3: Specifying the number of iterations, it is considered 50 steps for all datasets in this study. All main processes were placed into this loop. This is known as the named outer loop.

Step 4: A loop is created for the first to the last dataset in which all the main instructions can be placed. This loop is the inter loop.

Step 5: At this stage, the distances of cluster centres which have been previously considered from all members of the dataset are calculated. To calculate the distance, the coordinates of the cluster centre in one array and attributes of a row as dataset in another array are placed, and then the distance between these two arrays is calculated using the following formula. This operation is carried out for all cluster centres in one step.

Step 6: In this step, the distances of all cluster centres from one of the datasets are calculated separately and the minimum distance is taken into consideration. Now, members of datasets are placed in the cluster with the minimum distance.

Step 7: In this step, some variables are defined to represent summation of distances between cluster centre and its members. The number of define variables should be equal to the number of clusters. For instance, if there are 3 clusters, three variables s1, s2 and s3 are defined in which si is summation of distances among ith cluster centre to its member. (i=1,2,3).

Step 8: This step is the end of inter loop. It means that steps 4 to 7 are run until the ending condition of inter loop.

Step 9: Variable S which is intra cluster distance is defined as summation of s1, s2, s3and so on. From converging of S it is deducted that algorithm has stabilized. Generally, S should be tried to minimized as far as possible.

Step 10: The means of any cluster should be determined separately. Then, at the end of any step, the determined means are considered as cluster centres for the next step.

Step 11: This step is the end of outer loop. It means steps 3 to 8 are run until the ending condition of outer loop.

4.4 THE FORMULA

The steps shown in section 4.3 were used to calculate the error rate. It was required to calculate two measures; the number of error patterns and the total number of patterns, which was used to find the error rate in the improved K-means clustering algorithm and the K-means clustering algorithm in all data sets of this study. In next section, it will be seen that the RER-K-means algorithm reduced the error rate and iteration. In this algorithm, additional operations that have a negative effect on the calculation must be avoided. In all the data sets, the K-means clustering and the RER-K-means algorithms implementation were similar and only the data set name and data set coordinates were changed by the algorithms.

5 Experimental results

The clustering results are compared with K-means and improved K-means algorithm. These are implemented with the number of clusters as equal to the number of classes. Meanwhile, the number of data sets selected to solve the problem in the next section can be fully expressed. To check the results, two important criterions are used to error rates and iteration of running.

5.1 DATA SET

Experiments have been performed on seven data sets which consist of Balance, Blood, Breast, Glass, Iris, Pima and Wine that were selected from standard data set UCI. Each of them is described in the following:

Balance Scales (Balance): Balance Scale data set is composed of 625 instances, 4 attributes and 3 classes. Each example is classified as having the balance scale tip to the right, tip to the left, or balanced. Balance dataset contains 46.08% of class L, 7.84% of class B and 46.08% of class R.

Blood Transfusion Service Centre (Blood): This data set adopted the donor database of Blood Transfusion Service Centre in Hsin-Chu City in Taiwan. Blood data set has 748 samples which are 748 donors selected at random from the donor database. This data set has 5 attributes which include R (Recency - months since last donation), F (Frequency - total number of donations), M (Monetary - total blood donated in c.c.), T (Time - months since first donation), and a binary variable representing whether donor donated blood in March 2007 (1 stand for donating blood; 0 stands for not donating blood). The dataset contained 76% no (0) and 24% yes (1). Breast Cancer Wisconsin, Original (Breast): The Breast Cancer

Wisconsin dataset has 699 instances of cytological analysis of fine needle aspiration of breast tumors. In this data set each instance contains 10 attributes that are computed from a digitized image of a fine needle aspiration of a breast mass. Attributes of this data set include radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. The dataset contains 241 (34.48%) malignant instances and 458 (65.52%) benign instances [18].

Glass Identification (Glass): This data set has 214 samples and seven classes. Every sample in this data set has 10 attributes. Seven kinds of glass are in the data sets including building windows float, building windows non-float, vehicle windows float, vehicle windows non-float, containers, tableware and headlamps.

Iris: This data set is based on Iris flowers recognition with three different classes each consisting of 50 samples. Every sample has four attributes. It presents 150 instances containing width and length measures of the sepals and petals of three species of the flower Iris: 'Setosa', 'Versicolor' and 'Virginical'. With 4 attributes and 3 classes, each containing 50 objects, the aim is to cluster similar species based on their measurements [18, 49].

Pima: This data set is allocated to recognize diabetic patients. A total of 768 samples are classified into two groups consisting of 500 and 268 samples, respectively. Every sample in this data set has 8 attributes [49].

Wine: The Wine dataset has 178 instances and 13 attributes, which correspond to the results of chemical analyses performed with three types of wines produced in the same region of Italy, but from different cultivations. Attributes include alcohol content, acidity, alkalinity, color intensity, among others. The dataset has 59 instances of the first class, 71 instances of the second class and 48 instances of the third [18]. The databases used were obtained from the UCI data warehouse [4].

The relevant datasets are implemented in the clustering algorithm and proposed algorithm and compared in depth. In this study, two measures are used to compare the names of the error rates and number of iterations. In the next section, these two measures will be discussed above data sets.

5.2 RESULTS OF ERROR RATE

In this section, results concerning the number of errors of the proposed K-means clustering algorithm on the data sets are reviewed. In the previous section, it was noted that in this study, seven data sets have been selected to analyse the proposed K-means algorithm. These data sets are standard and are selected from UCI data sets. The proposed K-means algorithm is applied to the respective data sets to determine the results; the number of errors and the graphs and charts can then be fully expressed. In this study, our main objective is to improve the K-means clustering algorithms. To verify the improved algorithm, the improved clustering algorithm will be tested on the some data sets to answer the question of whether this algorithm is improved or not. Thus, the seven data sets (Balance, Blood, Breast, Glass, Iris, Pima and Wine) are implemented separately in the MATLAB software of the proposed algorithm and the results are discussed in this study. In Figure 2, clustering of seven clusters in Glass data set with improved K-means clustering algorithm is displayed. This data set has seven regular clusters, indicating that the clusters are not merged.

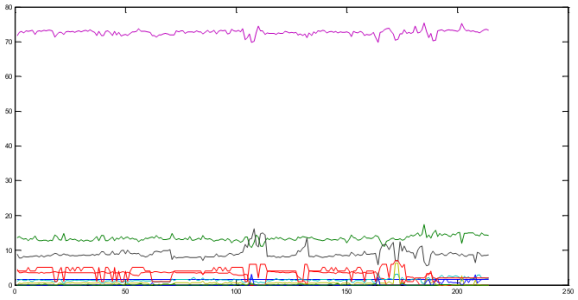


FIGURE 2 Display clustering the Glass data set with improved K-means algorithm

The proposed K-means algorithm was first applied to the Balance dataset. The specifications of this data set are described in the previous section, but will be mentioned briefly here. The Balance data set has 625 instances and 3 classes. In the MATLAB software, code programming proposed K-means algorithm is implemented and Balance data set is loaded.

Secondly, the proposed K-means algorithm was applied to the Blood dataset. The specifications of this data set are described in the previous section, but will be mentioned here briefly. The Blood data set has 748 instances and 2 classes. In the MATLAB software, code programming proposed K-means algorithm is implemented and Blood data set is loaded. The scattering diagram shows improved K-means clustering algorithm on the Blood data set. This diagram indicates the 748 members and the distance among members in this data set. In the diagram it is shown that a small number of members are scattered, mostly in the one level. In this data set are two clusters that are not regular. This means that the first and second clusters are merged; the Blood data set is such that the first and second clusters are not completely separated. In Figure 3, clustering of three clusters in Iris data set with improved K-means clustering algorithm is displayed. This data set has three clusters (first cluster, the first to fifty members; second cluster, members fifty one to one hundred; third cluster, members one hundred one to one hundred fifty). It can be seen that, in the first cluster, there are no errors after clustering, which means an error rate for the first cluster of zero. In the second cluster, there is a small error rate, but in the third cluster, the error rate is higher than both previous clusters. In general, the graphs display the clustering in the Iris data set using improved K-mean clustering algorithm.

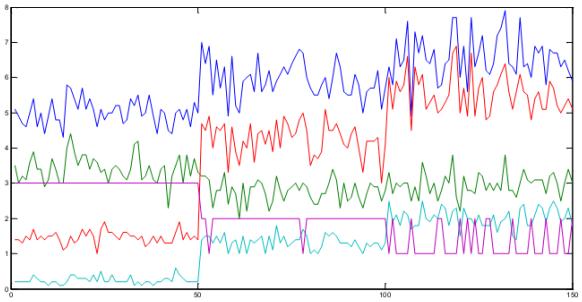


FIGURE 3 Display clustering the Iris data set with improved K-means algorithm

The proposed K-means algorithm was applied to the Breast dataset. The specifications of this dataset are described in the previous section but will, however, be mentioned briefly here. The Breast data set has 699 instances and 2 classes. In the MATLAB software, code programming the proposed K-means algorithm is implemented and Breast data set is loaded. The scattering diagram is shows an improved K-means clustering algorithm on the Breast data set. This diagram indicates the 699 members and the distance among members in this data set. Two clusters of Breast data set clustered with improved K-means clustering algorithm are displayed. In this data set are two clusters that are not regular. This indicates that the first and second clusters are merged; the Breast data set is such that the first and second clusters are not separated. Each algorithm was run twenty times and each time, the algorithm was repeated 50 times to achieve stability. In general, the algorithm is executed 1000 times for each data set. Five factors (number of true, number of errors, intra cluster distance, iteration and error rate) are compared to the improved K-means clustering algorithm and K-means clustering algorithm on the Breast data set. In all factors, the proposed algorithm is much better than previous algorithms.

5.3 EVALUATION OF RESULT

In this section the results of the experiments conducted in section 4 are compared to the results of the improved K-means clustering algorithm and the K-means clustering algorithm. One important factor for the clustering algorithm is intra cluster distance that will be reviewed first. For better comparison, both algorithms are run 20 times on all data sets. All diagrams that can be seen in the intra cluster distance of the proposed clustering algorithm have been improved and in all cases the intra cluster distance is reduced. Also, the intra cluster distance is constant during program execution, indicating the algorithm is stable. It can be seen that the proposed algorithm 20 times the intra cluster distance, meaning that the algorithm is stable. One of the main problems in the K-means clustering algorithm is stability, which the proposed algorithm has almost solved.

In this section, the results of the experiments obtained in section 4 are discussed and evaluated. The results were discussed with the three criteria, intra cluster distance (average), intra cluster distance (standard deviation) and error rate (average) for seven data sets. In the three comparison criteria between the improved K-means clustering algorithms and K-means clustering algorithm it can be seen that the improved K-means clustering algorithm is the best in each case. In general, the proposed algorithm reduces the error rate and intra cluster distance, and it will lead the clustering algorithm to become stable.

6 Conclusions

This paper focuses on a disadvantage of the K-means clustering algorithm, which is that the clustering algorithm has a high error rate. It also referred to one of the main problems in the K-means algorithm which is that the K-means clustering algorithm is not always stable. In this study, an algorithm was proposed to solve this problem in order to reduce the error rate in K-means clustering algorithms and to stabilize the algorithm. In this paper, examining the improved K-means clustering algorithm with the K-means clustering algorithm involved the consideration of five factors (average, standard deviation, best and worst) and four criteria (numbers of true, numbers of errors, intra-cluster distance, iterations and error rate). For comparing the improved K-means algorithm and K-means algorithm seven data sets were used (Balance, Blood, Breast, Glass, Iris, Pima and Wine) and the proposed algorithm shows better performance in all these data sets. In summary, the proposed algorithm has better efficiency than the K-means clustering algorithm in the all measures used in this study, the intra cluster distance and error rate was reduced in the proposed algorithm and the improved algorithm is closer to being stable. In this study a method is proposed to solve one of the main problems of K-means clustering algorithm, which is that the algorithm is not always consistent. In this survey, one of the best ways to calculate the distance it to use the Euclidean distance calculation in the MATLAB software to calculate the members distance from the cluster centre. Also, it should be noted that when calculating the Euclidean distance, additional operations that have a negative effect on the calculation must be avoided. Thus, the purpose of this paper is to improve the calculation of the members distance from the centre of the cluster, which this will help to stabilize algorithm clustering and reduce the error rate. Future work related to this paper can be done as a continuation as other problems of clustering algorithms can be studied using these data sets. Also, the proposed algorithm in this paper can be examined on other data sets and clustering algorithms and the obtained results compared. Finally, other criteria can be studied with the proposed algorithm on the new data set and data sets in this article.

References

[1] Al-Daoud M d B, Roberts S A 1996 New methods for the initialisation of clusters *Pattern Recognition Letters* **17** 451-5
[2] Anderberg M R 1973 Cluster analysis for applications *DTIC Document*
[3] Arthur D, Vassilvitskii S 2007 K-means++ The advantages of careful seeding *in Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* 1027-35
[4] Asuncion A, Newman D 2007 UCI Machine Learning Repository *University of California School of Information and Computer Science Irvine CA* ed 24
[5] Ball G H, Hall D J 1967 A clustering technique for summarizing multivariate data *Behavioral science* **12** 153-5
[6] Bayat F, et al. 2010 A non-parametric heuristic algorithm for convex and non-convex data clustering based on equipotential surfaces *Expert Systems with Applications* **37** 3318-25
[7] Bradley P S, Fayyad U M 1998 Refining Initial Points for K-Means Clustering *in ICML* 91-9
[8] Chang D, et al. 2012 A genetic clustering algorithm using a message-based similarity measure *Expert Systems with Applications* **39** 2194-202
[9] Chau M, et al. 2005 Uncertain data mining: a new research direction *in Proceedings of the Workshop on the Sciences of the Artificial Hualien Taiwan* 199-204
[10] Cormack R M 1971 A review of classification *Journal of the Royal Statistical Society Series A (General)* 321-67
[11] Cox D R 1957 Note on grouping *Journal of the American Statistical Association* **52** 543-7
[12] Duda R O, et al. 2001 Pattern classification *2nd Edition. New York*
[13] Emre C M, et al. 2012 A comparative study of efficient initialization methods for the K-means clustering algorithm *Expert Systems with Applications*
[14] Engelman] L, Hartigan J A 1969 Percentage points of a test for clusters *Journal of the American Statistical Association* **64** 1647-8
[15] Fisher W D 1958 On grouping for maximum homogeneity *Journal of the American Statistical Association* **53** 789-98
[16] Forgy E W 1965 Cluster analysis of multivariate data: efficiency versus interpretability of classifications *Biometrics* **21** 768-9
[17] Gordon A 1999 Classification. 1999 *Chapman&Hall CRC Boca Raton FL*
[18] Gorgônio F L, Costa J A F PartSOM A Framework for Distributed Data Clustering Using SOM and K-Means
[19] Hastie T 2001 The elements of statistical learning *Springer New York*
[20] Huang H, et al. 2013 Adaptive Correction Forecasting Approach for Urban Traffic Flow Based on Fuzzy-Mean Clustering and Advanced Neural Network *Journal of Applied Mathematics*
[21] Hyvarinen A 1999 Fast and robust fixed-point algorithms for independent component analysis *Neural Networks IEEE Transactions on* **10** 626-34
[22] S. Institute and P. S. Publishing, SAS/STAT 9.2 User's Guide The Glimmix Procedure (Book Excerpt) SAS Institute 2008
[23] Jain A K 2010 Data clustering: 50 years beyond K-means *Pattern Recognition Letters* **31** 651-66
[24] Jain A K, Dubes R C 1988 Algorithms for clustering data *Prentice-Hall Inc*
[25] Jancey R 1966 Multidimensional group analysis *Australian Journal of Botany* **14** 127-30
[26] Jiang D, Tang C, Zhang A 2004 *Knowledge and Data Engineering, IEEE Transactions on* **16**(11) 1370-86
[27] Jiawei H, Kamber M 2001 Data mining: concepts and techniques *San Francisco, CA, itd: Morgan Kaufmann* **5**
[28] Johnson R A, Wichern D W 2002 Applied multivariate statistical analysis *Prentice hall Upper Saddle River, NJ* **5**
[29] Ju C, Xu C 2013 A New Collaborative Recommendation Approach Based on Users Clustering Using Artificial Bee Colony Algorithm *The Scientific World Journal*
[30] Katsavounidis I, Kuo J C-C, Zhang Z 1994 *Signal Processing Letters IEEE* **1**(10) 144-6
[31] Kogan J, et al. 2006 Grouping multidimensional data *Springer*
[32] Lattin J M, et al. 2003 Analyzing multivariate data *Thomson Brooks/Cole Pacific Grove, CA*
[33] Leuski A 2001 Evaluating document clustering for interactive information retrieval *in Proceedings of the tenth international conference on Information and knowledge management* 33-40
[34] Lloyd S 1982 *Information Theory, IEEE Transactions on* **28**(2) 129-37
[35] Lu J 2008 Hierarchical initialization approach for K-Means clustering *Pattern Recognition Letters* **29** 787-95
[36] MacQueen J 1967 Some methods for classification and analysis of multivariate observations *in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 14
[37] Meilă M 2006 The uniqueness of a good optimum for K-means *in Proceedings of the 23rd international conference on Machine learning* 625-32
[38] Norusis M J 2012 IBM SPSS statistics 19 statistical procedures companion *Prentice Hall*
[39] Onoda T 2012 Careful Seeding Method based on Independent Components Analysis for K-means Clustering *Journal of Emerging Technologies in Web Intelligence* **4** 51-9
[40] Pizzuti C 19999 A divisive initialisation method for clustering algorithms *in Principles of Data Mining and Knowledge Discovery ed Springer* 484-91
[41] Sebestyen G S 1962 Decision-making processes in pattern recognition (*ACM monograph series*)
[42] Steinhaus H 1956 Sur la division des corp materiels en parties *Bull. Acad. Polon. Sci* **1** 801-4
[43] Steinley D 2006 K-means clustering: A half-century synthesis *British Journal of Mathematical and Statistical Psychology* **59** 1-34
[44] Su T, Dy J G 2007 In search of deterministic methods for initializing K-means and Gaussian mixture clustering *Intelligent Data Analysis* **11** 319-38
[45] Thorndike R L 1953 Who belongs in the family? *Psychometrika* **18** 267-76
[46] Timm N H 2002 Applied multivariate analysis *Springer*
[47] Tou J T, Gonzalez R C 1974 Pattern recognition principles
[48] Wang T, Hung W N 2013 Reliable Node Clustering for Mobile Ad Hoc Networks *Journal of Applied Mathematics*
[49] Yazdani D, et al. 2010 A new hybrid approach for data clustering *in Telecommunications (IST), 2010 5th International Symposium on* 914-9

Author

Jinguo Zhao, born in June, 1965, Shaodong, Hunan Province, China

Current position, grades: associate professor in School of Computer and Information Science, Hunan Institute of Technology, China.
University studies: Database.
Scientific interest: semantic web and database.
Publications: 15 papers.

Research on transition priorities in group based on CPN

Hong Wang[1, 2, 3*], Tao Zhang[2]

[1]*Academy of OPTO-Electronics, Chinese Academy of Sciences, Deng Zhuang Rd. 9, Distinct Haidian, Beijing, China*

[2]*Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Deng Zhuang Rd. 9, Distinct Haidian, Beijing, China*

[3]*University of Chinese Academy of Sciences, Yuquan Rd. 19, Shijingshan District, Beijing, China*

Received 1 March 2014, www cmnt.lv

Abstract

Transition priorities might be a useful mechanism when modelling using Petri nets. Accordingly, the newest CPN Tools, widely used for modelling and simulating the Coloured Petri net, implements transition priorities. Whereas, the algorithms compute enabling for all transitions in a highest-priority-first order. In the real system, it is usually that there are priorities relationships not for all transitions but only some ones. Based on the above analysis, this paper put all the transitions, having priority relations, into one group and advances relative theoretical definitions of transition priorities in group, such as absolute preset of transition, key place set, key colour set, etc. Furthermore, it proposes new algorithms when the systems have different key place set and key colour set, and construct the model of the interrupt priorities to solve the problem of software model checking for interrupt system.

Keywords: Petri net, CPN, transitions, transition priorities in group, counter place

1 Introduction

In the process of development of Petri network, in order to better describe realistic system, researchers advanced various high-level Petri net models, such as: time Petri nets [1,2], Stochastic Petri nets [3,4], Colored Petri nets [5,6], Priorities Petri nets [7], Hybrid Petri nets [8], and so on. These extended concepts enhanced Petri net description of capabilities, furthermore, modelling various Petri net had been widely applied in the synchronous system, asynchronous system, simulation [9, 10] and analysis [11, 12] process. However, the actual applications need a better tool to simulate and verity generated models. CPN Tools widen the usage of Petri nets in practical work [13-15].

CPN Tools version 4.0 [16], supporting transition priorities, could specify priority P_HIGH, P_NORMAL or P_LOW [17] for transitions, and could also directly give an integer. Nevertheless, this kind of priorities is static and absolute priority for all transitions, precedence relations, in the actual system, exist among not all transitions but some of them.

Reference [18] proposed to increase one Anti-Place to address the transition priorities, shown in FIGURE 1. The Anti-Place is effective, only if the place had tokens in initial state, however, most systems could not meet the mentioned requirements. Reference [19] used the concept of generalized complement position (Place), changing Petri net system with dynamic priority into the Petri net system with non-priority, but the algorithm was too complex even for constructing a simple system.

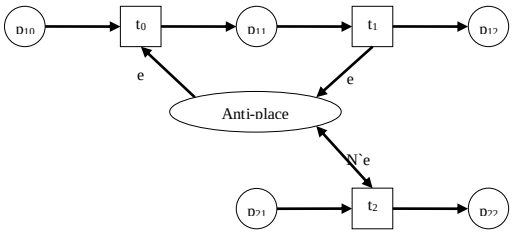


FIGURE 1 Anti-place for transitions with priorities

The remainder of this paper is structured as follows: in the next section, it presented background material and in Section III it provides the definitions of transition priorities in group, key place set and key colour set, etc., and In Section IV, it presents algorithms for increasing one "counter place" to calculate the change of tokens in key place set. Furthermore, the next section, the article solve the practical work interrupt relational model by using the above definitions and algorithms. In Section VI, it concludes and provides directions for future works.

2 Petry net and coloured Petri net

2.1. PETRI NET (PLACE/TRANSITION NET OR P/T NET).

A Petri net is one of several mathematical modelling languages for the description of concurrent systems. A Petri net is a directed bipartite graph, in which the nodes represent transitions and places. The directed arcs described which places are pre- and/or post- conditions, for which transitions (signified by arrows).

***Corresponding author** e-mail: neiep_wh@163.com

A Petri net is a 5-triple $PN=(P,T,F,W,M_0)$ [20], where:

(i) $P=\{p_1,p_2,p_3,...,p_n\}$ is a finite set of places.

(ii) $T=\{t_1,t_2,t_3,...,t_n\}$ is a finite set of transitions.

(iii) $F\subseteq(P\times T)\cup(T\times P)$ is a set of arcs (flow relation).

(iv) $W:F\to\{1,2,3,...\}$ is a weight initial marking.

(v) $M_0:P\to\{1,2,3,...\}$ is the initial marking.

$P\cap T=\Phi$ and $P\cup T\neq\Phi$.

The **preset** of a transition or place x is the set of its input places or transition: ${}^\bullet x=\{y\mid(y,x)\in A\}$; its **postset** is the set of its output places or transition: $x^\bullet=\{y\mid(x,y)\in A\}$.

The **preset** of a set $X_1(X1\subseteq P\vee X_1\subseteq T)$ is the set of its input sets, ${}^\bullet X_1=\bigcup_{x\in X_1}{}^\bullet x$; its **postset** is the set of its output sets $X_1^\bullet=\bigcup_{x\in X_1}x^\bullet$.

However, basic Petri nets have limitations. One of the limitations is 'homogenous' tokens: the tokens inside place represent resources; then in Petri nets, all these resources are of the same type and cannot be differentiated. Another limitation of Petri net is that it is not possible to impose additional logical functions ('firing conditions') for a transition.

2.2 COLOURED PETRI NETS (CP-NETS OR CPN)

Coloured Petri nets is one of popular extensions of basic Petri net. In a standard Petri net, tokens are indistinguishable. Nevertheless, in a Coloured Petri Net, each token has a value, which could be descripted with colour set.

A CP-net is a 9-tuple $CPN=(\Sigma,P,T,A,N,C,G,E,I)$ [21], where:

(i) Σ is a finite set of non-empty types, also called colour sets;

(ii) P is a finite set of places;

(iii) T is a finite set of transitions;

(iv) A is a finite set of arcs such that:

$P\cap T=P\cap N=T\cap A=\Phi$;

(v) N is a node function. It is defined from A into $P\times T\cup T\times P$;

(vi) C is a colour function. It is defined from P into E;

(vii) G is a guard function. It is defined from T into expressions such that:

$\forall t\in T:[Type(G(t))=Boolean\wedge Type(Var(G(t)))\subseteq\Sigma]$.

(viii) E is an arc expression function. It is defined from A into expressions such that:

$\forall a\in A:[Type(E(a))=C(p)_{MS}\wedge Type(Var(E(a)))\subseteq\Sigma]$,

where p is the place of $N(a)$;

(ix) I is an initialisation function. It is defined from P into closed expressions.

Such that:

$\forall p\in P:[Type(I(p))=C(p)_{MS}]$.

In popular tools for Coloured Petri nets, such as CPN Tools, the values of tokens are typed, and could be tested (using guard functions) and manipulated with a functional programming language.

3 Definitions of transitions priorities in group

In order to solve transitions priorities in group, the relative definitions listed as follows.

3.1. $P_S(t)$ (TRANSITION PRIORITIES IN GROUP)

When one transition t is enabled, there is another transition t_k, $(t_k\neq t)$ could also be enabled, but they have different priorities. A group set is composed of all transitions, which could be enabled at the same time and have different priorities. The group set $S=\{t,t_k\}$, the transition propriety of t in group set S is called $P_s(t)$. Similarly, the transition propriety of t_k in group set S is $P_s(t_k)$.

3.2 T_p (GROUP OF TRANSITIONS WITH PRIORITIES)

When multiple transitions could be enabled at the same time, there is relationship of priorities among transitions in the group. The group is an ordered collection of the sort, according to the priorities from high to low. Defined as follows:

A CP-net with Transitions priorities is a 10-tuple $CPN^*=(\Sigma,P,T,A,N,C,G,E,I,S)$, where:

S *is* a finite set of group of enabled transitions with priorities. For each, $T_p\in S$,

(i) $T_p\subseteq T, and\mid T_p\mid>1$;

(ii) $\forall t_i,t_j\in T_p,(j>i)$, transition t_i and t_j could be enabled at the same time, and have different priorities. The priorities of t_i in group T_p is $P_{T_p}(t_i)$, and t_j is

$P_{T_p}(t_j),\ P_{T_p}(t_i)\geq P_{T_p}(t_j)$;

(iii) In group of transitions with priorities $T_p=\{t_1,t_2,t_3,...\}$, at least one t_i, making $P_{T_p}(t_i)>P_{T_p}(t_{i+1})$.

3.3 $[{}^{\bullet}t_i]$ (ABSOLUTE PRESET OF TRANSITION t_i)

Absolute preset of transition t_i is composed of the places, which are elements of preset of t_i, but not be preset of other transitions in the same group T_p.

$$[{}^{\bullet}t_i] = \{p \mid p \in {}^{\bullet}t_i \wedge (\forall t_k \in T_p (k \neq i), p \notin {}^{\bullet}t_k)\}. \quad (1)$$

3.4 P_k (KEY PLACE SET)

P_k (Key Place Set) is a set of places, effecting transition priorities in same group. Search algorithm of P_k as follows:

STEP 1: Identify Group of Transitions with Priorities T_p.

STEP 2: Find Key Place Set P_k according to the following Equation:

$$P_k = \{\bigcup_{i=1}^{n} {}^{\bullet}t_i - \bigcap_{i=1}^{n} {}^{\bullet}t_i \mid t_i \in T_p\} - [{}^{\bullet}t_n]. \quad (2)$$

Firstly, identify preset for each transition t_i, $(t_i \in T_p)$ in group T_p, then, removed public preset of them. Finally, expurgated the absolute preset of transition t_n (with the lowest priority in group T_p).

3.5 $C_k(p)$ (KEY COLOUR SET OF KEY PLACE P) AND $C_k(P_k)$ (KEY COLOUR SET OF KEY PLACE SET P_k)

$C_k(p)$ (Key Colour Set of Key Place p, for short Key Colour Set) is a multi-set of colour set of key place $p(p \in P_k)$. They effected the change of tokens in the key klace p.

Usually, if the key place p has simple colours set $C(p)$, the key colour set of p is same with $C(p)$. Otherwise, p has compound colour set, the key colour set is a multi-set, composed with elements in compound colour set.

$C_k(P_k)$ (Key Colour Set of Key Place Set P_k) is a multi-set of key colour set for each key place p in group P_k.

$$C_k(P_k) = \sum_{k=1}^{n} C_k(p). \quad (3)$$

Supposing that:

(i) The key place set is $P_k = \{p_1, p_2\}$.

(ii) The colour set of p_1, a simple colour set, is $C(p_1) = c_1$, and the key colour set of p_1 is $C_k(p_1) = c_1$. Length of $C_k(p_1)$ is $|C_k(p_1)| = 1$.

(iii) The colour set of p_2, a compound colour set $C(p_2) = c_1 \cdot c_2 \cdot c_3$, and the key colour set of p_2 is $C_k(p_2) = \{c_1, c_2\}$. Length of $C_k(p_2)$ is $|C_k(p_2)| = 2$.

Then, $C_k(P_k) = \{c_1, c_1, c_2\}$ is a multi-set, and the length of $C_k(P_k)$ is $|C_k(P_k)| = 3$.

3.6 ${}^{\bullet}\overline{p}$ (RELATIVE PRESET OF P) AND $\overline{p}^{\bullet}$ (RELATIVE POSTSET OF P)

p is a key place. The relative preset and postset of p must be set of transitions.

$$ {}^{\bullet}\overline{p} = \{t \mid t \in {}^{\bullet}p \wedge (Type(E(a_{in})) \cap C_k(p) \neq \Phi)\}, \quad a_{in} = (t, p) \in A \quad (4)$$

$$ \overline{p}^{\bullet} = \{t \mid t \in p^{\bullet} \wedge (Type(E(a_{out})) \cap C_k(p) \neq \Phi)\}, \quad a_{out} = (p, t) \in A \quad (5)$$

4 Algorithms of transitions priorities in group

4.1 ALGORITHM 1: MODEL FOR $|P_k| = 1$ AND $|C_k(P_k)| = 1$

Example 1: FIGURE 2 is a typical of concurrent systems with priority relationships. When t_1 and t_2 are enabled at the same time, t_1 has a higher priority. However, when t_{10} and t_1 are enabled simultaneously, they have no priorities, randomly trigger.

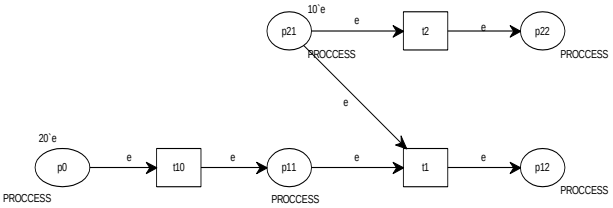


FIGURE 2 Source Model,with $|P_k| = 1$ and $|C_k(P_k)| = 1$

STEP 1: Identify group of transitions with priorities T_p.

In Figure 2, transition set is $T = \{t_1, t_2, t_{10}\}$. t_1 and t_2 have priorities between them, but have no priorities with other transitions. Furthermore, the priority of t_1 is higher than t_2. The group of enabled transitions with priorities is T_p, $T_p = \{t_1, t_2\}$ and $P_{T_p}(t_1) > P_{T_p}(t_2)$.

STEP 2: Identify P_k, the key place set of T_p, and $C_k(P_k)$, the key colour set of P_k.

$\bullet t_1 = \{p_{11}, p_{21}\}$,

$\bullet t_2 = \{p_{21}\}$,

$So, P_k = \{\bigcup_{i=1}^{2} \bullet t_i - \bigcap_{i=1}^{2} \bullet t_i \mid t_i \in T_p\} - [\bullet t_2] = \{p_{11}\}$,

$C_k(p_{11}) = \{PROCCESS\}$,

$C_k(P_k) = \{PROCCESS\}$.

STEP 3: Modify the CPN model, increasing one "counter place" to calculate tokens in key place set.

STEP 3.1: Increase variable for "counter place".

var n:INT;

STEP 3.2: Increase "counter place" for key place set $P_k = \{p_{11}\}$ in CP-net model, with definition of colour set and initial value.

Increase "counter place" *count*, with colorset *INT*. Because the initial count of the key colour PROCCESS is 0, the initial value of "counter place" *count* is 0.

STEP 3.3: If the relative preset of the key place p is not empty, added two arcs from "counter place" *count* to $\bullet\overline{p}$; and if the relative postset of the key place p is not empty, add two arcs from "counter place" *count* to $\overline{p}\bullet$. Added arcs describe the change of tokens in key place.

Transition t_{10} triggered, place p_{11}. Increase one *PROCCESS*, while t_1 triggered, place p_{11} decreased one *PROCCESS*. So, added two arcs from *count* to its relative preset $\bullet\overline{p_{11}} = \{t_{10}\}$, one is outcoming arc with function "n", the other is incoming arc with function "n+1". Similarly, added two arcs from *count* to its relative postset $\overline{p_{11}}\bullet = \{t_1\}$, one is outcoming arc with function "n", the other is incoming arc with function "n-1".

STEP 3.4: added one bi-directional arc with function "0", from "counter place" *count* to other transitions with lower priorities in the same group.

For $T_p = \{t_1, t_2\}$ in the model, extracted one bi-directional arc from *count* to t_2.

The finished model is shown in Figure 3.

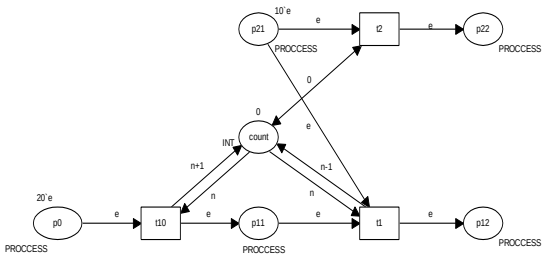


FIGURE 3 Model with priorities $|P_k| = 1$ and $|C_k(P_k)| = 1$

4.2 ALGORITHM 2: MODEL FOR $|P_k| \geq 1$ AND $|C_k(P_k)| > 1$

Example 2: Added two nodes, p_{21} and t_{20}, in FIGURE 2. Transition t_1 could enable, only when p_{11}, p_{21} and p_{31} have tokens in place. Furthermore, t_1 has a higher priority than t_3. However, t_1 and t_3 couldn't plunder privilege of other transitions, t_1 should have same priority with t_{10} and t_{20}, as shown in FIGURE 4.

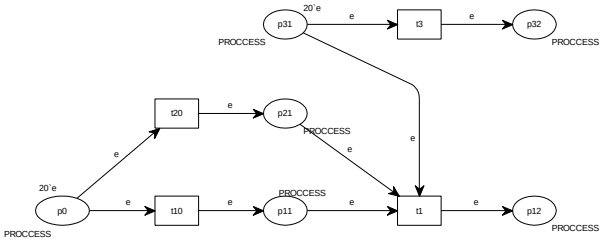


FIGURE 4 Source Model with $|P_k| \geq 1$ and $|C_k(P_k)| > 1$

STEP 1: Identify group of transitions with priorities T_p.

In Example 2, transition set is $T = \{t_1, t_3, t_{10}, t_{20}\}$. The priority of t_1 is higher than t_3. The group of transitions with priorities is T_p, $T_p = \{t_1, t_3\}$, and $P_{T_p}(t_1) > P_{T_p}(t_3)$.

STEP 2: Identify P_k, the key place set of T_p, and $C_k(P_k)$, the key colour set of P_k.

$\bullet t_1 = \{p_{11}, p_{21}, p_{31}\}$,

$\bullet t_3 = \{p_{31}\}$,

$P_k = \bullet t_1 \cup \bullet t_3 - \bullet t_1 \cap \bullet t_3 - [\bullet t_3] =$
$\{p_{11}, p_{21}, p_{32}\} - \{p_{31}\} - \Phi = \{p_{11}, p_{21}\}$.

Furthermore:

$C_k(p_{11}) = \{PROCCESS\}$,

$C_k(p_{21}) = \{PROCCESS\}$.

So: $C_k(P_k) = \{PROCCESS, PROCCESS\}$.

STEP 3: Modify the CPN model, Increase a "counter place" to calculate tokens in key place set.

STEP 3.1: Increase variables for "counter place" and trigger function for transition with lower priorities.

According to $C_k(P_k) = \{PROCCESS, PROCCESS\}$ and $|C_k(P_k)| = 2$, added a "counter place" with 2-tuples in the model for p_{11} and p_{21}. Defined variables and trigger function by CPN ML as follows:

```
//new colorset for "counter place" with two-tuples
    colset INTLIST = product INT * INT;
    var n1,n2:INT;
    fun tran(counts: INTLIST) =
```

```
let
  val count1 = #1 (counts);
  val count2 = #2 (counts);
in
//modify expression according actural condition
  not(count1 <>0 andalso count2 <>0)
end
```

STEP 3.2: Increase "counter place" for key place $P_k = \{p_{11}, p_{21}\}$ in CP-net model, with definition of colour set and initial value.

Increase "counter place" *count*, with colour set INTLIST. Because the initial count of the key colour PROCCESS in p_{11} and p_{21} are all 0, the initial value of "counter place" *count* is (0,0).

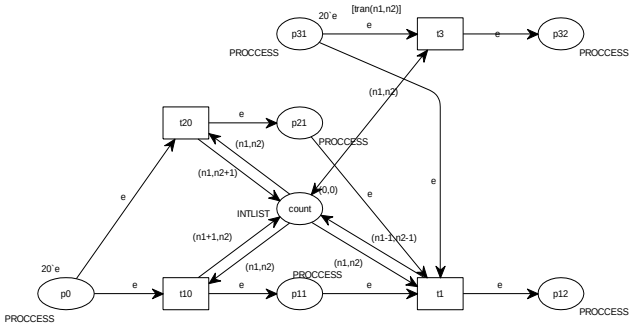


FIGURE 5 Model with priorities with $|P_k| \geq 1$ and $|C_s(P)| > 1$

STEP 3.3: Do as follows for each $\forall p \in P_k$: if the relative preset of the key place p is not empty, add two arcs from "counter place" *count* to ${}^{\bullet}\overline{p}$; and if the relative postset of the key place p is not empty, add two arcs from "counter place" *count* to $\overline{p}^{\bullet}$. Added arcs describe the change of tokens in key place *Count*, the "counter place", looks like the format (n_1, n_2) . And n_1 showed the change of p_{11} , n_2 records the change of p_{21} . Furthermore, changed in adding arcs and the arc functions (as shown in Figure 5).

STEP 3.4: If there is no outcoming arc from "counter place" *count* to other transitions with lower priorities in same group, added one bi-directional arc with function (n_1, n_2) .

STEP 3.5: Added trigger function in all transitions with lower priorities.

In this example, added trigger function $tran(n_1, n_2)$ for transition t_3 .

5 Practise example

Example 3: The systems have a main program Proccess, and two external interrupts $INTR_1$ and $INTR_2$, which have the priorities relationships that the priorities of $INTR_1$ is higher than $INTR_2$, namely $P_{INTR1} > P_{INTR2}$. The question is how to construct a CPN model to meet interrupts priorities.

The main process *Proccess* set interrupts on or off. The place *INTR* managed all the interrupts that is, when *Proccess* set the interrupt on, it send semaphore $intr_1$ to transition on_intr_1 in order to enable the active interrupt. Otherwise, when it set the interrupt off, it remove $intr_1$ from transition on_intr_1, thus disabled the interrupt, as shown in Figure 6.

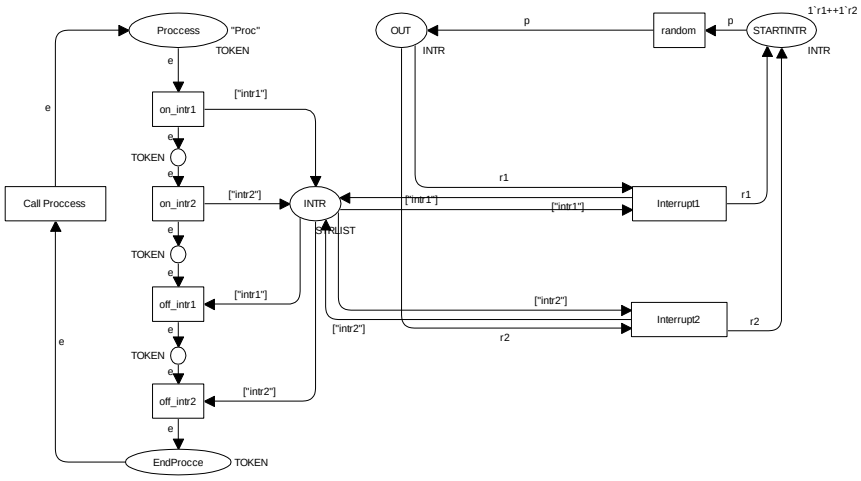


FIGURE 6 Interrupt model M_3 without priorites

The place *STARTINTR* simulated the external interrupts, the transition *random* control the happening of interrupts randomly, and the place *Out* represented the current enabled interrupts. The interrupt conditions are:

(i) $INTR_1$ had been enabled;

(ii) The existence of the external interrupt $INTR_1$, then $Interrupt_1$ could run. However, the model could be shown as FIGURE 6, although given the current disconnect, turn control to an external interrupt occurs, but not showing the priorities of relations between $INTR_1$ and $INTR_2$.

If omitted other unrelated contents, the interrupt model could be minimized, as shown in Figure 7.

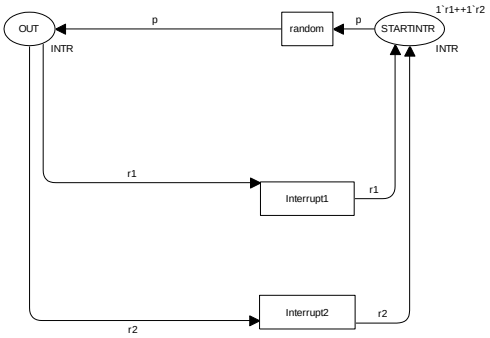


FIGURE 7 Minimized interrupt model M_4

The minimized interrupt Model M_4 use the above algorithm, as show as Section IV.

STEP 1: Identify group of transitions with priorities T_p .

In Figure 7, transition set is $T = \{random, Interrupt_1, Interrupt_3\}$. The group of enabled transitions with priorities is T_p .

$T_p = \{Interrupt_1, Interrupt_2\}$,

So, $P_{T_p}(Interrupt_1) > P_{T_p}(Interrupt_2)$.

STEP 2: Identify P_k , the key place set of T_p , and $C_k(P_k)$, the key colour set of P_k .

$^{\bullet}Interrupt_1 = \{out\}$,

$^{\bullet}Interrupt_2 = \{out\}$.

Thus,

$P_k = {}^{\bullet}Interrupt_1 \cup {}^{\bullet}Interrupt_2 -$
$^{\bullet}Interrupt_1 \cap {}^{\bullet}Interrput_2 - [{}^{\bullet}Interrupt_2] = \Phi$.

Namely, it became an unsolvable problem by using the above algorithm.

Secondary transitions, *call_intr1*, *call_intr1*, and places, *OUTING1*, *OUTING2*, could be introduced to the previous model M_4. Furthermore, place *OUT* would be divided into two places OUT_1 and OUT_2. Thus, the model M_4 could be changed as M_4', shown in Figure 8.

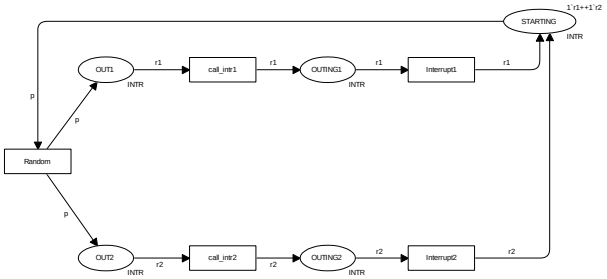


FIGURE 8 Auxiliary Model M_4'

$^{\bullet}Interrupt_1 = \{OUTING_1\}$,

$^{\bullet}Interrupt_2 = \{OUTING_2\}$.

Thus,

$P_k = {}^{\bullet}Interrupt_1 \cup {}^{\bullet}Interrupt_2 - {}^{\bullet}Interrupt_1 \cap {}^{\bullet}Interrput_2 -$
$[{}^{\bullet}Interrupt_2] = \{OUTING_1, OUTING_2\}$.

Next,

$C_k(OUTING_1) = \{INTR\}, C_k(OUTING_2) = \{INTR\}$

So, $C_k(P_k) = \{INTR, INTR\}$.

In the model M_4', when the places, $OUTING_1$, and $OUTING_2$, being changed, they could affect trigger priorities of transitions $Interrupt_1$ and $Interrupt_2$.

STEP 3: Modify the CPN model, increasing one "counter place" to calculate tokens in key place set, as shown in Figure 9.

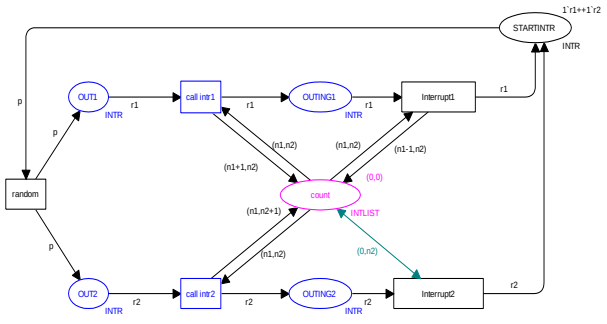


FIGURE 9 Partial interrupt model with priorities

Add the modified partial model to the original model to form a complete interrupt model with priorities M_5, as shown in Figure 10.

In the complete model, the main process *Proccess* could send semaphore to the place *INTR* to control whether the interrupts are enabled. Moreover, the transition $Interrupt_1$ or $Interrupt_2$ could be tangled only when outing interrupts signal is enabled. Meantime, only the transition $Interrupt_1$ could be triggered although it and $Interrupt_2$ could be enabled simultaneously, namely, $Interrupt_1$ have higher priorities than $Interrupt_2$.

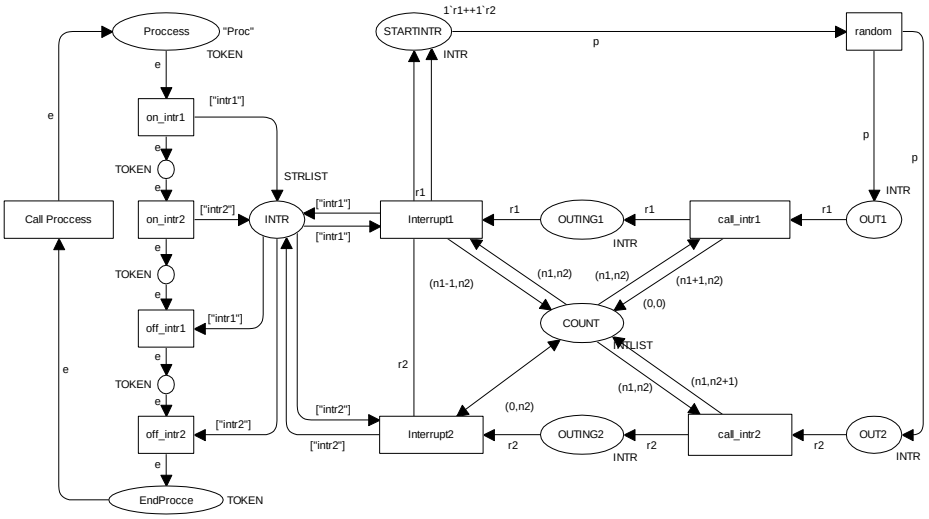


FIGURE 10 Complete interrupt model with priorities

6 Working outlook

This paper researched the CPN model with transition priorities in Group. When the original model has lower complexity, that was, $|P_k|$ and $|C_k(P_k)|$ is small, algorithm for transition with priorities is better. However, this algorithm only addressed transition priorities under enabling condition, not under certain conditions. After that, it could expand the research work in this regard. Furthermore, due to "count place" need to connect bi-directional arc to multiple transitions, making the more complex model structure, reducing the complexity of the model is also the focus of future research.

References

[1] Bożek A 2012 Using Timed Coloured Petri Nets for Modelling, Simulation and Scheduling of Production Systems *Production Scheduling ed Righi R D R: Production Scheduling* 24

[2] Louchka P-Z 2013 Timed Petri Nets. *Time and Petri Nets Springer Berlin Heidelberg* 72

[3] Rogge-Solti A, Weske M 2013 Prediction of remaining service execution time using stochastic petri nets with arbitrary firing delays *Service-Oriented Computing Springer Berlin Heidelberg* 15

[4] Lohmann N, Song M, Wohed P 2014 Discovering Stochastic Petri Nets with Arbitrary Delay Distributions from Event Logs *BPM 2013 ed Lohmann N et al Springer International Publishing Switzerland* 13

[5] Jensen K, Kristensen L M, Lisawells 2007 Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems. *International Journal on Software Tools for Technology Transfer* **9**(3) 213-54

[6] Gehlot V, Nigro C 2010 An introduction to systems modelling and simulation with colored Petri nets *2010 IEEE Proceedings of Simulation Conference* 104-18

[7] Boukhentiche H, Abbas-Turki A, Moudni A E 2011 On the bus priority dilemma: A Petri Net model with resource sharing and inhibitor arc *2011 IEEE International Conference on Industrial Technology* 88-93

[8] Balduzzi F, Giua A, Menga G 2000 First-Order Hybrid Petri Nets: A Model for Optimization and Control *IEEE Transactions on Robotics and Automation* **16**(4) 382-99

[9] Javan A, Akhavan M, Moeini A 2013 Simulating Turing Machines Using Colored Petri Nets with Priority *Transitions. International Journal on Recent Trends in Engineering & Technology* **9**(1) 75-81

[10] Van der Aalst W M P, Stahl C, Westergaard M 2013 Strategies for Modelling Complex Processes Using Colored Petri Nets *Transactions on Petri Nets and Other Models of Concurrency VII, ed Jensen K et al: Springer Berlin Heidelberg* 50

[11] Karamti W, Mahfoudhi A, Kacem Y H 2012 Hierarchical Modelling with Dynamic Priority Time Petri Nets for Multiprocessor Scheduling Analysis *2012 international conference on embedded systems and applications* 114-21

[12] Jensen K, Kristensen L M, Lisawells 2007 Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems *International Journal on Software Tools for Technology Transfer* 9 213-54

[13] Asmuss J, Lauks G and Zagorskis V 2012 Application of CPN Tools for Simulation and Analysis of Brandwidth Allocation *World Academy of Science, Engineering and Technology* 6 72-5

[14] Westergaard M, Fahland D, Stahl C 2013 Grade/CPN a Tool and Temporal Logic for Testing Colored Petri Net Models in Teaching *Transactions on Petri Nets and Other Models of Concurrency VIII, ed M. Koutny et al: Springer Berlin Heidelberg* 23

[15] Westergaard M, Kristensen L M 2009 the Access CPN Framework-A Tool for Interacting with the CPN Tools Simulator *Applications and Theory of Petri Nets Springer Berlin Heidelberg* 10

[16] Home Page of the CPN Tools, http://cpntools.org. Accessed on September 22th, 2013.

[17] Westergad M, Verbeek H M W 2011 Efficient Implementation of Prioritized Transitions for High-level Petri Nets *PNSE'11 - Petri Nets and Software Engineering* 27-41

[18] Sheng-De W, Wang-Bin H, Zhong-Chang X 2008 Research On Simulation Model of Petri Net with Priority by CPN Tools *Journal of System Simulation* **20**(3) 814-6

[19] Wen-Jun L, Xiao-Chong Z, Shi-Xian L, Jian M 2001 Petri Net Semantics of Dynamic Priority Systems *Chinese J. Computers* **24**(10) 1085-94

[20] Murata T 1989 Petri Nets: Properties, Analysis and Applications *Proceedings of the IEEE* **77**(4) 541-80

[21] Jensen K 1994 An Introduction to the Theoretical Aspects of Coloured Petri Nets *A Decade of Concurrency Reflections and Perspectives Springer Berlin Heidelberg* 230-72

Authors

Hong Wang, born in March, 1978, Tieling, Liaoning Province, China

Current position, grades: PhD candidate at Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, a lecturer in the School of Control and Computer Engineering, North China Electric Power University.
University studies: computer science technology and application.
Scientific interest: high reliability software testing and model checking.
Publications: 3 papers.

Tao Zhang, born in December, 1972, Zhaozhuang, Shandong Province, China

Current position, grades: professor at Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences.
University studies: PhD at Tsinghua University in China.
Scientific interest: high reliability software testing and validation, high reliability electronic information system analysis, design methods, complex system simulation.
Publications: more than 20 papers.

Mechanical characteristics and form-finding analysis of iced transmission lines

Yuejun Liu*, Aiping Tang, Ketong Liu, Jiewen Tu

School of Civil Engineering, Harbin Institute of Technology, Harbin, Heilongjiang province, China

Received 1 October 2014, www.cmnt.lv

Abstract

Form-finding is the important problem to be solved in the cable structure analysis, to the different forms of loads, the direct iterative method is used to determine the initial configuration of the cable structures. Horizontal tension or cable tension is used as iterative convergence condition, the form-finding of the cable is researched under its own gravity, uniform ice and non-uniform ice load. As for the multi-span transmission lines, two conditions of uniform ice and non-uniform ice loads on the whole span were analysed. The results of initial configuration are consistent with the analytical method, which verified the correctness of the direct iterative method, under the condition of non-uniform ice load, the stress of conductor is larger than the maximum stress, which is very dangerous in the actual operation.

Keywords: mechanical characteristics, form-finding, uniform ice, non-uniform ice, finite element method

1 Introduction

In the design of transmission lines, the sag, stress and length of transmission lines are very important parameters, which are the main contents to the mechanics research of transmission lines, this is because the sag and stress directly affect the safety of operation, the small changes and error of cable length will make considerable change to the sag and stress, form-finding is the primary problem to be solved before the analysis and calculation of the cable structures, which is the fundamental prerequisites for dynamic response. Currently, there are four methods for the form-finding analysis of cable structures:

1) The nonlinear finite element method [1,2];

2) The force densities method [3];

3) The dynamic relaxation analysis of form-finding [4, 5];

4) The exact element method [6].

The nonlinear finite element method is most widely used for form-finding. In previous study, the dynamic analysis of cable structures is often assumed to the initial curve with simple shape, the uniform ice and some other non-uniform distributed loads are considered as the ideal uniformly distributed loads, but in many cases, the thickness of ice is not same along the conductors, especially in some micro topography and micro climate area, transmission lines have large elevation difference and big spans, the conductor suffers serious non-uniformly distributed load [7]. Such as the jumping induced by transmission line ice-shedding [8], the research for the dynamic response of ice-shedding on bundled transmission lines [9], and vibration of bundled conductors following ice shedding [10]. The dynamics analysis of the overhead transmission lines with non-uniformly distributed load is accurate based on form-finding, which will be very important to the design of the cable structures.

2 The load distribution of cable structures

The transmission lines belong to one kind of cable structures, the theory of cable structures are based on two assumptions:

1) The cable is flexible: it neither suffers the pressure nor bending;

2) The material of the cable structures follows Hooke's law.

Figure 1 shows an infinitesimal piece of cable, H is horizontal component of the tension T in the tangential direction. V is the vertical component of the T, q_x is the distributed load along the x direction, q_y is the distributed load along the y direction.

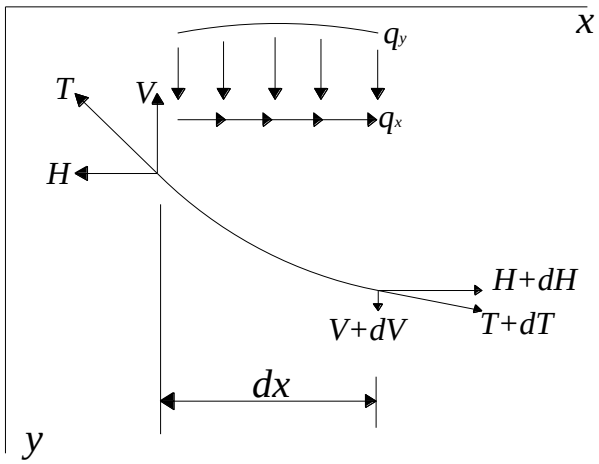


FIGURE 1 Equilibrium of an infinitesimal piece of cable

*Corresponding author e-mail: liuyuejun@163.com

The equilibrium equation of the horizontal direction is:

$$\frac{dH}{dx}+q_x=0\,. \tag{1}$$

The equilibrium equation of the vertical direction is:

$$\frac{dV}{dx}+q_y=0\,. \tag{2}$$

As $V=H\dfrac{dy}{dx}$ Equation (1) and (2) can be transformed to:

$$\frac{dH}{dx}+q_x=0\,, \tag{3}$$

$$\frac{d}{dx}\left(H\frac{dy}{dx}\right)+q_y=0\,. \tag{4}$$

Equation (3) and (4) are the governing differential equations of transmission line shape curve. The task for form-finding is to determine the transmission line shape curve, which is to obtain the equilibrium of the applied pretension and external loads by adjusting the form of cable under the given boundary condition.

The uniformly distributed load on the cable structure is generally divided into two forms:

1) uniformly distributed load along chord line of the cable, the shape of cable is parabola;

2) uniformly distributed load along arc length of cable, the shape of cable is catenary, such as the shape of the cable under the action of gravity.

According to the theoretical analysis, the smaller the sag of cable is, the smaller difference of two forms is, to the actual transmission lines, the sag of cable is very small, when the uniformly distributed load is along chord line of the cable, the error of sag is small, it can be received by engineering. The gravity of transmission lines is uniformly distributed along arc length of cable, therefore, under the action of its own gravity, the shape of the transmission line is catenary, rather than a parabola. When the span is small, the sag-span ratio is smaller, the difference is the smaller [11], generally in engineering, if the sag-span ratio is less than 1/8 [12], it can get enough precision of calculating with a parabolic curve.

3 The overall analysis of cable structures

Figure 2 shows a typical joint *J*, Fzie is vertical force of the endpoint i in the element *e*, (*i* changes with different element), *PJ* is the concentrated force acted on the joint *J*.

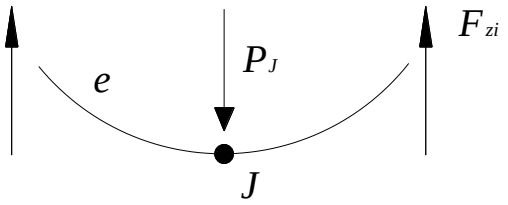


FIGURE 2 Joint J

The vertical equilibrium equation can be written as:

$$\sum_{e_J}F_{zi}^{\,e}-P_J=0\,, \tag{5}$$

where, e_J presents the element connecting with joint *J*, Equation (1) can be written as:

$$\sum_{e_J}\left(k^e_{\,i1}z^e_{\,1}+k^e_{\,i2}z^e_{\,2}\right)=P_J+\sum_{e_J}P_{Ei}^{\,e}\,. \tag{6}$$

According to the overall integration method, the equation of the global stiffness matrix is:

$$KZ=P\,, \tag{7}$$

where *K* is the global stiffness matrix, *P* is the resultant force from P_J and P_{Ei}^e, *Z* is vector of vertical coordinates.

The state equation of transmission lines is:

$$\sigma_{02}-\frac{E\gamma_2^{\,2}l^2\cos^3\beta}{24\sigma^2_{\,02}}=\sigma_{01}-\frac{E\gamma_1^{\,2}l^2\cos^3\beta}{24\sigma^2_{\,01}}-\alpha E\left(t_2-t_1\right), \tag{8}$$

σ_{01}, σ_{02} - the stress of transmission lines at the lowest of sag under two kinds of state, MPa;
γ_1, γ_2 - the specific load of transmission lines under two kinds of state;
t_1, t_2 - the temperature of the transmission lines under two kinds of state;
l, β - the span and the angle of differential elevation;
α, E – the temperature expansion coefficient and elastic modulus.

4 The basic principles of form-finding for transmission line

The direct iterative method is used to determine the initial configuration of the cable structure, the basic principles of the direct iterative method is connecting the chord line as the model, using actual material properties and real constants, and set a small initial strain, and then applying the gravity which distributed along the arc length, and gradually update the finite element model, the horizontal tension or cable tension is chose as iterative convergence conditions, the final result is initial deformation of cable structures under its own gravity, the basic process is showed as Figure 3:

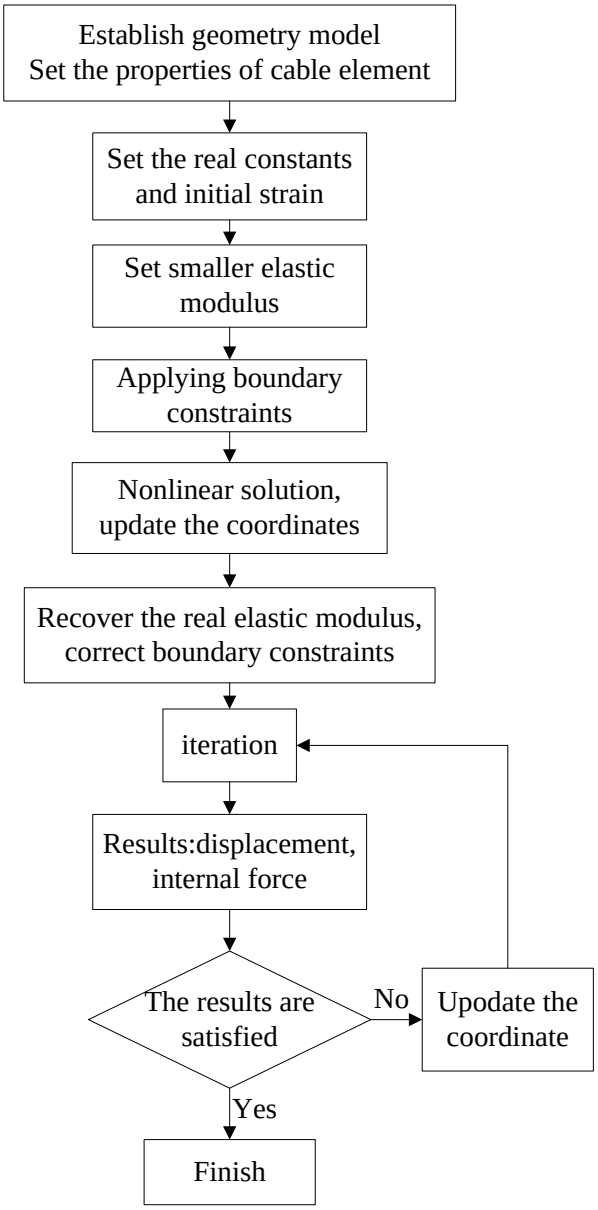


FIGURE 3 The flow chart of form-finding

5 The finite element model of the single span transmission conductor

5.1 THE ESTABLISHMENT OF THE CONDUCTOR MODEL

The finite element method of conductors is transforming the infinite degrees of freedom into a limited degree of freedom. The conductor should be discreted, the stiffness matrix is established through the appropriate shape function of link element, the global stiffness matrix is integrated, the basic equation is established according to the node displacement vector and the node load vector, applying the boundary conditions, the node displacement can be got by solving the basic equations.

The transmission line belongs to a kind of flexible cable structures, which has strong geometric nonlinearity, the geometric characteristics of small strain and large rotation will make stiffness matrix change nonlinearly, the stiffness matrix of structure becomes the function of geometric distortion, the form-finding of transmission lines should open large deformation options in the ANSYS and set the time step, the stress effect should be considered in order to ensure the accuracy of the results.

The form-finding for the transmission lines is analysed by the software ANSYS, the link 10 unit is chose to simulate the lines, the model is connecting a straight line between two points, the straight line is divided into a number of straight link 10 units, the two points are imposed fixed constraints, the gravity loads are imposed along the arc length. The equilibrium configuration of the transmission lines has got under the action of the gravity load, the form-finding of iced transmission lines is based on the equilibrium configuration under gravity. Set nonlinear solution selection. The equilibrium configuration of iced transmission lines is solved, the form-finding of transmission lines is finished under the action of ice load.

Select an overhead transmission line with different height of suspension points, the span is l = 1000 m, the elevation difference between the two suspension points is h = 40 m, the conductors choose LGJ – 400/35, the technical parameters are shown in Table 1. The ice thickness is 20 mm, the weight per unit length is 15.02 N/m, and the ice density is 0.9 g/cm^3. The mechanical parameters of ice are shown in Table 2.

TABLE 1 the technical parameters of conductor

Number of conductor	Cross-sectional area (mm^2)	Tensile force of calculation (N)	Mass per unit length (kg/km)	allowable stress [σ0] (Mpa)	Diameter (mm)
one	425.24	103900	1349	92.8	26.82

TABLE 2 the mechanical parameters of ice

Variables	Numerical value
The thickness of ice (mm)	20
Density (kg/mm^3)	9×10^{-7}
Poisson's ratio	0.3
The damping ratio	0.1
Coefficient of Comprehensive elasticity (Gpa)	10
Coefficient of Comprehensive expansion (1/°C)	50×10^{-6}

5.2 THE FORM-FINDING OF TRANSMISSION LINES UNDER THE ACTION OF GRAVITY

The conductor is divided into 100 units, the horizontal tension or cable force is the convergence criterion for iteration, the displacement vector diagram of conductor under gravity load as shown in Figure 4, it can be seen that

the biggest displacement vector of conductor is 0.654 m, the conductor tends to equilibrium state.

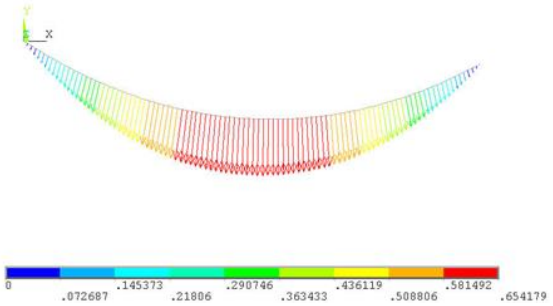


FIGURE 3 The displacement vector diagram of conductor under gravity load (m)

5.3 THE FROM-FINDING OF ICED TRANSMISSION LINES

In many cases, the ice on the conductor is not uniform, it is assumed that the weight is equal with uniform ice and non-uniform ice, the weight of ice on the conductor is measured by the ellipse method or the total weight method, the thickness of uniform ice is 20mm, distributing along the span of the conductors, the ice thickness of different position of is different, which makes the changes of sag and stress are also different, the transmission line is divided into five sections under the action of non-uniform ice load, the horizontal projection length of each section is 200m, the distribution of ice is shown in Table 3.

TABLE 3 The distribution of ice

the case of ice		**the section of ice distribution (mm)**				
		1	*2*	*3*	*4*	*5*
uniform ice		20	20	20	20	20
non-uniform ice	A	30	25	20	15	10
	B	20	30	25	15	10
	C	10	20	30	25	15

Through calculation, the displacement (namely the sag) and stress can be got, analytical values and nonlinear finite element simulation values of representative nodes are listed in Tables 4 and 5 under the condition of self-gravity. It can be seen that the absolute value of the difference between the simulation value with the analytical value is very small, which shows that application of the finite element method to find form of transmission lines is correct and feasible.

TABLE 4 The sag of conductor

		11	**21**	**31**	**41**	**51**	**61**	**71**	**81**	**91**
self-gravity	analytical value	24.73	32.23	37.69	40.97	41.53	41.41	38.13	33.34	26.45
	simulation value	24.82	32.08	37.34	40.60	41.85	41.07	38.25	33.41	26.53
uniform ice	analytical value	25.87	32.74	38.86	41.78	42.94	42.75	39.44	34.35	26.73
	simulation value	25.11	32.63	38.55	41.53	42.86	42.08	39.18	34.15	26.98
the absolute value of the difference	self-gravity	0.09	0.15	0.35	0.37	0.32	0.34	0.12	0.07	0.08
	uniform ice	0.76	0.11	0.31	0.25	0.08	0.67	0.26	0.2	0.25
non-uniform ice	A	25.31	32.92	38.38	41.69	42.88	41.95	38.94	33.86	26.74
	B	25.25	32.92	38.48	41.85	43.04	42.09	39.01	33.90	26.76
	C	25.06	32.64	38.22	41.75	43.14	42.34	39.35	34.21	26.95
	mean value	25.21	32.83	38.36	41.76	43.02	42.13	39.10	33.99	26.82
	Variance	0.016	0.026	0.016	0.006	0.018	0.039	0.049	0.036	0.013

TABLE 5 The stress of conductor

		11	**21**	**31**	**41**	**51**	**61**	**71**	**81**	**91**	**101**
self-gravity	analytical value	57.21	56.64	56.23	55.98	55.92	55.98	56.23	56.64	57.21	57.94
	simulation value	57.19	56.62	56.21	55.97	55.93	55.99	56.24	56.67	57.25	57.99
uniform ice	analytical value	88.4	87.46	86.79	86.38	86.25	86.39	86.80	87.48	88.41	89.6
	simulation value	88.47	87.51	86.82	86.40	86.27	86.50	86.98	87.65	88.56	89.67
the absolute value of the difference	self-gravity	0.015	0.019	0.014	0.009	0.017	0.007	0.012	0.027	0.042	0.05
	uniform ice	0.073	0.047	0.027	0.015	0.017	0.11	0.185	0.172	0.152	0.072
non-uniform ice	A	89.34	88.18	87.43	87.01	86.9	87.07	87.46	88.07	88.85	89.82
	B	91.67	90.68	89.85	89.37	89.25	89.44	89.85	90.47	91.25	92.22
	C	92.61	91.79	91.09	90.62	90.41	90.55	91.0	91.77	92.65	93.74
	mean value	91.21	90.22	89.46	88.99	88.85	89.02	89.44	90.1	90.92	91.93
	Variance	2.83	3.41	3.47	3.37	3.19	3.15	3.26	3.51	3.69	3.89

The sag of analytical and simulation value are shown in Table 4, the absolute value of difference is very small, It illustrates that the finite element method is feasible for simulating from-finding of transmission lines. The sags of iced transmission lines at all the points are larger than bare conductors. It can be seen from that the sag and stress of transmission lines changed with different kinds of the non-uniform ice conditions. Judging from its variance, the change of stress at all points is different with different non-uniform ice conditions, the biggest change appears at both ends and the central span of the conductor.

The deformation diagrams of transmission lines under uniform ice and non-uniform ice are shown in Figure 5-8, the sag and stress have changed a lot, with the increase of elevation difference, the difference of sag and stress becomes larger and larger. The allowable stress of conductor LGJ – 400/35 is 92.8 Mpa, maximum stress does not exceed the allowable stress under uniform ice and condition A, B. But to the condition C, the maximum stress

of the iced transmission line at the node 101 is larger than the allowable stress, which will have risk in the actual operation.

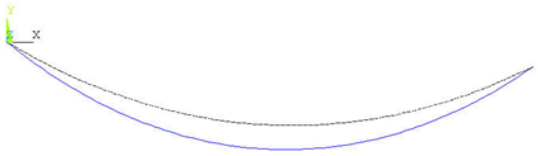


FIGURE 5The deformation diagrams of transmission lines under uniform ice

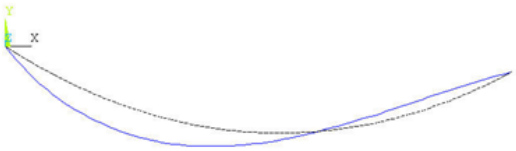


FIGURE 6 The deformation diagrams of transmission lines under ice condition A

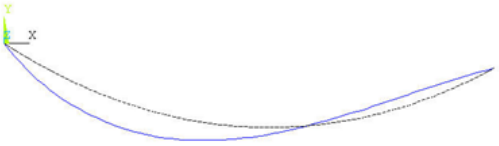


FIGURE 7 The deformation diagrams of transmission lines under ice condition B

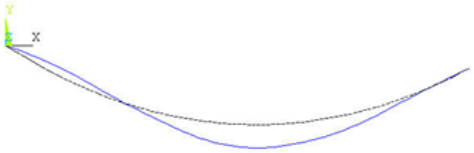


FIGURE 8 The deformation diagrams of transmission lines under ice condition C

6 The form-finding of multi-span transmission lines

6.1 THE MECHANICAL CHARACTERISTICS OF MULTI-SPAN TRANSMISSION LINES

Figure 9 shows the multi-span transmission lines, the force of point A is shown in Figure 10, it does not consider all the friction in the transition zone, it can be obtained the following equilibrium equation without considering the friction.

$$T_1 = T_2. \tag{9}$$

In general $\theta_1 \neq \theta_2$, therefore:

$$H_1 = T_1 \cos\theta_1 \neq T_2 \cos\theta_2 = H_2.$$

Applying the nonlinear finite element method to find form of transmission lines, the horizontal tension is the iterative condition (horizontal tension=the average tension stress by the cross-sectional area), for the multi-span transmission lines, the horizontal tension is unequal. The horizontal tension to each adjacent span can be calculated by Equation (9). If the error of horizontal tension is smaller than 5%, update the model with smaller numerical value, or update with a large numerical value.

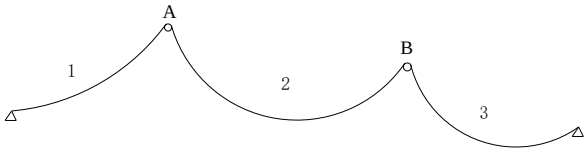


FIGURE 9 The multi-span transmission lines

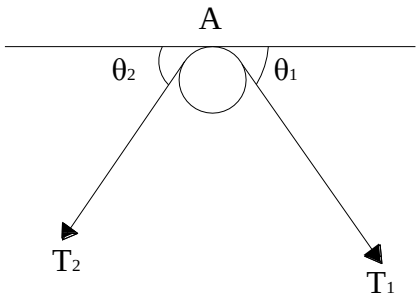


FIGURE 10 The force of point A

6.2 THE EXAMPLE OF FORM-FINDING FOR MULTI-SPAN TRANSMISSION LINES

With reference to the above model parameters of single span transmission line, the model of three consecutive transmission lines was set up, in the actual transmission lines, the adjacent spans are connected by insulators, so the joints of two middle spans are set as hinge constraint, the ends of the lines are used fixed constraint, the Y direction of degrees of freedom at the joints of two middle spans are restrained. Ice conditions of the transmission lines are divided into four kinds of working conditions, which is shown in the Table 6. The thickness of uniform ice is 20 mm. The deformation diagrams of transmission lines under uniform ice and non-uniform ice are shown in Figure 11-14. The distribution of sag has changed a lot with uniform and non-uniform ice, the sag increased in the condition of 2-3, which make the sag of the adjacent span changed, which has a significant impact on the normal operation of transmission lines.

TABLE 6 The ice conditions of three consecutive transmission lines

1	2	3	4
all spans	first span	second span	third span

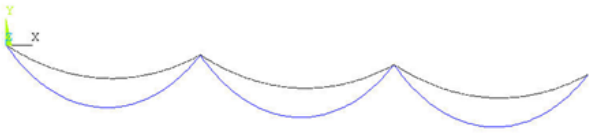


FIGURE 11 The deformation diagrams of three spans under uniform ice

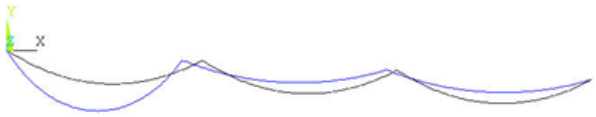


FIGURE 12 The deformation diagrams of the first spans under uniform ice

FIGURE 13 The deformation diagrams of the second spans under uniform ice

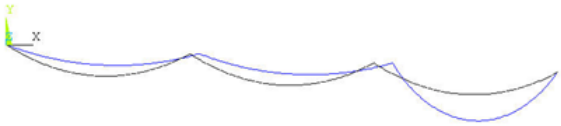


FIGURE 14 The deformation diagrams of the third spans under uniform ice

7 Conclusion

1) According to the different forms of load on cable structure, the form-finding of the cable is researched under its own gravity, uniform ice and non-uniform ice load. As for the multi-span transmission lines, two conditions of uniform ice and non-uniform ice loads on the whole span were analysed, which provide premise for the dynamics analysis of transmission lines, the accuracy of form-finding will seriously affect the results of dynamic response.

2) It can be seen that the absolute value of the difference between the simulation and the analytical value is very small, which shows that application of the finite element method to find form of transmission lines is correct and feasible. The sags of iced transmission lines at all the points are bigger than bare conductors. For the non-uniform ice conditions of A, B and C, it can be seen from that the sag and stress of transmission lines changed with different kinds of the non-uniform ice conditions. Judging from its variance, the change of stress at all points is different with different non-uniform ice conditions, the biggest change appears at both ends and the central span of the conductor.

3) The sag and stress have changed a lot between the conditions of uniform ice and non-uniform ice, with the increase of elevation difference, the difference of the sag and stress becomes larger and larger. The allowable stress of conductor LGJ – 400/35 is 92.8 Mpa, the maximum of the iced transmission line stress does not exceed the allowable stress under uniform ice and condition A, B. But to the condition C, the maximum stress at the node. 101 MPa is larger than the allowable stress, which will have risk in the actual operation.

References

[1] Xiang Y, Shen S 1997 Initial form-finding analysis of cable structures *Journal of Harbin University of C.E. & Architecture* **30**(3) 29-33 *(in Chinese)*

[2] Yan J, Gao R 2002 Comparison and analysis between two form-finding methods for cable net structures *Steel Construction* **17**(1) 4- *(in Chinese)*

[3] Schek H K 1974 The force densities method for form-finding and computation of general nets *Computer Methods in Applied Mechanics Engineering* **3**(1) 115-34

[4] Barnes M R 1974 Dynamic relaxation analysis of tension networks *Preceedings of the International Conference on Tension Structure* London

[5] Barnes M R 1988 Form-finding and analysis of prestressed nets and membranes *Computers and Structures* **30**(3) 685-95

[6] Yuan S, Cheng D, Ye K 2005 Exact element method for form-finding analysis of cable structures *Journal of Building Structures* **26**(2) 46-50 *(in Chinese)*

[7] Kálmán T, Farzaneh M, McClure G 2007 Numerical analysis of the dynamic effects of shock-load-induced ice shedding on overhead ground wires *Computers and Structures* (85) 375-84

[8] Yin P, Li L, Zhang H 2008 Investigation on control of jumping induced by transmission line ice-shedding *Advanced of power system & Hydroeletric Engineering* **24**(3) 3-7 *(in Chinese)*

[9] Roshan F M, McClure G 1998 Numerical Modeling of the dynamic Response of Ice-Shedding on Electrical Transmission lines *Atmospheric Research* **46**(1) 1-11

[10] Kollár L E, Farzaneh M 2008 *IEEE Transactions on Power Delivery* **23**(2) 1097-104

[11] Shen S, Xv Co, Zhao C, et al. 2006 Design of Cable Structures 2nd Edition *Beijing: China Architecture & Building Press (in Chinese)*

[12] Irvine H M 1983 Cable Structures *The MIT Press* Cambridge, Massachusetts and London England

Authors

Yuejun Liu, born in December, 1983, Jilin Province, Tonghua City, China

Current position, grades: PhD student in Civil Engineering at Harbin Institute of Technology.
University studies: Harbin Institute of Technology.
Scientific interests: the iced high voltage transmission tower - line system and de-icing method.

Aiping Tang, born in September, 1968, Hunan Province, Liling City, China

Current position, grades: professor and doctoral supervisor in Civil Engineering at Harbin Institute of Technology.
University studies: Harbin Institute of Technology.
Scientific interests: lifeline earthquake engineering, geotechnical earthquake engineering, geographic information system.

Ketong Liu, born in May, 1982, Shandong Province, Jining City, China

Current position, grades: PhD student in Civil Engineering at Harbin Institute of Technology.
University studies: Harbin Institute of Technology.
Scientific interests: bridge engineering, CFD.

Jiewen Tu, born in October, 1984, Jiangxi Province, Nanchang City, China

Current position, grades: PhD student in Civil Engineering at Harbin Institute of Technology.
University studies: Harbin Institute of Technology.
Scientific interests: the stability of rock slope.

Mobile ECG QRS detection algorithm and implementation

Ying Sun[1*], Meikui Deng[1], Shenghua Ye[2]

[1]*Medical Instrument and Food Engineering College, University of Shanghai for Science and Technology, Shanghai, China, 200093*

[2]*Division of Engineering Science, University of Toronto, Toronto, Canada, M5S 2E4*

Received 12 May 2014, www.cmnt.lv

Abstract

The ECG signal contains a lot of interference in the mobile ECG monitoring system. Reasonable selection of signal filtering and QRS wave detection method are the key to ECG signal analysis. According to the characteristics of ECG signals, the design of low-pass filtering and the improved median filter that can filter the interference has been conducted in the paper. Meanwhile, addressing to the limitation generated by selecting and fixing the median threshold using the traditional differential slope method, the paper has proposed the adaptive dynamic threshold and used quadratic difference algorithm to process signal in order to obtain R wave, and then locate Q and S wave based on R wave location. In addition, combining the characteristics of the QRS wave group, judgments on the missing and over detection are conducted, so that the algorithm is robust and has good fault-tolerant ability. The experimental results show that, this algorithm can not only satisfies the need of the real-time QRS wave detection, but is also more suitable for the transmission and processing of ECG signals in mobile ECG monitoring.

Keywords: mobile ECG signal, QRS detection, algorithm

1 Introduction

Mobile ECG monitoring device can quickly detect the real-time dynamic ECG signal to provide emergency ambulance, disease surveillance, medical advice and guidance services for users. Its application prospect is thus very promising [1].

Taking into account that the mobile ECG detection method shall be easily used by the users, the detection of the lead ECG signal is conducted by pressing the metal conductor with both thumbs. The Dynamic ECG signals obtained by this method contain a large amount of 50HZ interference, baseline drift, EMG interference and motion artifacts, as well as noise interference [2]. Various noises degrade the accuracy of diagnosis, and thus suitable methods must be used to conduct the filtering prior to the calculation of the heart rate data [3].

In this paper, the original data is received by the mobile phone from the Bluetooth transmission module. Raw data is then filtered by FIR low pass filter. Low frequency interferences, such as baseline drift, are removed through improved median filter. The smoothed data is finally returned. The said process is shown in Figure 1.

Since the mobile ECG monitoring equipment requires the real-time processing and transmission of ECG information [4], it is very important to design an accurate and fast heart rate detection algorithm.

At present, there are various ECG detection algorithms, such as difference vector analysis method [5], template matching method [6], wavelet transform method [7, 8] and neural network method [9, 10], etc. The differential threshold method has a simple algorithm which is fast in the processing speed and easy to be implemented in engineering [11, 12]. The template matching method is simple in principle, but it is very sensitive to high frequency noise and baseline drift [13, 14]. Wavelet transform method has excellent characteristics of time frequency localization, as well as high detection accuracy, but the amount of calculation is huge, which is not suitable for real-time processing [15, 16]. The neural network method can achieve a very good effect of discrimination, but its training time is long, and its real-time performance is poor [17, 18].

The said methods have both advantages and disadvantages. Addressing to the limitations caused by selecting and fixing the threshold in the traditional difference method, an adaptive dynamic threshold is set in this paper according to the signal characteristics. First, the initial data is intercepted and its first order difference is calculated, from which the initial threshold and amplitude threshold are obtained. Then, dynamic threshold adjustment is made by self-learning method. Finally, R wave is located according to the threshold. Q, R, S wave is then accurately localized based on QRS waveform characteristics.

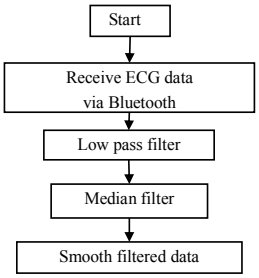


FIGURE 1 The flow diagram of signal filtering

[*]***Corresponding author*** e-mail: yysyysun@163.com

2 Filtering algorithm

2.1 LOW-PASS FILTER DESIGN

There are two ways to design low-pass filter design: one is IIR - infinite impulse response filter, and the other is FIR - finite impulse response filter. The difference lies in: IIR filter can get very high selectivity with the lower order, and it uses less memory, less calculations, and therefore it is not only economic but high in efficiency. However, high efficiency costs nonlinearity of the phase; hat is the better the selectivity, the more serious the nonlinearity is. In contrast, FIR filter can obtain a strictly linear phase. However, a higher order is required to get a certain selectivity, which asks for more memory and longer operation. As well, the signal delay is relatively large. In view of the fact that the mobile ECG signal has a strict requirement on the linear phase, and its stability in operation must be good, and noise power of the output signal caused by calculation error shall be small, as well as its nature of real-time shall be strong and operation speed shall be high [19], in this paper the window function method is adopted to design the filter [20].

Let the expected approximation response function of filter frequency be $H\left(e^{j\omega}\right)$, $h_d\left(n\right)$ be the corresponding unit impulse response, $H\left(z\right)$ be a system function. Then we can have the following equation:

$$H\left(e^{j\omega}\right)=\sum_{n=0}^{N-1}h_d(n)e^{-j\omega\omega}, \tag{1}$$

$$h_d\left(n\right)=\frac{1}{2\pi}\int_{-\pi}^{\pi}H\left(e^{jw}\right)e^{jwn}d\omega, \tag{2}$$

$$H\ (\mathrm{z})=\sum_{n=0}^{N-1}h(n)z^{-n}. \tag{3}$$

Generally speaking, $h_d\left(n\right)$ is infinitely long, and thus it needs an approximation of $H\left(e^{j\omega}\right)$. When the window function design method is adopted, the ideal filter unit sampling response and window function can be used to conduct the filter design:

$$h(n)=h_d(n)w(n), \tag{4}$$

where $\omega\left(n\right)$ is a window with a finite length, with a value of 0 outside the interval of $0 \le n \le N$, and it is symmetric with the middle point:

$$\omega\left(n\right)=\omega\left(N-1-n\right). \tag{5}$$

The frequency response can be obtained by the convolution theorem:

$$H\left(e^{j\omega}\right)=\frac{1}{2\pi}\int_{-\pi}^{\pi}H_d(e^{j\theta})W\left(e^{j(\omega-\theta)}\right)d\theta. \tag{6}$$

The degree of approximation of the frequency response of the filter designed using the window function method to the ideal response is determined by two factors: $\omega\left(e^{i\omega}\right)$ the width of the main lobe, and $\omega\left(e^{i\omega}\right)$ the amplitude of the side lobe.

In order to satisfy the requirement of stopband attenuation index, and at the same time to ensure the effective denoising while making the calculation time be the shortest, the low-pass filter of 34 orders and cutoff frequency of 35Hz is selected in this paper to remove the interference noise. The amplitude frequency characteristics is shown in Figure 2, and its denoising effect is shown in Figure 3.

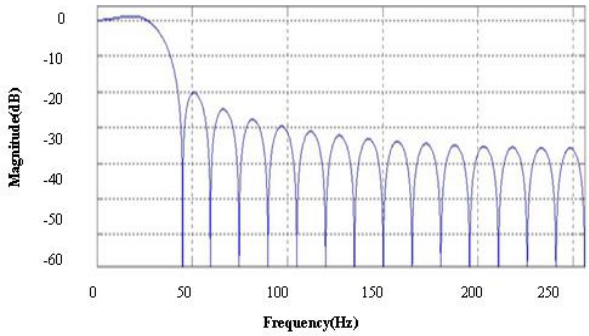


FIGURE 2 The low-pass filter characteristics

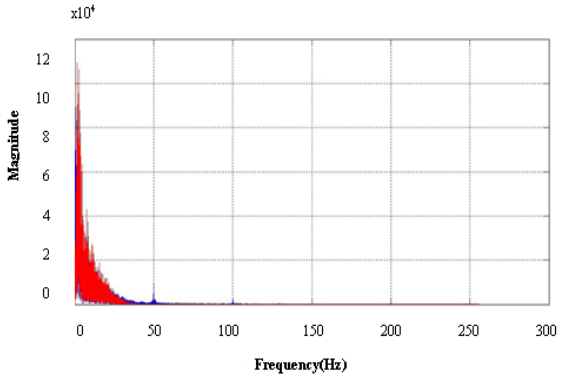


FIGURE 3 Low-pass filtering effect diagram

The red part is filtered spectrum, and the blue one is the original signal spectrum in Figure 3, from which we can see that the design of the low-pass filter can well remove the 50HZ frequency interference and quadratic harmonic of 100HZ frequency.

2.2 MEDIAM FILTER DESIGN

In addition to the filtering of the high frequency interference, a filter must be designed to remove the baseline drift, and the DC component.

Since the band mobile ECG baseline drift is less than 2HZ, if the filter cutoff frequency is too low, it cannot properly eliminate the baseline drift, but if the cutoff frequency is chosen too high, it would make the S-T segment definition waveform distort, as shown in Figure 4.

In Figure 4, the Q wave passing a high pass filter weakens downward, and the S wave rises up and T wave also weakens very significantly. Spectral contrast diagram before and after the filtering by the high-pass filter is shown in Figure 5, where the red part is filtered signal spectrum, and the blue part is the original signal spectrum. The spectrum chart shows that the filtering effect lower than 10Hz band is very poor. If the 0-2HZ band is to be filtered, higher order high pass filter would be needed, but the resulting high time-consumption is not suitable for mobile ECG detection. Hence, we turn to adopt the improved median filtering method.

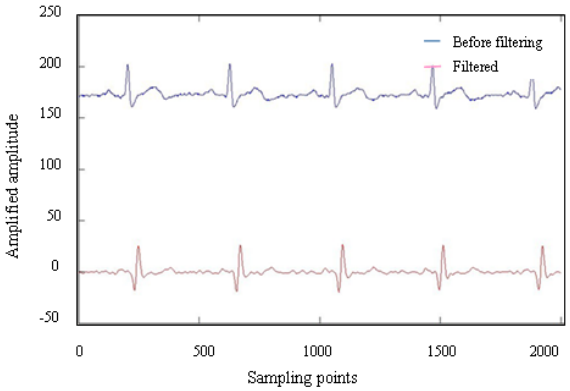


FIGURE 4 The filtered comparison chart in high-pass filter

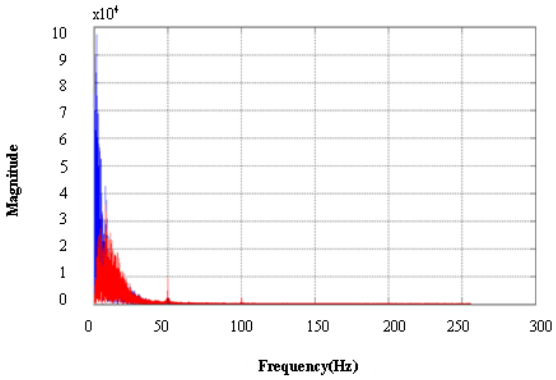


FIGURE 5 High-pass filter effect spectrum diagram

The traditional median filtering method is based on a value of $X(i)$ as the center and the odd point nearby as the window. The points of odd numbers are sorted, and then the median is taken as the new value of $X(i)$, and so on. The window length is generally in odd numbers like 5 or 7.

The improved median filtering method is: the window is taken at the centre point whose length is marked by an odd number larger than or equal to 127 The data of window lengths shall be sorted to obtain the median,, and then the new median is subtracted from the original median point, thus removing the DC component. It is also quite inhibiting against baseline drift.

The specific algorithm is as follows:

$X(i)$ is data of No. i sampling point of the original signal. Through the $X(i)$, values from the $X(i\text{-}63)$ to $X(i+63)$ are sorted to take the median value, *mid*. Then $Y(i)$ is take as $X(i)$ - mid (i.e. $X(i)$ minus *mid*), where $Y(i)$ is No. i sampling signal after the median filtering.

Spectrum comparison chart before and after the median filtering is as shown in Figure 6.

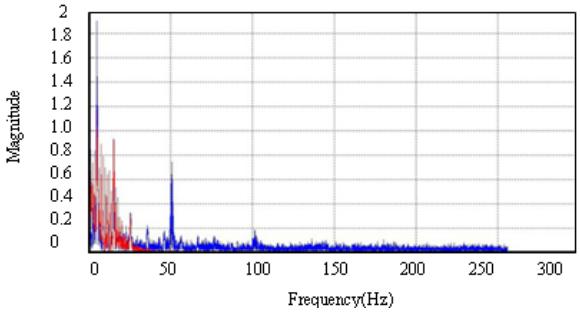


FIGURE 6 Median filter effect spectrum diagram

The filtering results of low-pass filter and median filter as shown in Figure 7.

As the figures show, the improved median filtering method can effectively eliminate the DC component, and the location of the baseline in Figure 7 effectively returns to zero. The algorithm is simple, and less time-consuming, which is suitable for low frequency filtering of the mobile ECG signal.

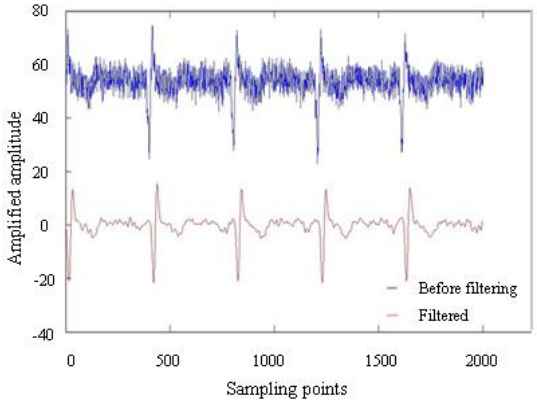


FIGURE 7 The filter effect results

2.3 SIGNAL SMOOTHING

The average smoothing filter is suitable for filtering the signals that have random noise, and the smoothing effect is obvious. However, if the window length selected is too large, the peak clipping phenomenon would be serious, because the average obtained in the apex of R wave and the surrounding points can reduce R wave characteristics. On the other hand, if the window length is too small, it will greatly reduce the smoothing effect. In order to improve the method of average smoothing filter, the average of several points around a point is taken as the new value of this point:

$$Y(i)=\big(Y(i-3)+Y(i-2)+Y(i-1)+Y(i)+ Y(i+1)+Y(i+2)+Y(i+3)\big), \tag{7}$$

where the window length is determined according to the 7 sampling points.

3 QRS wave group detecting

3.1 DETECTION ALGORITHM

The key point to use the differential threshold method is to determine a reasonable detection threshold, but the R wave form and amplitude will vary due to detection of different objects. Therefore, it is difficult to find a unified detection threshold that is suitable for various objects to be detected. In order to solve this problem, this paper adopts the self-learning algorithm to build the detection threshold.

Set the divergent ECG signal as $x(n)$, the length of the signal is L. An improved differential algorithm is adopted to detect the QRS wave group:

$$y(n) = x(n+2) + x(n+1) - x(n-1) - x(n-2), \tag{8}$$

where, $n = 2, 3, ..., L-2$.

Take the first order difference of $X(n)$:

$$d(n) = x(n+1) - x(n), \tag{9}$$

where, $n = 1, 2, L-1$

It is known from Equation (8) that, obtaining the signal $y(n)$ from the five point difference will prominently display the characteristics of R wave. The judgment on the R wave is made by using the self-learning difference threshold conditions against $y(n)$. Combining the above with self-adaptive threshold set would preliminarily locate the R wave.

Equation (9) shows that $d(n)$ signal is obtained by conducting the first-order differential of the $y(n)$ signal. Corresponding relationship between the singular points of $y(n)$ in the $d(n)$ is established, and thus effectively confirm the R wave. The differential of zero crossing as well as the initial average amplitude thresholds left and right of R wave are calculated to accurately position the ECG QRS wave group.

In the actual detection, when a human body first contacts the electrode, an unstable phenomenon may occur. Through the pilot test, it is determined that the initial 6 seconds is the unstable period. In this paper, the ECG signal sampling frequency is 512HZ. Considering the data processing speed and data buffering problem, real-time computing takes the data collected within 2 seconds as the dynamic cycle. Figure 8 is the heart rate calculation flow chart, where DDmax and DDmin, AAmax are the initial maximum, minimum and maximum threshold, TH is the initial amplitude drop.

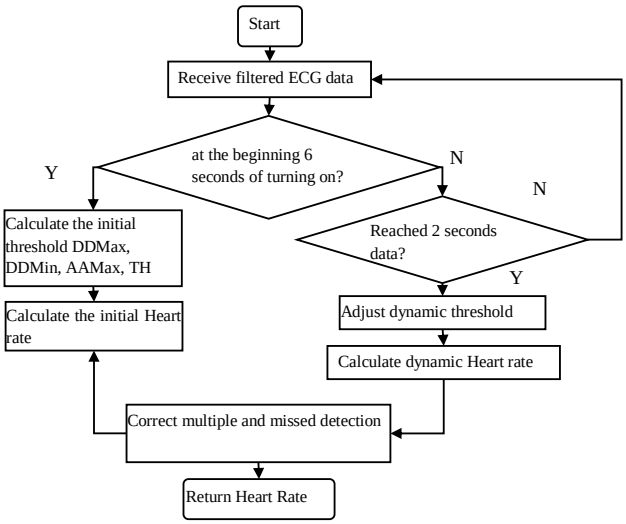


FIGURE 8 Heart rate calculation flowchart

3.2 ALGORITHM DESCRIPTION

3.2.1 Calculate the various initial thresholds

In this paper, X_i, the data in the initial 6 seconds of filtered $x(n)$, is divided into 3 sections, with the unit cycle taken as 2 seconds. Differential operation is conducted according to Equation (8) on each section to obtain the maximum difference value $D\max i$ and the minimum negative difference value $D\min i$ of its ECG data, as well as the maximum and minimum amplitude of the Xi signal. The arithmetic mean of $D\max i$, $D\min i$, $A\max i$ and amplitude difference ($A\max i$ - $A\min i$) is taken to obtain the initial thresholds of $DD\max$, $DD\min$, $AA\max$ and initial amplitude drop threshold TH. These thresholds can be expressed as the following equation:

$$\left.\begin{aligned} DD\max &= D\max/TH = 1/TH\left\{\frac{1}{3}\sum_{i=1}^{3} D\max i\right\} \\ DD\min &= D\min/TH = 1/TH\left\{\frac{1}{3}\sum_{i=1}^{3} D\min i\right\} \\ AA\max &= A\max/TH = 1/TH\left\{\frac{1}{3}\sum_{i=1}^{3} A\max i\right\} \\ TH &= (A\max - A\min)/TH = 1/TH\left\{\frac{1}{3}\sum_{i=1}^{3}(A\max i - A\min i)\right\} \end{aligned}\right\}, \tag{10}$$

where $D\max i = \max\left(Y\left[k\right]\right)$, $D\min i = \min\left(Y\left[k\right]\right)$, $A\max i = \max\left(X\left[k\right]\right)$,

$$Y\left[k\right] = X\left[k\right] + X\left[k-1\right] - X\left[k-3\right] - X\left[k-4\right].$$

The threshold of each unit cycle is jointly determined by the initial threshold, the maximum difference, the minimum difference and the maximum amplitude of the current section, *TH* is the parameter of the threshold. It can be adjusted according to the ECG R wave amplitude, and its value shall be an exponent with 2 as the base number.

3.2.2 Dynamic threshold adjustment

After the R wave is detected in the 6 seconds data using the initial threshold, the detection threshold is modified using the sliding average method. This paper takes every 2s as the dynamic threshold adjustment range. The newest threshold is jointly determined by the current segment threshold and the initial threshold. The update form of new threshold is:

$$\left.\begin{aligned} DD\max &= DD\max\cdot 2/3 + D\max\left[i\right]/3 \\ DD\min &= DD\min\cdot 2/3 + D\min\left[i\right]/3 \\ AA\max &= AA\max\cdot 2/3 + A\max\left[i\right]/3 \end{aligned}\right\}. \quad (11)$$

In the dynamic range, if the measured R wave amplitude *A*max[*i*] increases, the new threshold of the *AA*max new will increase. The threshold is dynamically adjusted, so the stability of the system is enhanced.

3.3 QSR WAVE DETECTION

3.3.1 R wave detection

After the detection threshold is obtained, the R wave detection shall be conducted on the signal of *y*(*n*). When the current difference value and the difference value of the next point is larger than the positive difference threshold *DD*max, and the current amplitude is larger than the amplitude threshold value *AA*max, this point shall be used as a starting point for the 160ms window. If there is a point that has difference value less than the negative difference threshold DDmin, the maximum peak point within the window shall be found as the peak of the R wave. Then the threshold in each sub cycle shall be adjusted according to the Equations (10) and (11), and the initial R wave is detected again with the new threshold. According to the "refractory period" principle in medical field [21], it is believed that the R wave won't appear again within 200ms, so we can skip the refractory period to conduct the detection. The R wave detection process is illustrated in Figure 9.

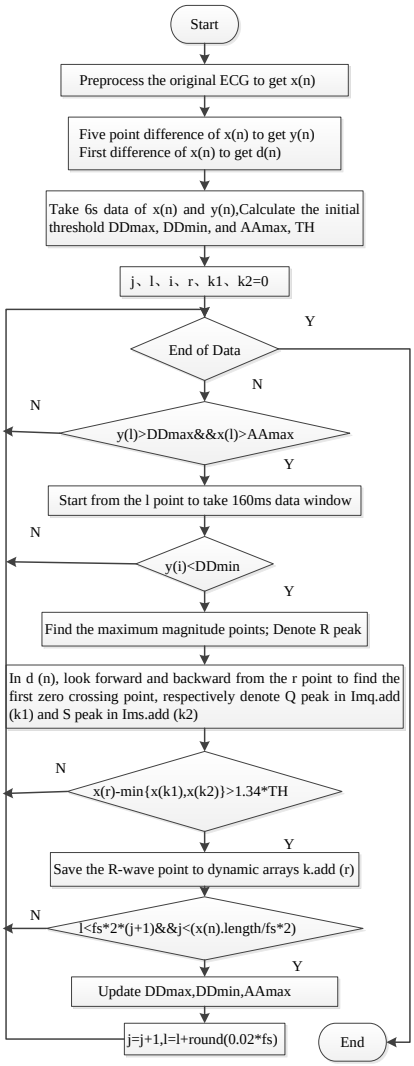


FIGURE 9 The R wave detection process

3.3.2 Q and S wave detection

After the detection of R wave is conducted, use a peak value of R wave point as the centre, forward and backward search the *Q* and *S* points.

In this paper, the first-order differential *d*(*n*) is used to detect the initial position detection of *Q*, *S* wave. The relation between the R wave and its corresponding *Q*, *S* wave position is: if the R wave value is the down zero (its value is negative) corresponding to the QRS wave group in the *d*(*n*). The *Q* wave shall appear in the first upward zero (its value is positive) before the R wave location; the *S* wave is the first upward zero (its value is positive) after the R wave position.

Because of the time - delay relation, the R wave position corresponds to the extreme position, k.get(*i*) in the *y*(*n*), and corresponds to the k.get(*i*)+3 position in the *d*(*n*). Similarly we can calculate the *Q* and *S* wave. The algorithm is as follows:

1) In the *d*(*n*), starting from k.get(*i*)-3, from back to front, search the first upward zero crossing point; mark its location as *Imq*(*i*).

2) In the $d(n)$, starting from k.get(i)+3, from front to back, search first upward zero; mark its location as $Ims(i)$.

3) Search the $Imq(i)$ and $Ims(i)$ in the $X(n)$; identify them as the positions of the Q and S wave, and calculate its amplitude.

4) Calculate the values of $Imq(i)$-3, $Imq(i)$-2, $Imq(i)$ corresponding to that in the $X(n)$, the minimum amplitude is Q wave position of $X(n)$ signal.

5) Calculate the values of $Imq(i)$-3, $Imq(i)$-2, $Imq(i)$ corresponding to that in the $X(n)$, the minimum amplitude is the S wave position of the ECG signal.

3.3.3 Processing the over detection and missing detection

In order to improve the detection accuracy, before the detection is completed, each RR interval (heartbeat interval) is judged. If the current RR interval is less than 0.6*RR, it is believed that there may be over-detections that need further treatment. If it is greater than 0.6*RR, and the sum of two adjacent RR clearances is less than 1.5*RR, the existence of over-detection would be confirmed. If over-detection is the case, the amplitude shall be re-adjusted, and the R wave is detected with the algorithm again to revise the detections. The process of fixing over-detection is shown in Figure 10.

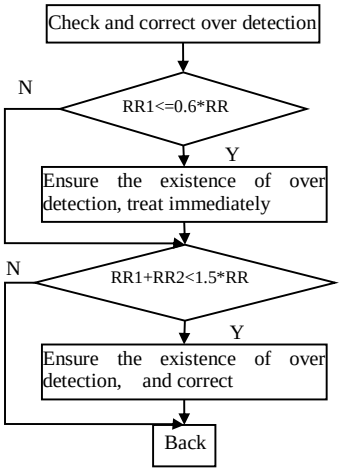


FIGURE 10 Over-detection process

In addition to over-detection, there may also be missing detection. The missing detection judging method is as follows: check whether the RR interval is greater than 1.6*RR or not. If the interval is between 1.6*RR and 2.6*RR, it shows that there is likely to be missed detections. Further, if there exists a data in the interval of 0.7*RR to 1.2*RR and with amplitude of 0.5HR (Heart Rate) to 1.5HR interval [22], it can be judged that an R wave is missed in the detection, and the missing R wave would be filled it into the R wave sequence automatically. If the said conditions are not met, it shall be determined whether the R wave is inverted or not. The process of fixing R wave missing-detection is shown in Figure 11. If none of the above cases occurs, it is believed that there is no over-detection or missing detection of the R wave, and then the detection of the next R wave can be conducted.

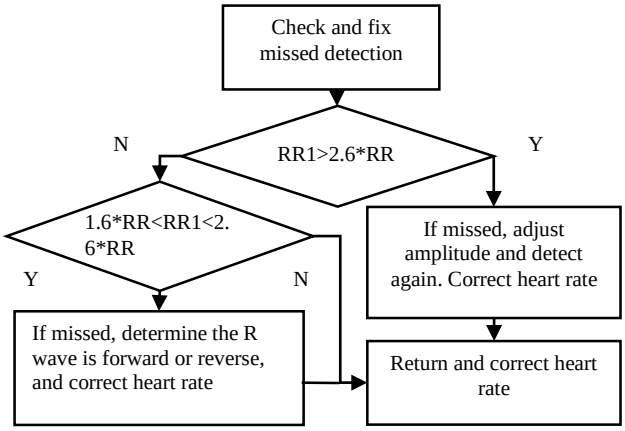


FIGURE 11 The process of R wave detection missed

4 Results

4.1 ALGORITHM VERIFICATION

The multiple groups of the ECG data from the MIT-BIH are adopted to validate the algorithm. After comparing the algorithm test results with the R wave and QRS wave group position provided by the MIT-BIH library [23], it shows that the method can accurately locate the position of the R wave. Tests show that the algorithm can have a detection rate above 99% of normal ECG R waveforms, but the detection rate for the disease ECG waveform (QRS deformation, high P wave and T wave etc.) is over 96%. The average accuracy rate can exceed 97.8%. The following table lists the experimental results respectively. In Table 1, A means atrial premature beat; / means delayed take-off; L means left bundle branch is blocked; a means abnormal atrial premature beat; V means premature ventricular contraction; F means ventricular fusion beat; N means normal; R means right bundle branch is blocked.

TABLE 1 MIT-BIH database library QT signal detection results

MIT-BIH Data record number	Signal Characteristics	False detection rate	Missed detection rate	Accuracy
100	A	0	0	100%
107	/	0	0	100%
108	V	1	3	96%
111	L	0	0	100%
119	VV	0	0	100%
122	N	0	0	100%
201	a	1	0	100%
208	FV	1	1	99%
212	R	0	0	100%
221	V	0	0	100%
234	N	0	0	100%

4.2 HEART RATE DETECTION

The algorithm is adopted to implement the mobile ECG real-time sampling. The mobile ECG detector consists of an integrated precision operational amplifier, a 12 bit digital to analogue converter, a high performance processor and Bluetooth module. It can realize the detection of ECG signals from the fingertips of both hands in the environment of strong background noise and high output impedance. Under the different experimental conditions, the authors have conducted the real-time ECG detection on 12 normal people. The data are as shown in Table 2.

TABLE 2 Mobile ECG monitoring clinical effects

Mobile ECG monitor file name	Visual QRS wave group number	Over detection number	Missed detection number	Error detection rate	Accuracy
201306261321	1992	4	1	2	99.65%
201306261407	1878	1	5	2	99.37%
201306281533	1893	0	2	1	99.84%
201306281607	1589	0	1	2	99.81%
201307011007	1678	3	2	3	99.53%
201307011110	1788	1	2	2	99.72%
201307051310	1684	0	0	3	99.83%
201307051443	1774	3	1	1	99.72%
201307101002	1678	0	1	3	99.77%
201307101120	1567	0	0	0	100%
201401071114	1956	3	0	1	99.58%
201401071207	1556	2	3	1	99.61%
201401071312	1782	1	4	0	99.71%
201401071618	1685	2	3	1	99.64%
201401081110	1856	1	3	2	99.67%
201401081211	1825	2	1	0	99.83%
201401081321	1736	0	0	0	100%
201401081505	1592	2	1	3	99.62%
201401081612	1683	0	1	0	99.94%

The experimental results in Table 2 show that the overall accuracy of the algorithm is above 99%. Experiments show that, in cases of the strong interference, the EMG interference or the conductive body being contacted, the algorithm can accurately calculate the QRS wave. Its detection accuracy is high; its operation speed is fast, and its memory consumption is small, which is suitable for mobile ECG signal processing and application.

5 Conclusion

This paper puts forward the improved detection method based on the difference threshold detection, and has solved the problem of the limitation caused by selecting and fixing the threshold in the traditional differential threshold method. The algorithm is simple, fast in operation, accurate and easy to realize. It can extract and identify the main characteristic parameters of the ECG signals, and lay a solid foundation for the further recognition and diagnosis. The algorithm is tested in the Android intelligent mobile phone, and the test results are satisfactory.

In addition, the system only realizes the lead measuring. The algorithm can be generalized to multiple lead detections, and thus the complete and further detection of the ECG signals can be conducted based on the characteristics of the *P* wave and *T* wave, so as to realize the analysis on the characteristics of more heart diseases.

References

[1] Gavalas D, Economoun D 2011 *IEEE Software* **28**(1) 77-86

[2] Arzeno N M, Deng Z, Poon C 2008 *IEEE Transactions on Biomedical Engineering* **55**(2) 478-84

[3] Ravindran S, Dunbar S, Nisarga B 2009 Real-time, low-complexity, low-memory solution to ECG-based heart rate detection *The 31st Annual International Conference of the IEEE EMBS Minneapolis Minnesota.USA* 1371-4

[4] Talha M, Guettouche M A, Bousbia-Salah A 2010 Combination of a FIR filter with a Genetic algorithm for the extraction of a fetal ECG *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers* 1756-9

[5] ter Haar C C, Maan A C, Schalij M J, Swenne C A 2013 Improved electrocardiographic detection of hyperacute ischemia by difference vector analysis *Computing in Cardiology Conference (CinC)* 9-12

[6] Chin F J, Fand Q, Zhang T, Cosic I 2010 A fast Critical Arrhythmic ECG waveform identification method using cross-correlation and multiple template matching *Engineering in Medicine and Biology Society (EMBC) 2010 Annual International Conference of the IEEE* 1922 - 5

[7] Wu D, Bai Z 2012 An improved method for ECG signal feature point detection based on wavelet transform *2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA)* 1836-41

[8] Alfaouri M, Daqrouq K 2008 ECG signal denoising by wavelet transform thresholding *American Journal of Applied Sciences* **5**(3) 276-81

[9] Alizadeh R 2010 A Dynamic Cellular Automaton Model for Evacuation Process with Obstacles *Safety Science* **49**(1) 315-23

[10] Wang L, Shen M, Dong J 2009.An Uncertainty Reasoning Method for Abnormal ECG Detection *Proceedings of the 2nd IEEE International Symposium on IT in Medicine & Education* 1091-6

[11] Sadhukhan D, Mitra M 2012 Detection of ECG characteristic features using slope thresholding and relative magnitude comparison *2012 Third International Conference on Emerging Applications of Information Technology (EAIT)* 1226

[12] Zheng D, Stevens S, Langley P, Wang K 2008 T-wave Alternans: A Comparison of Different Measurement Techniques *Computers in Cardiology* **35**(1) 597-600
[13] Zeng Z, Pantic M, Roisman G I, Huang T S 2009 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(1) 39-58
[14] Lei S 2013 Research and Implementation of Portable ECG Monitor Detection Algorithm *Xi'an Technological University* 16-24 *(in Chinese)*
[15] Sorensen J S, Johannesen L, Grove U 2010 A Comparison of IIR and Wavelet Filtering for Noise Reduction of the ECG *Computing in Cardiology* **37**(1) 489-92
[16] Omid S, Shamsoilahi M B 2007 Multi-adaptive bionic wavelet transform: Application to ECG denoising and baseline wandering reduction *EURASIP Journal on Advances in Signal Processing* 1-11
[17] Pal S, Mitra M 2010 QRS Complex detection using Empirical Mode Decomposition based windowing technique *2010 International Conference on Signal Processing and Communications (SPCOM)* 1-5
[18] Liu S, Yang L, Fang T 2009 Evacuation from a Classroom Considering the Occupant Density Around Exits *Physica A* **388** 1921-8
[19] Merzougui R, Feham M 2011 Design and implementation of an algorithm for cardiac pathologies detection on mobile phone *Int J Wireless Inf Networks* **18**(1) 11-23
[20] Lin Z, Wang J, Lin B 2011 ECG signal preprocessing based on morphological filtering *Journal of Biomedical Engineering* **28**(2) 365-70
[21] Zheng X, Li Z, Shen L, Ji Z 2008 Detection of QRS Complexes Based On Biorthogonal Spline Wavelet *2008 International Symposium on Information Science and Engineering* 502-6
[22] Lakhwani R, Singh A, Ayub S, Saini J P 2012 Comparison of Different Digital Filters for QRS Complex Extraction from Electrocardiogram *2012 Fourth International Conference on Computational Intelligence and Communication Networks* 276-82
[23] http://www.physionet.org/physiobank/database/mitdb/

Authors

Ying Sun, born in March, 1960, Beijing, China

Current position: Deputy Director at the Institute of Medical Informatics Engineering, Medical Instrument and Food Engineering College, University of Shanghai for Science and Technology.
University studies: Beijing Institute of Science and Technology.
Scientific interests: biomedical signal processing and medical information integration.
Publications: 10 patents, 30 papers.

Meikui Deng, born in January, 1990,Hunan, China

Current position, grades: Second year student at the Medical Instrument and Food Engineering College, University of Shanghai for Science and Technology.
University studies: University of Shanghai for Science and Technology.
Scientific interests: medical signal processing
Publications: 1 patent.

Shenghua Ye, born in December, 1990, Shanghai, China

Current position, grades: Third year student in the Division of Engineering Science, University of Toronto.
University studies: University of Toronto.
Scientific interests: physics, quantum optics and databases.
Publications: 3 papers.

Concurrent quantum key distribution using quantum teleportation and time division multiplexing

Xie Wu[1*], Ouyang Shan[1, 2], Hailin Xiao[2]

[1]*School of Electronic Engineering, Xidian University, 710071, Xi'an, China*

[2]*School of Information and Communication, Guilin University of Electronic Technology, 541004, Guilin, China*

Received 1 May 2014, www.cmnt.lv

Abstract

In view of the problem that it is difficult to use only one typical channel to deal with many quantum channels during multi-channel QKD (Quantum Key Distribution), the method of TDM (Time Division Multiplexing) in typical channel is introduced to construct a new CQKD (concurrent QKD) system. Using multi-channel quantum teleportation, every concurrent quantum channel is mapped to a time slot of TDM. The results of case study with EPR (Einstein-Podolsky-Rosen) pairs show that this problem can be solved with the CQKD methods. Moreover, this CQKD scheme also has the advantage of unconditional security, while the QKD bit rates can be improved. It opens a new way to develop large-capacity long-distance quantum secure communications.

Keywords: quantum information security, concurrent quantum key distribution, quantum teleportation, time division multiplexing

1 Introduction

QKD (Quantum Key Distribution) is a newly emerging information security technique to get quantum keys by sending and receiving unknown quantum bits. It is a very important approach to implement quantum information security based on the interdisciplinary fields of quantum cryptology, quantum computation, quantum mechanics, quantum communication and traditional communication. As one of the strongest guarantees to transmit the secure information of unknown quantum bits for quantum keys, it is also becoming the core part of modern cryptology. Since Bennett and Brassard presented the first QKD protocol in 1984 [1], scholars from many countries have dealt with the theories and technologies of QKD, and a lot of important fruits have been achieved. Some valuable research progresses have been achieved in such fields as BB (Bennett-Brassard) 84 protocol [1], Bennett 92 protocol [2], BBM (Bennett-Brassard-Mermin) 92 protocol [3], Ekert 91 protocol [4] and the QKD protocol via quantum teleportation [5-6], etc.

These developments have promoted greatly the QKD techniques to the actual application. Since the quantum cryptogram technology via QKD is a crucial kind of physical cryptography in the world, the QKD systems have a wide range of competitive strategies applications in such fields as military affairs or wars, business, communication, finance, government, etc. Compared with traditional secure communications, the quantum channel capacities of QKD are far smaller, and the corresponding quantum bit rates of QKD are far lower. In recent years, to improve the performances of quantum communication, the multi-channel or parallel QKD techniques were introduced to the fields of quantum information security. For instance, in 2012, Antonio et al [7] combined the WDM (Wavelength Division Multiplexing) and SCM (Subcarrier Multiplexed) techniques to implement microwave photonics parallel QKD. The multiple keys can be delivered, and the final key rate can be increased, which affords the way to multi-QKD in optical fibre networks. In 2013, Zhao et al [8] utilized the forward spectral filtering structure and presented a scheme of parallel QKD, and the shared key generation rate could be increased enhanced several times in theory. Fang et al presented a CV (continuous variable) parallel QKD scheme with the subcarrier multiplexing technique in 2014 [9], and the total secret key rate was increased. In addition, MIMO (Multiple-Input Multiple-Output)-QKD [10-12] and subcarrier multiplexing QKD [9, 13-16] have also been developed to improve quantum channels. These QKD researches above are the early work of multi-channel QKD system with lots of active exploration and significant progresses. These QKD schemes not only remain the advantage of the unconditional security, but also have great capacity of quantum channels. Multi-channel QKD is becoming novel potential footholds of the most powerful quantum cryptography among diverse information security ways, providing much theoretical and technical support for the developments and designs of long-distance large-capacity QKD systems.

Unfortunately, the quantum channels in these multi-channel QKD schemes are considered more, while the indispensable typical channels are not emphasized much. Only one classical channel is not enough to be mapped to every quantum channels. The quantum key transmission in each quantum channel needs the help of the busy essential

[*]***Corresponding author*** e-mail: xiewu588@126.com

classical channel, which are worth improving. In addition, the quantum computations processes of these multi-channel QKD schemes are very complicate lacking unified explicit quantum channel models. As a result, there are some troubles in practical applications and universal generalizations. From the development perspective of the current quantum technologies, the physical mechanisms of this multi-channel QKD system are so intricate that it has been still in the primary stage. The corresponding successful physical experiments of multi-channel QKD system are not many.

Therefore, motivated by ideas of the parallel QKD [7-9], MIMO-QKD [10-12], subcarrier multiplexing QKD [9, 13-16], QKD via quantum teleportation [5, 6], this paper introduces a novel CQKD (Concurrent QKD) system by applying TDM (time division multiplexing) method in the classical channel and concurrent quantum teleportation as multiple quantum channels. Differently from the QKD schemes [7-16], the quantum key string is transmitted currently by dividing the string into several sub-strings with CQKD. At a certain moment, one of concurrent quantum channels is processed, and the time slot resources of TDM typical channel are utilized fully.

This remainder of this paper is arranged as follows. In Section 2, the CQKD system is constructed using quantum entangled states to teleport multiple unknown quantum states, and Section 3 provides the case study of CQKD system by teleporting unknown quantum states of keys with several EPR (Einstein-Podolsky-Rosen) pairs. The performances of the CQKD are analysed in section 4. Finally in section 5, a brief summary is followed.

2 CQKD system using quantum teleportation and TDM

This CQKD system involves the process that legal sender (Alice) and receiver (Bob) send, transmit and obtain quantum keys through multiple quantum channels. The quantum states of the transmission performance by quantum mechanics are used as quantum information carriers. The core approaches to deal with the steps of CQKD are the quantum keys to transmit in every quantum channels with unknown quantum states. The CQKD system can be implemented using the quantum codebook of concurrent quantum key transmissions.

2.1 THE WHOLE SCHEMATIC CONSTRUCTION OF CQKD SYSTEM USING TDM AND QUANTUM TELEPORTATION

CQKD itself is a highly complex quantum system with the transmission of unknown quantum bits via quantum channels. It is also a difficult structural quantum project for physical experiments. From QKD to CQKD system, a potential problem occurs that it is very difficult to arrange and combine the complicate photonic components for the given QKD system, e.g. BS (Beam Splitter), PBS (polarizing beam splitter), BSM (Bell-state measurement), BBO (Beta Barium Borate) crystal, detectors and etc. In contrast to single channel QKD, it is meaningful to construct the CQKD for the unavoidable fact that the quantum components of CQKD are worth exploring.

To take advantage of high key rates of multi-channel quantum communication and explore novel approaches to transmit quantum key information in several quantum channels, the TDM technology is adopted in the classical channel, and a new CQKD system scheme is constructed as shown in Figure 1 by using quantum teleportation as concurrent quantum channels with quantum entangled states.

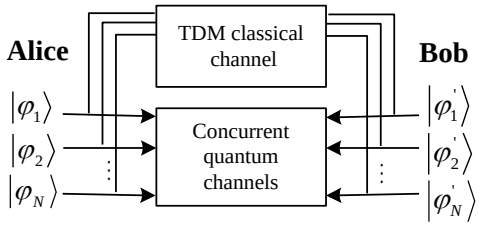


FIGURE 1 the whole schematic construction of CQKD system

Quantum key information of unknown quantum bits is transmitted through concurrent quantum channels. The concurrent processes of QKD can be implemented with N (N is a positive integer) quantum channels. The inputs are the unknown quantum states, while the outputs are the corresponding duplications of these inputs by teleportation with entangled states.

The quantum sources in Figure 1 are discrete with the representations as $|\varphi_1\rangle$, $|\varphi_2\rangle$, … , $|\varphi_N\rangle$, while the destinations are denoted as $|\varphi_1'\rangle$, $|\varphi_2'\rangle$, … , $|\varphi_N'\rangle$. The distances between any two senders (or receivers) are enough without mutual influences or interactions. The number of the senders and receivers of this CQKD system is less than the maximum amount of TDM slots.

2.2 QUANTUM CHANNELS OF CQKD SYSTEM VIA QUANTUM TELEPORTATION

It is feasible to construct the CQKD system via multi-channel quantum teleportation. The characteristics of quantum entangled states with some advantages from microscopic particle world as quantum entanglement resource can be utilized to construct the CQKD system. Therefore, it is necessary to utilize these good properties of quantum entangled particles to design and describe the CQKD system scheme for long-distance key transmission. The concurrent quantum channels can be mapped by diverse quantum entanglement sources, such as EPR states, GHZ (Greenberger-Horne-Zeilinger) states, W states, etc. The transmission distances of quantum bits for quantum key are not limited any more in theory. In quantum channels, no information carriers of quantum key and channels of physical systems for CQKD system interact and influence mutually, and different quantum systems for EPR pairs, GHZ states or W states are often handled differently.

For a CQKD system, the quantum bits are sent and received through multiple quantum channels for a period of time. The channels include TDM typical channel and multiple quantum entangled channels with entangled states. The latter is devised as Figure 2 to illustrate the concurrent processes for QKD with quantum entangled states via concurrent quantum teleportation.

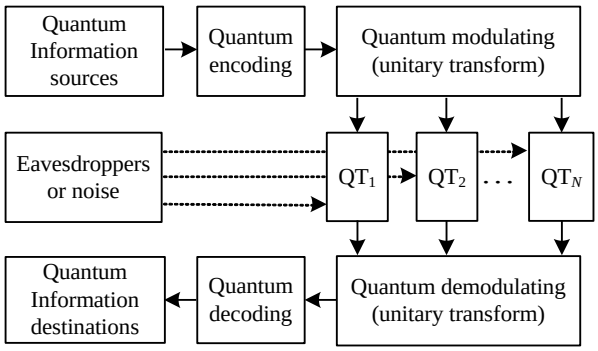


FIGURE 2 Concurrent quantum channels of the CQKD system via quantum teleportation (QT)

The CQKD system in Figure 2 consists of quantum information sources, multiple quantum channels and information destinations. The antennas used in traditional multi-channel communication systems are replaced with several quantum channels where the concurrent quantum bits information is transmitted from the sender to the receiver. The quantum information sources and quantum information destinations are unknown quantum states for quantum keys with the corresponding quantum encoding and quantum decoding processes. The information sources are some quantum bits as the inputs of the CQKD system for Alice, while the outputs are the information destinations as the transmitted results of those quantum bits for Bob. During teleporting of quantum unknown states, the quantum bits of quantum keys in Figure 2 needs modulating and demodulating with unitary transforms.

It is necessary to construct the input-output model of CQKD system using quantum teleportation. Although there are no unified quantitative quantum channel models for the existing multi-channel QKD schemes. Most researchers' schemes for these kinds of QKD system are oriented to their own complex quantum systems, lacking the common formulae of quantum channels. The derivation processes of current theoretical formulae of multi-channel QKD system are exceedingly complicated. It is hard to calculate and compare general quantitative performance indices for QKD. The quantum channels with different quantum carriers are too complicate to describe just like common communications models. Yet, CQKD system emphasizes its multiple inputs and multiple outputs since the inner interacts are too complicate to operate and compute without the unified concept of multi-channel QKD. To abstract the problem of CQKD system with mathematical methods and describe the transmission process of quantum key bits in theory, the CQKD system of N quantum channels is modelled with the abstract input-output relationship between the senders and the receivers.

If none of quantum channels are disturbed by various factors (such as environmental influence, etc.), the general explicit input-output model with quantum entangled states in Figure 1 can be represented as:

$$\begin{bmatrix} |\varphi_1'\rangle \\ |\varphi_2'\rangle \\ \vdots \\ |\varphi_N'\rangle \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}(1|1) & \boldsymbol{P}(1|2) & \cdots & \boldsymbol{P}(1|N) \\ \boldsymbol{P}(2|1) & \boldsymbol{P}(2|2) & \cdots & \boldsymbol{P}(2|N) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{P}(N|1) & \boldsymbol{P}(N|2) & \cdots & \boldsymbol{P}(N|N) \end{bmatrix} \begin{bmatrix} |\varphi_1\rangle \\ |\varphi_2\rangle \\ \vdots \\ |\varphi_N\rangle \end{bmatrix}, \quad (1)$$

where $\boldsymbol{P}(i|j)$ means the operation $\boldsymbol{P}_{ij}|\varphi_i\rangle \to |\varphi_j'\rangle$ in terms of a series of unitary transforms $(\boldsymbol{P}_{ij})$ $i, j = 1, 2, \ldots, N$.

If there is no disturbs in each ideal quantum channel with some independent irrelevant discrete quantum entanglement sources, the input-output model can be simplified as follows:

$$\begin{bmatrix} |\varphi_1'\rangle \\ |\varphi_2'\rangle \\ \vdots \\ |\varphi_N'\rangle \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}(1|1) & 0 & \cdots & 0 \\ 0 & \boldsymbol{P}(2|2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{P}(N|N) \end{bmatrix} \begin{bmatrix} |\varphi_1\rangle \\ |\varphi_2\rangle \\ \vdots \\ |\varphi_N\rangle \end{bmatrix}. \quad (2)$$

In a certain quantum channel, $|\varphi_j'\rangle$ is the input unknown quantum bit which carries the quantum key information to the output $|\varphi_j'\rangle$. The input-output relations of multiple quantum channels are determined by the variable $\boldsymbol{P}(i|j)$. According to quantum mechanics, $\boldsymbol{P}(i|j)$ are the results of a series of unitary transforms which are implemented in physics, which depends on the physical characteristics of quantum channels for CQKD.

2.3 EAVESDROPPING ANALYSIS OF CQKD SYSTEM IN QUANTUM CHANNELS AND TDM CLASSICAL CHANNEL

For the CQKD system, eavesdropping detection is essential in every quantum channels. The error quantum bits for CQKD can be from eavesdroppers and environment noise, which need to be distinguished. Although the quantum key transmission processes for CQKD are affected by environmental noise. For instance, the quantum key transmission in free space may be interfered by high buildings, mountains, dust, atmosphere and noise, etc. Yet concurrent quantum teleportation of unknown quantum bits in free space are affected little by environmental factors. The quantum bits information during teleporting in multiple quantum channels almost has little decay if the quantum entanglement sources are not affected. Potential noise in quantum channels for entangled states differs in free space or optical fibre. So, eavesdropping analysis for CQKD is crucial to obtain correct quantum bits of quantum keys.

Alice often transmits much quantum key information to Bob, and Bob then detects and determines the correctness of the received quantum bits. Alice and Bob need to identify the input and output parameters of quantum states in concurrent quantum channels. They also have to check whether the quantum key bits in concurrent quantum channels are influenced by the eavesdroppers during transmitting the keys. According to the theory of quantum signal detection, the behaviours of eavesdropping can be detectable in quantum systems. For a CQKD system, any eavesdropping in quantum channels will leave traces, providing the chance to detect the error quantum bits information. The quantum keys during concurrent quantum teleportation may be affected by potential eavesdroppers. According to the no-cloning quantum principle, none of them can copy accurately unknown quantum states which are adopted as information sources of this CQKD system. So, whenever the quantum channels between Alice and the Bob for the CQKD system are affected or eavesdropped on, Eve can be detected. In addition, data correction and privacy enhancing process are also required to realize CQKD.

3 Case study of CQKD system by concurrent quantum teleportation with EPR pairs and TDM

The quantum computing processes of a CQKD system need much work, and it is difficult to abstract the unified multiple channel models just like multi-channel quantum communications and the typical communication schemes [7-16]. It is also hard to achieve and solve universal general computational formulas and variables.

In order to simplify and derive the complex computation processes for the CQKD system, the quantum teleportation with EPR pairs are taken as an example of concurrent quantum channels to research the CQKD system according to Equation (2). So, for simply, it is supposed that there are no interacts in multiple quantum channels. Such quantum entanglement states as EPR states are taken as the transmission medium of the unknown quantum state to deal with quantum key bits by concurrent quantum teleportation just for the CQKD system.

3.1 SCHEME OF CQKD SYSTEM VIA CONCURRENT QUANTUM TELEPORTATION WITH EPR PAIRS

For the implementation for the detailed multiple quantum transmission of quantum keys with unknown quantum states, the antennae used in traditional multi-channel communications are replaced with multiple independent purified EPR pairs by the corresponding relationships between the quantum channels and quantum entangled states. These EPR pairs are mapped into the quantum channels for the CQKD system. The classical channel used in the quantum teleportation is divided into many time slots with the way from classical communications (such as a telephone), and the CQKD system scheme with EPR pairs is constructed and shown in Figure 3. Alice and Bob identify, negotiate and determine the unknown quantum bits of QKD through TDM typical channel.

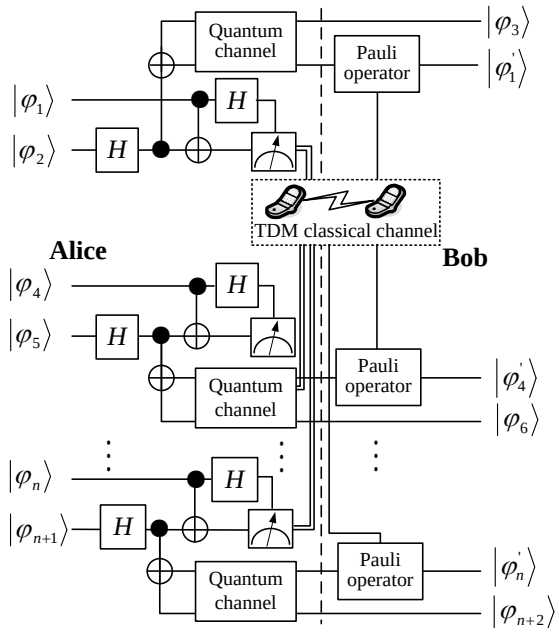


FIGURE 3 CQKD system by quantum teleportation with EPR pairs

H is a Hadamard gate, and $\oplus$ is a control-NOT gate. The quantum bits carrying quantum key information to be transmitted are denoted as the input unknown quantum particles $|\varphi_1\rangle$, $|\varphi_4\rangle$, …, $|\varphi_n\rangle$, while $|\varphi_1'\rangle$, $|\varphi_4'\rangle$, …, $|\varphi_n'\rangle$ is the corresponding outputs.

The EPR pairs in concurrent quantum channels are represented as $|\varphi_2\rangle$ and $|\varphi_3\rangle$, $|\varphi_5\rangle$ and $|\varphi_6\rangle$,… ,$|\varphi_{n+1}\rangle$ and $|\varphi_{n+2}\rangle$, which are from corresponding purified homotype independent quantum entangled sources. n is a positive integer.

Alice and Bob can implement remote quantum correlations with quantum entanglement states. To construct concurrent quantum channels, they can select purified entanglement photon sources to produce EPR pairs with polarization photos as the multiple quantum channels for the CQKD system.

3.2 PROCESSES OF CQKD SYSTEM VIA QUANTUM TELEPORTATION WITH EPR PAIRS

For simplification, three quantum channels for quantum communication are taken as examples to describe the principle of a CQKD system scheme to obtain the input-output transformations relations during the concurrent quantum key transmissions. The process of this scheme via concurrent teleportation with EPR pairs can be divided into the following steps as the main contents of the protocol [5, 6] for the CQKD system.

Firstly, unknown quantum states and EPR entangled pairs are prepared for the CQKD system. Assuming some

independent purified quantum entangled photon sources are prepared in three quantum channels, and these EPR pairs are transmitted to Alice and Bob in the opposite direction respectively. As a result, Alice has quantum particles $|\varphi_2\rangle$, $|\varphi_5\rangle$ and $|\varphi_8\rangle$, while Bob holds the responding quantum particles $|\varphi_3\rangle$ and $|\varphi_6\rangle$, $|\varphi_9\rangle$. The initial quantum states of these EPR pairs are supposed as the following Bell states:

$$|\beta_{00}\rangle_{23} = |\Phi^+\rangle_{23} = (|00\rangle_{23} + |11\rangle_{23})/\sqrt{2}, \tag{3}$$

$$|\beta_{01}\rangle_{56} = |\Psi^+\rangle_{56} = (|01\rangle_{56} + |10\rangle_{56})/\sqrt{2}, \tag{4}$$

$$|\beta_{10}\rangle_{89} = |\Phi^-\rangle_{89} = (|00\rangle_{89} - |11\rangle_{89})/\sqrt{2}, \tag{5}$$

where 23, 56 and 89 are denoted as three pairs of quantum entangled particles respectively.

The quantum key information is contained in three unknown quantum particles $|\phi_1\rangle$, $|\phi_4\rangle$ and $|\phi_7\rangle$ which initial states are represented as:

$$|\varphi_1\rangle = a|0_1\rangle + b|1_1\rangle, \tag{6}$$

$$|\varphi_4\rangle = c|0_4\rangle + d|1_4\rangle, \tag{7}$$

$$|\varphi_7\rangle = e|0_7\rangle + f|1_7\rangle, \tag{8}$$

where *a*, *b*, *c*, *d*, *e* and *f* are complex coefficients.

Theoretically, Alice and Bob are entangled through any quantum channels no matter how far away they are. So, the CQKD through quantum teleportation in every sub-channel is ready and practicable, and the quantum bit information is prepared to be transmitted and gained with a certain probability via quantum entangled states.

Secondly, unknown quantum states through EPR states are teleported and duplicated across three concurrent quantum channels. Alice and Bob teleport unknown quantum bit information for quantum keys in each quantum channel during different time slots of the TDM classical channel. The classical channel is divided into several discrete periodic time slots, so that Alice and Bob can observe and measure quantum entangled states to determine legal quantum bits for the CQKD system with the same physical connection of different TDM time slots.

During these time slots, Alice chooses unknown particles (such as $|\phi_1\rangle$, $|\phi_4\rangle$ and $|\phi_7\rangle$) to perform the joint measurement operations with the corresponding EPR pairs $|\phi_2\rangle$ and $|\phi_3\rangle$, $|\phi_5\rangle$ and $|\phi_6\rangle$, $|\phi_8\rangle$ and $|\phi_9\rangle$ respectively, that is, the tensor products are performed:

$$\begin{bmatrix} |\varphi_{123}\rangle \\ |\varphi_{456}\rangle \\ |\varphi_{789}\rangle \end{bmatrix} = \begin{bmatrix} (a|0_1\rangle + b|1_1\rangle) \otimes (|00\rangle_{23} + |11\rangle_{23})/\sqrt{2} \\ (c|0_4\rangle + d|1_4\rangle) \otimes (|01\rangle_{56} + |10\rangle_{56})/\sqrt{2} \\ (e|0_7\rangle + f|1_7\rangle) \otimes (|00\rangle_{89} - |11\rangle_{89})/\sqrt{2} \end{bmatrix}. \tag{9}$$

According to quantum mechanics, these joint quantum measurement operations will result in the instant collapse changes of each EPR pairs during the corresponding time slot, and the quantum entanglement between Alice and Bob are disentangled, while $|\phi_1\rangle$ and $|\phi_2\rangle$, $|\phi_4\rangle$ and $|\phi_5\rangle$, $|\phi_7\rangle$ and $|\phi_8\rangle$ are entangled separately. So, these quantum information of $|\phi_3\rangle$, $|\phi_6\rangle$ and $|\phi_9\rangle$ for Bob need separating, and the results are the following.

$$\begin{bmatrix} |\varphi_{123}\rangle \\ |\varphi_{456}\rangle \\ |\varphi_{789}\rangle \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} (a|00\rangle_{12} + b|10\rangle_{12})|0\rangle_3 + (a|01\rangle_{12} + b|11\rangle_{12})|1\rangle_3 \\ (c|01\rangle_{45} + d|11\rangle_{45})|0\rangle_6 + (c|00\rangle_{45} + d|10\rangle_{45})|1\rangle_6 \\ (e|00\rangle_{78} + f|10\rangle_{78})|0\rangle_9 - (e|01\rangle_{78} + f|11\rangle_{78})|1\rangle_9 \end{bmatrix}. \tag{10}$$

Using the method of dual items, we have:

$$\begin{bmatrix} |\varphi_{123}\rangle \\ |\varphi_{456}\rangle \\ |\varphi_{789}\rangle \end{bmatrix} = \frac{1}{2\sqrt{2}} \begin{bmatrix} \begin{pmatrix} (a(|00\rangle_{12} + |11\rangle_{12}) + b(|01\rangle_{12} + |10\rangle_{12}) + a(|00\rangle_{12} - |11\rangle_{12}) + b(-|01\rangle_{12} + |10\rangle_{12}))|0\rangle_3 \\ +(b(|00\rangle_{12} + |11\rangle_{12}) + a(|01\rangle_{12} + |10\rangle_{12}) + b(-|00\rangle_{12} + |11\rangle_{12}) + a(|01\rangle_{12} - |10\rangle_{12}))|1\rangle_3 \end{pmatrix} \\ \begin{pmatrix} (d(|00\rangle_{45} + |11\rangle_{45}) + c(|01\rangle_{45} + |10\rangle_{45}) + d(-|00\rangle_{45} + |11\rangle_{45}) + c(|01\rangle_{45} - |10\rangle_{45}))|0\rangle_6 \\ +(c(|00\rangle_{45} + |11\rangle_{45}) + d(|01\rangle_{45} + |10\rangle_{45}) + c(|00\rangle_{45} - |11\rangle_{45}) + d(-|01\rangle_{45} + |10\rangle_{45}))|1\rangle_6 \end{pmatrix} \\ \begin{pmatrix} (e(|00\rangle_{78} + |11\rangle_{78}) + f(|01\rangle_{78} + |10\rangle_{78}) + e(|00\rangle_{78} - |11\rangle_{78}) + f(-|01\rangle_{78} + |10\rangle_{78}))|0\rangle_9 \\ -f(|00\rangle_{78} + |11\rangle_{78}) - e(|01\rangle_{78} + |10\rangle_{78}) + f(|00\rangle_{78} - |11\rangle_{78}) + e(-|01\rangle_{78} + |10\rangle_{78}))|1\rangle_9 \end{pmatrix} \end{bmatrix}. \tag{11}$$

Four Bell states are used for Equation (11), and then we can get:

$$\begin{bmatrix} |\varphi_{123}\rangle \\ |\varphi_{456}\rangle \\ |\varphi_{789}\rangle \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \left(|\beta_{00}\rangle_{12}(a|0\rangle_3 + b|1\rangle_3) + |\beta_{01}\rangle_{12}(b|0\rangle_3 + a|1\rangle_3) + |\beta_{10}\rangle_{12}(a|0\rangle_3 - b|1\rangle_3) + |\beta_{11}\rangle_{12}(-b|0\rangle_3 + a|1\rangle_3)\right) \\ \left(|\beta_{00}\rangle_{45}(d|0\rangle_6 + c|1\rangle_6) + |\beta_{01}\rangle_{45}(c|0\rangle_6 + d|1\rangle_6) + |\beta_{10}\rangle_{45}(-d|0\rangle_6 + c|1\rangle_6) + |\beta_{11}\rangle_{45}(c|0\rangle_6 - d|1\rangle_6)\right) \\ \left((e|\beta_{00}\rangle_{78} + f|\beta_{01}\rangle_{78} + e|\beta_{10}\rangle_{78} - f|\beta_{11}\rangle_{78})|0\rangle_9 + (-f|\beta_{00}\rangle_{78} - e|\beta_{01}\rangle_{78} + f|\beta_{10}\rangle_{78} - e|\beta_{11}\rangle_{78})|1\rangle_9\right) \end{bmatrix}. \quad (12)$$

The Equations (6)-(8) of the unknown particles are substituted into Equation (12), and the three-channel CQKD system via teleportation is expressed as Equation (13):

$$\begin{bmatrix} |\varphi_{123}\rangle \\ |\varphi_{456}\rangle \\ |\varphi_{789}\rangle \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \left(|\beta_{00}\rangle_{12}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}|\varphi_3\rangle + |\beta_{01}\rangle_{12}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}|\varphi_3\rangle + |\beta_{10}\rangle_{12}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}|\varphi_3\rangle + |\beta_{11}\rangle_{12}\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}|\varphi_3\rangle\right) \\ \left(|\beta_{00}\rangle_{45}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}|\varphi_6\rangle + |\beta_{01}\rangle_{45}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}|\varphi_6\rangle + |\beta_{10}\rangle_{45}\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}|\varphi_6\rangle + |\beta_{11}\rangle_{45}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}|\varphi_6\rangle\right) \\ \left(|\beta_{00}\rangle_{78}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}|\varphi_9\rangle + |\beta_{01}\rangle_{78}\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}|\varphi_9\rangle + |\beta_{10}\rangle_{78}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}|\varphi_9\rangle + |\beta_{11}\rangle_{78}\begin{bmatrix}0 & -1\\ -1 & 0\end{bmatrix}|\varphi_9\rangle\right) \end{bmatrix}. \quad (13)$$

Using inverse unitary transform, Bob can duplicates the unknown quantum states $|\phi_1'\rangle$, $|\phi_4'\rangle$, $|\phi_7'\rangle$ according to Equation (13), and $\boldsymbol{P}(i|j)$ in Equation (2) can be implemented. Consequently, the concurrent unknown quantum bits for quantum keys are transmitted from Alice to Bob.

Thirdly, quantum keys are obtained from these teleported unknown quantum states for CQKD system. Through the typical channel, Alice and Bob compare and find out those transmitted quantum bits where the measurement bases are the same during the time slot of the classical channel. They choose effective unknown quantum bits by identical measurement bases to establish the codebook for QKD. The QKD protocols [5, 6] of quantum teleportation can be adopted, and Eve can be checked whether they exist in each quantum channel or not. When the quantum channel is not eavesdropped on, the quantum bits for the corresponding QKD are legal, or the transmitted quantum bits have to be cancelled. By transmitting constantly unknown quantum states, the quantum bit string can be gained and determine whether the transmission process is effective to carry out the collection of quantum key information. After quantum error correcting code and privacy amplification, Alice and Bob combine the effective unknown quantum bits from concurrent quantum channel to build the codebook for the quantum key of the CQKD system.

Thus, results of three-channel CQKD system by concurrent quantum teleportation with EPR pairs are obtained as Equation (13). It is worth mentioning that the CQKD systems of four or more concurrent quantum channels are similar through those quantum channels of EPR pairs and TDM classical channel. The input-output relationship in Equation (13) is also in accordance with the quantum channel model of Equation (2), and the former can be generalized to the CQKD of N quantum channels under the ideal condition of no environmental impacts.

4 Discussion

Similarly, with the existing multi-channel QKD schemes, the CQKD system in this paper also has unconditional security according to the Heisenberg uncertainty principle and the quantum no-cloning theorem. Moreover, this CQKD scheme in our work brings two other advantages.

Firstly, QKD can be implemented through concurrent quantum channels in the CQKD system by TDM methods. Every quantum channel can be mapped with a time slots of classical channel for Alice and Bob to exchange the information of quantum states and measure bases. From the concurrent quantum channels of the instance with EPR pairs, not only the input-output models of are simpler than the multi-channel QKD schemes, but also the concurrent quantum teleportation process of unknown quantum bits can be implemented. So, the CQKD system has the crucial factors of TDM time slots for successful quantum keys, and the time slot resource is utilized fully.

Secondly, this CQKD system improves the performance of QKD bit rates. For the QKD scheme of only one typical channel, the error (from eavesdroppers) probability during transmitting unknown quantum bits is denoted as p_e, then the ideal maximum successful probability of QKD for Alice and Bob is $1-p_e$ at a time. $0 \le p_e \le 1$. In contrast, for this CQKD system with TDM typical channel, Alice can transmit the quantum key information by teleporting multiple unknown quantum states to Bob when one or more of concurrent quantum channels are available. For instance, for a CQKD system of N irrelevant quantum channels, if the classical channel is divided into N time slots, then the ideal maximum successful probability of QKD is $1-p_e^N$ at a time. So, the

maximum probability for Alice and Bob to implement successful transmissions of unknown quantum bits for CQKD rises greatly with the increasing number of concurrent quantum channel. So, the high QKD bit rates for the CQKD system can be obtained by TDM technology, which was not be guaranteed with only one classical channel for every quantum channels before.

5 Conclusion

In this paper, the problem of only one typical channel to multiple quantum channels for multi-channel QKD schemes has been solved by the CQKD system. The TDM tool is used to ensure the one-to-one mapping of a time slot in classical channel and each quantum channel. The case study of EPR pairs illustrate that the model of CQKD system can be constructed via concurrent quantum teleportation. Performance analyses show that the CQKD system not only has the advantage of unconditional security, but also the QKD bit rates can be enhanced in contrast with the single-channel QKD. With the development of quantum information security, the model of CQKD system can be generalized with GHZ states, W states, etc. This CQKD scheme opens a novel way to develop large-capacity long-distance quantum secure communications.

Acknowledgments

We acknowledge the National Natural Science Foundation of China (Nos. 60972084 and 61261018).

References

[1] Bennett C H, Brassard G 1984 Quantum cryptography: public key distribution and coin tossing *Proceeding IEEE Intenational Conference on Computers, Systems and Signal Processing* Bangalore India 175-79
[2] Bennett C H 1992 Quantum cryptography using any two nonorthogonal states *Physical Review Letters* **68**(21) 3121-4
[3] Bennett C H, Brassard G, Mermin N D 1992 *Physical Review Letters* **68**(5) 557-9
[4] Ekert A K 1991 Quantum cryptography based on Bell's theorem *Physical Review Letters* **67**(6) 661-3
[5] Song H C, Gong L H, Zhou N R, 2012 Continuous-variable quantum deterministic key distribution protocol based on quantum teleportation *Acta Physica Sinica* **61**(15) 154206 *(in Chinese)*
[6] Zhou N R, Wang L J, Gong L H, Zuo X W, Liu Y 2011 Quantum deterministic key distribution protocols based on teleportation and entanglement swapping *Optics Communications* **284**(19) 4836-4842
[7] Ruiz-Alba A, Mora J, Amaya W, Martinez A, García-Muñoz V, Calvo D, Capmany J 2012 *IEEE Photonics Journal* **4**(3) 931-942
[8] Zhao G H, Zhao S H, Yao Z S, Meng W, Wang X, Zhu Z H, Liu F 2013 Forward spectral filtering parallel quantum key distribution system *Optics Communications* **298-299** 254-9
[9] Fang J, Huang P, Zeng G H 2014 Multichannel parallel continuous-variable quantum key distribution with Gaussian modulation *Physical Review* A **89**(2) 022315
[10] Gabay M, Arnon S 2006 Quantum key distribution by a free-space MIMO system *Journal of Lightwave Technology* **24**(8) 3114-20
[11] Cui G Q, Lu Y, Zeng G H 2009 A new scheme for quantum key distribution in free-space *Preceeding in 15th Asia-Pacific Conference on Communications* Shanghai China 637-40
[12] Xiao H L, Ouyang S, Nie Z P 2009 Capacity of multiple-input-multiple-output quantum key distribution channels *Acta Physica Sinica* **58**(10) 6779-6785 *(in Chinese)*
[13] Ortigosa-Blanch A, Capmany J 2006 Subcarrier multiplexing optical quantum key distribution *Physical Review* A **73**(2) 024305
[14] Mora J, Ruiz-Alba A, Amaya W, Martinez A, García-Muñoz V, Calvo D, Capmany J 2012 Experimental demonstration of subcarrier multiplexed quantum key distribution system *Optics Letters* **37**(11) 2031-3
[15] Zhao G H, Zhao S H, Yao Z S, Meng W, Wang X, Zhu Z H, Liu F 2012 Subcarrier multiplexing quantum key distribution based on polarization coding *Acta Physica Sinica* **61**(24) 240306 *(in Chinese)*
[16] Bhattacharya S, Krishnamurthy P K 2013 Decoy-state method for subcarrier-multiplexed frequency-coded quantum key distribution *Journal of the Optical Society of America B-Optical Physics* 30(4) 782-7

Authors

Xie Wu, born June, 1979, Yichun, Jiangxi Province, P.R. China

Current position, grades: PhD candidate at the School of Electronic Engineering, Xidian University, Xi'an, China. Senior experimentalist at Guilin University of Electronic Technology since 2011.
University studies: BS and MS degrees at Guilin University of Electronic Technology (GUET), China, in 2002 and 2005.
Scientific interests: quantum information security distribution and group theory.
Publications: more than 5 papers.

Ouyang Shan, born in September, 1960, Anfu, Jiangxi Province, P.R. China

Current position, grades: professor at the School of Information and Communications, Guilin University of Electronic Technology (GUET), China.
University studies: MS and PhD degrees in electronic engineering at Xidian University, Xi'an, China, in 1992 and 2000.
Scientific interests: signal processing for communications and radar, adaptive filtering, and neural network learning theory and applications.
Publications: more than 70 papers.

Hailin Xiao, born in June, 1975, Machen, Hubei Province, P.R. China

Current position, grades: professor at the School of Information and Communications, Guilin University of Electronic Technology (GUET), China.
University studies: PhD degree at the University of Electronic Science and Technology of China (UESTC) in 2007.
Scientific interests: MIMO wireless communications, cooperative communications and smart antenna techniques.
Publications: more than 50 papers.

A few expanding integrable models of WKI hierarchy and their Hamiltonian structures

Kaishi Sun*, Fasheng Liu[2]

[1]*College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266510, China*

[2]*College of Transportation, Shandong University of Science and Technology, Qingdao 266510, China*

Received 1 October 2014, www.cmnt.lv

Abstract

The integrable coupling of the WKI hierarchy is obtained by the perturbation approach. With the help of a higher dimensional loop algebra, the coupling integrable couplings of the WKI hierarchy are obtained, respectively. Their Hamiltonian structures are worked out by employing the component-trace identities and variational identity.

Keywords: coupling integrable couplings, component-trace identities, perturbation equation

1 Introduction

The notion on integrable couplings was introduced when study of Virasoro symmetric algebras [1, 2]. To find as many new integrable systems and their integrable couplings as possible and to elucidate in depth their algebraic and geometric properties are of both theoretical and practical value. During the past few years, some interesting integrable couplings and associated properties of some known interesting integrable hierarchies, such as the AKNS hierarchy, the KN hierarchy, the Burger hierarchy, etc. were obtained [3-13] In order to get Hamiltonian structures of integrable couplings, Guo and Zhang proposed the quadratic-form identity [14]. After this, Ma and Chen [15, 16] built the variational identity and generalized the quadratic-form identity and obtained some integrable couplings and their Hamiltonian structures. Recently, Ma and Zhang [17] proposed the notion on component-trace identities. They are very effective to construct the Hamiltonian structure of the perturbation equation. In Ref. [18], Ma and Gao proposed the notion called coupling integrable couplings of the nonlinear Schr¨odinger equation and associated symmetry properties, etc. Based on this, Zhang and Tam [19] constructed a few higher dimensional Lie algebras to obtain the coupling integrable couplings of the AKNS hierarchy and the KN hierarchy.

In the paper, we first give the first-order perturbation equation of the WKI hierarchy and its Hamiltonian structure is worked out by employing the component-trace identities. Then we use the way presented in Ref. [18-19] to investigate the coupling integrable couplings of the WKI hierarchy. In Refs. [18-19], the author didn't obtain the Hamiltonian structure of the coupling integrable couplings, while in the paper the Hamiltonian structures of the coupling integrable couplings of the WKI hierarchy will be worked out by using the variational identity.

2 The perturbation equation of the WKI hierarchy and its Hamiltonian structure

Yao and Zhang [20] utilized Tu scheme to obtain the multi-component WKI hierarchy. In this section, we take the perturbation way to deduce the integrable coupling of the WKI hierarchy and employ the component-trace identities to generate its Hamiltonian structure.

Consider the isospectral problem of the WKI hierarchy

$$\varphi_x = U\varphi, U = \begin{pmatrix} -i\lambda & u_1\lambda \\ u_2\lambda & i\lambda \end{pmatrix}, \varphi = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}, \qquad (1)$$

where u_1, u_2 are potentials and λ is the spectral parameter. By means of constructing a proper time evolution

$$\varphi_{t_n} = V^{[n]}\varphi, V^{[n]} = \sum_{m=0}^{n} \begin{pmatrix} \lambda a_m & b_{mx} + i\lambda u_1 a_m \\ c_{mx} + i\lambda u_2 a_m & -\lambda a_m \end{pmatrix} \lambda^{n-m}, \qquad (2)$$

and using the zero-curvature equation, we have the WKI hierarchy:

$$u_{t_n} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}_{t_n} = \begin{pmatrix} 0 & -\partial^2 \\ \partial^2 & 0 \end{pmatrix} L^n \begin{pmatrix} \frac{u_2}{\sqrt{1-u_1u_2}} \\ \frac{u_1}{\sqrt{1-u_1u_2}} \end{pmatrix} = JL^n \begin{pmatrix} \frac{u_2}{\sqrt{1-u_1u_2}} \\ \frac{u_1}{\sqrt{1-u_1u_2}} \end{pmatrix}, \qquad (3)$$

**Corresponding author* e-mail: 343927567@qq.com

where:

$$L=\begin{pmatrix} -\frac{i}{2}\partial-\frac{i}{4}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 & \frac{i}{4}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 \\ -\frac{i}{4}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 & \frac{i}{2}\partial+\frac{i}{4}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 \end{pmatrix}. \tag{4}$$

Accordingly, when $n=0$, we can get the WKI equation. By using the race identity, we can obtain the Hamiltonian structure of the WKI hierarchy

$$u_{t_n}=J\frac{\delta H_{n-1}}{\delta u}, \tag{5}$$

where $H_{n-1}=\int\frac{a_{n-1}+u_2b_{n-1}-u_1c_{n-1}}{n-1}dx.$

Next, we will construct a integrable coupling of the WKI hierarchy by the perturbation approach and its Hamiltonian structure by the component-trace identities.

Let us take a matrix Lie algebra g consisting of the following matrices:

$$A=\begin{bmatrix} A_0 & A_1 & \cdots & \cdots & A_N \\ & A_0 & A_1 & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & A_0 & A_1 \\ 0 & & & & A_0 \end{bmatrix}, \tag{6}$$

where $A_i,\ 0\le i\le N,$ are square matrices of the same order. For convenience, we rewrite an element of the Lie algebra g as a vector of matrices: $A=(A_0,A_1,\cdots,A_N)$, where the components $A_i,1\le i\le N$, are defined by:

$$A_i=\frac{1}{i!}\frac{\partial^i}{\partial\varepsilon^i}|_{\varepsilon=0}\ A(\hat{u}_N),\hat{u}_N=u+u+\sum_{i=1}^{N}\varepsilon^i\eta_i,1\le i\le N. \tag{7}$$

The enlarged zero-curvature equation by perturbation

$$\hat{U}_{Nt}-\hat{V}_{Nt}+[\hat{U}_N,\hat{V}_N]=0,$$

where $\hat{U}_N=u+\sum_{i=0}^{N}\varepsilon^i\eta_i$, $\hat{V}_N=V+\sum_{i=0}^{N}\varepsilon^i\eta_i,1\le i\le N$ give rise to the perturbation equation of the Nth order:

$$\hat{\eta}_{Nt}=\hat{K}_N(\hat{\eta}_N)=(K^T(u),\frac{1}{1!}\frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\ K^T(\hat{u}_N),\cdots,$$
$$\frac{1}{N!}\frac{\partial^N}{\partial\varepsilon^N}|_{\varepsilon=0}\ K^T(\hat{u}_N))^T,$$

where the column vector $\hat{\eta}_N$ of dependent variables is $\hat{\eta}_N=(u^T,\eta_1^T,\cdots,\eta_N^T)$. In what follows, we focus on the perturbation equation of the first order. We consider an isospectral problem as follows:

$$\bar{U}(\bar{u},\lambda)=\hat{U}_1=\begin{bmatrix} U_0 & U_1 \\ 0 & U_0 \end{bmatrix},\bar{V}(\bar{v},\lambda)=\hat{V}_1=\begin{bmatrix} V_0 & V_1 \\ 0 & V_0 \end{bmatrix}, \tag{8}$$

where U_0 and V_0 are defined by Equations (1) and (2), U_1 and V_1 are showed as follows:

$$U_1=\frac{1}{1!}\frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\ U(\hat{u}_1)=\frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\begin{bmatrix} -i\lambda & u_1\lambda+\varepsilon\lambda u_3 \\ u_2\lambda+\varepsilon\lambda u_4 & i\lambda \end{bmatrix}= \tag{9}$$
$$\begin{bmatrix} 0 & \lambda u_3 \\ \lambda u_4 & 0 \end{bmatrix},$$

$$V_1=\frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\begin{bmatrix} \lambda a+\varepsilon\lambda d & B \\ C & -\lambda a-\varepsilon\lambda d \end{bmatrix}= \tag{10}$$
$$\begin{bmatrix} \lambda d & f_x+i\lambda u_4a+i\lambda u_1d \\ g_x+i\lambda u_5a+i\lambda u_2d & -\lambda d \end{bmatrix},$$

where:

$$B=b_x+i\lambda u_1a+\varepsilon(f_x+i\lambda u_4a+i\lambda u_1d),$$
$$C=c_x+i\lambda u_2a+\varepsilon(g_x+i\lambda u_5a+i\lambda u_2d).$$

The enlarged stationary zero-curvature equation $\hat{V}_x=[\hat{U},\hat{V}]$ equivalently yields:

$$\begin{cases} V_x=[U,V], \\ V_{1x}=[U,V_1]+[U_1,V]. \end{cases} \tag{11}$$

A direct calculation leads to:

$$\begin{cases} a_{mx} = u_1 c_{mx} - u_2 b_{mx}, \\ i(u_1 a_{m+1})_x + b_{mxx} = -2ib_{m+1x}, \\ i(u_2 a_{m+1})_x + c_{mxx} = 2ic_{m+1x}, \\ d_{mx} = u_3 c_{mx} + u_1 f_{mx} - u_2 e_{mx} - u_4 b_{mx}, \\ f_{mxx} + i(u_1 d_{m+1})_x + i(u_3 a_{m+1})_x = -2if_{m+1x}, \\ g_{mxx} + i(u_2 d_{m+1})_x + i(u_4 a_{m+1})_x = 2ig_{m+1x}, \\ a_0 = \alpha_1, b_0 = \alpha_2, c_0 = \alpha_3, d_0 = \alpha_4, g_0 = \alpha_5, f_0 = \alpha_6, \\ a_1 = \dfrac{2}{\sqrt{1-u_1u_2}}, b_1 = \dfrac{-u_1}{\sqrt{1-u_1u_2}}, c_1 = \dfrac{u_2}{\sqrt{1-u_1u_2}}, \\ d_1 = \dfrac{1}{\sqrt{1-u_1u_2}} + \dfrac{u_1u_4 + u_2u_3}{(1-u_1u_2)^{3/2}}, \\ f_1 = -\dfrac{u_1}{2\sqrt{1-u_1u_2}} - \dfrac{u_1(u_1u_4 + u_2u_3)}{2(1-u_1u_2)^{3/2}} - \dfrac{u_3}{\sqrt{1-u_1u_2}}, \\ g_1 = \dfrac{u_2}{2\sqrt{1-u_1u_2}} + \dfrac{u_2(u_1u_4 + u_2u_3)}{2(1-u_1u_2)^{3/2}} + \dfrac{u_4}{\sqrt{1-u_1u_2}}. \end{cases} \tag{12}$$

From the recursion relation in Equation (12), we have a recursive formula for determining f_n, g_n:

$$\begin{bmatrix} g_n \\ -f_n \end{bmatrix} = L\begin{bmatrix} g_{n-1} \\ -f_{n-1} \end{bmatrix} + L_1\begin{bmatrix} c_{n-1} \\ -b_{n-1} \end{bmatrix} = L\begin{bmatrix} g_{n-1} \\ -f_{n-1} \end{bmatrix} + \begin{bmatrix} L_{11} & L_{12} \\ L_{13} & L_{14} \end{bmatrix}\begin{bmatrix} c_{n-1} \\ -b_{n-1} \end{bmatrix}, \tag{13}$$

where:

$$L_{11} = -\frac{i}{4}\frac{u_4}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 - \frac{i}{8}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^{-1}\Big(\frac{u_1}{\sqrt{1-u_1u_2}}\partial\frac{u_4}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_2}{\sqrt{1-u_1u_2}}\partial\frac{u_3}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_3}{\sqrt{1-u_1u_2}}\partial\frac{u_2}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_4}{\sqrt{1-u_1u_2}}\partial\frac{u_1}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 + 2i\frac{u_3}{\sqrt{1-u_1u_2}}\partial^2\Big),$$

$$L_{12} = \frac{i}{4}\frac{u_4}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 + \frac{i}{8}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^{-1}\Big(\frac{u_1}{\sqrt{1-u_1u_2}}\partial\frac{u_4}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_2}{\sqrt{1-u_1u_2}}\partial\frac{u_3}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_3}{\sqrt{1-u_1u_2}}\partial\frac{u_2}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_3}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_4}{\sqrt{1-u_1u_2}}\partial\frac{u_1}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 + 2i\frac{u_4}{\sqrt{1-u_1u_2}}\partial^2\Big),$$

$$L_{13} = -\frac{i}{4}\frac{u_3}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 - \frac{i}{8}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^{-1}\Big(\frac{u_1}{\sqrt{1-u_1u_2}}\partial\frac{u_4}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_2}{\sqrt{1-u_1u_2}}\partial\frac{u_3}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_3}{\sqrt{1-u_1u_2}}\partial\frac{u_2}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_4}{\sqrt{1-u_1u_2}}\partial\frac{u_1}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^2 + 2i\frac{u_3}{\sqrt{1-u_1u_2}}\partial^2\Big),$$

$$L_{14} = \frac{i}{4}\frac{u_3}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 + \frac{i}{8}\frac{u_1}{\sqrt{1-u_1u_2}}\partial^{-1}\Big(\frac{u_1}{\sqrt{1-u_1u_2}}\partial\frac{u_4}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_2}{\sqrt{1-u_1u_2}}\partial\frac{u_3}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_3}{\sqrt{1-u_1u_2}}\partial\frac{u_2}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_3}{\sqrt{1-u_1u_2}}\partial^2 + \frac{u_4}{\sqrt{1-u_1u_2}}\partial\frac{u_1}{\sqrt{1-u_1u_2}}\partial^{-1}\frac{u_2}{\sqrt{1-u_1u_2}}\partial^2 + 2i\frac{u_4}{\sqrt{1-u_1u_2}}\partial^2\Big).$$

Then the enlarged zero-curvature equation $\hat{U}_{t_n} - \hat{V}_x^{[n]} + [\hat{U}, \hat{V}^{[n]}] = 0$ yields the hierarchy of the first-order perturbation equation:

$$\bar{U}_{t_n} = (u_1, u_2, u_3, u_4)_{t_n}^T = \begin{bmatrix} 0 & J \\ J & J_1 \end{bmatrix} \left(g_{n-1}, -f_{n-1}, c_{n-1}, -b_{n-1}{}^T \right), \quad (14)$$

where

$$J_1 = \frac{1}{1!} \frac{\partial}{\partial \varepsilon} |_{\varepsilon=0} \; J(\hat{u}_1) = \begin{bmatrix} 0 & -\partial^2 \\ \partial^2 & 0 \end{bmatrix}. \quad (15)$$

Hence, we can get the following integrable couplings of Equation (3) as follows:

$$\begin{cases} u_{1t} = \left(\dfrac{iu_1}{\sqrt{1-u_1u_2}} \right)_{xx}, \\ u_{2t} = \left(-\dfrac{iu_2}{\sqrt{1-u_1u_2}} \right)_{xx}, \\ u_{3t} = \left(-\dfrac{u_1}{2\sqrt{1-u_1u_2}} - \dfrac{u_1(u_1u_4+u_2u_3)}{2(1-u_1u_2)^{\frac{3}{2}}} - \dfrac{u_3}{\sqrt{1-u_1u_2}} \right)_{xx}, \\ u_{4t} = \left(\dfrac{u_2}{2\sqrt{1-u_1u_2}} + \dfrac{u_2(u_1u_4+u_2u_3)}{2(1-u_1u_2)^{\frac{3}{2}}} + \dfrac{u_4}{\sqrt{1-u_1u_2}} \right)_{xx}. \end{cases} \quad (16)$$

In order to generate the Hamiltonian structure of the first-order perturbation equation of the WKI hierarchy, we introduce the following theorem:

Theorem 1 [17] Let g be a matrix Lie algebra consisting of block matrices defined by Equation (6). For a given spectral matrix $U = U(u,\lambda) = (U_0, U_1, \cdots, U_N) \in g$, we have the variational identity:

$$\begin{aligned} &\frac{\delta}{\delta u} \int \sum_{k=0}^{N} \alpha_k \operatorname{tr}\left(\sum_{i+j=k} V_i \frac{\partial U_j}{\partial \lambda} \right) dx = \\ &\lambda^{-\gamma} \frac{\partial}{\partial \lambda} \lambda^{\gamma} \sum_{k=0}^{N} \alpha_k \operatorname{tr}\left(\sum_{i+j=k} V_i \frac{\partial U_j}{\partial u} \right), \end{aligned} \quad (17)$$

where $V = V(v,\lambda) = (V_0, V_1, \cdots, V_N) \in g$ satisfies the zero-curvature equation, all α_k's are arbitrary constants with $\alpha_N \neq 0$ and γ is the constant determined by $\gamma = -\frac{1}{2} \frac{d}{d\lambda} ln |< V, V >|$. This variational identity Equation (17) is called the component-trace identity. For a general integer N, we have:

$$\frac{\delta}{\delta u} \int \operatorname{tr}\left(\sum_{i+j=N} V_i \frac{\partial U_j}{\partial \lambda} \right) dx = \lambda^{-\gamma} \frac{\partial}{\partial \lambda} \lambda^{\gamma} \operatorname{tr}\left(\sum_{i+j=N} V_i \frac{\partial U_j}{\partial u} \right), \quad (18)$$

which is called the last-component-trace identity. Then, the generating function of Hamiltonian functions for the perturbation equation of N-th order is computed as follows: $\hat{H}(\hat{\eta}_N) = \operatorname{tr}\left(\sum_{i+j=N} V_i \frac{\partial U_j}{\partial \lambda} \right)$, where $\hat{\eta}_N = (u^T, \eta_1^T, \cdots, \eta_N^T)^T$. This implies that the last-component-trace identity provides the generation function of Hamiltonian functions for the perturbation equation. Then, basing on the generating function of Hamiltonian functions for the original equation $H_g(u) = \operatorname{tr}\left(V \frac{\partial U}{\partial \lambda} \right)$, we can get the generating function of Hamiltonian functions for the perturbation equations of N-th order as follows:

$$\begin{aligned} &\hat{H}_{g,N}(\hat{\eta}_N) = \frac{1}{N!} \frac{\partial^N}{\partial \varepsilon^N} |_{\varepsilon=0} \; H_g(\hat{u}_N) = \\ &\operatorname{tr}\left(\frac{1}{N!} \frac{\partial^N}{\partial \varepsilon^N} |_{\varepsilon=0} \; V(\hat{u}_N) \frac{\partial}{\partial \lambda} U(\hat{u}_N) \right) = \\ &\operatorname{tr}\left(\frac{1}{N!} \sum_{i+j=N} \binom{i}{N} \frac{\partial^i}{\partial \varepsilon^i} |_{\varepsilon=0} \; V(\hat{u}_N) \frac{\partial^j}{\partial \varepsilon^j} |_{\varepsilon=0} \; \frac{\partial}{\partial \lambda} U(\hat{u}_N) \right) = \\ &\operatorname{tr}\left(\sum_{i+j=N} \frac{1}{i!} \frac{\partial^i}{\partial \varepsilon^i} |_{\varepsilon=0} \; V(\hat{u}_N) \frac{\partial}{\partial \lambda} (\frac{1}{j!} \frac{\partial^j}{\partial \varepsilon^j} |_{\varepsilon=0} \; U(\hat{u}_N) \right) = \\ &\operatorname{tr}\left(\sum_{i+j=N} V_i \frac{\partial U_j}{\partial \lambda} \right). \end{aligned} \quad (19)$$

By using the above results, a direct calculation reads

$$\begin{aligned} &\operatorname{tr}\left(\sum_{i+j=1} V_i \frac{\partial U_j}{\partial \lambda} \right) = \operatorname{tr}\left(V_1 \frac{\partial U_0}{\partial \lambda} + V_0 \frac{\partial U_1}{\partial \lambda} \right) = \\ &(-2i\lambda d - 2iu_2 f + 2iu_1 g - 2iu_4 b + 2iu_3 c), \end{aligned} \quad (20)$$

$$\begin{aligned} &\operatorname{tr}(V_1 \frac{\partial U_0}{\partial u_1}) = 2i\lambda g, \operatorname{tr}\left(V_1 \frac{\partial U_0}{\partial u_2} \right) = -2i\lambda f, \\ &\operatorname{tr}\left(V_0 \frac{\partial U_1}{\partial u_3} \right) = 2i\lambda c, \operatorname{tr}\left(V_0 \frac{\partial U_1}{\partial u_4} \right) = -2i\lambda b. \end{aligned} \quad (21)$$

Basing on the last-component-trace identities, we have

$$\begin{aligned} &\frac{\delta}{\delta u} \int (-2id_{n-1} - 2iu_2 f_{n-1} + 2iu_1 g_{n-1} - \\ &2iu_4 b_{n-1} + 2iu_3 c_{n-1}) dx = 2i(2+\gamma-n)[g, -f, c, -b]^T. \end{aligned} \quad (22)$$

Take $n = 2$ in above equation gives $\gamma = -1$. Thus, the Hamiltonian structure of the perturbation equation of the WKI hierarchy is derived as follows:

$$\hat{U}_{t_n} = (u_1, u_2, u_3, u_4)_{t_n}^T = \begin{bmatrix} 0 & J \\ J & J_1 \end{bmatrix} \frac{\delta H_{n-1}}{\delta u}, \quad (23)$$

where:

$$H_{n-1} = \frac{1}{n-1} \int (d_{n_1} + u_2 f_{n-1} - u_1 g_{n-1} + u_4 b_{n-1} - u_3 c_{n-1}) dx.$$

3 The coupling integrable couplings of the WKI hierarchy and its Hamiltonian structure

The coupling of the WKI hierarchy is given in the above. In the section, we will construct the coupling integrable couplings by following the way in Ref. [18], which is introduced as follows.

Given two integrable couplings of the integrable equation $u_t = K(u)$:

$$\bar{u}_{1,t} = \bar{K}_1(\bar{u}_1) = \begin{bmatrix} K(u) \\ S(u,v) \end{bmatrix}, \bar{u}_1 = \begin{bmatrix} u \\ v \end{bmatrix}, \tag{24}$$

$$\bar{u}_{2,t} = \bar{K}_2(\bar{u}_2) = \begin{bmatrix} K(u) \\ T(u,v) \end{bmatrix}, \bar{u}_2 = \begin{bmatrix} u \\ w \end{bmatrix}, \tag{25}$$

a new bigger system is formed as follows:

$$\hat{u}_t = \hat{K}(\hat{u}) = \begin{bmatrix} K(u) \\ S(u,v) \\ T(u,v) \end{bmatrix}, \hat{u} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}. \tag{26}$$

We call Equation (26) coupling integrable couplings of Equations (24) and (25).

First, we will construct a 9-dimensional vector-Lie algebra and its corresponding loop algebra. Consider a vector space [18]:

$$R^9 = span\{a = (a_1, \cdots, a_9)^T, a_i \in R, i = 1, 2, \cdots, 9\}.$$

For $\forall a = (a_1, \cdots, a_9)^T$, $b = (b_1, \cdots, b_9)^T$, define a commutation operation:

$$\begin{aligned}[a,b] = &(a_2b_3 - a_3b_2, 2(a_1b_2 - a_2b_1), 2(a_3b_1 - a_1b_3), a_2b_6 - a_6b_2 + a_5b_3 - a_3b_5, 2(a_1b_5 - a_5b_1 + a_4b_2 - a_2b_4),\\ &2(a_6b_1 - a_1b_6 + a_3b_4 - a_4b_3), a_2b_9 - a_9b_2 + a_8b_3 - a_3b_8 + a_8b_9 - a_9b_8, 2(a_1b_8 - a_8b_1 + a_7b_2 - a_2b_7 + a_7b_8 - a_8b_7),\\ &2(a_9b_1 - a_1b_9 + a_3b_7 - a_7b_3 + a_9b_7 - a_7b_9))^T\end{aligned} \tag{27}$$

It is verified that R^9 is a Lie algebra if equipped with Equation (27). Take a basis of R^9 as follows:

$$e_i = (e_{i1}, \cdots, e_{i9})^T, e_{ij} = \begin{cases} 1, i = j, \\ 0, i \neq j, 1 \leq i, j \leq 9. \end{cases} \tag{28}$$

A loop algebra $\tilde{R}^9$ corresponding to the Lie algebra R^9 is defined as:

$$\tilde{R}^9 = span\{e_i(n)\}_{i=1}^9, e_i(n) = e_i\lambda^n, [e_i(m) - e_j(n)] = [e_i, e_j]\lambda^{m+n}, 1 \leq i, j \leq 9, m, n \in Z. \tag{29}$$

By employing the loop algebra $\tilde{R}^9$, we consider the following Lax pair:

$$\begin{cases} U = -ie_1(1) + u_1e_2(1) + u_2e_3(1) + u_3e_5(1) + u_3e_6(1) + u_5e_8(1) + u_6e_9(1), \\ V = \sum_{m=0}^{\infty}[a_me_1(1-m) + iu_1a_me_2(1-m) + b_{mx}e_2(-m) + iu_2a_me_3(1-m) + c_{mx}e_3(-m)n + \\ d_me_4(1-m) + f_{mx}e_5(-m) + i(u_3a_m + u_1d_m)e_5(1-m) + g_{mx}e_6(-m) + i(u_4a_m + u_2d_m)e_6(1-m) + . \\ h_me_7(1-m) + p_{mx}e_8(-m) + i(u_5a_m + u_5h_m + u_1h_m)e_8(1-m) + q_{mx}e_9(-m)n + \\ i(u_6a_m + u_6h_m + u_2h_m)e_9(1-m)] \end{cases}$$

A solution to $V_x = [U, V]$ exhibits that:

$$\begin{cases} a_{mx} = u_1c_{mx} - u_2b_{mx}, \\ i(u_1a_{m+1})_x + b_{mxx} = -2ib_{m+1x}, \\ i(u_2a_{m+1})_x + c_{mxx} = 2ic_{m+1x}, \\ d_{mx} = u_1g_{mx} - u_2f_{mx} + u_3c_{mx} - u_4b_{mx}, \\ i(u_3a_{m+1} + u_1d_{m+1})_x + f_{mxx} = -2if_{m+1x}, \\ i(u_4a_{m+1} + u_2d_{m+1})_x + g_{mxx} = 2ig_{m+1x}, \\ h_{mx} = u_1q_{mx} - u_2p_{mx} + u_5c_{mx} - u_6b_{mx} + u_5q_{mx} - u_6p_{mx}, \\ i(u_5a_{m+1} + u_5h_{m+1} + u_1h_{m+1})_x + p_{mxx} = -2ip_{m+1x}, \\ i(u_6a_{m+1} + u_6h_{m+1} + u_2h_{m+1})_x + q_{mxx} = 2iq_{m+1x}. \end{cases} \tag{30}$$

Note:

$$V_+^{(n)} = \sum_{m=0}^{n}[a_m e_1(n+1-m) + iu_1 a_m e_2(n+1-m) + b_{mx} e_2(n-m) + iu_2 a_m e_3(n+1-m) +$$
$$c_{mx} e_3(n-m) + d_m e_4(n+1-m) + f_{mx} e_5(n-m) + i(u_3 a_m + u_1 d_m) e_5(n+1-m) + g_{mx} e_6(n-m) +$$
$$i(u_4 a_m + u_2 d_m) e_6(n+1-m) + h_m e_7(n+1-m) + p_{mx} e_8(n-m) + i(u_5 a_m + u_5 h_m + u_1 h_m) e_8(n+1-m) +$$
$$q_{mx} e_9(n-m) + i(u_6 a_m + u_6 h_m + u_2 h_m) e_9(n+1-m)].$$

Set $V_-^{(n)} = \lambda^n V - V_+^{(n)}$, we have:

$$-V_{+x}^{(n)} + [U, V_+^{(n)}] = -\lambda b_{n-1xx} e_2(0) - \lambda c_{n-1xx} e_3(0) - \lambda f_{n-1xx} e_5(0) - \lambda g_{n-1xx} e_6(0) - \lambda p_{n-1xx} e_8(0) - \lambda q_{n-1xx} e_9(0). \quad (31)$$

Therefore, the zero-curvature equation

$$U_t - V_x^{(n)} + [U, V^{(n)}] = 0, \quad (32)$$

admits the following bigger integrable system:

$$U_t = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{pmatrix}_t = \begin{pmatrix} 0 & 0 & 0 & -\partial^2 & 0 & 0 \\ 0 & 0 & \partial^2 & 0 & 0 & 0 \\ 0 & -\partial^2 & 0 & 0 & 0 & -\partial^2 \\ \partial^2 & 0 & 0 & 0 & \partial^2 & 0 \\ 0 & 0 & 0 & -\partial^2 & 0 & -\partial^2 \\ 0 & 0 & \partial^2 & 0 & \partial^2 & 0 \end{pmatrix} \begin{pmatrix} c_{n-1} + g_{n-1} + q_{n-1} \\ -b_{n-1} - f_{n-1} - p_{n-1} \\ c_{n-1} \\ -b_{n-1} \\ c_{n-1} + q_{n-1} \\ -b_{n-1} - p_{n-1} \end{pmatrix}. \quad (33)$$

Substituting $u_5 = u_6 = 0$ in Equation (33) reduces to the Equation (14), which is an integrable coupling of the WKI hierarchy; when taking $u_3 = u_4 = 0$ in Equation (33) reduces to another integrable coupling of the WKI hierarchy. So we call Equation (33) the coupling integrable couplings of the WKI hierarchy.

In order to deduce to the Hamiltonian structure of Equation (33), we rewrite Equation (30) as follows:

$$[a, b] = a^T R(b), \quad (34)$$

where:

$$R(b) = \begin{pmatrix} 0 & 2b_2 & -2b_3 & 0 & 2b_5 & -2b_6 0 & 2b_8 & -2b_9 & 0 \\ b_3 & -2b_1 & 0 & b_6 & -2b_4 & 0 & b_9 & -2b_7 & 0 \\ -b_2 & 0 & 2b_1 & -b_5 & 0 & 2b_4 & -b_8 & 0 & 2b_7 \\ 0 & 0 & 0 & 0 & 2b_2 & -2b_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_3 & -2b_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -b_2 & 0 & 2b_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2(b_2 + b_8) & -2(b_3 + b_9) \\ 0 & 0 & 0 & 0 & 0 & 0 & b_3 + b_9 & -2(b_1 + b_7) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -(b_2 + b_8) & 0 & 2(b_1 + b_7) \end{pmatrix}.$$

Solving the matrix equation for the constant matrix F:

$$R(b)F = -(R(b)F)^T, F^T = F,$$

$$F=\begin{pmatrix}2&0&0&2&0&0&2&0&0\\0&0&1&0&0&1&0&0&1\\0&1&0&0&1&0&0&1&0\\2&0&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0\\2&0&0&0&0&0&2&0&0\\0&0&1&0&0&0&0&0&1\\0&1&0&0&0&0&0&1&0\end{pmatrix}. \tag{35}$$

Then in terms of F, define a linear functional: $\{a,b\}=a^TFb$. Rewrite the Lax pair as follows:

$$\begin{cases}U=(-i\lambda,\lambda u_1,\lambda u_2,0,\lambda u_3,\lambda u_4,0,\lambda u_5,\lambda u_6)^T,\\ V=(\lambda a,b_x+i\lambda u_1a,c_x+i\lambda u_2a,\lambda d,f_x+i\lambda(u_3a+u_1d),g_x+i\lambda(u_4a+u_2d),\lambda h,\\ p_x+i\lambda(u_5a+u_5h+u_1h),q_x+i\lambda(u_6a+u_5h+u_1h))^T.\end{cases} \tag{36}$$

By using the above linear functional, we have:

$$\left\{V,\frac{\partial U}{\partial\lambda}\right\}=-2i(a+d+h)+2iu_1(c+g+q)-2iu_2(b+f+p)+2iu_3c-2iu_4b+2iu_5(c+q)-2iu_6(b+p),$$

$$\left\{V,\frac{\partial U}{\partial u_1}\right\}=2i\lambda(c+g+q),\left\{V,\frac{\partial U}{\partial u_2}\right\}=-2i\lambda(b+f+p),\left\{V,\frac{\partial U}{\partial u_3}\right\}=2i\lambda c,\left\{V,\frac{\partial U}{\partial u_4}\right\}=-2i\lambda b,$$

$$\left\{V,\frac{\partial U}{\partial u_5}\right\}=2i\lambda(c+q),\left\{V,\frac{\partial U}{\partial u_6}\right\}=-2i\lambda(b+p).$$

According the variational identity, we have

$$\frac{\delta}{\delta u}\int(-2i(a+d+h)+2iu_1(c+g+q)-2iu_2(b+f+p)+2iu_3c-2iu_4b+2iu_5(c+q)-2iu_6(b+p))dx=$$

$$\lambda^{-\gamma}\frac{\partial}{\partial\lambda}\lambda^{\gamma}(2i\lambda(c+g+q),-2i\lambda(b+f+p),2i\lambda c,-2i\lambda b,2i\lambda(c+q),-2i\lambda(b+p))^T.$$

Comparing the coefficients of λ^{-n+1} yields

$$\frac{\delta}{\delta u}\int(-2i(a_{n-1}+d_{n-1}+h_{n-1})+2iu_1(c_{n-1}+g_{n-1}+q_{n-1})-2iu_2(b_{n-1}+f_{n-1}+p_{n-1})+2iu_3c_{n-1}-$$

$$2iu_4b_{n-1}+2iu_5(c_{n-1}+q_{n-1})-2iu_6(b_{n-1}+p_{n-1}))dx=2i(2+\gamma-n)\begin{pmatrix}-c_{n-1}-g_{n-1}\\b_{n-1}+f_{n-1}\\-c_{n-1}\\b_{n-1}\end{pmatrix}.$$

Taking $n=2$ in above equation gives $\gamma=-1$.

Hence, the coupling integrable couplings of WKI hierarchy Equation (33) can be written as a Hamiltonian form:

$$U_t = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{pmatrix}_t = \begin{pmatrix} 0 & 0 & 0 & -\partial^2 & 0 & 0 \\ 0 & 0 & \partial^2 & 0 & 0 & 0 \\ 0 & -\partial^2 & 0 & 0 & 0 & -\partial^2 \\ \partial^2 & 0 & 0 & 0 & \partial^2 & 0 \\ 0 & 0 & 0 & -\partial^2 & 0 & -\partial^2 \\ 0 & 0 & \partial^2 & 0 & \partial^2 & 0 \end{pmatrix} \begin{pmatrix} c_{n-1} + g_{n-1} + q_{n-1} \\ -b_{n-1} - f_{n-1} - p_{n-1} \\ c_{n-1} \\ -b_{n-1} \\ c_{n-1} + q_{n-1} \\ -b_{n-1} - p_{n-1} \end{pmatrix} = J \frac{\delta H_{n-1}}{\delta u}, \tag{37}$$

where

$$H_{n-1} = \frac{1}{n-1} \int ((a_{n-1} + d_{n-1} + h_{n-1}) - u_1(c_{n-1} + g_{n-1} + q_{n-1}) + u_2(b_{n-1} + f_{n-1} + p_{n-1}) -$$
$$u_3 c_{n-1} + u_4 b_{n-1} - u_5(c_{n-1} + q_{n-1}) + u_6(b_{n-1} + p_{n-1}))dx,$$

J is a Hamiltonian operator.

4 Conclusions

In Ref.[19], the coupling integrable couplings of KN and AKNS hierarchy are obtained, but their Hamiltonian structures aren't given. In the paper, however, the integrable coupling of WKI hierarchy is obtained by the perturbation approach and its Hamiltonian structure is given by using the component-trace identities. Meanwhile, basing on a 9-dimensional Lie algebra, we discuss the coupling integrable couplings of the WKI hierarchy and obtain its Hamiltonian structure by the variational identity.

In the future, we will discuss the perturbation equation and coupling integrable couplings of other hierarchies and their Hamiltonian structures.

Acknowledgements

This work was supported by Nature Science Foundation of Shandong Province of China (No.ZR2013AQ017), Open Fund of the Key Laboratory of Ocean Circulation and Waves, Chinese Academy of Science (No. KLOCAW1401).

References

[1] Ma W X, Fuchssteiner B 1996 *Chaos, Solitons and Fractals* **7**(8) 1227-50
[2] Ma W X 2000 *Methods and Applications of Analysis* **7**(1) 21-56
[3] Xia T C 1999 *Acta Math Phys* **19** 507
[4] Fan E G, Zhang Y F 2005 *Chaos, Solitons and Fractals* **25**(2) 425
[5] Zhang Y F, Guo F K 2006 *Communications in Theoretical Physics* **46**(5) 812-8
[6] Xia T C, You F C 2006 *Chaos, Solitons and Fractals* **26**(2) 605-13
[7] Guo F K, Zhang Y F 2006 *Communications in Theoretical Physics* **45**(5) 799-801
[8] Dong H H, Liang X Q 2008 *Chaos, Solitons and Fractals* **38**(2) 548-55
[9] Zhang Y F, Zhang H Q 2002 *Journal of Mathematical Physics* **43**(1) 466-72
[10] Xia T C, You F C, Chen D Y 2005 *Chaos, Solitons and Fractals* **24**(3) 877-83
[11] Li Z, Zhang Y J, Dong H H 2007 *Modern Physics Letters B* **21**(10) 595
[12] Fan E G, Zhang Y F 2006 *Chaos, Solitons and Fractals* **28**(4) 966-71
[13] Ma W X, Xu X X, Zhang Y F 2006 *Physics Letters A* **351** 125-130
[14] Guo F K, and Zhang Y F 2005 *Journal of Physics A: Mathematical and General* **38** 8537
[15] Ma W X and M.Chen 2006 *Journal of Physics A: Mathematical and General* **39** 10787-801
[16] Ma W X 2009 *Nonlinear Analysis: Theory, Methods & Applications* **71** e1716-26
[17] Ma W X and Zhang Y F 2010 *Applicable Analysis* **89** 457-72
[18] Ma W X, Gao L 2009 *Modern Physics Letters B* **23**(15) 1847-60
[19] Zhang Y F, Tam H W 2011 *Communications in Nonlinear Science Numerical Simulations* **16**(1) 76-85
[20] Yao Y Q, Zhang Y F 2005 *Chaos, Solitons and Fractals* **26**(4) 1087-9

Authors

Kaishi Sun, born in August, 1981, Qingdao, China

Current position, grades: lecturer in Shandong University of Science and Technology.
University studies: Shandong University of Science and Technology d in 2004.
Scientific interests: control theory and control engineering.

Fasheng Liu, born in June, 1957, Qingdao, China

Current position, grades: professor and doctoral supervisor at Dalian University of Technology. Director of the Computer Society in Shandong Province and in charge of the traffic discipline in Shandong University of Science and Technology
University studies: PhD in advance professional learning and cybernetics at Dalian University of Technology
Scientific interests: entropy optimization methods and applications, intelligent transportation.

Automated unit-level testing of java memory leaks

Lijuan Hong*, Ju Qian, Jifeng Cui

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Received 15 December 2013; Accepted 29 August 2014, www.cmnt.lv

Abstract

Java programs may suffer from serious memory leak bugs. To resolve these bugs, various leak diagnosing and even fixing techniques have been proposed. However, in literature, there is very few work, which focuses on memory leak testing. Without revealing leak phenomenon by testing in advance, even excellent leak diagnosing and fixing techniques can be hard to work. In software testing, unit testing is a technique to avoid faults at early stage of software development. This paper proposes an automated unit-level memory leak testing approach to find potential leak bugs in Java methods. It firstly identifies the methods with high leaking risks. Then, strengthened unit tests are generated accordingly to check whether those risky modules can really cause leaks. Cases studies show that our method could be valuable for real programs.

Keywords: Java, memory leak, unit testing, test generation

1 Introduction

Even though with garbage collection supports, memory leak still remains a problem for Java programs. The leaks usually occur when a Java program unnecessarily maintains references to objects that are no longer required. Memory Leaks may degrade runtime performance and even lead to crashes due to out of memory exceptions.

To resolve these leak bugs, various techniques have been proposed [1-7], and there are also a lot of supporting tools [11, 12]. The previous work mainly focuses on leak diagnosing and fixing, which find out the causes of leaks after memory leak phenomenon occur [1-7] and fix the leaking code [8-10]. However, in literature, there is very few work concerning on how to discover those potential memory leak phenomenon. Without a discovered leak phenomena, in most cases, it will be hard to trigger a diagnosis process and eliminate the leak error.

Software testing is a promising technique to discover potential failures. But approaches for testing memory leaks are hard to see. In this paper, we present an automated testing method to find memory leak phenomenon at unit level. The approach firstly finds out the leak risky methods and then generates normal unit tests for them. We then strengthen these unit tests to detect memory leaks. By testing leaks at unit-level, memory leaks can be avoided as early as possible in the whole lifecycle of software.

In the work, we classify the leak risky modules into three categories: modules directly creating and leaking large number of objects, modules that accumulate new objects created by themselves and may lead to leaks after repeated calls to them, modules that absorb their arguments and may lead to leaks after repeated calls. These modules can be identified with dependency, points-to, and escape information.

For those leak risky modules, we firstly use some existing approaches [13] to generate normal unit tests as start points. Then, these tests are augmented with large input data, repeated calls, and other techniques to strengthen the memory usage and monitor mechanisms to observe the leak. Finally, we can get the leak revealing unit tests. We studied usage of our approach on several already found memory leak bugs from JDK bug database. The results show that the approach is effective in revealing real leaking modules. This indicates that it can be valuable for practical uses.

2 Technique backgrounds

2.1 DEPENDENCE ANALYSIS

Dependency between program statements can be categorized into two types: control dependence caused by control structures in the program and data dependence caused by reads and writes of memory locations [14].

We can get control and data dependences between nodes in control flow graph via dependence analysis [14]. In this paper, we need to know whether a loop's control condition is influenced by method parameters. By dependence analysis, we can get the data dependence relationships between program nodes, and then judge whether the loop's condition node directly or indirectly depends on method parameter by checking whether there is a path from the method parameter to the condition node. If there exist such kinds of paths, it indicates that the loop's control condition is potentially influenced by method's parameters.

2.2 POINTER AND ESCAPE ANALYSIS

Pointer analysis determines all the possible memory locations that a pointer may point to at runtime. In Java, a pointer is a variable of reference type and what it points to is an object on the heap [15].

This paper uses a context-insensitive algorithm implemented in Soot [17] to do the pointer analysis.

*Corresponding author e-mail ljhong307@163.com

Escape analysis tracks the lifetime of objects and determine whether it may escape from some given scopes. An object can be directly created by a new instruction or indirectly created by a wrapper method of some new instruction. It is considered to possibly escape from the scope of a method if a reference to the object is returned from the method, or if a reference to the object is assigned to a field of an external object [16].

This paper extends the escape analysis presented in [16] to do risky method analysis. In addition to that in [16], the extended analysis also adds loops as the analysed scopes. We firstly identify all objects created in the loops. Then, a constraint system is built according to the statements in the relevant method following the constraint-based approach in [16]. Finally, we can determine whether the objects in the loop may escape loop scope by the extended escape analysis.

3 Leak risky modules: a classification

As discussed in the introduction part, the leak risky modules can be classified into three categories. We will introduce them in more detail in this section.

3.1 MODULES DIRECTLY LEAKING OBJECTS

Modules directly creating and leaking objects have the following characteristics:

1) Creating objects repeatedly in loops inside the module;

2) The number of rounds that these loops can execute is unbounded and determined by the module inputs;

3) The objects created by the loops cannot be released on time. Such modules may directly consume large memory when the inputs are large. The memory requirement may be caused by improper object allocation and release mechanism and is unnecessary. It may cause memory leaks.

```
class Test1{
1    static Vector cache = new Vector();
2    public void foo(int n){
3       for( int i = 1; i < n; i ++ ){
4          Data d = new Data();
5          cache.add(d);
6          doSth(d);
7       }
8    }
}
```

FIGURE 1 An example of modules directly leaking objects

Figure 1 demonstrates an example for the modules that directly creating and leaking objects. In Figure 1, method foo may directly cause memory leaks. In foo, there is a loop which creates new objects inside it. The execution rounds of the loop are unbounded and determined by the method's input parameter *n*. During each round, object d created in the loop is added into an external cache. The cached objects are not freed on time. Given a very large input, the method may directly consume too much memory and lead to out of memory error.

3.2 MODULES ACCUMULATING NEWLY CREATED OBJECTS

Modules that accumulate newly created objects and can lead to memory leak usually have the following characteristics:

1) Creating objects inside the module;

2) The objects escape from module's scope and get stored through a way other than the method return value and parameters. It may unconsciously consume large memory after repeated calls and thereby cause memory leaks.

```
class Test2{
1    static Vector cache = new Vector();
2    public Data bar(){
3       Data d = new Data();
4       cache.add(d);
5       return d;
6    }
  }
```

FIGURE 2 An example of modules accumulating newly created objects

Figure 2 demonstrates an example for the modules of the second category. In Figure 2, method bar creates a new object d and accumulates it into container cache. If the cache is not cleaned on time, after a large number of calls to the bar() method, there will be too many Data objects stored in the container cache, which may cause memory leaks.

3.3 MODULES ABSORBING ARGUMENT OBJECTS

A module that absorbs argument objects and can lead to memory leak usually has at least one of its parameter objects potentially escaping out of the module's scope through a way other than the method return value and parameters. The escaped parameter object can be long-termly absorbed by the module. It may lead to memory leaks after many calls to the module.

```
class Test3{
1    static Vector cache = new Vector();
2    public void zar(Data d){
3       cache.add(d);
4       doSth(d);
5    }
  }
```

FIGURE 3 An example of modules absorbing argument objects

Figure 3 shows an example for the modules of this category. In Figure 3, a reference type parameter is passed into method zar in line 2. In line 3, parameter d is absorbed by method zar to an external cache. The method may unconsciously absorb too many parameter objects after repeated calls, which may causes leaks and finally lead to out of memory error.

4 Identifying Leak Risky Modules

This section presents the methods for identifying three kinds of leak risky modules, respectively.

4.1 IDENTIFYING MODULES DIRECTLY LEAKING OBJECTS

We firstly search the loops that create new objects in a method on the control flow graph. Then whether these loops' control conditions can be influenced by the method inputs are determined. Finally, we use pointer and escape analyses to determine whether the new objects' created in those loops can live beyond the loop scope. If the condition of a loop that creates new objects is potentially influenced by the method inputs and the created objects can live beyond the loop scope, it indicates that the module has a high risk in directly leaking huge memory.

Algorithm 1: Identifying modules that directly leaking objects

```
Input: m: Method
Output: Boolean
   Dependence analysis;
pointer and escape analyses;
let L_m be the set of all loops in m
foreach l∈L_m do
     if hasNewInstruction(l) then
      if isInputDependent(l) then
if isNewObjectEscape(l) then
return true;
end
        end
    end
  end
return false;
```

The algorithm is shown in Algorithm 1. It returns a Boolean value for each input method m to show whether the method may cause leaks. The algorithm firstly collects Lm, the set of all loops in m, by loop analysis. Then, each loop l is processed. We firstly check whether loop l can introduce new objects by predicate hasNewInstruction(l). The new objects include the ones created directly by new instructions and the ones created by other callee methods. Then we determine whether loop l is input dependent on the method's parameters by predicate isInputDependent(l). Finally, we use predicate isNewObjectEscape(l) to check whether the newly introduced objects may escape from the loop scope. When all the above conditions are satisfied, it indicates that method m is a leak risky module.

For isInputDependent(l), we firstly obtain the dependence node corresponding to the loop condition. Starting from this node, we traverse the program dependence graph to get a set of nodes that the loop condition node depends on. If the set contains any node corresponding to the method's parameter, it indicates loop l's condition depends on method inputs.

For isNewObjectEscape(l), we firstly check whether the objects created in the loop escape from the method scope by existing escape analysis. If they escape, of course the objects escape from the loop scope. Otherwise, we will check whether the objects created by the loop may escape from the loop scope by our extended escape analysis.

4.2 IDENTIFYING MODULES ACCUMULATING NEWLY CREATED OBJECTS

We firstly obtain all the objects newly created in a module. Then, pointer and escape analysis are used to determine lifetime of these objects and whether they may escape from the method scope via internal leak sources and thereby be accumulated.

The algorithm is shown in Algorithm 2. It returns a Boolean value for each input method m to show whether the method may cause leaks after repeated calls. We firstly obtain a collection of internal escape sources excluding the return value and parameters by function getInternalLeakSources(). If method m accumulates its created objects, the new objects will escape from these sources. With these special escape sources, we do escape analysis for the method. Having got the escape information, we check each newly created object in the method, and finally determine whether there is any object escaping from the method scope by predicate isEscape(o). If such objects exist, the method may potentially cause memory leaks.

Algorithm 2: Identifying modules that accumulating newly created objects

```
Input: m:method
Output: Boolean
   escape_sources := getInternalLeakSources();
   pointer and escape analysis;
   newObjects := getAllNewObjects();
   foreach o∈newObjects do
     if isEscape(o) then
        return true;
    end
end
return false;
```

4.3 IDENTIFYING MODULES ABSORBING ARGUMENT OBJECTS

We firstly check whether a module's parameters are reference types and regard the reference typed parameters as newly created objects in the modules. Then, pointer and escape analyses can be used to find out the lifetime of these objects and determine whether they may escape from method scope.

Algorithm 3: Identifying modules that absorbing argument objects

```
Input: m:method
Output: Boolean
     newObjects := markParamsAsNewObject (m);
pointer and escape analyses;
foreach o∈newObjects then
if isEscape(o) then
return true;
          end
    end
return false;
```

The algorithm is shown in Algorithm 3. It is similar to Algorithm 2. We firstly set method's reference parameters as newly created objects by markParamsAsNewObject(m). With these special newObjects, we do pointer and escape analyses for the method. Based on the escape information,

we check each newly created object in the method to determine whether there is any object escaping from method scope by predicate isEscape(o). If such objects exist, the method is a leak risky module.

5 Creating unit tests

For the leak risky modules, the users can pick up the modules that they believe should not cause continuously memory growth to do the unit testing.

Our unit test generation approach firstly uses some existing methods to generate normal unit tests. Then, we use enlarged input data, repeated method calls, and other techniques to strengthen their memory use, and insert memory monitor mechanisms to observe the leaking behaviors. Finally a collection of unit tests for memory leak testing purpose can be obtained.

The approach creates JUnit format test cases. In the current implementation, we use a tool named CodeProAnalytix [13] to generate normal JUnit test cases for the risky modules. The augmenting methods are discussed below.

5.1 CREATING UNIT TESTS FOR MODULES DIRECTLY LEAKING OBJECTS

For the modules that could directly cause leaks, we use large input data to strengthen memory use and insert memory growth assertions to determine whether the leaks can really occur.

```
public class UnitTest1 {
1       @Before
2       public void setUp() throws Exception {
3       }
4       @Test
5       public void testFoo(){
6           Test1 test = new Test1();
7           int n = MemoryTester.LARGE_INT;
8           long memoryBefore = MemoryTester.getUsed-
Memory();
9           test.foo(n);
10                  long memoryAfter =
MemoryTester.getUsedMemory();
11      MemoryTester.assertMemoryGrowth(memoryBefore,
memoryAfter, NO_SIGNIFICANT_GROWTH);
12  }
}
```

FIGURE 4 Unit testing with large input data and memory assertions

Figure 4 shows a unit test generated for the example in Figure 1. The normal unit test generated by the existing tools only contains the creation of Test1 object and a call to its method foo. In the normal unit test, it passes a random initial value to the tested method. To test memory leaks, in statement 7, we set a large input data MemoryTester.LARGE_INT for the tested method foo to strengthen memory use. Our approach currently supports several different types of large data, including the primitive types, such as int, long, float, and so on, and some object types, such as String. For the primitive types, we just use some previously defined huge value. For String type, we generate a pool of large strings and randomly select one of them. We obtain the memory consummation before and after foo by calls to method MemoryTester.getUsedMemory() and determine whether the method cause leaks by assertion MemoryTester.assertMemoryGrowth(memoryBefore, memoryAfter, NO_SIGNIFICANT_GROWTH). The assertion checks whether the memory growth is in the normal range. It uses a predefined value NO_SIGNIFICANT_GROWTH to set the allowed growth range. The constant means only small memory growth is allowed. If the memory grows over the allowed value, we consider the risky method really causes memory leaks. With the above strengthen, the unit test can validate whether the leak risky method actually lead to a noticeable leak.

5.2 CREATING UNIT TESTS FOR MODULES ACCUMULATING NEWLY CREATED OBJECTS

For the second kind of leak risky modules, we use repeated method calls to strengthen the memory usage, and insert memory growth assertions to check whether the leak risky modules can really lead to leaks.

```
public class UnitTest2 {
1       @Before
2       public void setUp() throws Exception {
3       }
4       @Test
5       public void testBar(){
6           Test2 test = new Test2 ();
7           long memoryBefore = MemoryTester.getUsedMemory();
8           for(int i=0;i<MemoryTester.LARGE_LOOPNUM;i++){
9               test.bar();
10          }
11          long memoryAfter = MemoryTester.getUsedMemory();
12          MemoryTester.assertMemoryGrowth(memoryBefore,
memoryAfter, NO_SIGNIFICANT_GROWTH);
13      }
}
```

FIGURE 5 Unit testing with repeated method calls and memory assertions

Figure 5 shows a unit test for the example in Figure 2. We use the existing tools to generate the normal unit test. In the normal unit test, there only have the creation of Test2 object and a call to method bar. We generate the test code on the base of the normal unit test. Statement 8 puts method bar into a loop and sets a large number for the loop to strengthen memory use. Then, the memory growth assertions checks whether the memory growth is normal. If the growth is abnormal, it indicates the leak risky method may be leaking the memory.

5.3 CREATING UNIT TESTS FOR MODULES ABSORBING ARGUMENT OBJECTS

We use a weak reference based mechanism to determine whether the third kind of leak risky modules can cause problems.

```
public class UnitTest3{
1   @Before
2   public void setUp() throws Exception {
  }
  @Test
public void testZar(){
   Test3 test = new Test3();
   Object obj = new Data();
   WeakReference<Object>  ref= MemoryTester.prepareArgu-
ment(obj);
   test.zar(obj);
   obj = null;
   MemoryTester.assertArgumentNotLeaked (ref);
}
}
```

FIGURE 6 Unit testing with weak reference based leak detection mechanism

Figure 6 shows a unit test generated for the example in Figure 3. The paper firstly generates a normal unit test only containing the instantiation of Test3, the creation of an argument object, and a call to its method zar. Then, it generates strengthened unit test based on weak reference mechanism. A weak reference will be garbage collected when its referee is disconnected from other references. By checking whether a weak reference is broken, we can know whether an object is hold by other references. In the unit test, statement 8 uses MemoryTester.prepareArgument (objs) to add a weak reference ref to method zar's argument obj. It sets reference obj to null in statement 10. After that, if the argument obj is not absorbed by method zar, then the weak reference should be broken after some round of garbage collections, since there is no other reference to the argument object. Statement 11 does some GC and check the referee of the weak reference ref to judge whether object obj is absorbed and there can be leak source. By the weak reference checking mechanism, the unit test can determine whether the leak risky method may cause leaks.

6 Case studies

To validate the proposed approach, we implemented our approach as an Eclipse plugin and conduct case studies on several memory leak bugs in JDK which are typical examples of the risky modules introduced in section 3.

6.1 MODULES DIRECTLY LEAKING OBJECTS

The modules directly leaking objects are a little hard to find in the JDK memory leak bugs. But many existing bugs can easily be turn into this kind. For example, in Figure 7, we can easily get a representative directly leaking case based on a real memory leak bug JDK-6942989. In the case, there is a loop calling leaking method getAnonymousLogger() (the body of getAnonymousLogger() can be found in Figure 9). The execution rounds of the loop are unbounded and determined by the method's input. In method getAnonymousLogger(), the newly created Logger objects are added into an external container. In other words, the new objects escape from the loop scope. Therefore, the case matches the characteristics of the first kind of leak risky modules.

```
public class Worker {
 public static void doLoggedOperation(int n){
  for(int i = 0; i< n; i++){
       Logger logger = getAnonymousLogger();
       logger.log(record);
       doSth();
  }
}
}
```

FIGURE 7 A case for modules directly leaking objects

```
public class LoggerTest {
@Before
public void setUp() throws Exception {
}
@Test
public void testDoLoggedOperation () throws Exception{
Worker fixture = new Worker();
int n = MemoryTester.LARGE_INT;
long memoryBefore = MemoryTester.getUsedMemory();
fixture. doLoggedOperation(n);
long memoryAfter = MemoryTester.getUsedMemory();
          MemoryTester.assertMemoryGrowth(memoryBefore,
memoryAfter, NO_SIGNIFICANT_GROWTH);
}
}
```

FIGURE 8 Unit test for the case shows in Figure 7

Our approach firstly finds out the loop in method doLoggedOperation. Then, it identifies the loop is input dependent on method parameters. Finally, the approach determines that the newly created objects escape from loop scope. Therefore, the approach identifies the leak risky method doLoggedOperation and then generates the unit test. The unit test is shown in Figure 8. In the unit test, a large input data MemoryTester.LARGE_INT is set for tested method doLoggedOperation to strengthen memory use. When running the unit test, the results show that method doLoggedOperation consumes abnormal amount of memory and hence causes memory leaks. The case indicates our approach is effective in revealing the first kind of leak risky modules.

6.2 MODULES ACCUMULATING NEWLY CREATED OBJECTS

For this kind of leak risky modules, we use memory leak bug JDK-6942989: Memory leak of java.lang.ref.WeakReference objects as the studied case. The bug affects JDK version 4.2u27, 5.0u25, and 6. Its relevant code is briefly shown in Figure 9. In the case, a new WeakReference object created in the method doSetParent (indirectly called by method getAnonymousLogger()) is added into external container kids. Although the weak references are finally broken, there are still references from the external container to the new WeakReference objects. These objects in kids are not released on time. Therefore, the case matches the characteristics of the second kind of leak risky modules, and method getAnonymousLogger() has risk in leaking memory.

```
public class Logger {
   private boolean anonymous;
   private static Object treeLock = new Object();
   private Logger parent;
private ArrayList<LogManager.LoggerWeakRef> kids;
public static Logger getAnonymousLogger() {
  return getAnonymousLogger(null);
}
public static synchronized Logger getAnonymousLogger(String re-
sourceBundleName) {
……;
Logger result = new Logger(null, resourceBundleName);
result.anonymous = true;
Logger root = manager.getLogger("");
result.doSetParent(root);
return result;
}
private void doSetParent(Logger newParent) {
    ……;
    if (parent != null) {
            for (Iterator iter = parent.kids.iterator(); iter.hasNext(); ) {
              WeakReference ref = (WeakReference) iter.next();
Logger kid = (Logger) ref.get();
              if (kid == this) {
                  iter.remove();
                  break;
}
            }
 }
parent = newParent;
……;
parent.kids.add(new WeakReference(this));
   updateEffectiveLevel();
}
}
```

FIGURE 9 Code relevant to bug JDK-6942989

In our approach, we firstly do escape analysis with the escape sources in method getAnonymousLogger() excluding the return value and the parameters by Algorithm 2. Then, we can obtain all the escape objects. Finally, it identifies that the escape objects contain the newly created objects. Therefore, the approach considers method getAnonymousLogger() as risky. The unit test generated by our approach is shown in Figure 10. In the test, we put method getAnonymousLogger() into a loop and set a large upper bound for the loop to strengthen memory use. When running the unit test, the results show that the method getAnonymousLog-ger() can lead to memory leaks and indicate our approach can identify this kind of risky modules.

```
public class LoggerTest {
@Before
public void setUp() throws Exception {
}
@Test
public void testGetAnonymousLogger throws Exception (){
        long memoryBefore = MemoryTester.getUsedMemory();
        for(int i=0;i<MemoryTester.LARGE_LOOPNUM;i++)
Logger.getAnonymousLogger();
        long memoryAfter = MemoryTester.getUsedMemory();
        MemoryTester.assertMemoryGrowth(memoryBefore,
memoryAfter, NO_SIGNIFICANT_GROWTH);
    }
}
```

FIGURE 10 Unit test for case JDK-6942989

6.3 MODULES ABSORBING ARGUMENT OBJECTS

We use memory leak bug JDK-6525563: Memory leak in ObjectOutputStream as the studied case for the third category of risky modules. The bug affects JDK version 6. Its relevant code is briefly shown in Figure 11. In the case, a reference argument obj is passed to method writeObject0. The parameter obj is then added into an external array field objs of the receiver object in method insert which is indirectly called by method writeUnshared. Finally, the argument object will be absorbed inside method writeUnshared. The case matches the characteristics of the third kind of leak risky modules.

```
public class ObjectOutputStream{
private int size;
private int threshold;
private int[] spine;
private int[] next;
private Object[] objs;
public void writeUnshared(Object obj) throws IOException {
writeObject0(obj, true);
}
private void writeObject0(Object obj, boolean unshared) {
int h;
Object orig = obj;
Class cl = obj.getClass();
ObjectStreamClass desc = null;
writeOrdinaryObject(obj, desc, unshared);
}
private void writeOrdinaryObject(Object obj,  ObjectStreamClass
desc,  boolean unshared)
{
        assign(unshared ? null : obj);
}
int assign(Object obj) {
        insert(obj, size);
        return size++;
    }
private void insert(Object obj, int handle) {
        int index = hash(obj) % spine.length;
        objs[handle] = obj;
        next[handle] = spine[index];
        spine[index] = handle;
    }
private int hash(Object obj) {
        return System.identityHashCode(obj) & 0x7FFFFFFF;
    }
}
```

FIGURE 11 Code relevant to bug JDK-6525563

Our approach firstly marks the parameter object passed to method writeUnshared(Object obj) as a special new object by Algorithm 3. Then, we obtain all the escape objects by escape analysis. Finally, it can be found that the escape objects contain the special new object. In another words, the parameter object escapes from the method scope. Therefore, the approach recognizes method writeUnshared(Object obj) as a leak risky module and then gener-ates the final unit test for it (see Figure 12). In the unit test, it uses MemoryTester.prepareArgument(obj) to add a weak reference to the method's reference parameter and use MemoryTester.assertArgumentNotLeaked(ref) to test whether the weak reference have been broken to check memory leaks. While running the unit test, the results show that the test method writeUnshared(Object obj) causes

memory leaks. The case indicates our approach is able to test memory leaks for the third category of risky modules.

```
public class ObjectOutputStreamTest {
@Before
public void setUp() throws Exception {
}
@Test(expected = java.io.NotSerializableException.class)
public void testwriteUnshared() throws Exception{
ObjectOutputStream fixture = new ObjectOutputStream(new ByteAr-
rayOutputStream());
Object obj = new Data();
WeakReference<Object> ref = MemoryTester.prepareArgument(obj);
fixture.writeUnshared(obj);
obj = null;
    MemoryTester.assertArgumentNotLeaked (ref);
}
}
```

FIGURE 12 Unit test for JDK-6525563

The above cases show that the approach can identify and test three kinds of leak risky modules effectively. It can reveal real leaking methods, which could be helpful for practical use.

7 Conclusions

The paper proposes an approach for discovering memory leak phenomenon from testing perspective. The approach automatically generates unit tests to find potential memory leaks in Java methods. It firstly identifies three kinds of leak risky modules. Then, leak-oriented unit tests are generated from normal unit tests to strengthen the ability in finding leaks. We conduct case studies on real bugs. The results show that the approach is effective in revealing real leak modules. The paper focuses on unit testing. In the future, we also plan to do leak testing at system level to further support the discovering of memory leak bugs.

Reference

[1] Park J, Choi B 2012 Automated Memory Leakage Detection in Android Based Systems *International Journal of Control & Automation* **5**(2) 35-42

[2] Pienaar J A, Hundt R 2013 JSWhiz: Static analysis for JavaScript memory leaks *IEEE/ACM International Symposium on Code Generation and Optimization (CGO)*

[3] Jump M, McKinley K S 2010 Detecting memory leaks in managed languages with Cork *Software: Practice and Experience* **40**(1) 1-22

[4] Maxwell E K, Back G, Ramakrishnan N 2010 Diagnosing memory leaks using graph mining on heap dumps *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)* 115-24

[5] Aftandilian E E, Guyer S Z 2009 GC assertions: using the garbage collector to check heap properties *Proceedings of the 2009 ACM SIGPLAN conference on programming language design and implementation (PLDI)* 235-44

[6] Xu G, Bond M D, Qin F, Rountev A 2011 LeakChaser: helping programmers narrow down causes of memory leaks *Proceedings of the 32nd ACM SIGPLAN conference on Programming Language Design and Implementation (PLDI)* 270-82

[7] Yan D, Xu G, Yang S, Rountev A 2014 LeakChecker: Practical Static Memory Leak Detection for Managed Languages *International Symposium on Code Generation and Optimization (CGO)*

[8] Qian J, Zhou X 2012 Inferring Weak References for Fixing Java Memory Leaks *The 28th IEEE International Conference on Software Maintenance (ICSM)* ERA Track 571-4

[9] Kim D, Nam J, Song J, Kim S 2013 Automatic patch generation learned from human-written patches *The 2013 International Conference on Software Engineering (ICSE)*

[10] Zhang S, Lü H, Ernst M D 2013 Automatically repairing broken workflows for evolving GUI applications *International Symposium on Software Testing and Analysis (ISSTA)*

[11] JProbe. http://www.quest.com/jprobe/

[12] O'Hair K 2004 HPROF: A Heap / CPU Profiling Tool in J2SE 5.0. http://java.sun.com/developer/technicalArticles/Programming/HPROF.html

[13] CodeProAnalytix. https://developers.google.com/java-dev-tools/codepro/doc/

[14] Xu B, Qian J, Zhang X, Wu Z, Chen L 2005 A brief survey of program slicing *ACM SIGSOFT Software Engineering Notes* **30**(2) 10-45

[15] Hind M 2001 Pointer analysis: Haven't we solved this problem yet? *Proceedings of the ACM SIGPLAN-SIGSOFT Workshop on Program Analysis for Software Tools and Engineering*

[16] Gay D, Steensgaard B 2000 Fast escape analysis and stack allocation for object-based programs *Proceedings of the 9th International Conference on Compiler Construction* 82-93.

[17] Vallée-Rai R, Co P, Gagnon E, Hendren L, Lam P, Sundaresan V 1999 Soot – a Java bytecode optimization framework *Proceeding CASCON '99 Proceedings of the 1999 conference of the Centre for Advanced Studies on Collaborative research* 13-23

Authors

Lijuan Hong, born in March, 1989, Hefei, China

University studies: MS degree in Computer Science and Technology at Nanjing University of Aeronautics and Astronautics, China, 2012.
Scientific interests: program analysis, program tests.

Ju Qian, born in 1981, Nanjing, China

Current position, grades: associate professor at Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China.
University studies: PhD degrees in Computer Science and Technology at Southeast University, China.
Scientific interests: program analysis, program tests, program diagnostics.
Publications: 4 papers.

Jifeng Cui, born on March 16, 1988, Nanjing, China

University studies: MS degree in Computer Science and Technology at Nanjing University of Aeronautics and Astronautics, China, 2011.
Scientific interests: data mining, big data analysis.
Publications: 3 papers.

Improved PSO clustering routing algorithm for WSN

Minsheng Tan, Huan Zhou*, Yangwei Li, Jianxue Liu

School of Computer Science and Technology, University of South China, Hengyang Hunan, 421001, China

Received 1 March 2014, www.tsi.lv

Abstract

For cluster head selection randomness of clustering algorithm, and PSOC algorithm were not considered the distance from cluster head to base station, an improved particle swarm optimization (I-PSOC) routing algorithm was brought out. The improved algorithm particle swarm fitness function was improved by considering the node residual energy, nodes' distance and the distance between nodes and base station. At the same time, the optimal nodes were selected as the cluster head and the cluster head has transmitted the data to base station in a single or more jumps through searching right path in the improved algorithm. The simulation results show that the nodes energy consumption of network was reduced significantly and the network life cycle was extended

Keywords: PSO, Clustering Routing Algorithm, cluster head, Wireless Sensor Network

1 Introduction

Wireless Sensor Network (WSN) is relatively large-scale, and has the distribution density of numerous nodes, especially limited energy of sensor node. Therefore, the overall energy consumption of the network must be considered to achieve efficient data transfer, extending the network lifetime. For large-scale WSN, generally being used cluster-based routing algorithm, the network is divided into different clusters, select the cluster head for data integration, and reduce data redundancy [1] is a method for effectively reducing network energy consumption, enhancing network robustness and scalability, and prolonging network lifetime.

The execution of cluster-based routing algorithm is generally divided into four steps: cluster head selection, formation of clusters, intra-cluster and inter-cluster communication. Cluster head nodes not only receive and fuse the date transmitted from other nodes, but also transmit the processed data to the distant station in the same time. So it needs to consume more energy. However, most clustering algorithms used with equal probability select cluster head randomly without taking into account the energy and location information of cluster head, which leads to uneven distribution of cluster head in the network, imbalance in network energy consumption, energy consumption of data communication, premature node death, impact on the entire network life cycle [3].

Particle Swarm Optimization (PSO) is a swarm intelligence optimization algorithm and a community-based collaborative random search algorithm [4, 5], developing by simulating birds foraging behaviour. It has fast search speed, high efficiency, simple algorithm. First, a group of particles is initialized, and then search out the global optimum target area by their own experience and social collaboration. Research shows, PSO find the global optimal solution with relatively high probability and accuracy, so it can be applied to the WSN which has numerous nodes and limited energy for cluster head selection. In this paper, it mainly optimizes the selection of cluster head in clustering routing by improving fitness function, thus, forming the optimal cluster, minimizing sensor node energy consumption, extending the network lifetime [6-8].

2 Analysis of PSO clustering routing algorithm

2.1 PSO INTRODUCTION AND PROCESS

PSO is to solve optimal problems by simulating birds foraging behaviour. Every bird is abstracted as a "particle", and has its own position and velocity to determine the search direction and distance, as well as a value of fitness function to measure particle properties. It initiates a group of particles firstly, and then particles find the global optimal solution in the search space by iteration. Suppose: the community size is M; the *i*-th particle in the search space is defined as $X_i = (x_{i1}, x_{i2}, ..., x_{im})$; the particle velocity is defined as: $V_i = (v_{i1}, v_{i2}, ..., v_{im})$; the optimal location searched by particle i can be expressed as $P_i = (p_{i1}, p_{i2}, ..., p_{im})$; the current optimal location in the group can be expressed as $P_g = (p_{g1}, p_{g2}, ..., p_{gm})$. Each iteration particle updates the information of position and velocity by tracing P_i and P_g. The Equations for updating information of position and velocity are following:

$$V_{id}(t+1) = \omega V_{id}(t) + c_1 r_1 (P_{id} - X_{id}(t)) + c_2 r_2 (P_{gd} - X_{id}(t)) , \quad (1)$$

*Corresponding author's e-mail: tanminsheng65@163.com

$$X_{id}(t+1)=X_{id}(t)+V_{id}(t+1)\ . \quad (2)$$

Parameters and description in the above Equation are shown in Table 1.

TABLE 1 Update the parameters and describe the Equation

V	Velocity of particle
X	location of particle
t	Number of iteration
ω	Weighting coefficient
$c1$,$c2$	Factor of studying
$r1$,$r2$	Random Num in(0,1)
P_{id}	extreme value of Every particle
P_{gd}	extreme value of Every particle
i	The positive integer between [1,M]
d	The positive integer between [1,D]

In Table 1, P_{id} is Individual optimum solution and extreme value of every particle, P_{gd} is all particle group's optimum solution and extreme value of every particle, M is the amount of the particle, D is the biggest dimension in searching zoom.

PSO process is as shown in Figure 1. Its basic steps are following as:

1) Initiate particles, every particle has velocity V_i and position X_i, set the number of iterations as K, and the learning factors as $c1$ and $c2$.

2) According to the fitness function, figure out each particle fitness function value, the compare with P_{id} and P_{gd}, update individual and global optimal solution.

3) Update particle velocity and position by the Equations (1) and (2).

4) Until the end of iterations, repeat the steps (2) and (3).

In large-scale WSN, routing algorithms are used to adopt clustering methods. While PSO mainly optimize the cluster head selection for clustering algorithms and form optimal cluster, and then balance network energy consumption and prolong the network lifetime.

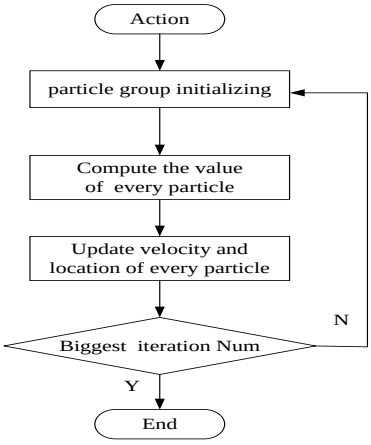


FIGURE 1 Particle swarm algorithm flow chart

2.2 CLUSTERING ALGORITHM

LEACH algorithm is a low-power, self-organizing adaptive clustering routing algorithm. Its execution is cyclical, and it involves two phases in every cycle: establishment of clusters and stable data transmission. LEACH algorithm randomly selected cluster head, so the probabilities of each node becoming a cluster head are equal. This makes the energy consumption of each node is relatively balanced, and then prolong the network lifetime. The two stages process is as follows.

2.2.1 Establishment of clusters

The sensor node generate a random number during [0,1]. Set the threshold as $T(n)$. If node has be selected as cluster head, $T(n)$ is 0, and can't be cluster head any more. While unselected nodes will be elected with equal probabilities $T(n)$. $T(n)$ calculate is as follows:

$$T_{(n)}=\begin{cases}\dfrac{p}{1-p\cdot(r \bmod \frac{1}{p})} & n\in G\\ 0 & Other\end{cases}. \quad (3)$$

In the Equation (3) P is the proportion of cluster head odes number in the network as whole, r is cycles, r *mod* $(1/p)$ is the nodes number. These nodes have been cluster head, G is a set, expressing all unselected nodes in the current cycle.

After selecting cluster head, the cluster head node should inform other non-cluster head node by releasing messages. These nodes select and join to the cluster, which the nearest cluster head belong to, while sending the message to inform its selected cluster head. After all member nodes send messages to inform the cluster head, the cluster head will establish TDMA slot table and broadcast messages to inform all members. Thus, all nodes can send data in each time slot and avoid conflicts when intra-cluster nodes are working.

2.2.2 Stable data transmission

The network will enter a stable data transmission phase after clusters establishment end. Intra-cluster member nodes send the collected monitoring data to the cluster head in its own distributed time slot. Member nodes will be in a sleeping state to save energy during some time that its own time slot is yet to come. While cluster head will always be in working state, sending the data is processed through the fusion to the base station. To avoid signal interference between nearby clusters, different intra-clusters all member nodes can adopt different coding way of CMDA. LEACH algorithm process is described on Figure 2. The algorithm has the following advantages:

1) Randomly select cluster head nodes with equal probability. Relatively balance overall network energy

consumption, save energy, avoid excessive energy consumption of nodes to death and prolong the lifetime of the node.

2) The entire network is divided into multiple clusters, and data communication is involved the intra-clusters and the inter-cluster. So it is more suitable for large-scale WSN.

3) Cluster head fuses the data transmitted from others, and sends them to the base station. Then reduce data redundancy of transmission.

4) Intra-cluster all member nodes has a sleep mode. It will be working condition only when send or receive data.

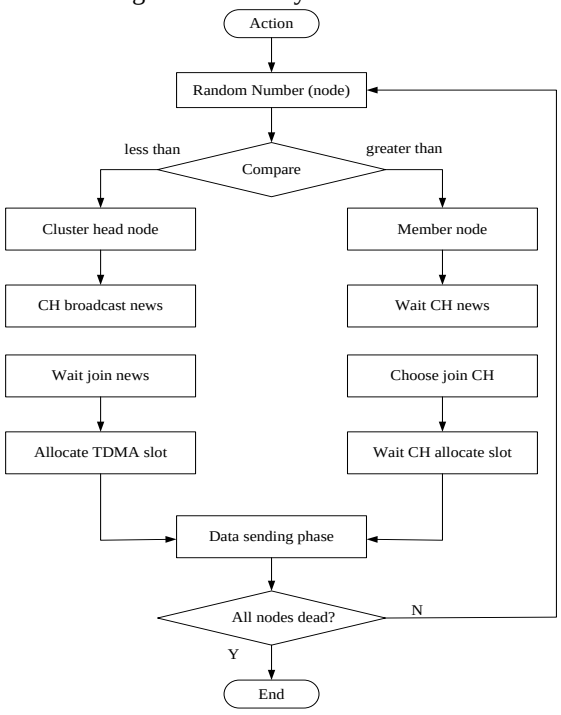


FIGURE 2 Flow chart of the LEACH algorithm

But there are also many disadvantages of LEACH algorithm:

1) Not take into account the position of the cluster head.

2) Not take into account the residual energy of cluster head, cluster head need greater energy consumption for conducting data fusion.

3) Communication between cluster head and sink node use single-hop routing. When aggregation node farther away from the cluster head in the cluster, because of excessive power consumption will prematurely die, and then affect network connectivity.

4) After the network operates stably for a period of time, a new cycle cluster head selection should be conducted again.

2.3 DESIGN REQUIREMENTS OF PSO CLUSTERING ROUTING ALGORITHM

In the design of particle swarm clustering algorithm, we must take into account the following aspects:

1) With an appropriate fitness function, the fitness function setting has an important impact on the algorithm.

2) With the appropriate updating algorithm, to update the particle velocity and position information.

3) With adjustable optimization parameters, make the results do not deviate from the correct solution.

Many scholars at home and abroad research PSO algorithm to implement network optimization clustering problem. In spite of the particle swarm optimization algorithm has many advantages, but there are also easy to fall into local optimum. We mainly improve the parameters of the fitness function in algorithm, the topology structure, the inertia weight, learning factors and mixed strategy, such as paper 47 28 49 and 50, of which more representative is 50 PSOC algorithm, effectively balance the network energy consumption, although it considers the nodes energy and the distance between nodes within cluster, but it without considering the effect of cluster heads and base station distance. If the distance is larger, cluster heads are very easy to premature deaths due to energy consumption too fast.

This article, mainly, improve the situation aimed at selecting cluster heads random in clustering algorithm and PSOC algorithm problems.

3 Improvement of particle swarm cluster routing algorithm

3.1 ALGORITHM IMPROVEMENT SCHEME

Firstly, network has been divided into different clusters by using clustering algorithm, set a threshold energy, and then form a candidate cluster heads set, then combined with improved fitness function, according to the candidate set of cluster heads within each node's residual energy and location, using PSO algorithm, select the global optimal node as the cluster head, cluster heads broadcast information, transmit and receive within the cluster node's residual energy and location information, and to establish a time slot table, for each of the distribution of cluster nodes when the TDMA slots. The main function of Cluster member nodes is testing the object of data collection, and transmitting data to the cluster head, cluster heads not only collect the data, but also transmit the data fusion in the cluster member nodes, finally, through single hop or multiple hops they send the date to the base station (gathering node).

The method can efficiently prolong the network life cycle, the network node have the opportunity to become a cluster head when the network node energy is higher than the energy threshold set. if the cluster head suddenly die or energy is lower than the threshold, the network node will be back for cluster heads selection, The shortage for clustering algorithms when it used data transmission in one hop paths which lead to far from the base station of the premature death of cluster nodes, in the data communication stage, the cluster heads sent to the distances from the base station or base station of cluster nodes in the cluster based on decision fusion data. Cluster heads choose the appropriate path to base station. This

method can effectively reduce the network energy consumption and prolong the network life cycle.

3.2 NETWORK MODEL

Network model of this paper is shown in Figure 3, and the research of wireless sensor network makes the assumption.

1) Node and base station position is fixed, base station far away from node distance, cluster heads selection is not affected by the environment.

2) The nodes know their position information and residual energy.in addition to the base station, the rest of the initial energy of nodes and processing power are equal, the same communication range.

3) The node transmission power can be adjusted according to the communication distance.

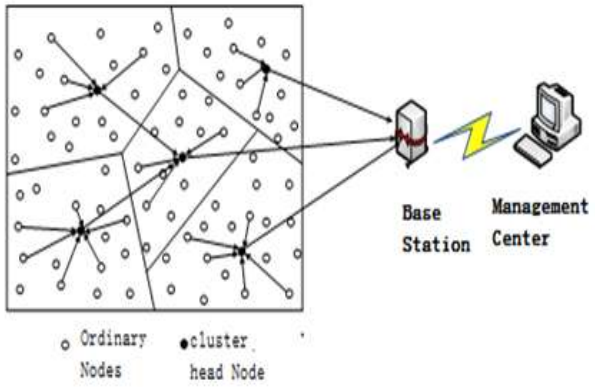


FIGURE 3 The structure of the network

3.3 ENERGY MODEL

In this paper we use the first-order energy consumption of wireless communication model, as shown in Figure 4. The model is based on the following assumptions:

1) The node energy limited.

2) The node energy consumption is basically the same.

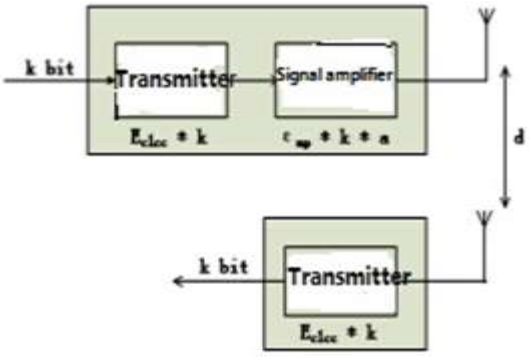


FIGURE 4 Wireless communication energy model

Form of the energy consumption according to the attenuation of power amplifier can be divided into two parts: free space model and the multipath attenuation model. Attenuation model, power amplifier based on node spacing selection for d0 set threshold method, the distance between nodes is d, if d < d0, multipath attenuation model is adopted, therefore, when Kbit node transmission data information, the energy consumption between sensor nodes can be represented as:

$$E_{Tx}(k,d)=\begin{cases} E_{elec}\cdot k+\varepsilon_{fs}\cdot k\cdot d^2 (d<d_0) \\ E_{elec}\cdot k+\varepsilon_{mp}\cdot k\cdot d^4 (d\geq d_0) \end{cases}. \tag{4}$$

$$E_{Rx}(k)=E_{elec}\cdot k \tag{5}$$

Among the equation above, the $E_{Tx}(k,d)$ represents the energy consumed by the transmission of k bits data, the $E_{Rx}(k)$ represents the energy consumed by the reception of k bits data, the E_{elec} represents the energy consumed by the transmission of 1 bit data, respectively, ε_{fs} and ε_{mp} represent the power amplifier circuit to model consumption coefficient of multi-path model and the free space attenuation, d is the distance that data transmitted, d_0 is the threshold distance set.

3.4 IMPROVEMENT OF FITNESS FUNCTION

Evaluate the quality of the selected cluster head based on the fitness function, we mainly consider the three aspects of content. Firstly is the cluster head energy evaluation factor f_1. Cluster energy takes the inverse proportion of the total energy S_f the nodes within the cluster. Cluster heads should not only accept the quantity of node data but also transmitted the data to base station. The largest energy consumption, therefore, cluster heads residual energy is larger, the better. Secondly, evaluation factor f_2, stand for the average distance of the node to the cluster head, the smaller the average distance, the less data transmission energy consumption between nodes. Thirdly, evaluation factor f_3, stand for cluster heads to the base station distance, the smaller the best path to the base station, the smaller consume energy cluster head used, hypothesis, network consists of M nodes, divided into k clusters, the improvement of fitness function is as follows:

$$\mathrm{f}=\alpha f_1+\beta f_2+\gamma f_3, \tag{6}$$

$$f_1=\sum_{j\in(C_i)} E(n_j) \Big/ E(CH_i), \tag{7}$$

$$f_2=\sum_{j\in(C_i)} \mathrm{d}(n_j,CH_i) \Big/ |C_i|, \tag{8}$$

$$f_3=\min_{k=1,2,\ldots,K}\{\mathrm{d}(CH_i,BS)\}. \tag{9}$$

According to (6), it calculates particle fitness function value, select the global optimal node as the cluster head. Fitness function value is smaller, the residual energy of the

cluster heads is much more or the distance between the cluster member nodes and cluster heads is closer, the distance between cluster and the base station is closer.

3.5 DETAILED STEPS OF ALGORITHM IMPROVEMENT

Network nodes could be divided into different clusters, this paper uses PSO algorithm to select the global optimal node as the cluster head, and find the optimal cluster. The cluster heads through single hop or multiple hops, and then transmit data after fusion methods to the base station and balance network energy consumption and prolong the network life cycle. I-PSOC algorithm in this paper and the cluster head selection cycles, each cycle is divided into clusters of establish and stability of the data transmission phase. Include the cluster initialization of the candidate, according to the improvement of the fitness function to select the optimal cluster heads and the formation of optimal cluster; Data transmission phase including the cluster-heads data transmission and data transmission between clusters. Its algorithm steps are as follows:

Step 1. Initialize clustering of network, set up energy threshold E_T, when the node energy is greater than E_T, it have a chance to become cluster head. Assume the $CH_{candidate}$ represents the set of candidate cluster:

$$CH_{candidate} = \{n_i \mid E_i > E_T\},$$

and the $E_T = \sum_{i=1}^{M} E_i / M$.

E_i represents the remaining energy of current node, M represents the total number of nodes in the network. Although, determined the candidate cluster head basically at this stage, only consider the residual energy, ignoring the wasted energy that cluster head caused for the uneven distribution, therefore need to improve the fitness function, select the optimal cluster head from the candidate cluster head collection.

Step 2. Improve the fitness function, select the optimal cluster head. Initialize candidate of cluster particles, calculate the particle's fitness, adopt iterative method to update particle's speed and position, until the cessation of reaches maximum number of iterations of the algorithm, the global optimal solution will become the cluster head.

Basic steps are as follows:

Step 1. Initialize the particle swarm. initial speed and position of particles in the candidate set of cluster heads, calculate each particle's fitness by the Equation (6), and initialize the particle position as the individual optimal solution P, choose to adapt the value of the smallest particles as the global optimal solution of the whole particle P1.

Step 2. Update the particle's position and speed. According to the Equations (1) and (2) get each particle's new position and speed, and calculate a new adaptive value f_1 by the Equation (6).

Step 3. Update the individual optimal solution. Compare the i-th particle current fitness fi and individual optimal solution P. If f_i < P, it updates the P, on the other hand, keep P.

Step 4. Update the global optimal solution. Compare the smallest individual optimal solution in the particle swarm and the initial global optimal solution of the whole particle swarm P1. If less than P1, the individual optimal solution of particles become the global optimal solution of the whole particle P1, otherwise, keep P1.

Step 5. Repeat steps (2), (3) and (4), until the maximum number of iterations, and then end of the recycle. Select the global optimal solution as the cluster head.

Algorithm process is shown in Figure 5.

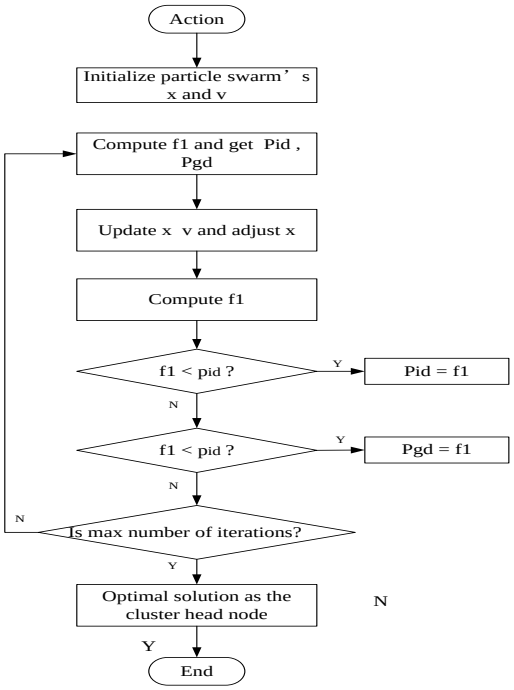


FIGURE 5 Improved cluster heads selection algorithm

3.6 THE FORMATION OF OPTIMAL CLUSTER

After using the I-PSOC algorithm to select cluster head, each node choose to join the cluster according to the sent message signal strength when it receives, and sent its energy and location information to cluster heads. The cluster head nodes allocated time slot for each members. Stable data transmission phase:

1) Data transmission phase in the clusters. After form optimal clusters, cluster member nodes will collect data transmission to the cluster head in its own slot, the rest of the time it dormant to save energy.

2) Data transmission between clusters. The stage mainly is cluster head accept the members of cluster nodes transmit data and fusion, and then choose the appropriate path to transmit data according to the distance of cluster heads to the base station.

Limited distance value d0, and generally take the average distance of the cluster heads to the base station. If the distance from the base station to cluster head d>d0, then the cluster head search the optimal path for transmitting the merged data to the base station by the multi-hop way. Else d<d0, cluster head will be directly to transmit the merged data to the base station by the single-hop way.

When the node and the base station are fixed, this cluster head selection process is not affected by environmental factors. If the cluster head energy is below the energy threshold E_T, then re-select for cluster head. The selected node as the cluster head global optimal way, balancing the network energy consumption, thus effectively extending the network life cycle.

4 Simulation and result analysis

4.1 THE PERFORMANCE OF I-PSOC ALGORITHM EVALUATION INDEX

In this paper the improved routing algorithm evaluation mainly consider the following aspects: network life cycle, the network energy consumption.

1) Network life cycle can be expressed as the relationship between the number of nodes in the network and the survival time (number of rounds) of running.

2) Energy consumption of the network can be expressed as the energy of all nodes in the network over time, of the remaining amount of energy consumed.

4.2 SIMULATION ENVIRONMENT AND PARAMETER SETTINGS

To verify the improved performance of the algorithm, use TOSSIM tools for simulation, to carry out large-scale network simulation, to imitate the behaviour of the hardware resources, through to the hardware simulation component changes, can provide a variety of different hardware environment performance, meet the requirements of different users.

TABLE 2 Energy simulation parameters

Parameter	Value
E_{abc}	50nJ/bit
E_{DA}	50nJ/bit/singal
$\varepsilon_{\varepsilon}$	10pJ/bit/m^2
ε_{mp}	0.0013Pj/bit/m^4
E_0	0.5J
Package's length	4000bits

In this article, through TOSSIM simulation tools from two aspects of the network life cycle and the network energy consumption compared with LEACH and PSOC algorithm Using Figure 4 wireless communication model and energy consumption is calculated using the Equation (4). Simulation environment: the 100 sensor nodes randomly distributed within the 100x100m square area, the base station is located in (= 50 X, Y=150). The simulation parameters are shown in Table 2. The parameters of the particle swarm optimization algorithm for: α=0.35, β=0.35, γ=0.3, ω=0.9, learning factor c1=2, c2=2.

4.3 EXPERIMENT AND RESULT ANALYSIS NETWORK LIFE CYCLE COMPARED

The I-PSOC algorithm simulation, record and with LEACH algorithm and PSOC compare the amount of surviving nodes. Suppose among the node does not appear data retransmission and the phenomenon of transmission errors, and the members of the cluster nodes without transferring data in a sleep states. Simulation in this paper, a total of 750 rounds, discrete simulation get round number and node number offered as shown in Tables 3 and 4.

TABLE 3 The relations between total number of rounds and surviving nodes

Rounds surviving nodes	0	100	200	300	400
LEACH	100	100	100	95	80
PSOC	100	100	100	100	90
I-PSOC	100	100	100	93	85

TABLE 4 The relations between total number of rounds and remaining energy

Rounds surviving nodes	500	600	650	700	750
LEACH	52	20	0	0	0
PSOC	68	40	20	0	0
I-PSOC	75	49	33	16	0

Surviving nodes in the network and simulation rounds relationship shown in Figure 6. The simulation results show that compared with LEACH and PSOC algorithm, with the number of rounds progressed, the I-PSOC algorithm in the number of surviving nodes significantly more. PSOC algorithm is not survive more node than the I-PSOC algorithm, but still want to survive more than LEACH algorithm nodes, if all node dies, the I-PSOC algorithm in the network experienced a number of rounds is clearly more than LEACH and PSOC algorithm, It can be seen, the paper improved PSOC fitness function selected as the cluster head global optimal solution, compared to LEACH algorithm randomly selected cluster head and PSOC algorithm does not consider the distance between cluster head and the base station method, network significantly less energy consumption, extended lifetime effectively.

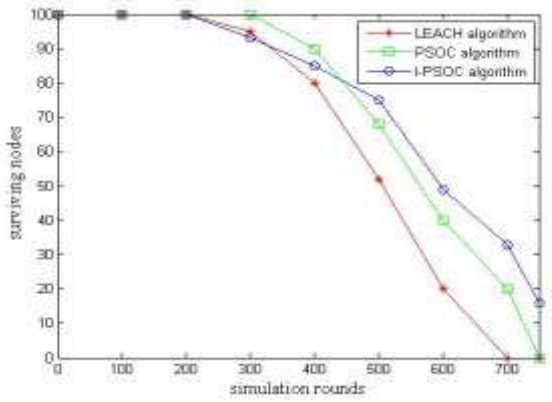


FIGURE 6 Network life time contrast

4.4 NETWORK ENERGY CONSUMPTION COMPARED

On improved particle swarm routing algorithm simulation, record the total residual energy of network nodes, and with LEACH and PSOC algorithm total residual energy of network nodes for comparison. This co-simulation 750, discrete number of rounds simulated node residual energy relationship shown in Tables 5 and 6.

TABLE 5The relations between total number of rounds and remaining nodes

Rounds remaining energy	0	100	200	300	400
LEACH	50	47	42	35	27
PSOC	50	48	44	38	32
I-PSOC	50	48	45	40	34.5

TABLE 6 The relations between total number of rounds and remaining energy

Rounds remaining	500	600	650	700	750
LEACH	17	6	0	0	0
PSOC	25	15	8	0	0
I-PSOC	28	19	14	7	0

Network nodes remaining energy relations with simulation round number is shown in Figure 7. The simulation results show that when going through the same number of rounds, the I-PSOC algorithm in network nodes significantly more total residual energy, PSOC algorithm followed, LEACH algorithm in network energy consumption the most. Therefore, the I-PSOC algorithm can effectively prolong the network lifetime. That is the cluster head selection algorithm in the first set when an energy threshold, the candidate cluster head constitute a collection through I-PSOC fitness function, selection from the candidate set of cluster heads the global optimal solution as the cluster head, forming optimal clusters. While inter-cluster communication, also set a distance d0, if the distance between the cluster head and the base station is smaller than d0, transmit data in a single jump directly, else, the cluster head will search for the most suitable path to transmit the merge data to the base station by multi-hop way, effectively balancing the network energy consumption and prolong the life cycle.

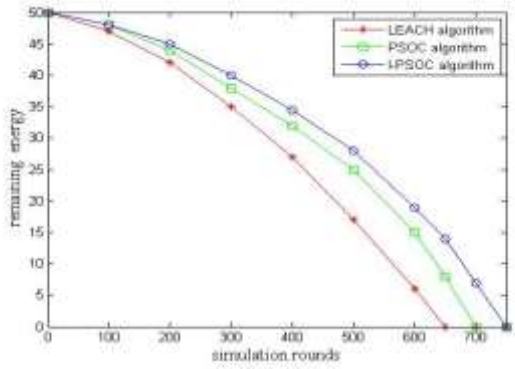


FIGURE 7 The remaining energy of nodes in the network contrast

5 Conclusions

This paper first introduces the basic principles of particle swarm algorithm and algorithm-step process, and made a careful analysis of particle swarm clustering routing algorithm. Then according to the clustering algorithm of cluster head selection of randomness and PSOC algorithm does not take into account the distance cluster head and base station, etc. From the node residual energy distance between nodes and the node and the base station distance to consider, improved particle swarm fitness function, select the globally optimal node as the cluster head, forming optimal cluster. Meanwhile, the intra-cluster communication using single-hop transmission of data, the inter-cluster communication from the base station in accordance with the distance of cluster heads choose the right path (single-hop or multi-hop) to transfer data to a base station. Finally introduces TOSSIM discrete simulation tools, build a simulation environment, set energy consumption parameters and use TOSSIM simulation tools. From two aspects of the network life cycle and energy consumption to evaluate the performance of the I-PSOC algorithm, compared with LEACH and PSOC algorithm simulation results, analysis shows that the I-PSOC algorithm significantly reduces the network energy consumption, effectively prolong the network life cycle.

Acknowledgements

This work was supported by Project 61300234 of the National Science Foundation of China, the construct program of the key discipline in University of South China, the construct program of the key laboratory in University of South China, Project 2012KG71 of the Hengyang City Science and Technology Plan, key Project 14A121 of the Hunan Provincial Education Office Science Research of China.

References

[1] Kennedy J, Eberhart R C 1995 Particle swarm optimization *IEEE International Conference on Neural Networks* **4** 1942-8

[2] Eberhart R, Kennedy J 1995 A new optimizer using particle swarm theory *Sixth International Symposium on Micro Machine and Human Science*

[3] Liang Y, Yu H, Zeng P 2006 Application of PSO optimization of wireless sensor network routing protocol based on clustering *Control and Decision* **21**(4) 453-6

[4] Xue Y, Yang Z 2008 WSN based on discrete particle swarm clustering routing algorithm *Wuhan University Natural Science version* **54**(1) 99-102

[5] Tillett J, Rao R, Sahin F 2002 Cluster-head identification in ad hoc sensor networks using particle swarm optimization *Proc IEEE Int Conf Personal Wireless Communications* [S.1.]: *IEEE Press* 201-5

[6] Latiff N M A, Tsimenidis C C, Sharif B S 2007 Energy-aware Clustering for Wireless Sensor Networks Using Particle Swarm Optimization *Proc 18th IEEE Int Symposium on Personal, Indoor and Mobile Radio Communications Athens Greece IEEE Press* 1-5

[7] Wang Y, Si H, Su Y, Xu P 2013 A new clustering routing algorithm for WSN based on PSO *Advanced Materials Research* **850–851**(2014) 689-92

[8] Yan X, Cheng G 2012 Balance energy consumption routing algorithm in WSN based on improved particle swarm optimization Computer Engineering and Design 33(10) 71-5

Authors

Minsheng Tan, China

Current position, grades: professor, master tutor of University of South China, China.
University study: Master degree in control engineering, Hunan University, Hunan Province, China, 2005.
Research activities: network and information security, mobile networks and sensor networks.
Professional Activities and Memberships: Vice President of Computer Users Association of Hunan Province, in China, Vice President of Committee of Institute of China Hunan province college education in Computer education profession, Senior member of China Computer Federation.

Huan zhou, China

Current position, grades: master of University of South China.
University study: Master degree in Computer Science and Technology, University of South Hunan Province, China, 2012.
Research activities: trusted network.
Professional Activities and Memberships: member of China Computer Federation.

Yangwei Li, China

Current position, grades: master of University of South China.
University study: Master degree in Computer Application Technology, University of South China, Hunan province, China, 2010 the PLA.
Research activities: wireless sensor networks.
Professional Activities and Memberships: member of China Computer Federation.

Jianxue Liu, China

Current position, grades: master University of South China.
University study: Master degree in Computer Science and Technology, University of South Hunan Province, China, 2013.
Research activities: wireless sensor networks.
Professional Activities and Memberships: member of China Computer Federation.

Theoretical analyses and numerical simulation of the interaction time and the separation time of two elastic bars after the loading of a triangular wave

Wei Zhong*, Ming Lei, Zhou Tian

Northwest Institute of Nuclear Technology, Xi'an 710024, Shaanxi Province, China

Received 1 October 2014, www.cmnt.lv

Abstract

With theoretical analyses, the stress-wave propagation and reflection between two elastic bars, whose left side achieved a triangular wave load, is mainly studied. Considering the effects of pulse width of the load, length and wave impedance of the bars and the relationship between the length of the load wave and the two bars, the regularities of the stress-wave propagation and reflection between the two bars is analysed, and the formulas for calculating the interaction time and the separation time of the two bars are derived, in three different conditions. The same problems in three different conditions are simulated by using AUTODYN, and curves of displacements of points at contact surface of the two bars varying the time are given. According these curves, the simulating results of the separation time can be obtained. By the comparison of the simulating results and the theoretical calculated results using formulas derived in this paper, the correctness of the theoretical analyses and formulas here is demonstrated.

Keywords: triangular wave load, elastic bar, separation time, theoretical analyses, AUTODYN

1 Introduction

With the rapid development of high-tech, the field of such relate to explosion and shock problem as national defence technology, aerospace industry, new material technology, etc, is more and more widely, especially in areas such as the arms penetration and explosion, high-speed impact, dynamic response materials, structural damage and other protective aspects [1]. As we know, the study of those problem mentioned above is related to the stress wave propagation. Stress wave propagation characteristics in multi-layer composite structure are a newly rising research field [2-4]. Although the current experimental study of homogeneous material for macroscopic dynamic mechanical behaviour and impulse response has developed many ripe methods, such as Taylor test, split Hopkinson pressure bar, flat impact and detonation technology [5-7], the structures of the actual project are usually layer structures stacked by a variety of materials with different physical properties. Because of light quality, strong design, good performance and strong impact absorbing characteristics of these kind of structures, they have become one of the hot issues of structural engineering applications in recent years [8]. Taking into account that the multi-layer dielectric stress wave propagation problem in the actual project is very complicated, as a theoretical analysis here, the research was focused on the stress wave propagation and reflection of two elastic bars after the loading of a triangular wave. Recently, the most widely application of experimental studies of such problem is the split Hopkinson pressure bar's (SHPB) [9, 10], also known as the Kolsky bar. SHPB technology is on the basis of one-dimensional stress wave theory, which requires the long-thin-bar is linear, isotropic and the dispersion effect can be ignored, meanwhile, the cross-section of the bar in the axial direction is assumed to be a constant and the bar to maintain flexibility and uniform stress state in the process of loading and unloading [11]. As a theoretical analysis of the situation above, the theoretical analysis of this article and the following derivation is also based on the same requirements above.

2 Theoretical analysis and calculation formula derivation

In Figure 1, two elastic bars contact with each other, and a input triangular wave loads with the wavelength of λ acts on the left of bar 1. In the following, the interaction time and the separation time of the two bars will be analyzed. L_1 and L_2 is respectively the length of bar 1 and bar 2, as well as c_1 and c_2 the triangular-wave propagation velocity in bar 1 and bar 2. We assume that the two bars are the same, that is to say, c_1 and c_2 are equal, and can be taken as c. Next, we will prove that the interaction time and the separation time of the two bars are determined by λ, L_2, c_2, and the relative relationship between λ and L_2.

*Corresponding author e-mail: lengshui222@163.com

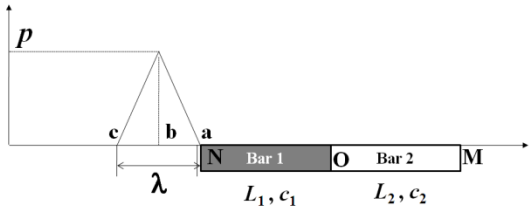


FIGURE 1 Model sketch of this problem.

2.1 CASE 1: $\lambda/2<L_2$

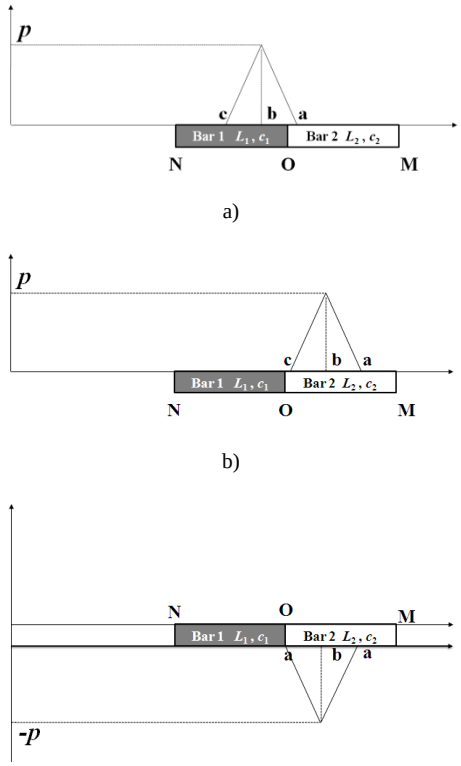


FIGURE 2 Diagram of the stress-wave propagation when $\lambda/2 < L_2$

Since the load is linear and the bars are elastic, the relationship between stress in bars and the input load is linear. Therefore, for easy discussion, the stress of bars will be expressed in the input load in following analysis.

When the front of the stress wave come location O, which is the contact surface of bar 1 and bar 2, the interaction force between the two bars occurs, and compressive stress appears in these two bars, as shown in Figure 2a. When the end of the stress wave passes location O, the interaction force disappears, and the two bars have a same velocity, as shown in Figure 2b. Then, the interaction time of bar 1 and bar 2 can be obtained by:

$$t_{\text{interaction}} = \frac{\lambda}{c}. \tag{1}$$

When the stress wave comes to location M, which is the right end of bar 2, stress wave reflection occurs at the free surface, and an extension wave propagation to the left appears [12]. Particles of the bar will get a velocity to the right after the pass of the extension wave. So when the extension wave comes to the location O, bar 1 and bar 2 will separate from each other, as shown in Figure 2(c). Then, the separation time of bar 1 and bar 2 can be obtained by:

$$t_{\text{separation}} = \frac{L_1 + 2L_2}{c}. \tag{2}$$

2.2 CASE 2: $\lambda/2 > L_2$ and $\lambda/2 < (L_1+L_2)$

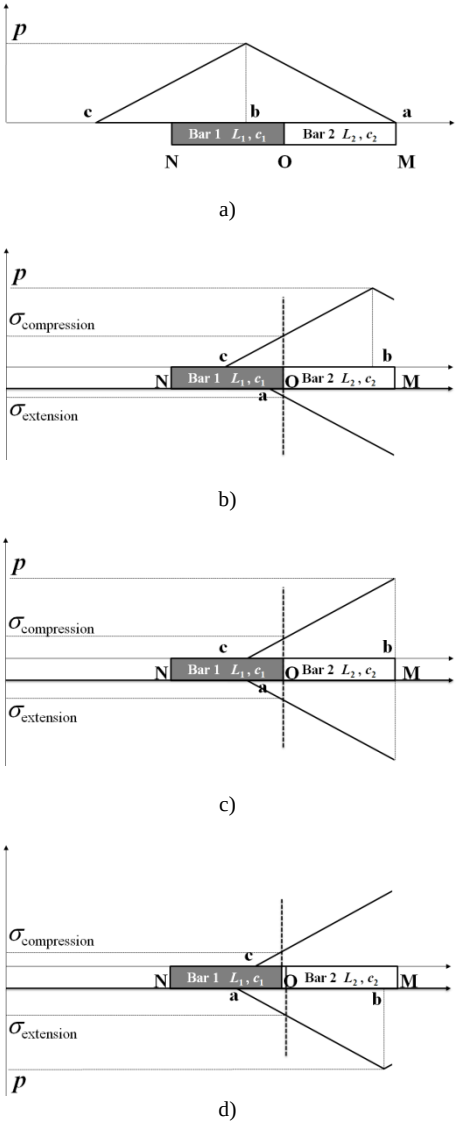


FIGURE 3 Diagram of the stress-wave propagation when $\lambda/2 > L2$ and $\lambda/2 < (L1+L2)$

When the front of the stress wave come location O, the interaction force between the two bars occurs, and compressive stress appears in these two bars, as shown in Figure 3a. When the front of the stress wave comes to the right end of bar 2 but the mid of the stress wave has not come to location O, stress wave reflection occurs at the

free surface, and an extension wave propagation to the left is formed. At the time of this extension wave comes to the location O, the end of the stress wave has not come to the location O yet, and meanwhile, the compressive stress is higher than the stress generated by the reflection extension wave, so the interaction force between bar 1 and bar 2 is always maintained, which will keep the two bars contacting with each other and moving together, as shown in Figure 3b. As the mid of the stress wave comes to the right end of bar 2 and the front of the tensile stress wave passes the location O, although the end of the stress wave has not come to the location O, the tensile stress at location O is equal to the input stress, which means the stress at location O is zero at this time, which can be seen in Figure 3c. As shown in Figure 3d, with the continuing wave propagation, bar 1 and bar 2 will separate from each other, and the separation time of bar 1 and bar 2 can be obtained by:

$$t_{\text{separation}} = \frac{0.5\lambda + L_1 + L_2}{c}. \tag{3}$$

According to the analysis above, it is easy to get the interaction time of bar 1 and bar 2 by:

$$t_{\text{interaction}} = \frac{0.5\lambda + L_2}{c}. \tag{4}$$

2.3 CASE 3: $\lambda/2 > (L_1+L_2)$

This is a very complex case, and the interaction time and separation time are closely related to the waveform of the input stress. In order to make the research more convenient, a special stress wave propagation graph was made, which was shown in Figure 4.

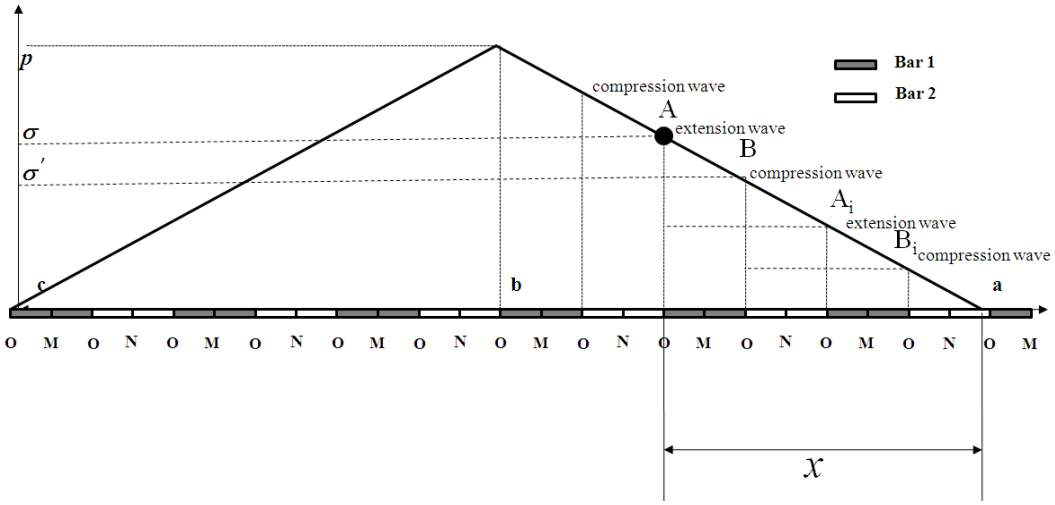


FIGURE 4 Diagram of the stress-wave propagation when $\lambda/2 > (L_1+L_2)$

Assume the distance between the front wave and the extension wave A is x at the time of the extension wave A comes to location O the first time. The relationship between the tensile stress and the compression stress were given in Figure 4. For example, the stress in location B appears, when the extension wave comes to the M free surface - that is, after the input stress wave reflected in the left free surface of bar 1, a reflect compression wave formed, with the wave propagation continuing a length of bar 1, the compressive stress in contact surface of bar 1 and bar 2 appears. Other cases have the similar analysis above.

It is easy to be proved that there is a range of x, and the range is not larger than half of the wavelength. If x is equal to half of the wavelength, the location A will be the top of the input wave when the reflection extension wave comes to location A at the first time. And this location A is where the tensile stress of the contact surface could rise to a maximum. The Equation of the tensile stress and the compressive stress will be derived next.

2.3.1 Tensile stress

The slope of the triangular wave is:

$$k = \frac{p}{\lambda / 2}. \tag{5}$$

From geometrical relations in Figure 4, the slope on the expression of stress and x is:

$$k = \frac{\sigma}{x}.$$

Then:

$$\sigma = k \cdot x. \tag{6}$$

According to Figure 4, it is easy to know the number of the tensile stress waves reflected back and forth at this point is:

$$N = \left[\frac{x}{2(L_1 + L_2)}\right] + 1 \quad 3 \tag{7}$$

The symbol【 】indicates that the number is rounded to the nearest whole number.

For the *i*-th reflection (*i*≥2), the tensile stress satisfies the following Equation:

$$k=\frac{\sigma-\sigma_i}{(i-1)\times[2(L_1+L_2)]}.$$

Then:

$$\sigma_i=\sigma-k\times(i-1)\times[2(L_1+L_2)]. \tag{8}$$

All the tensile stress at the contact surface is the sum of σ_i:

$$\sigma_{\text{tensile}}=\sigma+\sum_{i=2}^{N}\{\sigma-k\times(i-1)\times[2(L_1+L_2)]\}. \tag{9}$$

2.3.2 Compressive stress

The compressive stress at location B can be obtained from geometrical relations in Figure 4:

$$\frac{\sigma\text{-}\sigma'}{2L_1}=k.$$

Then:

$$\sigma'=\sigma-k\times 2L_1 \tag{10}$$

As the same, according to Figure 4, it is easy to know the number of the tensile stress waves reflected back and forth at this point is:

$$M=【\frac{x-2L_1}{2(L_1+L_2)}】+1. \tag{11}$$

For the *i*-th reflection (*i*≥2), the compressive stress satisfies the following equation:

$$k=\frac{\sigma'-\sigma'_i}{(i-1)\times[2(L_1+L_2)]}.$$

Then:

$$\sigma'_i=\sigma'-k\times(i-1)\times[2(L_1+L_2)] \tag{12}$$

So, all the tensile stress at the contact surface is the sum of σ'_i:

$$\sigma_{\text{compressive\_reflect}}=\sigma'+\sum_{i=2}^{M}\{\sigma'-k\times(i-1)\times[2(L_1+L_2)]\} \tag{13}$$

When $(\frac{\lambda}{2}-x)\geq 2L_2$, we get:

$$\frac{\sigma_{\text{compressive\_incident}}-\sigma}{2L_2}=k.$$

Then:

$$\sigma_{\text{compressive\_incident}}=\sigma+k\times 2L_2. \tag{14-1}$$

When $0<(\frac{\lambda}{2}-x)<2L_2$, we get:

$$\begin{cases}\dfrac{p-\sigma}{\delta}=k\\ \dfrac{p-\sigma_{\text{compressive\_incident}}}{2L_2-\delta}=k\end{cases}.$$

Where δ is the distance between location A and the mid of the wave, so:

$$\sigma_{\text{compressive\_incident}}=2p-2L_2k-\delta \tag{14-2}$$

From (13), (14-1) and (14-2), we obtained the total compressive stress:

$$\begin{aligned}\sigma_{\text{compressive}}&=\sigma_{\text{compressive\_reflection}}+\sigma_{\text{compressive\_incident}}\\&=\sigma'+\sum_{i=2}^{M}\{\sigma'-k\times(i-1)\times[2(L_1+L_2)]\}\end{aligned} \tag{15}$$

$$\begin{cases}+\sigma+k\times 2L_2 & (\frac{\lambda}{2}-x)\geq 2L_2\\ +2p-2L_2k-\sigma & 0<(\frac{\lambda}{2}-x)<2L_2\end{cases}$$

If tensile stress is larger than compressive stress on the contact surface, bar 2 and bar 1 will separate from each other.

Assume $\sigma_{\text{tensile}}=\sigma_{\text{compressive}}$, we get $x_{\text{separation}}$:

$$x_{\text{separation}}=\begin{cases}\dfrac{1}{[N-(M+1)]}\times\{2L_2-2L_1M+\\ \sum_{i=M+1}^{N}[k(i-1)\times 2(L_1+L_2)]\} & (\frac{\lambda}{2}-x)\geq 2L_2\\ \dfrac{1}{[N-(M-1)]}\times\{2p/k-2L_2\\ -2L_1M+\sum_{i=M+1}^{N}[k(i-1)\times 2(L_1+L_2)]\} & 0<(\frac{\lambda}{2}-x)<2L_2\end{cases} \tag{16}$$

and:

$$\begin{cases}k=\dfrac{p}{\lambda/2}\\ N=【\dfrac{x}{2(L_1+L_2)}】+1.\\ M=【\dfrac{x-2L_1}{2(L_1+L_2)}】+1\end{cases} \tag{17}$$

So the separation time of bar 2 and bar 1 can be got by:

$$t_{\text{separation}}=\frac{x_{\text{separation}}+L_1+2L_2}{c}. \tag{18}$$

And the interaction time:

$$t_{\text{interaction}} = \frac{x_{\text{separation}} + 2L_2}{c}. \quad (19)$$

Equations (16) and (18) shows that, the separation time and the interaction time of bar 1 and bar 2 are determined by the length of bar 1 and bar 2, the length of the stress wave and the wave resistance (equals to the product of stress wave propagation velocity and the density of the bars).

If the calculation results of Equation (16) is that $x_{\text{separation}}$ is larger than length of the wave, bar 1 and bar 2 will not separate, although the reflective tensile stress of the end of the input wave has arrived at the contact surface of the two bars. Then, for this situation, the two bars will always move together, and never separate from each other. This situation is actually a completely inelastic collision.

3 Numerical simulation and verification

In order to verify the correctness of the theoretical analysis above, numerical simulation of the three cases above has been carried out by AUTODYN, and results of the separation time were given. The comparison between these numerical results and the analytical solutions according to the theoretical analysis in this paper has been done also.

In the following simulation and computation, bar 1 and bar 2 are both the 20 # steel (E=210 GPa, ρ=7.82 g/cm^3, μ=0.28), as well as the length both 1 metre.

3.1 CASE 1: $\lambda/2 < L_2$

Conditions of the input triangle wave: σ_0=300 Mpa, λ=292.96 mm.

The numerical simulation has been done by AUTODYN, and the results of displacement of points at the right side of bar 1 and the left side of bar 2 varying the time was shown in Figure 5.

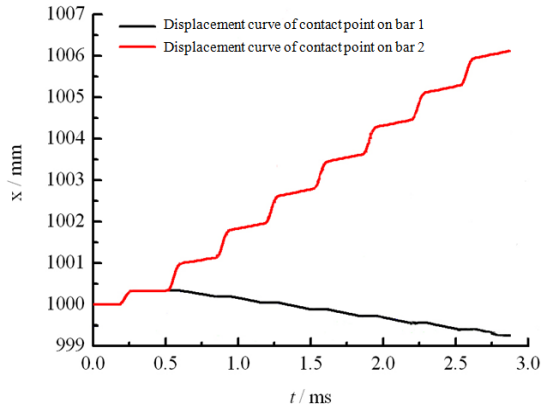


FIGURE 5 Curves of displacement of points at the right side of Bar 1 and the left side of Bar 2 varying the time

According to the simulations and Figure 5, we got the separation time of bar 1 and bar 2 was about 0.5117 ms.

As the elastic wave velocity in the bars is:

$$c = \sqrt{\frac{E(1-\mu)}{(1+\mu)(1-2\mu)\rho}} = 5859.24 \quad \text{mm/ms}.$$

According to Equation (2), we got:

$$t_{\text{separation}} = \frac{L_1 + 2L_2}{c} = 0.5120 \quad \text{ms}.$$

The relative error was:

$$\eta = \frac{0.5120 - 0.5117}{0.5117} = 0.058\%.$$

And the interaction time of bar 1 and bar 2 could also been obtained from Equation (1):

$$t_{\text{interaction}} = \frac{\lambda}{c} = 0.1 \quad \text{ms}.$$

3.2 CASE 2: $\lambda/2 > L_2$ and $\lambda/2 < (L_1+L_2)$.

Conditions of the input triangle wave: σ_0=300 Mpa, λ=1464.80 mm.

The numerical simulation has been done by AUTODYN, and the results of displacement of points at the right side of bar 1 and the left side of bar 2 varying the time was shown in Figure 6.

According to the simulations and Figure 6, we got the separation time of bar 1 and bar 2 was about 0.63 ms.

As the elastic wave velocity in the bars is:

$$c = \sqrt{\frac{E(1-\mu)}{(1+\mu)(1-2\mu)\rho}} = 5859.24 \quad \text{mm/ms}.$$

According to Equation (2), we got:

$$t_{\text{separation}} = \frac{0.5\lambda + L_1 + L_2}{c} = 0.5913 \quad \text{ms}.$$

The relative error was:

$$\eta = \frac{0.63 - 0.5913}{0.63} = 6.1\%.$$

The interaction time of bar 1 and bar 2 could also been obtained from Equation (1):

$$t_{\text{interaction}} = \frac{0.5\lambda + L_2}{c} = 0.4206 \quad \text{ms}.$$

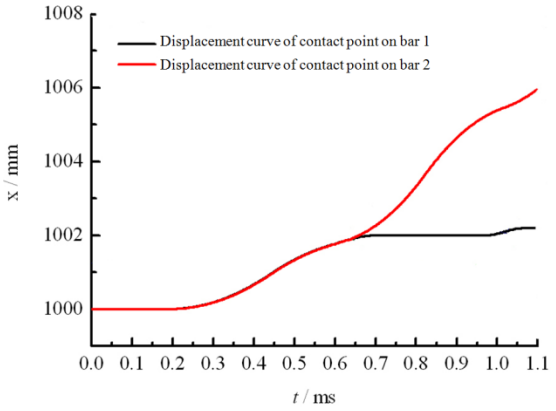


FIGURE 6 Curves of displacement of points at the right side of Bar 1 and the left side of Bar 2 varying the time

3.3 CASE 3: $\lambda/2 > (L_1+L_2)$

Conditions of the input triangle wave: σ_0=300 Mpa, λ=14648 mm.

The numerical simulation has been done by AUTODYN, and the results of displacement and velocity of points at the right side of bar 1 and the left side of bar 2 varying the time was shown in Figure 7 and Figure 8.

With reference to Figure 7 and Figure 8 we could see that bar 1 and bar 2 almost always move together, and don't separate from each other. Figure 8 showed that, during the first 5 ms, velocities of bar 1 and bar 2 have been increasing. After the first 5 ms, both bars travelled in a constant velocity.

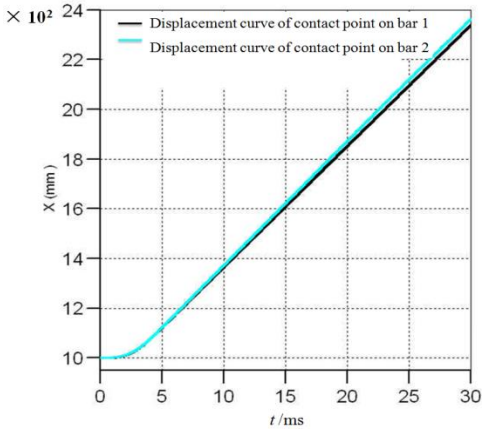


FIGURE 7 Curves of displacement of points at the right side of Bar 1 and the left side of Bar 2 varying the time

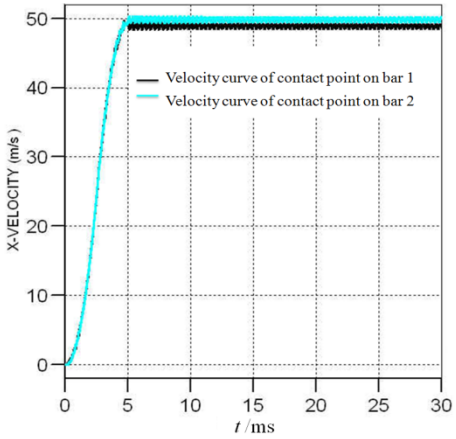


FIGURE 8 Curves of velocity of points at the right side of Bar 1 and the left side of Bar 2 varying the time

As the elastic wave velocity in the bars is:

$$c=\sqrt{\frac{E(1-\mu)}{(1+\mu)(1-2\mu)\rho}}=5859.24 \quad \text{mm/ms}.$$

According to the analysis in section II, we could get the following result by calculation:

$$x_{\text{separation}} > \lambda . \tag{20}$$

Solving the Equation (16) directly is a difficult work, but Equation (20) can be verified indirectly: substituting any $x_{\text{separation}}$ which was smaller than the wavelength λ " in Equation (9) and Equation (15), and solving these equations, we could obtain that $\sigma_{\text{compressive}}$ was always larger than σ_{tensile}. This result means that the compressive stress is always larger than the tensile stress at the contact surface of bar 1 and bar2. So the two bars didn't separate from each other. And this conclusion is consistent with the numerical simulation.

4 Conclusions

On the basis of the basic theory and laws of elastic wave propagation and reflection, stress wave propagation and reflection between two contact elastic bars, one of which was loaded by a triangle wave on the left, has been researched. For three different cases, formulas of separation time and interaction time have been derived. Numerical simulation of these three cases has also been done by using AUTODYN, and comparison of the numerical simulation and the results obtained by the method of this paper showed that the formulas derived here is correct.

Acknowledgments

The authors are grateful for the support of the Major Research plan of the National Natural Science Foundation of China (Grant No. 91330205).

References

[1] Wang L, Ren H, Yu J 2013 Development and application of the theory of nonlinear stress wave propagation *Chinese Journal of Solid Mechanics* **34**(3) 217-40

[2] Yang Z 2010 Research on Dynamic Mechanical Properties and Stress Wave Propagation Characteristics of Layered Composite Structures *Harbin China Harbin Institute of Technology* (*in Chinese*)

[3] Dong Y, Huang C, Duan Z 2005 Analysis on the influence of multi-layered media on stress wave propagation *Chinese Journal of High Pressure Physics* **19**(1) 59-65

[4] Gong M, Yu Y 1999 Experimental Study on stress wave propagation by a linear charge in anisotropic layered media *Engineering Blasting* **5**(1) 22-6

[5] Li Y L, Guo Y, Hu H 2009 A critical assessment of high-temperature dynamic mechanical testing of metals *International Journal of Impact Engineering* **36**(2) 177-84

[6] Sato Y, Komori M, Kikuchi W 1999 Application of the unsteady wave sensing system to the plate impact experiment *Journal of Materials Processing Technology* **85**(1-3) 43-7

[7] Stevens JB, Batra RC 1998 Adiabatic shear bands in the taylor impact test for a Wha rod *International Journal of Plasticity* **14**(9) 841-54

[8] Han C H 2000 Dynamic Response and Failure in Layered Structures and Composites *America Purdue University* 4-12

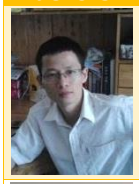
[9] Feng M, Peng Y, Liu Y 2006 Study on SHPB technique *Progress in Geophysics* **21**(1) 273-78

[10] Liu Y, Li R, Liu D 2007 Research on stress wave propagation of layered medium in SHPB *Chinese Science Paper Online* **2**(1)63-6

[11] Shan R, Jiang Y, Li B 2000 Obtaining dynamic complete stress strain curves for rock using the split hopkinson pressure bar technique *International Journal of Rock Mechanics and Mining Sciences* **37** 983-92

[12] Wang L 2010 *Foundation of Stress Waves* (*second edition*) *Beijing National Defense Industry Press* 41-2

Authors



Wei Zhong, China

Current position, grades: research assistant in Northwest Institute of Nuclear Technology, China.
University study: Master's degree in mechanical engineering, Northwest Institute of Nuclear Technology, Xi'an, China, 2012.
Research activities: mathematics, fluid dynamics, mechanical engineering, explosion mechanics, chemical reaction kinetics, numerical simulation methods.

Ming Lei, China

Current position, grades: research assistant in Northwest Institute of Nuclear Technology, China.
University study: Master's degree in mechanical engineering, Northwest Institute of Nuclear Technology, Xi'an, China, 2004.
Research activities: mathematics, finite element method, mechanical engineering, explosion mechanics, numerical simulation methods.

Zhou Tian, China

Current position, grades: professor in Northwest Institute of Nuclear Technology, China.
University study: PhD degree in nuclear science and technology, Northwest Institute of Nuclear Technology, Xi'an, China, 2003.
Research activities: mathematics, physics, mechanical engineering, detonation and shock waves, explosion mechanics, numerical simulation methods.

Research of numerical solutions of differential equations model based on the finite element method

Zhiyan Li[1]*, Baoxia Jin[2]

[1]*Handan College, Hebei, 056005, China*

[2]*Lushan College, Guangxi University of Science and Technology*

Received 15 May 2014, www.cmnt.lv

Abstract

Using the finite element method solving a class of second order ordinary differential equations, analyses the two-point boundary value problem of a class of second order ordinary differential equations, through numerical examples to validate its effectiveness.

Keywords: ordinary differential equations, finite element method, two-point boundary value problem

1 Introduction

Finite element method (fem) is a calculating method that booming in the 1960's, it has a wide range of applications, such as elasticity related problem [1-2], related problems in fluid mechanics [3], the heat conduction problem [4-5].

Using the finite element method solves the following two-point boundary value problem of second order ordinary differential equations:

$$\begin{cases} Ly := -\frac{d}{dx}(p(x)\frac{du}{dx}) + r(x)(\frac{du}{dx}) + q(x)u = f(x) \quad a < x < b & (1) \\ u(a) = u'(b) = 0 & (2) \end{cases}$$

Among them $p(x) \in C'[a,b], r(x), q(x), f(x) \in C[a,b], p(x) \geq p_{\min} > 0, q(x) \geq 0$.

2 Generalized solution

Through the following method, the equation (3) and equation (1) are equivalent, the solution of original problem is called the classical solution, and solution meet the integral form is called the generalized solution. To determine the generalized solution of the new equation is the starting point of the finite element method (fem).

Use it multiply equation (1) both ends, and find the x integrate on [a, b], using part of the integral, then we can get:

$$-p(x)u'(x)\Big|_{x=a}^{x=b} + \int_a^b p(x)u'(x)\phi'(x)dx + \int_a^b r(x)u'(x)\phi(x)dx + \int_a^b q(x)u(x)\phi(x)dx = \int_a^b f(x)\phi(x)dx.$$

By the boundary conditions (2) and $\phi(a) = 0$, then

$$\int_a^b p(x)u'(x)\phi'(x)dx + \int_a^b r(x)u'(x)\phi(x)dx + \int_a^b q(x)u(x)\phi(x)dx = \int_a^b f(x)\phi(x)dx. \quad (3)$$

Then the $u(x)$ is the general solution of boundary problem (1)-(2).

3 Element subdivision and interpolation

Let $a = x_0 < x_1 < x_2 < \cdots < x_n = b$, element e_i is region $[x_i, x_{i+1}]$, let $u(x_i) = u_i$ in $x = x_i$, u_0 is known, $u_1, \cdots, u_n$ is unknown. In finite element method (fem), when determining the subdivision, determine the specific form of interpolation polynomial on each small unit , and express it through the node function value, that means, in every element e_i , let $u(x) = N_i(x)u_i + M_i(x)u_{i+1} = [N]\{\delta\}e_i, \phi(x) = [N]\{\delta^*\}_{ei}$, and $N_i(x) = \frac{1}{L_i}(x_{i+1} - x), M_i(x) = \frac{1}{L_i}(x - x_i)$ is linear interpolation function.

$L_i = x_{i+1} - x_i, \phi(x)$ is virtual displacement $[N] = (N_i(x), M_i(x)), \{\delta\}_{e_i} = (u_i, u_{i+1})^T$ is nodal displacement in element e_i , $\{\delta^*\}_{e_i} = (u_i^*, u_{i+1}^*)^T$ is node virtual displacement vector in element e_i .

4 Computing units stiffness matrix and load vector

In this paper, the finite element method (fem) is given by specific physical instance and in the engineering, matrix $[K]_{e_i}$ reflects the unit's rigidity, $[K]_{e_i} \cdot \{\delta\}_{e_i} = \begin{pmatrix} -U_i^{e_i} \\ -U_{i+1}^{e_i} \end{pmatrix}$ show that in order to maintain deformation of the unit e_i , two endpoints nodes $x = x_i$, x_{i+1} in unit e_i need external force, to his balance, the two external force is the force

* ***Correspondign author*** e-mail: lzyzhiyanli@163.com

$-U_i^{e_i}$ and $-U_{i+1}^{e_i}$ through node, and the force known as the equivalent nodal force. While unit load vector $\{F\}e_i$ reflects the effect on the unit e_i displacement distribution of physical strength, its each component shows after the displacement the effect on the two end node $x = x_i, x_{i+1}$ of equivalent strength, therefore, at first, we need to calculate the element stiffness matrix.

Rewrite (3) as

$$\sum_{i=0}^{n-1}\int_{e_i}[p(x)u'(x)\phi'(x)+r(x)u'(x)\phi(x)+q(x)u(x)\phi(x)]dx=\sum_{i=0}^{n-1}\int_{e_i}f(x)\phi(x)dx\ . \tag{4}$$

Using u_i, u_{i+1} and $u_i^*, u_{i=1}^*$ represent $\int_{e_i}(pu'\phi'+ru'\phi+qu\phi)dx, \int_{e_i}f\phi dx$. If a number of a is regarded as a matrix of order, we can see $a^T=a$, then

$$\begin{aligned}\int_{e_i}(pu'\phi'+ru'\phi+qu\phi)dx&=\int_{x_i}^{x_{i+1}}(p(\phi')^Tu'+r\phi^Tu'+q\phi^Tu)dx\\&=\int_{x_i}^{x_{i+1}}p([B]\{\delta^*\}_{e_i})^T([B]\{\delta\}_{e_i})dx+\int_{x_i}^{x_{i+1}}r([N]\{\delta^*\}_{e_i})^T([B]\{\delta\}_{e_i})dx+\int_{x_i}^{x_{i+1}}q([N]\{\delta^*\}_{e_i})^T([N][\delta]_{e_i})dx\\&=\{\delta^*\}_{e_i}^T(\int_{x_i}^{x_{i+1}}p[B]^T[B]dx+\int_{x_i}^{x_{i+1}}[N]^T[B]dx+\int_{x_i}^{x_{i+1}}q[N]^T[N]dx)\{\delta\}_{e_i}dx\\&=\{\delta^*\}_{e_i}^T[K]_{e_i}\{\delta\}_{e_i}\end{aligned}\tag{5}$$

$[K]_{e_i}$ is called element stiffness matrix, its concrete form is:

$$\begin{aligned}[K]_{e_i}&=\int_{x_i}^{x_{i+1}}(p[B]^T[B]+[N]^T[B]+q[N]^T[N])dx\\&=\begin{pmatrix}\frac{p_i}{L_i^2}+\frac{r_i}{L_i}\int_{x_i}^{x_{i+1}}N_idx+\int_{x_i}^{x_{i+1}}qN_i^2dx & -\frac{p_i}{L_i^2}+\frac{r_i}{L_i}\int_{x_i}^{x_{i+1}}N_idx+\int_{x_i}^{x_{i+1}}qN_iM_idx\\-\frac{p_i}{L_i^2}+\frac{r_i}{L_i}\int_{x_i}^{x_{i+1}}N_idx+\int_{x_i}^{x_{i+1}}qN_iM_idx & \frac{p_i}{L_i^2}+\frac{r_i}{L_i}\int_{x_i}^{x_{i+1}}N_idx+\int_{x_i}^{x_{i+1}}qM_i^2dx\end{pmatrix}.\end{aligned}$$

Among it, $p_i=\int_{x_i}^{x_{i+1}}p(x)dx$.

Make the element stiffness matrix as $[K]_{e_i}=\begin{pmatrix}k_{i,i}^{e_i} & k_{i,i+1}^{e_i}\\ k_{i+i,i}^{e_i} & k_{i+1,i+1}^{e_i}\end{pmatrix}$, then stiffness coefficient $k_{j,m}^{e_i}$ is

$$k_{i,i}^{e_i}=\frac{p_i}{L_i^2}+\frac{r_i}{L_i}\int_{x_i}^{x_{i+1}}N_idx+\int_{x_i}^{x_{i+1}}qN_i^2dx$$

$$k_{i,i+1}^{e_i}=k_{i+1,i}^{e_i}=-\frac{p_i}{L_i^2}+\frac{r_i}{L_i}\int_{x_i}^{x_{i+1}}N_idx+\int_{x_i}^{x_{i+1}}qN_iM_idx,$$

$$k_{i+1,i+1}^{e_i}=\frac{p_i}{L_i^2}+\frac{r_i}{L_i}\int_{x_i}^{x_{i+1}}N_idx+\int_{x_i}^{x_{i+1}}qM_i^2dx\ . \tag{6}$$

$$\begin{aligned}\int_{e_i}f\phi dx&=\int_{e_i}\phi^Tfdx\\&=\int_{x_i}^{x_{i+1}}([N]\{\delta^*\}_{e_i})^Tfdx=\{\delta^*\}_{e_i}^T\{F\}_{e_i}\end{aligned}$$

$\{F\}_{e_i}$ is called Unit load vector, its concrete form is:

$$\{F\}_{e_i}=\int_{x_i}^{x_{i+1}}[N]^Tfdx=\begin{pmatrix}\int_{x_i}^{x_{i+1}}N_ifdx\\ \int_{x_i}^{x_{i+1}}M_ifdx\end{pmatrix}.$$

Expressed $\{F\}_{e_i}$ as $\{F\}_{e_i}=\begin{pmatrix}F_i^{e_i}\\ F_{i+1}^{e_i}\end{pmatrix}$, then, unit load coefficient is $F_i^{e_i}=\int_{x_i}^{x_{i+1}}N_ifdx, F_{i+1}^{e_i}=\int_{x_i}^{x_{i+1}}M_ifdx$.

5 Total stiffness matrix and the total load vector

Expand $[K]_{e_i}, \{F\}_{e_i}, \{\delta\}_{e_i}$ and $\{\delta^*\}$ for n + 1 order matrix and the n + 1 d vector, put (5) and (6) into (4), then

$$\begin{aligned}&\{\delta^*\}^T(\sum_{i=0}^{n-1}[K]_{e_i})\{\delta\}=\{\delta^*\}^T(\sum_{i=0}^{n-1}\{F\}_{e_i})\\&\{\delta^*\}^T([K]\{\delta\}-\{F\})=0\end{aligned}\ . \tag{7}$$

Among them, $[K]=\sum_{i=0}^{n-1}[K]_{e_i}, \{F\}=\sum_{i=0}^{n-1}\{F\}_{e_i}$ as the total stiffness matrix and the total load vector respectively.

6 Constraint handling

As on the endpoint x = 0, $u(x)$ satisfy the boundary conditions (2), therefore $u_0=0$, in addition $\phi(0)=0$, so

$u_0^* = 0$, and (7) has the form $\left\{\tilde{\delta}^*\right\}^T([\tilde{K}]\{\delta\} - \left\{\tilde{F}\right\}) = 0$, $\left\{\tilde{\delta}^*\right\}$, $\left\{\tilde{F}\right\}$ are respectively after the first element with 0 instead of n+1-dimensional vector in $\left\{\delta^*\right\}$, $\{F\}$, $\left[\tilde{K}\right]$ is the element in the first row and the first column in addition to the diagonal to 1, use 0 instead of the back rest of the n+1 matrix. Due to the arbitrary of $\left\{\tilde{\delta}^*\right\}$, we can obtain:

$$\left[\tilde{K}\right]\{\delta\} = \left\{\tilde{F}\right\}. \tag{8}$$

From the change in the overall stiffness matrix $[K]$ for the first line in the first column elements, forming process of $\left[\tilde{K}\right]$, known as the constraint processing. Main purpose is to require the desires of function satisfy the boundary constraint conditions $u_0 = 0$ on boundaries. After dealing with the constraints, a new overall stiffness matrix $\left[\tilde{K}\right]$ from positive semi-definite to positive definite. In fact, because: $\left\{\tilde{\delta}\right\}^T[K]\{\delta\} = (0, u_1, \cdots, u_n)$ $[K](0, u_1, \cdots, u_n)^T \geq 0$

So $\left[\tilde{K}\right]$ is symmetric positive definite, so a linear equations (8) solution is unique. Because above is our whole process in solving the equation (3), and because equation (3) and equation (1) are equivalent, so here equation (8) is the solution of the original problem (1) what we want.

7 A Numerical example and conclusion

Consider the following second order ordinary differential equation with two-point boundary value problems: $\begin{cases} -(x^2u'(x))' + 2u(x) = 2x^2, 0 < x < 1 \\ u(0) = u'(1) = 0 \end{cases}$.

Compare with equation (1), this example as $p(x) = x^2$, $r(x) = 0$, $q(x) = 2$, $f(x) = 2x^2$, now the interval [0, 1] is divided into four units, the five coordinates of nodes are $x_0 = 0, x_1 = \frac{1}{4}, x_2 = \frac{1}{2}, x_3 = \frac{3}{4}, x_4 = 1$ respectively.

Stiffness matrix:

$$[K]_{e_i} = \begin{pmatrix} 4(x_{i+1}^2 - x_{i+1}x_i + x_i^2) & \frac{4}{3}(x_{i+1}^2 + x_i^2) - \frac{20}{3}x_{i+1}x_i \\ \frac{4}{3}(x_{i+1}^2 + x_i^2) - \frac{20}{3}x_{i+1}x_i & 4(x_{i+1}^2 - x_{i+1}x_i + x_i^2) \end{pmatrix}.$$

Unit load vector:

$$[F]_{e_i} = \begin{pmatrix} \frac{2}{3}x_{i+1}(x_{i+1}^2 + x_{i+1}x_i + x_i^2) - \frac{1}{2}(x_i + x_{i+1})(x_i^2 + x_{i+1}^2) \\ \frac{1}{2}(x_i + x_{i+1})(x_i^2 + x_{i+1}^2) - \frac{2}{3}x_{i+1}(x_{i+1}^2 + x_{i+1}x_i + x_i^2) \end{pmatrix}.$$

The overall stiffness matrix is $[K] = \sum_{i=0}^{3}[K]_{e_i}$, total load vector is $[F] = \{F\} = \sum_{i=0}^{3}\{F\}_{e_i}$, the exact solution is $u(x) = x - \frac{1}{2}x^2$, the numerical results are shown in Table 1.

TABLE 1 Numerical result

node	Exact solution	Approximate solution	Absolute error	Relative error
1/4	0.21875	0.21637	2.38E-03	1.088E-02
1/2	0.375	0.37476	2.4E-04	6.4E-04
3/4	0.46875	0.46881	6.0E-05	1.28E-04
1	0.5	0.52475	2.475E-02	4.95E-02

This paper discusses the application of finite element method for solving a class of second order ordinary differential equation numerical solution, analyses a class of second order ordinary differential equation with two-point boundary value problem, the effectiveness of finite element method is verified by a numerical example for solving a class of second order ordinary differential equation numerical solutions algorithm.

References

[1] Chen M C, Ping XC 2009 A Novel Hybrid Finite Element Analysis of Plane Singular Elastic Field around Inclusion Corners in Elastic Media *International Journal of Solids and Structures* **46**(13) 2527-38

[2] Reddy R M, Rao B N 2008 Fractal Finite Element Method Based Shape Sensitivity Analysis of Mixed - mode Fracture *Finite Elements in Analysis and Design* **44**(15) 875-88

[3] Ammar A, Clermont J R 2005 A Finite -element Approach in Stream - tube Method for Solving Fluid and Solid Mechanics Problems *Mechanics Research Communication* **32**(1) 65-73

[4] Huiping L, Guoqun Z, Shanting N, et al 2006 Inverse Heat Conduction Analysis of Quenching Process Using Finite - element and Optimization Method *Finite Elements in Analysis and Design* **42**(12) 1087-96

[5] Nicolai B M, Egea J A 2011 Fuzzy Finite Element Analysis of Heat Conduction Problem with Uncertain Parameters *Journal of Food Engineering* **103**(1) 38-46

Authors

Zhiyan Li, born in 1982, Hebei Province of China

Current position, grades: lecturer
University studies: Master's degree was earned in major of computational mathematics, Zhengzhou University in 2008.
Scientific interest: the finite element method and its application

Baoxia Jin, born in 1981, Hebei Province of China

Current position, grades: lecturer
University studies: Master's degree was earned in major of the partial differential equation and its application, Zhengzhou University in 2008.
Scientific interest: the partial differential equation and its application

Design and application of iterative Monte Carlo localization for mobile wireless sensor networks based on MCL

Jing Cao[1*], Xuefeng Xing[2], Shan Liu[1]

[1]*School of Mathematics and Information Science & Technology, Hebei Normal University of Science & Technology, Hebei, 066004, China*

[2]*Northeast Petroleum University at Qinhuangdao, Hebei, 066004, China*

Received 1 October 2014, www.cmnt.lv

Abstract

In recent years, wireless sensor network had been more and more widely used in our daily life, and with the propose of Monte Carlo localization (MCL) algorithm, node localization of the mobile wireless sensor network had been solved effectively. But it needed to have a large number of samples if it used the Monte Carlo localization algorithm to obtain a high positioning accuracy. This paper proposed a new improved algorithm (iterative Monte Carlo localization algorithm) based on the Monte Carlo localization algorithm. In iterative Monte Carlo localization (IMCL) algorithm, each anchor node location information was forwarded by its neighbour nodes only once and preserved by the receiving node in each period. Then the next period, merge it and the sent/forwarding information into a packet and forward. Make sure that points have more observations for estimating previous location sets. IMCL, meanwhile, also can make full use of observation to filter out some samples that were far from the real position of node, so as to improve the accuracy of node localization. We finally confirmed by experiment: IMCL algorithm had higher positioning accuracy compared with other algorithm.

Keywords: IMCL, immune genetic algorithm, wireless sensor network, improve the accuracy of localization

1 Introduction

Wireless Sensor network (WSNs) has characteristics of self-organization, low cost, low power consumption, and it has broad application prospects in the environmental monitoring, remote medical monitoring, military and other fields. Generally, it is meaningful for monitoring information that the location has been determined. GPS devices can be installed on each sensor node to determine the location information, but it is very expensive. Therefore, the research of wireless sensor network node location technology is of great significance.

In recent years, many scholars at home and abroad devote to the research of node location technology, and put forward many effective algorithm Jennifer Yick [1]. During the process of positioning, according to whether there is need to measure the distance between nodes, node localization algorithm can be divided into two categories:

1) Positioning algorithm based on distance: these algorithms rely on the range information between the nodes of additional hardware measuring, such as the received signal strength (RSSI), the time of arrival (TOA) Kuang Xinghong [2], arrival time difference of the different signal (TDOA) and the signal arrive Angle (AOA). The shortcomings of this method are that the extra range of hardware increases volume and the cost of node.

2) Localization algorithm irrelevant with distance: this kind of algorithm achieves node localization mainly based on the connectivity between the nodes, and localization cost is low. Centroid algorithm is a simple localization algorithm which is irrelevant with distance, unpositioning node takes polygon centroid as its estimate which is composed by its neighbours of beacon nodes. DV - Hop algorithm is typical. In addition, many scholars also explore the positioning method based on the mobile beacon nodes Pubudu N Pathirana [3], Kuo-Feng Su [4], Luo Ji [5].

2 MCL algorithm

2.1 THE INTRODUCTION OF MCL ALGORITHM

In MCL algorithm time is divided into several discrete time periods and suppose that the unpositioning node and beacon nodes do not know their own direction and speed, but their maximum velocity $v_{\max}$ is given. Unpositioning node needs to relocation in each time period. $L_k = \{l_k^0, l_k^1, ..., l_k^{n-1}\}$ means the location of the sample set of unpositioning node at k time. At the beginning, unpositioning node has no any information about the location, so it randomly selects N points to form the initial location of sample set $L_0 = \{l_0^0, l_0^1, ..., l_0^{n-1}\}$ from decorate area. In the following time period, unpositioning node repeats the period of forecast and filtering to implement the position. Take the positioning process at k time for example. In the prediction stage, according to the location

[*]***Corresponding author*** e-mail: cjcaojing@163.com

of sample set at the k–1 $L_{k-1}=\{l_{k-1}^0, l_{k-1}^1, ..., l_{k-1}^{n-1}\}$ and its motion model, unpositioning node randomly collects N samples $L_k=\{(l_k^0)c, (l_k^1)c, ..., (l_k^{n-1})c\}$, namely $(l_k^i)c, (i=0,1,2,...,N-1)$ randomly selects from the circle which is centered in l_{k-1}^i and the radius is $v_{\max}$. In updating stage, unpositioning node uses the observed value $O_k=\{S_1, S_2, ..., D_1, D_2, ...\}$ received at k moment to filter samples that do not satisfy the conditions. S_j and D_j is the position of the beacon nodes of unpositioning node. Suppose receive all the information instantly at k time and the define $d(l_1, l_2)$ is the Euclidean distance between location l_1 and l_2, r is the wireless communication distance of a node. Filter conditions of samples $(l_k^i)c$ are:

$$\begin{matrix} Filter((l_k^i)c)=PS_j, d((l_k^i)c), S_j \\ rCPD_j, r<d((l_k^i)c, D_j)|2r \end{matrix} . \quad (1)$$

If the value of the filter conditions is false, $(l_k^i)c$ will be filtered. In particular, if the distance between $(l_k^i)c$ and jump of the beacon node is greater than r or the distance between $(l_k^i)c$ and two jump beacon node is not between r and 2r, then it means that $(l_k^i)c$ does not comply with the conditions, and should be filtered. On the other hand, $n(l_k^i)c$ should be stayed.

Due to the unfitted samples have been filtered, after filtering stage, there may be not enough samples, unpositioning node repeats prediction and filtering process until the samples are enough. Then, these samples form the location of the sample set $L_k=\{l_k^0, l_k^1, ..., l_k^{n-1}\}$ at k. And take all the sample average as position estimation of unpositioning node at k.

2.2 THE ANALYSIS OF MCL ALGORITHM

According to the positioning process of MCL algorithm, we can deduce that the following two factors may affect the point positioning accuracy of unpositioning node at k.

1) Observations O_k: the direct way to increase the observation is to improve the density of beacon node, but it will increase the cost of system, including hardware cost and communication consumption of beacon nodes. Another common method is to use location information of multiple hops beacon node, but it can lead nodes forwarding packets frequently in each time period to increase the node communication costs. In addition, multiple hops can also cause a large time delay of communication.

2) The location of the sample set L_{k-1} at k–1: if at k–1, the more closer to the actual location of unpositioning node $l_{k-1}^i, (i=0,1,2,...,N-1)$ is at k, the larger the probability that predicting samples $(l_k^i)c$ is close to the node's actual location is. And improve the positioning accuracy of unpositioning node at k.

3 IMCL algorithm

3.1 THE BASIC DEFINITIONS AND SYMBOLS

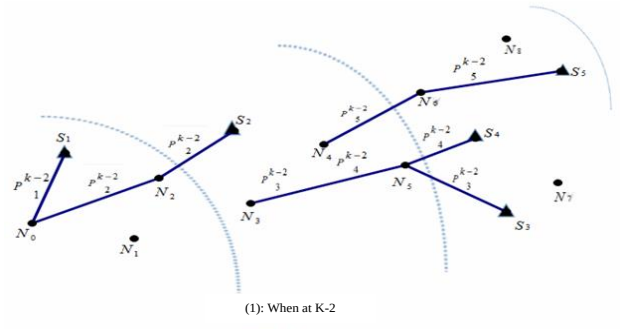
For convenience, this section will illustrate the related definitions and symbols.

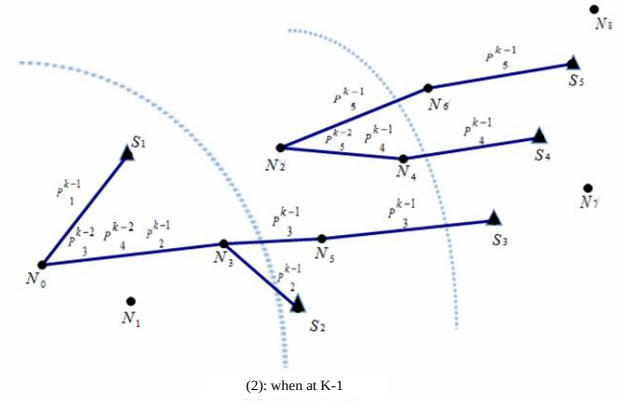
S_i means the beacon node I, i=0,1,2,…. N_i means the unpositioning node I, i=0,1,2,…. S_k^i is the position of beacon node S_i of unpositioning node at k. D_k^i is the position of beacon node S_i of unpositioning node at k. T_{k-1}^i is the position of beacon node S_i of unpositioning node at k–1. However, unpositioning node can receive information at k, but not at k–1; M_{k-1}^i is the position of multiple jumps beacon node S_i of unpositioning node at k–1(l\2). Similarly, the unpositioning node can receive information at k. O_{k-1}^k is the observation of unpositioning node at k which is used to calculate the position sample set at k–1, $l_{k-1}^i|k$ is a sample set during prediction stage, when the unpositioning node at k calculates the position sample set at k–1.

3.2 DESIGN OF ALGORITHM

Firstly, analyse how unpositioning node estimates observations of position sample sets in previous times. As shown in Figure 1, a circle means the moving unpositioning node, triangle means the moving beacon nodes. Linear between nodes means communication link. In order to simplify the graphic, Figure 1 only gives a few typical communication links. The line symbol p_j^t represents the position of the beacon node j at t. It is included in the node packets sent currently. Similar to the MCL algorithm, these packets which include position of beacon node can be forwarded at most once in each segment. However, the node, which receives the packet will save this information, and merge them into momentum/forwarding packets to forward in the next period of time. At k–2, for example, N_2 receives the packets which contains its beacon node location (p_2^{k-2}), and forward the packet immediately. At k–2, N_4 receives the packet which contains its beacon node location (p_5^{k-2}), but N_4 does not forward it immediately. At k–1, N_4 receives the packet which contains its beacon node location (p_4^{k-2}), and then forwards the packet, at this point; N_4 merges p_5^{k-2} into the packets and sends out. Obviously, IMCL algorithm does not increase the packet

number. Of course, the bytes number of each packet will increase. However, through proper fusion technology, we can minimize the number of bytes.





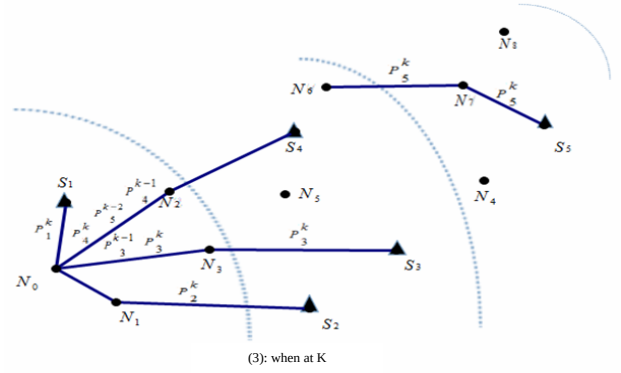


FIGURE 1 The communication link of nodes between k–2 and k, where N_0 is static

According to the analysis of Figure 1, from k–2 to k, the observations of position sample set that the unpositioning node N_0 saves to estimate at k–2 are:

$$\begin{aligned} O_{k-2}^{k-2} &= \left\{S_{k-2}^1, D_{k-2}^2\right\}, \\ O_{k-2}^{k-1} &= \left\{S_{k-2}^1, D_{k-2}^2, T_{k-2}^3, T_{k-2}^4\right\}, \\ O_{k-2}^{k} &= \left\{S_{k-2}^1, D_{k-2}^2, T_{k-2}^3, T_{k-2}^4, M_{k-2}^5\right\}. \end{aligned} \tag{2}$$

It shows that unpositioning node can obtain more observations of position sample set of k-i at k-j than at k-s. According to the analysis of the influence factors 1 in the section 2.2, the result of position sample set of k-i at k-j is better than at k-s. At k, therefore, it is necessary for unpositioning node to recalculate the periods location of the sample set to get a better L_{k-1}. At the same time, IMCL refers to MCB algorithm in order to collect better samples during forecast period. The definition $L_{t'}^{t}$ is the position sample set of t_c that is calculated by unpositioning set at t. Figure 2 is the flow diagram of IMCL algorithm. Consider unpositioning node reiterate location of the sample set of previous n–1 times at k, then L_k^k can be got from the process of n MCB iterative calculation, namely:

$$\begin{aligned} L_{k-n+1}^k &= MCB\left\{L_{k-n}^{k-1}, O_{k-n+1}^k\right\}, \\ L_{k-n+2}^k &= MCB\left\{L_{k-n+1}^k, O_{k-n+2}^k\right\}, \\ L_{k-n+3}^k &= MCB\left\{L_{k-n+2}^k, O_{k-n+3}^k\right\}, \\ L_{k-1}^k &= MCB\left\{L_{k-2}^k, O_{k-1}^k\right\}, \\ L_k^k &= MCB\left\{L_{k-1}^k, O_k^k\right\}. \end{aligned} \tag{3}$$

The average of all samples in L_k^k are the estimated position of the unpositioning node at k.

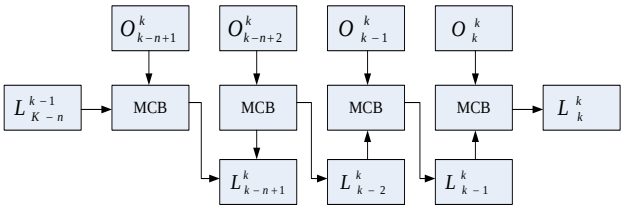


FIGURE 2 The diagram of IMCL algorithm

3.3 THE FILTERING CONDITIONS

In order to make full use of the observed value, we need to complete the filtering conditions of MCL and MCB algorithm. In IMCL, observation O_{k-1}^k contains four types of beacon node location: S_{k-1}^i, D_{k-1}^i, T_{k-1}^i and M_{k-1}^i. Therefore, in the process of each MCB, prediction sample $l_{k-i}^{i'} \mid k$ has four corresponding filter conditions:

$$filter1\left(l_{k-i}^{i'} \mid k\right) = PS_{k-1}^i \mid O_{k-1}^k, d\left(l_{k-1}^{i'} \mid k, S_{k-1}^i\right) \mid r, \tag{4}$$

$$filter2\left(l_{k-i}^{i'} \mid k\right) = PD_{k-1}^i \mid O_{k-1}^k, r < d\left(l_{k-1}^{i'} \mid k, D_{k-1}^i\right) \mid 2r, \tag{5}$$

$$\begin{aligned} &filter3\left(l_{k-i}^{i'} \mid k\right) = PT_{k-1}^i \mid O_{k-1}^k, \\ &2r < d\left(l_{k-1}^{i'} \mid k, T_{k-1}^i\right) \mid 2r + 2v_{\max}, \end{aligned} \tag{6}$$

$$filter4\left(l_{k-i}^{i'} \mid k\right) = PM_{k-1}^{i} \mid O_{k-1}^{k},$$
$$2r < d\left(l_{k-1}^{i'} \mid k, M_{k-1}^{i}\right). \quad (7)$$

In addition, the movement distance of the nodes within a period of time is not more than $v_{\max}$. This means that if N_j and S_i are not neighbor nodes k–l–1, but are neighbor nodes in k–l, then the distance between them is between r–$v_{\max}$ and r in k–l. On the contrary, the distance between them should be between r-v_{max} and r+v_{max}. This type of filtering is called Motion on Boundary, which can be written as the following form:

$$filter5\left(l_{k-i}^{i'} - 1 \mid k\right) = PS_{k-1}^{i} \mid O_{k-1}^{k} CS_{k-1}^{i} \mid O_{k-1}^{k},$$
$$r - 2v_{\max} < d\left(l_{k-1}^{i'} \mid k, S_{k-1}^{i}\right) \mid r, \quad (8)$$

$$filter6\left(l_{k-i}^{i'} - 1 \mid k\right) = PS_{k-l-1}^{i} \mid O_{k-l-1}^{k} CS_{k-1}^{i} \mid O_{k-1}^{k},$$
$$r - v_{\max} < d\left(l_{k-1}^{i'} \mid k, S_{k-l-1}^{i}\right) \mid r + v_{\max}, \quad (9)$$

So the filtering conditions of prediction sample $l_{k-1}^{i'} \mid k$ are:

$$Filter\left(l_{k-1}^{i'} \mid k\right) = Filter1\left(l_{k-1}^{i'} \mid k\right) C,\ filter6\left(l_{k-1}^{i'} \mid k\right). \quad (10)$$

We find that T_{k-1}^{i} and M_{k-1}^{i} may be useless O_{k-1}^{i}. Because node is random in movement speed and direction will lead nodes cannot be connected in a certain period of time. In k-l, for example, the distance between S_i and N_i is between r and 2r, but they lack a rely node, so N_l can't get the location information of S_i at a time of k-l, but can receive this information at the time of k, namely T_{k-1}^{i} or M_{k-1}^{i}. The process according to the sub-conditions of filter 3 or 4 is clearly not appropriate. Although the probability is low, in order to make more precise positioning process, in IMCL algorithm, set a limit. If the number of filtering prediction sample which does not meet filter3 or filter4 is more than the limit value, then remove T_{k-1}^{i} or M_{k-1}^{i} from O_{k-1}^{i}.

4 The simulated analysis

According to IMCL algorithm, we extend the MCL simulator. This section uses extended simulator to test the performance of IMCL algorithm. All nodes randomly distributed in a 500 m @ 500 m square area, and the communication radius of beacon node and unpositioning node is 50 m, the density of unpositioning node (the average number of neighboured node), number of samples are 10 and 50 respectively. We analyse the IMCL algorithm from several aspects:

1) With the change of time, the positioning error of different MCL algorithm.

2) Research the influence of beacon node density S_d to the positioning accuracy. In the rest tests, the density of beacon node is fixed, that is S_d=1.

3) Research the influence of the node maximum speed $v_{\max}$ to the positioning performance. In the rest tests, set $v_{\max}$=0.2r.

4) Analyse the influence of iterations n to the positioning accuracy and processing time. In other tests, get n=6.

All the simulation results are got from the average of 20 times independent experiments. Positioning error is the ratio between absolute positioning error and the radius of wireless communication. The average position error is the average of positioning error of all unpositioning nodes. In particular, the average positioning error of k3.2~k3.4 is measured in the stable stage of position.

4.1 POSITIONING ERROR

Figure 3 is the average positioning error curve of different MCL algorithm. During the positioning initialization phase, the average position error between IMCL and MCB is close; however, as time goes on, IMCL achieves better location performance. In the end, all MCL algorithm can achieve a stable state, i.e. the average positioning error fluctuates in a small scope. In stable phase, compared with the MCL and MCB, the average positioning error of IMCL algorithm has fallen about 40% and 40% respectively.

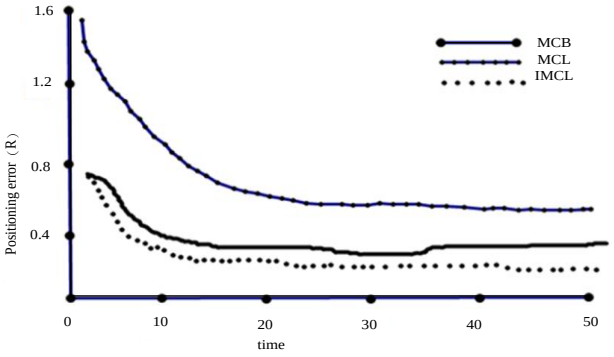


FIGURE 3 The average positioning error of different MCL $(S_d = 1, V_{\max} = 0.2x, n = 6)$

4.2. THE DENSITY OF BEACON NODE

Increase the density of beacon node can improve positioning accuracy, but also increase the network cost and energy consumption of communication. As the change of S_d, the average position error of different MCL algorithm is shown in Figure 4. With the increase of S_d, positioning accuracy of all MCL algorithms is increased, because the unpositioning node can obtain more observations. But it is worth noting that the increase of S_d has large influence on positioning accuracy of the MCL and MCB. In particular, when S_d is greater than 1.5, the average positioning error of IMCL is close to MCB. For how to get more observations is one of the inspired factors

of IMCL algorithm, so if the S_d is too high, there will exist redundant observation. This shows that under the condition of the low density of beacon node, IMCL algorithm has more advantages.

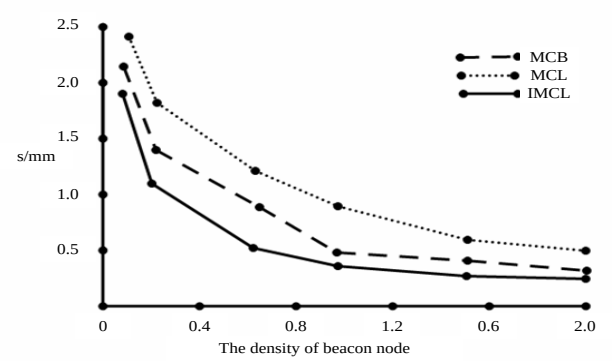


FIGURE 4 The influence of beacon node density to the positioning accuracy $(V_{\max}=0.2r, n=6)$

4.3 THE MAXIMUM SPEED OF NODE

Figure 5 is the average position error curve of different MCL algorithm with the change of v_{max}. Pending a node and beacon node's actual speed is uniformly distributed in the interval [0, v_{max}], so the node's average speed is proportional to the v_{max}. From the diagram, we can see that along with the increase of v_{max}, the difference of the average position error between IMCL with MCB (MCL) decreases. Because v_{max} affects the positioning process from two aspects. First of all, in the prediction stage, the sample is selected from the circle which is centred on l_{t-1}^{i} and the radius is v_{max}, so the increase of v_{max} will cause a decline in the accuracy of the sample collection. The negative effect on all MCL algorithms is same. In each time period, on the other hand, the greater the average speed of nodes is, the more observed values received are, which can filter out more samples do not comply with the conditions. However, one of the key technologies of IMCL algorithm is to make full use of the observed value to filter out worse samples, so the positive effect is not obvious.

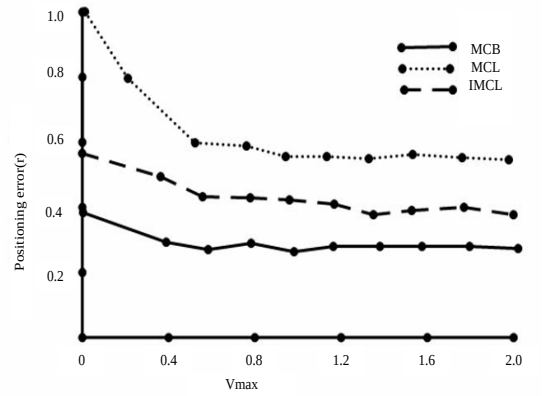


FIGURE 5 The influence of node maximum speed on the positioning accuracy

4.4 ITERATIONS

Change the number of iterations n, the average position error and the average processing time of IMCL (in each time period of the positioning process, the time of performing prediction and update phase) are shown in Figures 6 and 7 respectively. With the increase of the number of iterations, the positioning accuracy has improved, the processing time also increases. However, when the number of iterations reaches to a certain value, the number of iterations has less influence on positioning accuracy. Therefore, by adjusting the number of iterations cans the keep the relationship between the positioning accuracy and the processing time required balance. In addition, we can find an interesting result from Figure 7: the required processing time when the number of iterations is 1 is longer than that when the number of iterations is 2. Because when n=1, IMCL algorithm degenerates into MCB algorithm. In MCB algorithm, the accuracy of position of sample set in predict phase is low, so after filtering stage, reserved samples is less, leading that we have to constantly perform re-sampling/update process in place to ensure that there is enough samples.

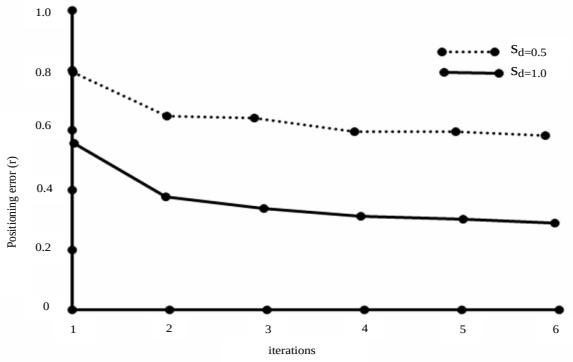


FIGURE 6 The influence of iterations on positioning accuracy $(V_{\max}=0.2r)$

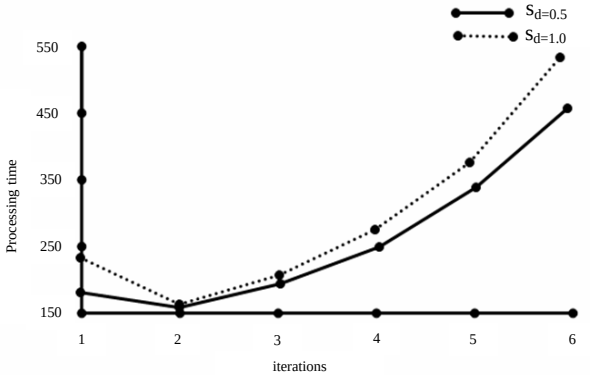


FIGURE 7 The relationship between iterations and processing time $(V_{\max}=0.2r)$

5 Conclusion

Based on the MCL positioning technology of mobile wireless sensor network, this paper puts forward an iterative Monte Carlo localization algorithm. Use extension MCL simulator to process simulation experiments, and analyse algorithm performance from beacon node density, node maximum velocity, number of iterations, etc. The simulation results show that compared with the MCB and MCL, IMCL algorithm is better in performance. In particular, when beacon node density is lower, the superiority of IMCL algorithm can be better reflected. In addition, by adjusting the number of iterations can keep the relationship between the positioning accuracy and the processing time required balance.

References

[1] Yick J, Mukherjee B, Ghosa D 2008 Wireless Sensor Network Research *Computer Network* **52**(12) 2292-330
[2] Kuang X, He S 2007 In Wireless Sensor Networks Distributed Node Localization Using Mobile Signal *China University of Post and Telecommunications* **14**(4) 7-12
[3] Pubudu N Pathirana, Nirupama Bulusu, Savkin AV 2005 *IEEE Portable Computer* **4**(3) 285-96
[4] Kuo-Feng Su, Chia-HoOu, Hewijin Christine Jiau 2005 *IEEE Trading Vehicle Technology* **54**(3) 1187-97
[5] Luo J, Zhang Q 2007 Mobile Sensor Network Localization Based on Relative Distance *IEEE Application of Telecommunications Assembly* 1076-80

Authors

Jing Cao, Hebei Province of China

Current position, grades: lecturer.
University studies: Master degree in Yanshan University in 2009.
Scientific interest: computer technology.

Xuefeng Xing, Hebei Province of China

Current position, grades: research assistant.
University studies: Master degree of computer application technology, Yanshan University in 2010.
Scientific interest: computer technology.

Shan Liu, Hebei Province of China

Current position, grades: lecturer.
University studies: Master degree of computer software and theory, Sichuan Normal University in 2006.
Scientific interest: computer technology.

Algorithm design and the application for cluster validity based on geometric probability

Jian-Wei Li[1*], Xiao-Wen Li[2]

[1]*College of physics and information engineering, Fuzhou University, Fuzhou 350108*

[2]*College of mathematics and computer science, Longyan University, Longyan 364012*

Received 20 May 2014, www.cmnt.lv

Abstract

Determining optimum cluster number is a key research topic included in cluster validity, a fundamental problem unsolved in cluster analysis. In order to determine the optimum cluster number, this article proposes a new cluster validity function for two-dimension datasets theoretically based on geometric probability. The function makes use of the corresponding relationship between a two-dimension dataset and the related two-dimension discrete point set to measure the cluster structure of the dataset according to the distributive feature of the point set in the characteristic space. It is designed from the perspective of intuition and thus easily understood. Through TM remote sensing image classification examples, compare with the supervision and unsupervised classification in ERDAS and the cluster analysis method based on geometric probability in two-dimensional square, which is proposed in literature 2. Results show that the proposed method can significantly improve the classification accuracy.

Keywords: Cluster validity, Geometric probability, optimum cluster number

1 Introduction

Clustering is an important method of multivariate statistical analysis; it is also an irreplaceable analysis tool in data mining and is widely applied in various research projects [1]. Complete cluster analysis includes: (1) clustering trend analysis; (2) clustering structure extraction; (3) clustering results evaluation. The task of clustering trend analysis is to determine whether a given data set has clustering structure. Clustering structure extraction is a narrow cluster analysis, is the process of applying some algorithm to get the clustering results. Clustering results evaluation is to determine the rationality of clustering results by some standards, whether the clustering results (including levels of classes, the number and the boundaries of classes on each level) are consistent with the clustering structure in the data set. It is also known as cluster validity. Cluster validity function is the basic method to measure the rationality of clustering results [2].

Many clustering structure extraction algorithms have been implemented [3-5], the adapting capabilities of same algorithm for different data sets and different algorithms for same data sets are different, in theory how to evaluate the adapting capabilities of a clustering algorithm and in application how to choose a applicative clustering algorithm for a particular data set is a problem must be solved. For this reason cluster validity is highly valued by scholars, much research focus on it. Currently many literatures preassign a range of optimum cluster number, then use exhaustive method to select the optimum cluster number. It is only fit for the data set that has less data. The method is lack of feasibility in the spatial cluster, which involves huge numbers of the objects (such as remote sensing images' automatic classification). This article proposes and implements a cluster validity function for huge data volume theoretically based on geometric probability [1].

2 Clustering validity classification based on geometric probability

2.1 ANALYTIC EXPRESSION OF THE FUNCTION

Assuming the minimum value and the maximum value of the first term's attribute is x_{min} and x_{max}, respectively; the minimum value and the maximum value of the second term's attribute is y_{min} and y_{max}, respectively, and let:

$$a = Min[(x_{max} - x_{min}), (y_{max} - y_{min})], \quad (1)$$

$$b = Max[(x_{max} - x_{min}), (y_{max} - y_{min})]. \quad (2)$$

Then S is mapped to a discrete point set (still named S) in a rectangular region of a*b. Based on the structural information indicator function and the structural information extraction algorithm about clustering point mode in square region, which is mentioned in reference [13], we design a function to measure the structural information of the point set's geometric distribution by geometric probability.

Do paired connection of n points in this discrete point set S and get a set of line segments: $L(s_i; r_i; \theta_i; i = 1,2,3,...)$, where s_i, r_i, θ_i denote midpoint coordinate, length and angle between vertical axis of the i^{th} line segment in L , respectively, and

* ***Corresponding author*** e-mail: lwticq@163.com

$$0 < r_i \le \sqrt{a^2+b^2}, 0 \le \theta_i \le \pi, i = 1,2,3,\ldots . \tag{3}$$

For L, construct H Function:

$$H(\theta,\Delta\theta,r,\Delta r) = \frac{N}{T} \times \frac{1}{P(\theta,\Delta\theta,r,\Delta r)} . \tag{4}$$

Here, T denotes the total number of line segments in L, and

$$T = C_n^2 = n(n-1)/2 . \tag{5}$$

N denotes the number of the line segments in L, which satisfies:

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2}\right] \bigcap \left[r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right]. \tag{6}$$

Function $P(\theta,\Delta\theta,r,\Delta r)$ means the probability of the line segments satisfy

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2}\right] \bigcap \left[r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right] \tag{7}$$

in random in rectangle region whose lengths of sides are $a, b(a \le b)$.

Function $H(\theta,\Delta\theta,r,\Delta r)$ means the ratio of the frequency of the number of the line segments in Set L is in the interval

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2}\right] \bigcap \left[r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right] \tag{8}$$

and the probability of line segment is in the same interval in completely random.

So we expect

$$E\left[H(\theta,\Delta\theta,r,\Delta r)\right] = 1. \tag{9}$$

If $r = \frac{\sqrt{a^2+b^2}}{2}, \Delta r = \sqrt{a^2+b^2}$, then

$$\begin{aligned} &P(\theta,\Delta\theta,r,\Delta r) \\ &= P(\theta,\Delta\theta,\frac{\sqrt{a^2+b^2}}{2},\sqrt{a^2+b^2}) = U(\theta,\Delta\theta) . \end{aligned} \tag{10}$$

So

$$\begin{aligned} &\int_0^a \int_0^{\pi} A(r,\theta)drd\theta + \int_a^b \int_{\arccos\frac{a}{r}}^{\pi-\arccos\frac{a}{r}} A(r,\theta)drd\theta + \int_b^{\sqrt{a^2+b^2}} \int_{\arccos\frac{a}{r}}^{\arcsin\frac{b}{r}} A(r,\theta)drd\theta \\ &= \frac{1}{3}a^3 + \frac{1}{3}b^3 + ab^2 \ln\frac{\sqrt{a^2+b^2}+a}{b} + a^2 b \ln\frac{\sqrt{a^2+b^2}+b}{a} - \frac{1}{3}(a^2+b^2)\sqrt{a^2+b^2} \end{aligned} , \tag{14}$$

$$\begin{aligned} &\int_{r_1}^{r_2} \int_{\theta_1}^{\theta_2} (b - r\sin\theta)(a - r\cos\theta)drd\theta \\ &= ab(\theta_2-\theta_1)(r_2-r_1) + \frac{1}{6}(r_2^3 - r_1^3)(\sin^2\theta_2 - \sin^2\theta_1) - \frac{1}{2}(r_2^2 - r_1^2)(b\sin\theta_2 - b\sin\theta_1 - a\cos\theta_2 + a\cos\theta_1) . \\ &= M(a,b,r_1,r_2,\theta_1,\theta_2) \end{aligned} \tag{15}$$

$$H(\theta,\Delta\theta,r,\Delta r) = \frac{N}{T} \times \frac{1}{U(\theta,\Delta\theta)} = H(\theta,\Delta\theta) \tag{11}$$

and we expect

$$E\left[H(\theta,\Delta\theta)\right] = 1 . \tag{12}$$

Because the structural information of points set S is stored in Set L, while the points in S gather to multi clusters, the line segments in L must be clustering. It means in completely random the value in some interval is greater than the value. The Figure of $H(\theta,\Delta\theta)$ function will get some peak value in the corresponding interval, the difference between the peak value and 1 indicates the level of clustering, and the number of the peak value indicates the number of clustering directions. A clear corresponding relation exists between the geometric distribution of the points set and the feature of the data set, also exists between Figure of $H(\theta,\Delta\theta)$ function and the structure feature of the point set, so this function is a cluster validity function, which can judge the cluster structure and the validity of clustering results in the data set. The difficulty in process is to derive the analytical expression of $P(\theta,\Delta\theta,r,\Delta r)$ function.

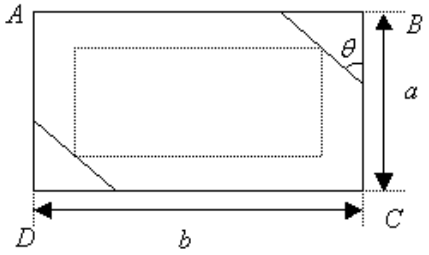


FIGURE 1 The measurement of line segments in the specific intervals and within a rectangle

As shown in Fig.1, if the side lengths of Rectangle $ABCD$ are $a, b(a \le b)$, $P(\theta,\Delta\theta,r,\Delta r)$ indicates the probability of the line segments got from a given rectangle region are in the interval:

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2}\right] \bigcap \left[r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right], \tag{13}$$

We measure the total of the line segment by measuring the area of the region where the points of line segments are included in. For a line segment whose length is $r(0<r\le\sqrt{a^2+b^2})$ and angle between Side CB in positive direction is $\theta(0\le\theta\le\pi)$, the probability of the line segment is in the Rectangle $ABCD$ is proportional to the area of region where the points of line segments are in(the rectangle with dashed lines in Fig. 4):

$$A(r,\theta)=(a-r|\cos\theta|)(b-r|\sin\theta|). \tag{16}$$

Let

$$r_1=r-\frac{\Delta r}{2},r_2=r+\frac{\Delta r}{2};\theta_1=\theta-\frac{\Delta\theta}{2},\theta_2=\theta+\frac{\Delta\theta}{2}, \tag{17}$$

Then

$$P(\theta,\Delta\theta,r,\Delta r)=P(\frac{\theta_1+\theta_2}{2},\theta_2-\theta_1,\frac{r_1+r_2}{2},r_2-r_1)=P, \tag{18}$$

$$P=\frac{\int_{r_1}^{r_2}\int_{\theta_{\min}}^{\theta_{\max}}A(r,\theta)drd\theta}{\int_0^a\int_0^{\pi}A(r,\theta)drd\theta+\int_a^b\int_{\arccos\frac{a}{r}}^{\pi-\arccos\frac{a}{r}}A(r,\theta)drd\theta+2\int_b^{\sqrt{a^2+b^2}}\int_{\arccos\frac{a}{r}}^{\arcsin\frac{a}{r}}A(r,\theta)drd\theta}. \tag{19}$$

Here

$$\theta_{\min}=f(r_1,r_2)\ge\theta_1,\theta_{\max}=g(r_1,r_2)\le\theta_2. \tag{20}$$

2.2 NUMERATOR VALUE OF THE FUNCTION

If r_1,r_2 are in the interval $[0,a]$, $f(r_1,r_2)=\theta_1,g(r_1,r_2)=\theta_2$,

$$\begin{aligned}&\int_{r_1}^{r_2}\int_{\theta_{\min}}^{\theta_{\max}}A(r,\theta)drd\theta=\int_{r_1}^{r_2}\int_{\theta_1}^{\theta_2}A(r,\theta)drd\theta\\&=\int_{r_1}^{r_2}\int_{\theta_1}^{\theta_2}(a-r|\cos\theta|)(b-r|\sin\theta|)drd\theta\end{aligned}. \tag{21}$$

Note

TABLE 1 The value of numerator in P Function

Angle	the numerator value
$0<\theta_1<\theta_2\le\frac{\pi}{2}$	$M(a,b,r_1,r_2,\theta_1,\theta_2)$.
$0<\theta_1<\frac{\pi}{2}<\theta_2\le\pi$	$M(a,b,r_1,r_2,\theta_1,\frac{\pi}{2})+M(a,b,r_1,r_2,\pi-\theta_2,\frac{\pi}{2})$.
$\frac{\pi}{2}\le\theta_1<\theta_2\le\pi$	$M(a,b,r_1,r_2,\pi-\theta_2,\pi-\theta_1)$

(1) $0\le r_1<r_2\le a$

1. $$\begin{aligned}&\int_{l_1}^{l_2}\int_{\arccos\frac{a}{r}}^{\arcsin\frac{b}{r}}A(r,\theta)drd\theta\\&=\int_{l_1}^{l_2}\int_{\arccos\frac{a}{r}}^{\arcsin\frac{b}{r}}(a-r\cos\theta)(b-r\sin\theta)drd\theta\end{aligned}, \tag{22}$$

$$\begin{aligned}&=abl_2(\arcsin\frac{b}{l_2}-\arccos\frac{a}{l_2})-abl_1(\arcsin\frac{b}{l_1}-\arccos\frac{a}{l_1})+\frac{1}{2}ab^2\ln\frac{l_2+\sqrt{l_2^2-b^2}}{l_1+\sqrt{l_1^2-b^2}}+\frac{1}{2}a^2b\ln\frac{l_2+\sqrt{l_2^2-a^2}}{l_1+\sqrt{l_1^2-a^2}}+\frac{1}{2}b(l_2\sqrt{l_2^2-a^2}-l_1\sqrt{l_1^2-a^2})\\&-\frac{1}{2}(a^2+b^2)(l_2-l_1)-\frac{1}{6}(l_2^3-l_1^3)+\frac{1}{2}a(l_2\sqrt{l_2^2-b^2}-l_1\sqrt{l_1^2-b^2})\\&=L(a,b,l_1,l_2)\end{aligned}, \tag{23}$$

$$\int_{l_1}^{l_2}\int_{\arccos\frac{a}{r}}^{\theta_2}A(r,\theta)drd\theta=\int_{l_1}^{l_2}dr\int_{\arccos\frac{a}{r}}^{\theta_2}d\left[ab\theta-r(b\sin\theta-a\cos\theta)+\frac{1}{2}r^2\sin^2\theta\right]$$

2. $$\begin{aligned}&=(ab\theta_2-\frac{1}{2}a^2)(l_2-l_1)-\frac{1}{2}(b\sin\theta_2-a\cos\theta_2)(l_2^2-l_1^2)+\frac{1}{2}a^2b\ln\frac{l_2+\sqrt{l_2^2-a^2}}{l_1+\sqrt{l_1^2-a^2}}+\frac{1}{2}b(l_2\sqrt{l_2^2-a^2}-l_1\sqrt{l_1^2-a^2}),\\&-\frac{1}{6}\cos^2\theta_2(l_2^3-l_1^3)-ab(l_2\arccos\frac{a}{l_2}-l_1\arccos\frac{a}{l_1})\\&=C(a,b,l_1,l_2,\theta_2)\end{aligned} \tag{24}$$

$$\begin{aligned}&\int_{l_1}^{l_2}\int_{\theta_1}^{\arcsin\frac{b}{r}}A(r,\theta)drd\theta\\&=\int_{l_1}^{l_2}dr\int_{\theta_1}^{\arcsin\frac{b}{r}}d\left[ab\theta-r(b\sin\theta-a\cos\theta)+\frac{1}{2}r^2\sin^2\theta\right]\end{aligned}$$

3. $$\begin{aligned}&=ab(l_2\arcsin\frac{b}{l_2}-l_1\arcsin\frac{b}{l_1})+\frac{1}{2}ab^2\ln\frac{l_2+\sqrt{l_2^2-b^2}}{l_1+\sqrt{l_1^2-b^2}}+\frac{1}{2}a(l_2\sqrt{l_2^2-b^2}-l_1\sqrt{l_1^2-b^2}).\\&-(ab\theta_1+\frac{1}{2}b^2)(l_2-l_1)-\frac{1}{6}\sin^2\theta_1(l_2^3-l_1^3)-\frac{1}{2}(a\cos\theta_1-b\sin\theta_1)(l_2^2-l_1^2)\\&=S(a,b,l_1,l_2,\theta_1)\end{aligned} \tag{25}$$

(2) $a\le r_1<r_2\le\sqrt{a^2+b^2}$

Firstly calculate three integral expression, let $0<a\le l_1<l_2\le\sqrt{a^2+b^2},0\le\theta_1<\frac{\pi}{2},0<\theta_2\le\frac{\pi}{2}$, then

When r is in the interval $\left[a,\sqrt{a^2+b^2}\right]$, $[f(r_1,r_2),g(r_1,r_2)]$ equals to $[\theta_1,\theta_2]\cap\left(\arccos\frac{a}{r},\pi-\arccos\frac{a}{r}\right)$. (26)

Now, calculate the following integral expressions: $\int_a^l\int_{\theta_1}^{\theta_2}A(r,\theta)drd\theta$. So $R(l,\theta_1,\theta_2)=\int_a^l\int_{\theta_1}^{\theta_2}A(r,\theta)drd\theta$.

TABLE 2 The value of $R(l,\theta_1,\theta_2)$

Angle	Value of l_1, l_2	Range of l	Value of $R(l,\theta_1,\theta_2)$
$0<\theta_1<\theta_2<\beta$	$l_1=a/\cos\theta_1$ $l_2=a/\cos\theta_2$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
		$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l,\theta_2)$
		$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l_2,\theta_2)$
$0<\theta_1<\alpha<\beta\le\theta_2\le\frac{\pi}{2}$	$l_1=a/\cos\theta_1$ $l_2=b/\sin\theta_2$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
		$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l_2,\theta_2)$
		$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l_2,\theta_2)+L(a,b,l_2,l)$
$\alpha\le\theta_1\le\beta$ $<\theta_2\le\frac{\pi}{2}$	$a\sin\theta_2-b\cos\theta_1>0$ $l_1=b/\sin\theta_2$ $l_2=a/\cos\theta_1$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
		$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+S(a,b,l_1,l,\theta_1)$
		$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+S(a,b,l_1,l_2,\theta_1)+L(a,b,l_2,l)$
	$a\sin\theta_2=b\cos\theta_1$ $l_1=a/\cos\theta_1$ $l_2=b/\sin\theta_2$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
		$l\ge l_1=l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+L(a,b,l_2,l)$
	$a\sin\theta_2-b\cos\theta_1<0$ $l_1=a/\cos\theta_1$ $l_2=b/\sin\theta_2$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
		$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l,\theta_2)$
		$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l_2,\theta_2)+L(a,b,l_2,l)$
$\beta<\theta_1<\theta_2\le\frac{\pi}{2}$	$l_1=b/\sin\theta_2$ $l_2=b/\sin\theta_1$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
		$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+S(a,b,l_1,l,\theta_1)$
		$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+S(a,b,l_1,l_2,\theta_1)$

Angle	Value of $R(l,\theta_1,\theta_2)$
$0<\theta_1<\frac{\pi}{2}<\theta_2\le\pi$	$R(l,\theta_1,\frac{\pi}{2})+R(l,\pi-\theta_2,\frac{\pi}{2})$
$\frac{\pi}{2}\le\theta_1<\theta_2\le\pi$	$R(l,\pi-\theta_2,\pi-\theta_1)$

The value of $R(l,\theta_1,\theta_2)$ is shown as Table 2, and $\alpha=\arccos\frac{a}{b},\beta=\arccos\frac{a}{\sqrt{a^2+b^2}}$.

When $a\le r_1<r_2\le\sqrt{a^2+b^2}$, the value of numerator is $R(r_2,\theta_1,\theta_2)-R(r_1,\theta_1,\theta_2)$.

3) $0\le r_1\le a\le r_2\le\sqrt{a^2+b^2}$

The numerator value is:

$$\int_{r_1}^{r_2}\int_{\theta_1}^{\theta_2}A(r,\theta)drd\theta=\int_{r_1}^{a}\int_{\theta_1}^{\theta_2}A(r,\theta)drd\theta+\int_{a}^{r_2}\int_{\theta_1}^{\theta_2}A(r,\theta)drd\theta=\int_{r_1}^{a}\int_{\theta_1}^{\theta_2}A(r,\theta)drd\theta+R(r_2,\theta_1,\theta_2). \quad (27)$$

3.2 CLUSTERING RESULTS AND COMPARATIVE ANALYSIS

Figure 2 is a synthetic false colour remote sensing images based on the TM data of Band 5, 4 and 3. After the clustering with our proposed algorithm, there are seven categories, including residential areas, shadows, close planting is, dilute vegetation, water, Cho and dry land (Shown as Figure 3).

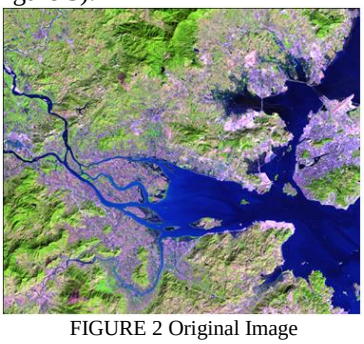


FIGURE 2 Original Image

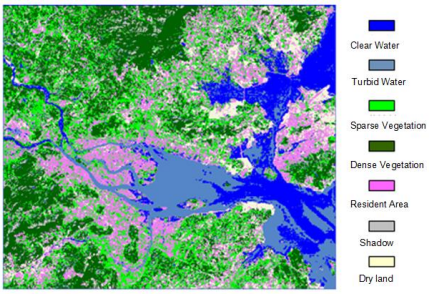


FIGURE 3 Classification Results

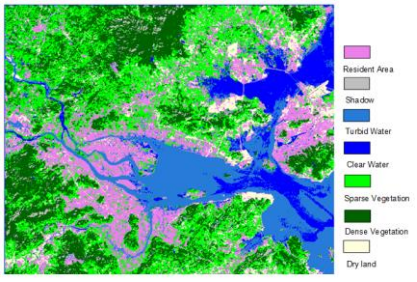


FIGURE 4 clustering results of literature

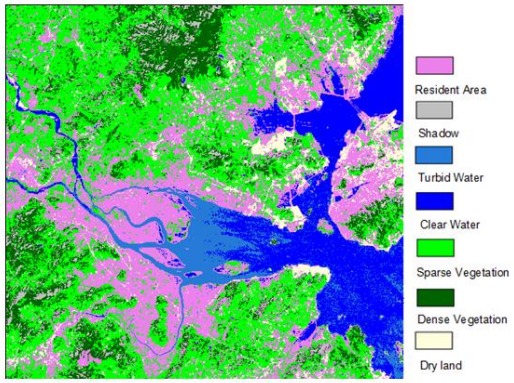


FIGURE 5 results of supervised classification

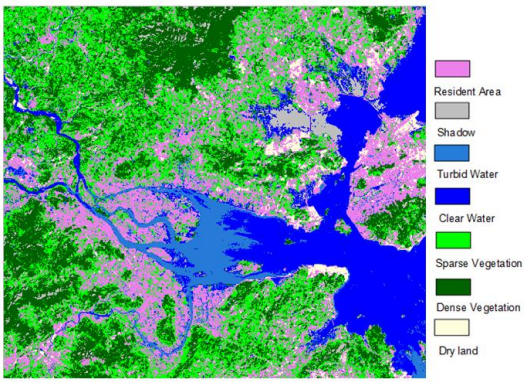


FIGURE 6 Results of Unsupervised Classification

According to the clustering results (Shown as Figure 4) of literature [2] and ERDAS, we conduct supervised classification and unsupervised classification to the same image separately. We pick up 250 pixels in the classification results to accuracy assessment, and compare with the clustering results of literature [2]. Upon examination, specific classification results of four clustering methods are shown as table 1, table 2, table 3 and table 4. The clustering methods include ERDAS supervised classification, ERDAS unsupervised classification, clustering method of literature [2] and clustering method of this paper. Bold figures in the table indicate the number of correct classification pixels for each class. Then calculate classification accuracy and the specific quantitative indicators using the figures in the four tables, results are shown in paper [1].

The paper [1] shows that the classification accuracy of the cluster validity method based on geometric probability is 85.20%, while the classification accuracy of the method, supervised classification and unsupervised classification, which are used in literature [2] are 81.60%, 62.80% and 58.00%, so this cluster validity classification method based on geometry probability is superior to the methods in literature [2].

4 Conclusions

According to the basic idea and clustering steps of geometric probability-based classification method of cluster validity, we select 1498 × 1281 pixels of Xiamen Jiulong River TM images in 2002 Winter, evaluate and compare the clustering results with the supervised classification and unsupervised classification results by ERDAS of the same image. Experimental results show that the classification method of cluster validity based on geometric probability is superior to the literature [2], and it is also superior to the methods of supervised classification and unsupervised classification.

Acknowledgment

This project is sponsored by the National Science Foundation of China (Grant No.31100415) and Education Foundation of Fujian Province (Grant No.Jb12210).

References

[1] Li X W, Mao Z Y, Li J W 2008 A Cluster Validity Function Based on Geometric Probability *Journal of Image and Graphics* **13**(12) 2351-6
[2] Huang L W, Mao Z Y, Li W Z, Wang X Q, Wu Sheng 2007 The Cluster Analysis Approaches Based on Geometric Probability and Its Application in the Classification of Remotely Sensed Images *Journal of Image and Graphics* **4**
[3] Gao X B 2004 *Fuzzy Cluster Analysis and Its Application* Xi'an Electronic Science and Technology University Press
[4] He Q 1998 Research of Fuzzy Clustering Theory and Applications *Fuzzy Systems and Mathematics* **12**(2) 89-94
[5] Gao X B, Xie W X 1999 Development and Application of Fuzzy Clustering Theory *Chinese Science Bulletin* **44**(21)
[6] Jain A K, Flynn P J 1996 *Image segmentation using clustering* In: Ahuja N, Bowyer K, eds. Advances in Image Understanding: A Festchrift for Azriel Rosenfeld. Piscataway: IEEE Press 65 − 83
[7] Cades I, Smyth P, Mannila H 2001 Probabilistic modeling of transactional data with applications to profiling, visualization and prediction, sigmod *In: Proc. of the 7th AC M SIGKDD. San Francisco: ACM Press, 2001. 37− 46* http://www.sigkdd.org/kdd2001/
[8] Jain A K, Murty M N, Flynn P J 1999 Data clustering: A review *ACM Computing Surveys* **31**(3) 264 − 323
[9] Gelbard R, Goldman O, Spiegler I 2007 Investigating diversity of clustering methods: An empirical comparison *Data & Knowledge Engineering* **63**(1) 155−66

Authors

Li Jianwei, born on October 27, 1979, Longyan, China

Current position, grades: lecturer, Master of Geography information system and Faculty of Fuzhou university
University studies: physics and information engineering in Fuzhou University
Scientific interest: Image Processing and Virtual reality
Publications: 20 Papers

Li Xiaowen, born on November 5, 1977, Longyan, China

Current position, grades: lecturer, Master of Geography information system and Faculty of Longyan university
University studies: physics and information engineering in Fuzhou University
Scientific interest: Image Processing and Virtual reality
Publications: 20 Papers

An algorithm for mining frequent itemsets on uncertain dataset

Si Tian[1], Shui Wang[1*], Yang Liu[2], Le Wang[1]

[1]*School of Information Engineering, Ningbo Dahongying University, Ningbo, Zhejiang, China, 315175*

[2]*School of Innovation Experiment, Dalian University of Technology, Liaoning, China, 116024*

Received 1 March 2014, www.cmnt.lv

Abstract

Mining frequent itemsets from uncertain transaction dataset is a research topic in data mining. Some algorithms are based on FP-Growth, but they construct the tree structure in a manner that cannot be as compact as the original FP-Tree, so the tree is easily developed to huge size and this hinders their performance. In this paper, we propose a new tree structure called IT-Tree (Itemset Tail-node Tree) to efficiently maintain probability information of itemsets in tail-nodes; we also propose a corresponding algorithm IT-Mine to mine frequent itemsets from IT-Tree without additional dataset scans. We evaluate our approach on real sparse and dense datasets with different minimum support numbers that can produce non-null frequent k-itemsets ($k \geq 2$); the results show that IT-Mine outperforms other algorithms in terms of execution time, especially for large dataset or small minimum expected support number.

Keywords: frequent itemset, frequent pattern, uncertain transaction dataset, data mining

1 Introduction

Uncertain transaction dataset describes the existential probability of each item in a transactional process. Many applications create uncertain datasets; for example, as indicated by [1,2] and [3], measurement errors of RFID, GPS and other sensors are part of the major sources of uncertain data, because the sensor readings are constantly fluctuating and can hardly be precise, such as the location of an object provided by RFID or GPS. Another kind of uncertainty comes from statistical laws; for example, in medical field, the illness or disease diagnosed for a patient cannot be completely determined by one or more symptoms; in market analysis, customer purchase behaviours, computed from basket data for predicting what a customer will buy in the future, are also statistical probabilities [4, 5]. With the development of applications using uncertain datasets, the issue of data mining over uncertain dataset has become a hot topic in data mining in recent years [4-12].

Table 1 shows an example of an uncertain transaction dataset. Each transaction in Table 1 represents that a customer buy a certain item with a probability. The decimal value associated with an item is called the existential probability of the item. For instance, the first transaction T_1 in Table 1 shows that the customer *A* might purchase products "a", "b", "c" and "d" with 60%, 20%, 30% and 50% chances in the future, respectively. The probability values in Table 1 may be obtained from the analysis of the customers' browsing online-shop history: if customer *A* visited an online-shop ten times in a certain period of time, out of which "a" product was clicked six times, then it might be established that customer *A* has a 60% probability to buy "a" in the future. The probability values may also be obtained from data mining results on the supermarket basket data.

TABLE 1 An example of uncertain transaction dataset

TID	Customer	Transaction itemset
T_1	*A*	(a: 0.6), (b: 0.2), (c: 0.3), (d,0.5)
T_2	*B*	(a: 0.7), (e: 0.25), (d,0.8)
T_3	*C*	(a: 0.3), (c: 0.8) ,(d,0.4)
T_4	*D*	(c: 0.7), (e: 0.2), (d,0.3)
T_5	*E*	(a: 0.5), (b: 0.3), (e: 0.3)

Mining frequent itemsets from uncertain transaction dataset is to discover those itemsets whose sum of probability values or occurring probability exceeds the user specified threshold. Because of its probablistic nature, frequent itemsets mining on uncertain transaction dataset is different from that on precise dataset, which has already been well defined and studied [13, 14], and many algorithms have been proposed, such as Apriori [13], FP-Growth [14], MAFIA [15], COFI [16], Pincer-search [17], CHARM [18], Index-BitTableFI [19] etc. Researchers usually extrapolate the existing algorithms on precise data to get their new algorithms on uncertain data. The papers [5, 7, 10-12] proposed algorithms based on Apriori. These algorithms bear the same bottleneck as Apriori: the generating & processing of the candidate itemsets. And with the increasing of the number of long transactions and the decreasing of the minimum expected support, their performance deteriorates rapidly. The algorithms proposed in papers [6,8,9] are based on FP-Growth; in building their UF-Tree, two items with the same name but different existential probabilities are considered as different nodes; this approach leads to excessive memory requirement to maintain the tree.

[*] ***Corresponding author's*** e-mail: seawan@163.com

Our approach is also inspired by FP-Growth, and tries to amend the defects of the above tree-based algorithms: we propose a more efficient tree structure, named IT-Tree (Itemset Tail-node Tree), to maintain the probability information of transaction itemsets, and give an algorithm, named IT-Mine, to mine frequent itemsets from the IT-Tree. An IT-Tree is created by two scans of the dataset, and IT-Mine mines frequent itemsets from the IT-Tree without additional scan of the dataset, and without generating candidate itemsets.

1.1 CONTRIBUTIONS OF THIS PAPER

1) A tree structure named IT-Tree is proposed for maintaining transaction itemsets of an uncertain dataset.
2) An algorithm named IT-Mine is proposed for discovering frequent itemsets from uncertain dataset based on IT-Tree.
3) Both real and synthetic datasets are used in our experiments to evaluate the performance of the proposed algorithm with the state-of-the-art algorithm MBP.

1.2 THE CONSTRUCTION OF THIS PAPER

The rest of this paper is organized as follows: Section 2 is the description of the problem and definitions; Section 3 is related works; Section 4 is our proposed algorithm IT-Mine; Section 5 is the experimental results; and Section 6 is the conclusion and discussion.

2 Problem definitions

An uncertain transaction dataset $D=\{T_1,T_2,\ldots,T_n\}$ contains n transaction itemsets and m distinct items i.e. $I=\{i_1,i_2,\ldots,i_m\}$, and each transaction itemset T_d ($1\le d\le n$) is represented as $\{i_1:p_1, i_2:p_2,\ldots, i_v:p_v\}$ or $\{\{i_1,i_2,\ldots,i_v\},\{p_1,p_2,\ldots,p_v\}\}$, where $\{i_1,i_2,\ldots,i_v\}$ is a subset of I, and p_u is the probability of the item i_u ($1\le u\le v$) in transaction itemset T_d. An itemset $X=\{i_1,i_2,\ldots,i_k\}$ is called a k-itemset, and k is the length of the itemset X.

We adopt definitions similar to those presented in the previous works [5, 11, 13].

Definition 1: According to the paper [13], the *support number* (*sn*) of an itemset X is the number of transaction itemsets containing X.

Definition 2: The probability of an item i_r in transaction itemset T_d is denoted as $p(i_r,T_d)$, and is defined by

$$p(i_r,T_d)=p_r. \tag{1}$$

For example, in Table 1, $p(\{a\},T_1)=0.6$, $p(\{d\},T_1)=0.5$.

Definition 3: The probability of an itemset X in a transaction itemset T_d is denoted as $p(X,T_d)$, and is defined by

$$p(X,T_d)=\prod_{i_r\in X, X\subset T_d} p(i_r,T_d). \tag{2}$$

For example, in Table 1, $p(\{a,d\},T_1)=0.6\times0.5=0.3$, $p(\{a, d\},T_2)=0.7\times0.8=0.56$, $p(\{a, d\},T_3)=0.3\times0.4=0.12$.

Definition 4: The *expected support number* (*exp*) of an itemset X in an uncertain transaction dataset is denoted as $E(X)$, and is defined by

$$E(X)=\sum_{T_d\in D} P(X,T_d). \tag{3}$$

For example, in Table 1, $E(\{a, d\})=p(\{a, d\},T_1)+p(\{a, d\},T_2)+p(\{a, d\},T_3)=0.3+0.56+0.12=0.98$.

Definition 5: According to the probability theory, the occurring probability of the itemset X occurring in k mutually independent transaction itemsets ($0\le k\le|D|$) is denoted as $P_k(X)$, and is defined by

$$P_k(X)=\sum_{S\subseteq D,|S|=k}\left(\prod_{T_d\in S} P(X,T_d)\cdot\prod_{T_d\in D-S}(1-P(X,T_d))\right). \tag{4}$$

For example, in Table 1,

$$\begin{aligned}P_2(\{a,d\})=&p(\{a,d\},T_1)\times p(\{a,d\},T_2)\times\\&(1-p(\{a,d\},T_3))+p(\{a,d\},T_1)\times p(\{a,d\},T_3)\times\\&(1-p(\{a,d\},T_2))+p(\{a,d\},T_2)\times p(\{a,d\},T_3)\times\\&(1-p(\{a,d\},T_1))=0.21072;\end{aligned}$$

$$P_3(\{a,d\})=p(\{a,d\},T_1)\times p(\{a,d\},T_2)\times p(\{a,d\},T_3)=0.02016.$$

Definition 6: According to the probability theory, the occurring probability of the itemset X occurring in more than k mutually independent transaction itemsets ($0\le k\le|D|$) is denoted as $P_{\ge k}(X)$, and is defined by

$$P_{\ge k}(X)=\sum_{S\subseteq D,|S|\ge k}\left(\prod_{T_d\in S} P(X,T_d)\cdot\prod_{T_d\in D-S}(1-P(X,T_d))\right). \tag{5}$$

For example, in Table 1,

$$P_{\ge2}(\{a,d\})=P_2(\{a,d\})+P_3(\{a,d\})+P_{\ge4}(\{a,d\})=0.21072+0.02016+0=0.23088.$$

Note $P_{\ge4}(\{a,d\})$ is 0 because there is only 3 transaction itemsets containing the itemset $\{a,d\}$.

3 Related work

The approaches of finding frequent itemsets from precise dataset can be classified into two categories: the level-wise approach and the pattern-growth approach. The algorithms Apriori [13] and FP-Growth [14] are representative ones for mining frequent itemsets from precise transaction dataset, and they are representative ones for the level-wise approach and pattern-growth approach, respectively. Apriori is to iteratively generate frequent (k+1)-itemsets using frequent k-itemsets ($k\ge1$):

1) a (k+1)-itemset X is a candidate itemset if its k sub k-itemsets are frequent itemsets;

2) the support number of X is calculated by one scan of dataset if X is a candidate itemset;

3) the itemset X is frequent if its support number is not less than the specified minimum support number.

One advantage is that it has a high time performance when the dataset is sparse, does not contain many long transaction itemsets, and the minimum support number is not small. Its main shortcoming is that it requires multiple scans of dataset, and generating candidate itemsets. Pattern-growth also employs the iteration approach:

1) it finds the set of frequent 1-itemsets (the set is denoted as F) under the condition of a k-itemset X ($k \geq 1$),

2) any itemset $X \cup f$ ($f \in F$) is a frequent (k+1)-itemset. It maintains all transaction itemsets on a tree by one scan of dataset, and generates a conditional sub-tree for each frequent itemset X. Thus, it will find all frequent itemsets under the condition of X by scanning this conditional sub-tree, instead of scanning the whole dataset.

An important difference between precise and uncertain transaction dataset is that each transaction itemset of the former only contains items, and that of the latter contains items and their existential probabilities. Thus, the existing algorithms of mining frequent itemsets from precise dataset cannot be used directly on uncertain transaction dataset. Recently, some algorithms have been proposed for mining frequent itemsets from uncertain transaction dataset.

U-Apriori [7] was proposed in 2007 to find frequent itemsets from uncertain transaction dataset, and it was a level-wise approach. The difference of U-Apriori and Apriori is that U-Apriori calculates the sum of probability of a candidate itemset/item X in all transaction itemsets while Apriori calculates the number of transaction itemsets containing X when they scan a dataset to judge whether the candidate itemset/item X is frequent. U-Apriori and Apriori have the same advantage and disadvantage. The time and memory performance may be worse with the increasing of the number of long transaction itemsets and the decreasing of the minimum expected support number.

In 2007, Leung et al. [8] propose a tree-based algorithm, named UF-Growth, for mining frequent itemsets from uncertain transaction dataset using the same definition of frequent itemsets as in paper [7]. UF-Growth is based on FP-Growth, and is a pattern-growth approach. UF-Growth also constructs a UF-Tree using the given uncertain transaction dataset. But, if two items have the same item name but different existential probabilities, they are considered as different items, and they cannot share the same node when they are added to a tree. For example, two sorted itemsets {a:0.90, b:0.70, c:0.73} and {a:0.95, b:0.85, c:0.70}, they will not share the same node "a" because the probabilities of item "a" in two itemsets are different. After the first UF-Tree is built, UF-Growth retrieves the frequent itemsets from the UF-Tree or sub UF-Trees recursively as the method of FP-Growth. However, the UF-Growth algorithm requires a lot of memory to store tree nodes and a large amount of computational time to process tree nodes.

In 2008, Leung et al. [9] proposed two improvements to boost the time and memory performance of UF-Growth. The first improved algorithm uses the idea of the co-occurrence frequent itemset tree [16] to avoid creating of sub UF-Trees. The second improved algorithm considers that the items, which have the same ***k***-digit value after the decimal point, have the same probability. For example, for two sorted itemsets {a:0.90, b:0.70, c:0.73} and {a:0.95, b:0.85, c:0.70}, both probabilities of item "a" are 0.9 and they will share the node "a". When they are inserted to a UF-Tree if k is set as 1; both probabilities of item "a" are 0.90 and 0.95 respectively and they will not share the node "a" if k is set as 2. This improved algorithm has a better performance than its original algorithm UF-Growth. However, the improved algorithm still does not build a UFP-Tree as compact as the original FP-Tree [14], and it may loss some frequent itemsets.

In 2009, Aggarwal et al. [6] proposed the two algorithms UH-mine and UFP-Growth respectively. These two algorithms employ the pattern-growth approach. Aggarwal also performed a comparison on three frequent itemsets mining algorithms U-Apriori, UH-Mine and UFP-Growth, and concluded that U-Apriori outperforms the other two algorithms.

In the papers [6-9], the proposed algorithms are based on the expected support number of an itemset. An itemset, whose *expected support number* is not less than the *user specified minimum expected support number*, is called as a frequent itemset.

In 2008, Zhang et al. [12] proposed an approximate algorithm for mining frequent itemsets from uncertain transaction dataset based on Definition 6, which defines an itemset to be a frequent itemset if its occurring probability is not less than the *user specified minimum probability*.

In 2009, Bernecker et al. [5] developed a dynamic-programming-based algorithm based on Definition 6, to mine frequent itemsets from uncertain transaction dataset, which employs the level-wise approach. This algorithm inherits the advantage and disadvantage of Apriori.

In 2010, Sun et al. [10] proposed algorithms p-Apriori and TODIS to mine frequent itemsets from uncertain transaction dataset. The algorithm p-Apriori is based on Apriori, and applies a divide-and-conquer approach in calculating the occurring probability of an itemset. The difference between TODIS and p-Apriori is that TODIS works in a top-down manner and p-Apriori works in a bottom-up manner. However calculating the occurring probability of an itemset requires a large amount computation, the time performance of the algorithms proposed in the papers [5, 10, 12] is low.

The problem of getting the occurring probability of an itemset is changed to the problem of getting the expected support number of an itemset in the paper [11], because a Poisson binomial distribution can be well approximated by a Poisson distribution [20] and calculating the occurring probability of an itemset requires a large amount computation. The paper [11] also proposed an algorithm MPB based on U-Apriori. MPB can fast identify frequent and non-frequent itemsets from candidate itemsets without scanning the whole dataset; it can achieve a better performance than U-Apriori in terms of time and space. In

this paper, we propose a mining algorithm based on the expected support number of an itemset.

4 Our proposed method

Our proposed algorithm IT-Mine mainly includes two procedures: firstly, create an IT-Tree or sub IT-Tree; secondly, mine frequent itemsets from the IT-Tree or sub IT-Tree. We descript the structure of IT-Tree in Section 4.1, give an example of the construction of IT-Tree in Section 4.2, and describe the process of mining frequent itemsets from the tree with an example in Section 4.3.

4.1 STRUCTURE OF IT-TREE

Definition 7: Let X be a sorted k-itemset $\{i_1, i_2, i_3, \ldots, i_k\}$, where i_k is the *tail-item*. When the itemset X is inserted into a tree T in this order, the node N on the tree that represents this *tail-item* is defined as a *tail-node* for itemset X. The itemset X is called *tail-node-itemset* for node N.

Definition 8: Let an itemset X containing itemset Y be added to a conditional sub-tree T of the itemset Y. On the tree T, the *base probability* of an itemset X is denoted as $BP(X,Y)$, and is defined by

$$BP(Y, X) = P(Y, X). \quad (6)$$

The structure of IT-tree is illustrated in Figure 1. There are two types of nodes on the IT-tree: one is normal node, as shown in Figure 1a, where N records the item name of each node, *Parent* records the parent node, *Children list* records all children nodes; and the other is tail-node, as shown in Figure 1b, where S records support number, *BP* is *base probability* value of a transaction itemsets, and *ItemsP* is an array which records probability values of all items in corresponding *tail-node-itemset*. We use a "list" to maintain *BP* and *ItemsP values* because there may be many itemsets sharing the same tail-node.

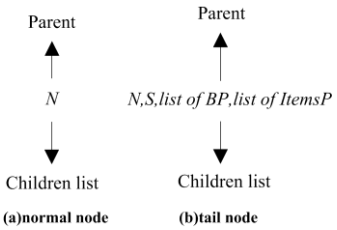


FIGURE 1 Structure of nodes on IT-Tree

4.2 CONSTRUCTION OF AN IT-TREE

The construction of an IT-Tree needs two scans of dataset. In the first scan, create a header table to maintain the support number and expected support of each item; delete those items whose expected support number is less than the user specified *minimum expected support number* from the header table; arrange remaining items of the header table in descending order of support numbers. In the second scan, all transaction itemsets are inserted into an IT-tree. The process is as follows:

1) Delete items that are not in the header table from the transaction itemset;

2) Sort remaining items in the transaction itemset according to the order of the header table;

3) Insert the modified transaction itemset into an IT-tree, and store support number and probability value of each item in each modified transaction itemset to the tail-node of the itemset.

To facilitate tree traversals, *links* in the header table are also maintained (not shown in the Figure 2 for simplicity).

For example, consider the transaction dataset in Table 1. Here the user specified *minimum expected support number* is set to 0.8.

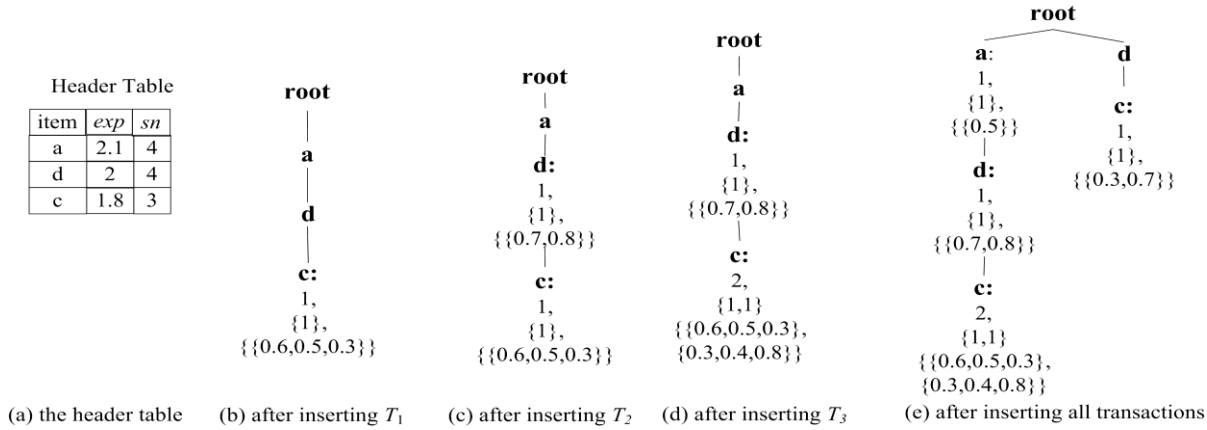


FIGURE 2 Construction of an IT-Tree

Figure 2a is a header table created in the first scan of dataset. Initially, the IT-Tree is created with a root R. When the transaction itemset T_1 is retrieved, the algorithm IT-Mine firstly deletes the item "b"; and then arranges the itemset in the order as the header table in Figure 2a; last, inserts the sorted itemset into the IT-tree. The resulting IT-Tree is shown in Figure 2b. On tail-node "c", "1" represents the support number of this transaction itemset, "{1}" represents the *base probability*, "{0.6,0.5,0.3}"

represents the probability values of items "a", "d" and "c" respectively.

Now insert the transaction itemset T_2 into the IT-Tree. Since the path "root-a-d" can be shared, the algorithm IT-Mine changes the node "d" to a tail-node for the transaction itemsets T_2. The resulting IT-Tree is shown in Figure 2c.

Afterwards, insert the transaction itemsets T_3 into the IT-Tree. Since the path "root-a-d-c" can be shared, only the probability information of this itemsets needs to be stored into tail node "c". The resulting IT-Tree is shown in Figure 2d. On the tail node "c", "{1,1}" represents the base probability values of two transaction itemsets, "{{0.6,0.5,0.3},{0.3,0.4,0.8}}" represents the probability values of items "a", "d" and "c" in two transaction itemsets respectively.

Figure 2e is the result after adding all transaction itemsets in Table 1 are added to the tree; this is the first IT-Tree, and we also call it the *global* IT-Tree. Since the conditional *base-itemset* of the first IT-Tree is null, the base probability of any tail node on the first IT-Tree is set to 1.

4.3 MINING FREQUENT ITEMSETS FROM THE TREE

Start processing from the last item (denoted as *Z*) in the header table (denoted as *HT*):

Step 1. Add item *Z* to the current *base-itemset* (which is initialized as null). Each new *base-itemset* is a frequent itemset.

Step 2. Let *Z.links* contain *k* nodes whose item name is *Z*; we denote these *k* nodes as $N_1,N_2,\ldots,N_k$ respectively; because item *Z* is in the last of the header table, all these *k* nodes are *tail nodes*, i.e., each of these nodes contains *probability information* (*S*, list of *BP*, list of *ItemsP*).

Substep 2.1. Since each one of these *k* nodes contains the probability of each item, create a sub header table through scanning these *k* branches whose paths are from these *k* nodes to the root.

Substep 2.2. Go to Step 3 if the sub header table is null.

Substep 2.3. Create a sub IT-tree for the current *base-itemset* {*Z*} according to the sub header table and these *k* branches. The probability values of items in the current *base-itemset* are stored to the corresponding tail-nodes.

Substep 2.4. Perform a recursive mining process on this sub IT-Tree.

Step 3. Remove the item *Z* from the *base-itemset*.

Step 4. For each of the *k* nodes (which we denote as N_i, $1\le i\le k$), modify its *tail-information*:

Substep 4.1. Delete item *Z*'s probability from list of *BP* of these *k* nodes.

Substep 4.2. Move the *probability information* (*S*, list of *BP*, list of *ItemsP*) on each node to the parent of each node.

Step 5. Process the next item in the header table *HT* using the same method.

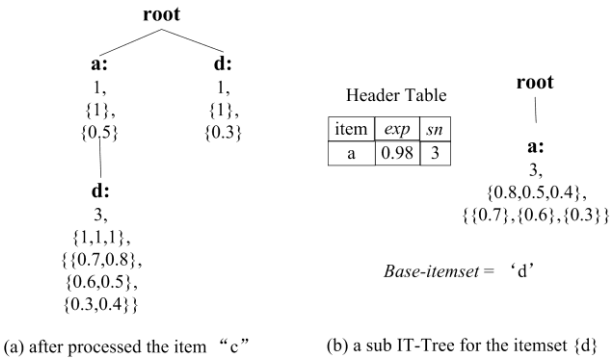


FIGURE 3 Mining the frequent itemsets

The following is the detailed explanation of the mining process illustrated in Figure 2 and Figure 3.

In Figure 2a, item "c" is the last item in the header table. Firstly, add item "c" to the *base-itemset*, and get a frequent itemset {c}; then create a header table for the current *base-itemset* by scanning the branches "root-a-d-c" and "root-d-c". For branch "root-a-d-c", under the condition of the current *base-itemset* {c}, the probability values of items "a" and "d" are calculated as 0.42 (=0.3*0.6+0.8*0.3) and 0.47 (=0.3*0.5+0.8*0.4); for branch "root-d-c", under the condition of the current *base-itemset* {c}, the probability value of item "d" is calculated as 0.21 (=0.7*0.3). The expected support numbers of items "a" and "d", 0.42 and 0.68 (=0.47+0.21), are less than the user specified *minimum expected support number* 0.8. So remove item "c" from the base-itemset, and remove the probability of each node "c" from *ItemsP* of each node, then pass the modified probability information to the parent node of this node. The resulting IT-Tree is shown in Figure 3a.

Afterwards, begin processing the item "d" in the header table in Figure 2a.

Add item "d" to the current *base-itemset*, we get a frequent itemset {d}; then create a header table for the current *base-itemset* {d} by scanning the branch "root-a-d" in Figure 3a; as the result, under the condition of the current *base-itemset* {d}, the expected support number of item "a" is calculated as 0.98 (= 0.5 * 0.6 + 0.4 * 0.3 + 0.8 * 0.7). Since this expected support number is not less than *minimum expected support number* 0.8, create a conditional sub IT-Tree and a sub header table for the current *base-itemset* {d}, as shown in Figure 3b. On tail-node "a" of Figure 3b, "{0.8,0.5,0.4}" represents the base probability values of 3 transaction itemsets containing the itemset {da}, "{{0.7},{0.6},{0.3}}" represents the probability value of the item "a" in 3 transaction itemsets respectively.

Now, the item "a" in header table in Figure 3b is processed.

Add item "a" to the current *base-itemset* {d}, and get a frequent itemset {da}. After the sub IT-tree in Figure 3b is processed, return to process previous tree in Figure 3a.

Go on processing the next item "a" of the header table in Figure 2a.

After finish processing item "a", we get a new frequent itemset {a}.

Lastly, we find 4 frequent itemsets from the dataset in Table 1: {c}, {d}, {da} and {a}.

5 Experimental results

In this section, we compare the performance of the proposed algorithm IT-Mine, the level-wise algorithm MBP and the pattern-growth algorithm UF-Growth using five datasets.

TABLE 2 Dataset characteristics

Dataset	|D|	|I|	ML	DS (%)	Type
retail	88162	16470	10.3	0.06	sparse
T20I6D100K	100000	980	20	2.03	sparse
mushroom	8124	119	23	19.33%	dense
connect	67557	129	43	33.33	dense
T20I6D200K	200000	980	20	2.03	sparse

Table 2 shows the characteristics of 5 datasets used in our experiments. "|D|" represents the number of transactions; "|I|" represents the number of distinct items; "ML" represents the mean length of all transaction itemsets; "DS" represents the degree of sparse or dense. The datasets *retail*, *mushroom* and *connect* are real-world and obtained from FIMI Repository [21]. The datasets *T20I6D100K* and *T20I6D200K* came from the IBM Data Generator [13**]**. Because the original datasets do not provide probability values for each item, to use these dataset as uncertain transaction dataset, we assign a randomly generated existential probability from range (0, 1] to each item of each transaction itemset. The runnable programs and testing datasets used in our experiments can be downloaded from the following address: http://code.google.com/p/it-tree/downloads/list.

The configuration of the testing platform is as follows: Windows 7 operating system, 3G memory, Intel(R) Core(TM) i5-2300 CPU @ 2.80 GHz; Java heap size is 1024M.

The minimum expected support number is set from 130 down to 40 with the decreasing of 10 on *retail*, from 160 down to 70 with the decreasing of 10 on *T20I6D100K*, from 550 down to 100 with the decreasing of 50 on *mushroom*, and is set from 9000 down to 7200 with the decreasing of 200 on *connect*. The more the number of frequent itemsets is, the smaller the minimum expected support number is set. Under the upper bound of the minimum expected number on our experiments, there are at least frequent 2-itemsets on those datasets. As shown in Figures 4-7, they show the distribution of frequent itemsets (FIs) under different minimum expected support number.

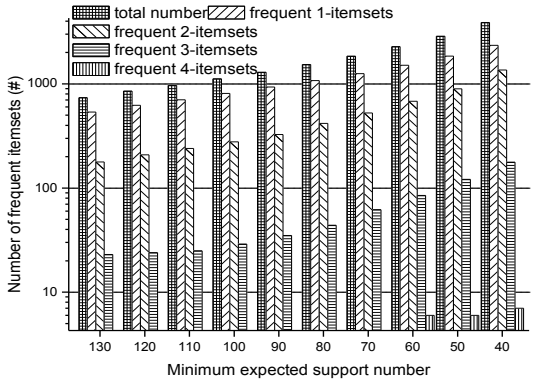


FIGURE 4 Distribution of FIs of retail

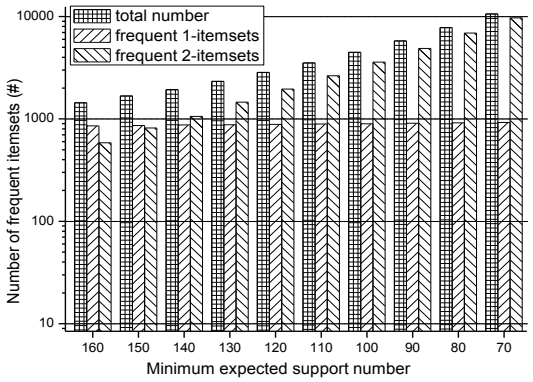


FIGURE 5 Distribution of FIs of *T20I6D100K*

TABLE 3 Number of candidate itemsets under varied minimum expected support number

retail		T20I6D100K		mushroom		connect	
Min_exp	*Candidate itemsets (#)*	*Min_exp*	*Candidate itemsets (#)*	*Min_exp*	*Candidate itemsets (#)*	*Min_exp*	*Candidate itemsets (#)*
130	143524	160	367899	550	2002	9000	7758
120	193308	150	375253	500	2403	8800	7829
110	246261	140	384624	450	2777	8600	8078
100	327889	130	392468	400	3533	8400	8524
90	435082	120	404535	350	4589	8200	9496
80	576580	110	417895	300	5692	8000	10761
70	786105	100	436599	250	7860	7800	11986
60	1142949	90	467162	200	12275	7600	13524
50	1713089	80	516153	150	19595	7400	15263
40	2756787	70	596506	100	40061	7200	17221

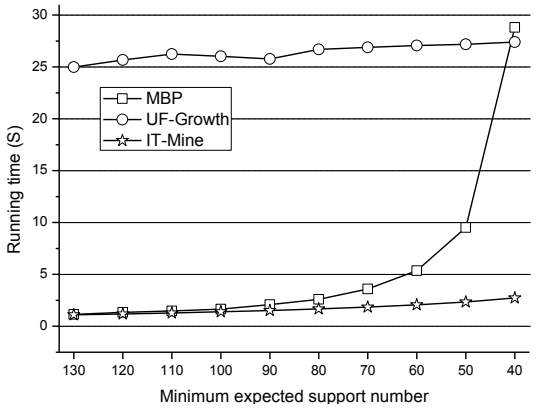


FIGURE 8 Running time on retail

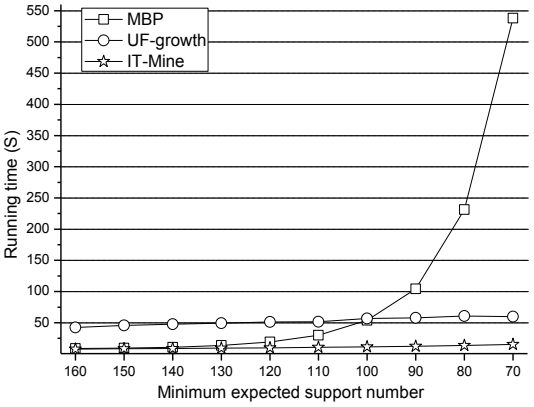


FIGURE 9 Running time on T20I6D100K

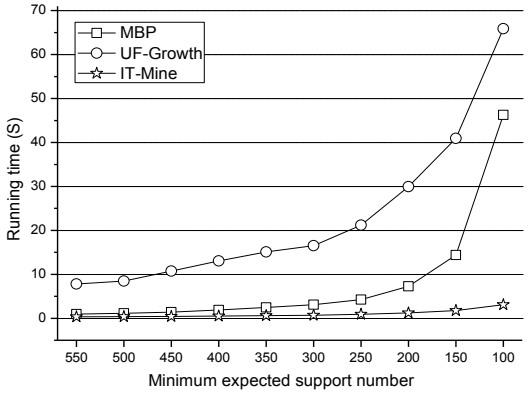


FIGURE 10 Ru3nning time on mushroom

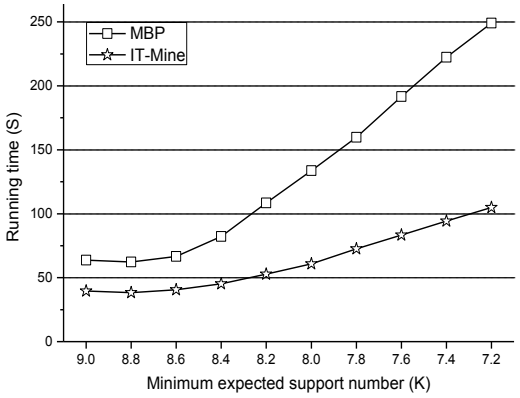


FIGURE 11 Running time on connect

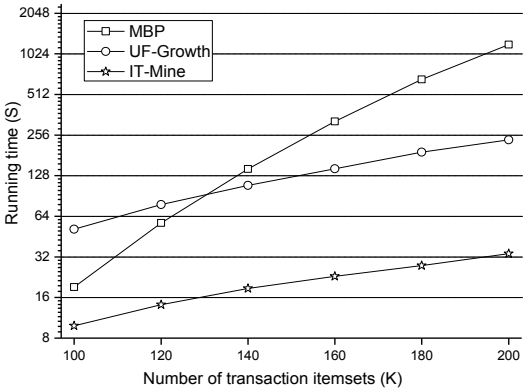


FIGURE 12 Scalability for the algorithms on the sparse dataset

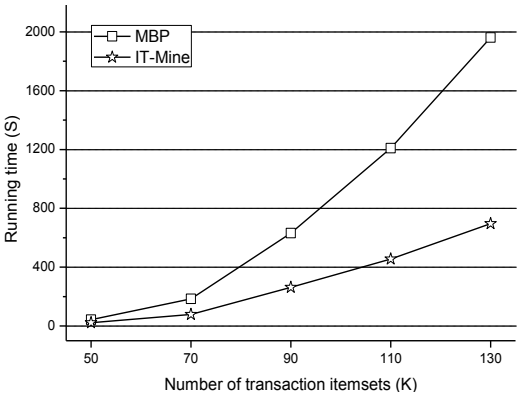


FIGURE 13 Scalability for the algorithms on the dense dataset

Figures 8-11 show the running time of 3 algorithms on four testing datasets under different minimum expected support number. On the dense dataset *connect*, UF-Growth is out of memory although the minimum expected support number is set 9000 because the dataset connect is very dense and the length of each transaction itemset is 43. The time performance of UF-Growth is low because UF-Growth generates too many tree nodes and requires too much time to process these tree nodes. The number of candidate itemsets generated by MBP increases with the decreasing of the minimum expected support number, as shown in Table 3. For example, the number of candidate itemsets is up to 2756787 from 143524 when the minimum expected support number is down to 40 from 130 on *retail*, thus the running time of MBP increases quickly with the decreasing of the minimum expected support number. Figures 8-11 shows that IT-Mine has a high time performance on those four dataset under different minimum expected support number.

In the following experiments, we evaluate the scalability of IT-Mine, UF-Growth and MBP on sparse datasets using the synthetic sparse dataset T20I6D200K; we vary the size of the dense dataset *connect* to evaluate the scalability of IT-Mine, UF-Growth and MBP on dense

datasets; The minimum expected support number is set 120 and 8000 on the sparse dataset and dense dataset respectively.

MBP requires generating candidates and identifying frequent itemsets from the candidates by scanning the dataset, thus its time performance is dependent on the numbers of candidates, the size of dataset and the length of transaction itemsets. As shown in Figure 12 and Figure 13, the time performance of MBP slows down drastically with the increasing of number of transaction itemsets. Both IT-Mine and UF-Growth have a stable time performance on the sparse dataset. On the dense dataset, because the length of each transaction itemset is long and the dataset is very dense, UF-Growth is out of memory in our experiment of dense dataset and the time performance of MBP also slows down with the increasing of the size of dataset. Figure 12 and Figure 13 show that the strong scalability of IT-Mine is obvious.

6 Conclusion and discussion

In this paper, we propose a novel tree structure, named IT-Tree, to represent transaction itemsets of an uncertain transaction dataset. An important feature of IT-Tree is that it maintains probability information of each transaction itemset into a tail-node of the transaction itemset. This way the IT-Tree is as compact as the original FP-Tree, while it does not lose information with respect to the distinct probability values for transaction itemsets. We also give an algorithm named IT-Mine to mine frequent itemsets from IT-Tree without additional scans of dataset.

Experiments were performed on both real sparse and dense dataset, and IT-Mine outperforms the algorithm MBP and UF-Growth in terms of running time; the performance of IT-Mine is also quite stable with the changing of the minimum expected support number.

There are two models in mining frequent itemsets from uncertain transaction dataset: one is based on the *expected support number* of an itemset; the other is based on *occurring probability* of an itemset. IT-Mine is developed on *expected support number*, but it can also be applied to the latter model, because after getting probability information from the tree, the data structure itself is irrelevant with the calculation of the *expected support number* or the *occurring probability* of an itemset.

Although IT-Mine has achieved better performance on testing dataset, construction of global IT-Tree and conditional IT-Trees consumes much time. We hope to improve the speed of IT-Tree construction in future work.

Acknowledgements

This work is partially supported by National Natural Science Foundation of P. R. China (Grant No. 61173163, 51105052, 2013A610115), Ningbo Natural Science Foundation (2013A610115, 2014A610073, 2014A610020), and General Scientific Research Fund of Zhejiang Provincial Education Department (Y201432717).

References

[1] Sistla A Prasad 2003 Querying the uncertain position of moving objects *Proceedings of Seminar Temporal Databases: Research and Practice Dagstuhl Germany*
[2] Khoussainova N, Balazinska M, Suciu D 2006 Towards correcting input data errors probabilistically using integrity constraints *MobiDE 2006: 5th ACM International Workshop on Data Engineering for Wireless and Mobile Access Chicago IL United states*
[3] Hung C, Peng W 2010 Model-driven traffic data acquisition in vehicular sensor networks *39th International Conference on Parallel Processing, ICPP San Diego CA United states*
[4] Aggarwal C C, Yu P S 2009 *Knowledge and Data Engineering IEEE Transactions* **21**(5) 609-23
[5] Bernecker T 2009 Probabilistic frequent itemset mining in uncertain databases *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Paris France*
[6] Aggarwal C C 2009 Frequent pattern mining with uncertain data *15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD Paris France*
[7] Chui C, Kao B, Hung E 2007 Mining frequent itemsets from uncertain data *11th Pacific-Asia Conference on Knowledge Discovery and Data Mining PAKDD Nanjing China*
[8] Leung C K, Carmichael C L, Hao B 2007 Efficient mining of frequent patterns from uncertain data *17th IEEE International Conference on Data Mining Workshops, ICDM Workshops Omaha NE United states*
[9] Leung CK, Mateo M A F, Brajczuk D A 2008 A tree-based approach for frequent pattern mining from uncertain data *12th Pacific-Asia Conference on Knowledge Discovery and Data Mining PAKDD Osaka Japan*
[10] Sun L 2010 Mining uncertain data with probabilistic guarantees *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Washington DC United states*
[11] Wang L, Cheung, D W-L, Cheng R, Sau D L, Yang X S 2012 *IEEE Transactions on Knowledge and Data Engineering* **24**(12) 2170-83
[12] Zhang Q, Li F, Yi K 2008 Finding frequent items in probabilistic data *Proceedings of the ACM SIGMOD International Conference on Management of Data Vancouver BC Canada*
[13] Agrawal R, Srikant R 1994 Fast algorithms for mining association rules in large databases *Proceedings of the 20th International Conference on Very Large Data Bases Santiago Chile*
[14] Han J, Pei J, Yin Y 2000 Mining frequent patterns without candidate generation *ACM SIGMOD - International Conference on Management of Data Dallas TX United states*
[15] Burdick D, Calimlim M, Flannick, J, Gehrke J, Yiu T 2005 *IEEE Transactions on Knowledge and Data Engineering* **17**(11) 1490-504
[16] El-hajj M, Zaïane O R 2003 COFI-tree mining: a new approach to pattern growth with reduced candidacy generation *IEEE. International Conference on Frequent Itemset Mining Implementations*
[17] Lin D I, Kedem Z M 1998 Pincer search: A new algorithm for discovering the maximum frequent set 1377 105-105
[18] [18] Zaki MJ, Hsiao C 2002 CHARM: An efficient algorithm for closed itemset mining *ACM SIGMOD - International Conference on Management of Data Dallas TX United states*
[19] Song W, Yang B, Xu Z 2008 Index-BitTableFI: An improved algorithm for mining frequent itemsets *Knowledge-Based Systems* **21**(6) 507-13
[20] Cam L L 1960 An approximation theorem for the Poisson binomial distribution *In Pacific Journal of Mathematics* **4**(10) 1181-97
[21] Goethals B 2011 Frequent itemset mining dataset repository http://fimi.cs.helsinki.fi/data/ Accessed 11

Authors

Si Tian, born on July 20, 1967, Xiangcheng Henan, China

Current position, grades: professor in Department of Scientific Research at Ningbo Dahongying University.
University studies: Ningbo Dahongying University.
Scientific interest: data mining and image processing.
Publications: 2 SCI, 7 EI.

Shui Wang, born on November 17, 1967, Nanyang Henan, China

Current position, grades: professor in the School of Information Engineering, Ningbo Dahongying University, China.
University studies: Lanzhou University.
Scientific interest: data mining and software engineering.
Publications number or main: 9 EI, 2 books.

Yang Liu, born on February 16, 1993, Huaian Jiangsu, China

Current position, grades: master in computer technology and application towards Ningbo Dahongying University, China.
University studies: M.S. degree in Dalian University of Technology.
Scientific interest: data mining
Publications: none

Le Wang, born on August 15, 1978, Nanyang Henan, China

Current position, grades: associate professor in the School of Information Engineering, Ningbo Dahongying University, China.
University studies: PhD degree in Computer Application from Dalian University of Technology, China, in 2013.
Scientific interest: frequent pattern mining, high utility pattern mining, data streams and big data.
Publications:2 SCI, 10 EI.

Simplification of 3D point cloud data based on ray theory

Changsu Liao[1], Xiaojing Niu[1], Meili Wang[1*], Dongjian He[2]

[1]*College of Information Engineering, Northwest A & F University, Yangling, Shaanxi 712100, China*

[2]*Mechanical and Electronic Engineering, Northwest A & F University, Yangling, Shaanxi 712100, China*

Received 1 July 2014, www.cmnt.lv

Abstract

To effectively reduce the amount of 3D point cloud data, whose shape is symmetrical or spherical, this paper proposes an efficient simplification algorithm based on ray theory. Meanwhile, a boundary retention method based on the distribution uniformity of neighbouring data points is used to keep the model complete. Avoiding time-consuming recursion and curvature estimation, the proposed method is much efficient and achieves good simplification results.

Keywords: 3D Point Cloud, Data Simplification, Ray Generation, Boundary Retention

1 Introduction

In recent years, with the decrease in cost and the increased precision of 3D scanners, 3D point clouds have become an important data representation form in graphics, reverse engineering, and industrial fields [1]. However, the huge of original 3D point cloud data causes great difficulties in 3D reconstruction. As a result, denoising and simplification of the 3D point cloud are important pre-processing steps [2]. Algorithms for simplifying 3D point cloud data include simplification based on clustering [3-5] and simplification based on curvature [6-8], amongst others. Although these methods can reduce the size of the point cloud, retain characteristics of the point cloud model, and increase the reconstruction efficiency of the model to some extent, most of them require recursive calculations and curvature estimation. Thus, these algorithms are quite time-consuming, especially for models with enormous point cloud data.

This paper aims at reducing 3D point cloud data whose shape is symmetrical or spherical, such as head, fruit, mechanical devices et al. In this paper, in order to simplify 3D point cloud data, we assume, based on the ray principle, that the centre of a 3D point cloud model generates several rays, and that data locates at a certain distance from the rays will be reduced. Meanwhile, a boundary retention method based on distribution uniformity of neighbouring data points is used to retain the boundaries of the model [9, 10], thereby guaranteeing the completeness of the data after simplification.

To conduct a rapid local search of data points, a kd-tree is constructed to establish the topological relations among the point cloud data. The organized point data structure of the kd-tree in a k-dimensional Euclidean space is a special binary tree. There are many studies on search of k nearest neighbours in a kd-tree, please refer to [11].

2 Foundation of the proposed algorithm

2.1 RAY PRINCIPLE

The ray principle is simple and easy to understand. First, suppose that the central point of the point cloud data generates rays evenly in all directions, as shown in Figure 1(a), with the rays filling the entire 3D space. Regarding the point cloud model (as shown in Figure 1(b, c)) in this space, if the distance from a certain point in the model to the nearest ray is smaller than a given distance, this point will be simplified. The denser the rays are, the greater the distance is, and the more data points in the 3D point cloud model can be reduced. Therefore, different degrees of simplification results can be achieved by controlling the number of rays and other parameters.

a. Ray in the space b. The cat point cloud model c. Schematic diagram of simplification method

FIGURE 1 Generation principle of rays

2.2 RAY GENERATION

Since the ray principle is the foundation of this paper, the simplification result of the point cloud model is directly related to the way that the rays are generated. Basically, two points make up one line, the centre of the 3D point cloud model is set as one point for all rays. Thus, the

[*] ***Corresponding author*** e-mail *meili_w@nwsuaf.edu.cn*

coordinates of the other points must be generated according to a specific rule, to form all rays. For convenience of generating coordinate points, a minimum cube V_m that encircles the point cloud model is first generated. Suppose that the length of this cube is L, and the maximum and minimum values of the scattered point cloud in the directions of the X-, Y- and Z-axes are $X_{\max}$, $Y_{\max}$, $Z_{\max}$, $X_{\min}$, $Y_{\min}$, and $Z_{\min}$, respectively. Then, the length of the cube is calculated as

$$L = Max((X_{\max} - X_{\min}), (Y_{\max} - Y_{\min}), (Z_{\max} - Z_{\min})). \quad (1)$$

A cube that entirely encircles the point cloud model is created by setting L as the length and $(X_{\min}, Y_{\min}, Z_{\min})$ as the vertex. Then, the required points are generated on the six sides of the cube. By taking a plane parallel with the *XOY* plane as an example as shown in Figure 2, suppose that the two sides of this plane parallel with the X- and Y-axes are $L1$ and $L2$, respectively, and the end vertexes of the two sides are $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, P_2 and $P_3(x_3, y_3, z_3)$, respectively. The coordinates of the generated point $P_i(x_i, y_i, z_i)$ are calculated as

$$x_i = (i/n) \times (x_1 - x_2) + x_2 \quad (i = 0,1,2...,n-1,n) \quad (2)$$

$$y_i = (i/n) \times (y_3 - y_2) + y_2 \quad (i = 0,1,2...,n-1,n) \quad (3)$$

$$z_i = z_1 = z_2 = z_3 \quad (i = 0,1,2...,n-1,n) \quad (4)$$

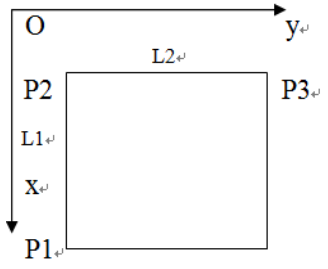


FIGURE 2 Schematic diagram of x-y plane of cube V_m

The number of points generated on each plane is $(n-1)^2$, and $(n+1)^2+4$ rays can be formed. In this way, the number of rays can be controlled by adjusting parameter n.

2.3 DETERMININATION OF THE DATA POINTS FOR SIMPLIFICATION

To increase efficiency, the equation of the line on which the ray is located is not calculated, instead, the distance from the point to the line is calculated using the relation between the vectors. Suppose that the direction vector of line L is $s = (n, m, q)$, M is a point on the line and p is a point beyond the line. Thus, the distance from p to L is $d = \frac{|\overrightarrow{MP} \times s|}{s}$. If d is less than the given distance dis, then p should be reduced; otherwise p will be retained. dis is used as the judgment distance to control the degree of simplification. In this paper, locations far away from the central point, lines will be sparse, and thus the simplification degree will be low. Therefore, dis should vary with a change in the distance from the data point to the central point. The greater the distance, the greater the value of dis should be. Suppose that the central point is Mid and the data point is p_i; dis_i is used as the judgment distance when carrying out the simplification operation for each data point, dis_i is computed by Equation (5), where x is an input, used to control the value of dis_i and the point cloud quantity is Max. Let

$$dis_i = x \times Dis\tan ce(Mid, p_i)^2, \quad (i = 0,1,2...Max), \quad (5)$$

where $Dis\tan ce(Mid, p_i)$ denotes the distance from Mid to p_i. The simplification degree can be controlled by x.

3 Boundary retention

It is important to retain the boundary of the model during the data simplification procedure in that the boundary information is vital to keep the final reconstructed model complete. Thus, keeping the boundary points is also discussed in this paper.

FIGURE 3 Result without boundary retaining treatment

FIGURE 4 Result with boundary retaining treatment

Figure 3 and Figure 4 illustrate the data simplification results with and without applying the boundary retention method. It can be seen that using the boundary retention method, the boundary is complete, thereby improving the effects of data simplification effectively. There are proposed several methods for retaining the boundary points of a model. For example, when determining whether a point is a boundary point [12], several neighbouring points must be searched firstly using the topological relations among the point cloud, and then a least squares plane is constructed in the space based on the neighbouring points. Whether a point is a boundary point is determined by the distribution uniformity of the neighbouring point projection on the least squares plane. Meanwhile, the boundary retention degree can be controlled through a proportionality coefficient e. In this method, when calculating the distribution uniformity of the neighbouring point projection, either the standard deviation of the included angle must be calculated or the coordinate values must be compared. Besides, a boundary point judgment must be made for each data point.

We proposed a method to improve the method for determining boundary points by a way of comparing point cloud coordinate values. In this method, boundary point determination is not carried out for every data point, if a certain point is a non-boundary point, then several neighbouring points of this point are probably not boundary points either. We define that Num denotes the number of point clouds in which neighbouring points are treated as non-boundary points without any judgment, if a certain point is detected as a non-boundary point. This value increases with an increase in the data size of the 3D point cloud model. Through combined control of the proportionality coefficient e and Num, this method has little influence on the boundary retention effect. Figure 5 and Figure 6 show the results of making a boundary judgment, respectively, for each point or only for certain points using the proposed method. The value of the proportionality coefficient e in Figure 5 and Figure 6 is 0.9 and 0.85, respectively, while the threshold value Num is set to 1. It can be seen that there are no obvious differences in the results. However, execution of the former method takes 3.75 s, while the latter only requires 2.70 s. Therefore, the proposed method effectively improves the performance.

FIGURE 5 Result of making a boundary judgment for each point

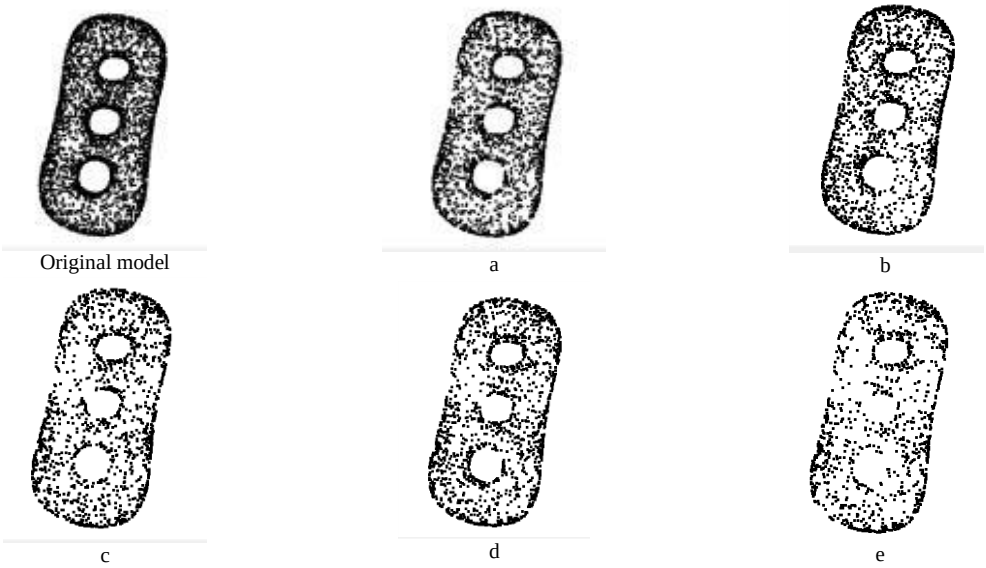
FIGURE 6 Result of making a boundary judgment for some points

4 Simplification results and analysis

The experiments were conducted on Intel CORE i5-3210M, 2.50 Hz processor and 4.00 GB memory. Table 1 gives the test data for simplification under different parameters, while Figure 7 shows the corresponding simplification results, The proportionality coefficient e is only used to control the boundary retention degree of the point cloud model, and therefore its value remains unchanged in general situations, e is set to be 0.85.

TABLE 1 Test results under different parameters

Test number	Model	Input points	d	x	Output points	Execution time (s)
test1	Test model	4000	10	0.11	2592	0.96
test2			15		2036	1.72
test3			20		1514	2.25
test4			10	0.23	1729	1.19
test5				0.35	1125	1.21



Original model a b

c d e

FIGURE 7 Simplification results

According to Table 1 and Figure 7(a, b and c), since the simplification degree is controlled by the number of generated rays, a more number of rays requires a longer execution time. According to Table 1 and Figure 7(a, d and e), the simplification degree is controlled by the value of x, a larger x varies the simplification degree, the execution time has not change much. However, a larger x will lead to holes in the model. Therefore, under the premise that the simplification effect is not affected, threshold x can be increased as much as possible to improve data simplification efficiency.

Table 2 presents the test results for a model using the proposed algorithm and the traditional data simplification algorithm, which involves curvature estimation and recursion. By adjusting the relevant parameters, the number of output points is approximately achieved. Figure 8 shows the simplification effects of Model1 of Table 2, respectively.

TABLE 2 Comparison of the proposed algorithm and traditional curvature estimation algorithm

Model	**Input points**	**Output points**		**Execution time(s)**	
		The traditional algorithm	The proposed algorithm	The traditional algorithm	The proposed algorithm
Model1	4102	1162	1196	2.02	0.94

a. Original model b. Effect of simplification involving curvature estimation and recursion c. Simplification effect using proposed algorithm

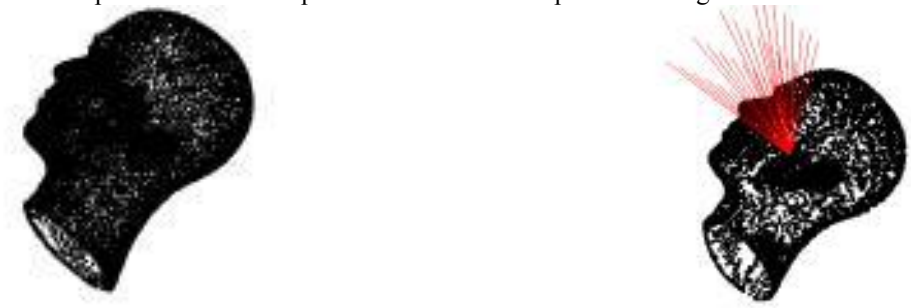
FIGURE 8 Comparison of simplification effects for Model1

It can be seen from Table 2 and Figure 8 that the proposed algorithm is more efficient than the traditional curvature estimation algorithm, while good simplification effects are guaranteed.

The proposed algorithm is based on the ray principle, and thus is more suitable for simplification under a ball model or point cloud model with symmetrical structure. Figure 9 and Figure 10 depict a schematic diagram of the simplification effects as well as some of the rays when applying the proposed algorithm to a ball model and a symmetrical model.

As shown in Figure 9 and Figure 10, according to the definition of the central point in the ray generation method used in this paper, all central points used to generate rays are approximately located in the central position of the entire model when data simplification is carried out for Model 1 of Figure 9 or Model 2 of Figure 10. In this situation, results of data simplification are more even. Regarding approximate ball models, the

distance from the central point to each surface of the model is almost the same, and thus every part of the model can be reduced by a similar degree. Besides, the central point will not be too close to a certain surface of the model when simplification is completed for ball models. According to the ray principle, rays are dense at places near the central point. If the central point is too close to a certain surface of the model, this part can be reduced more easily than other parts when the simplification degree is increased.



a. Original model b. Schematic diagram of simplification effects and some rays

FIGURE 9 Model 1(Ball model)

a. Original model b. Schematic diagram of simplification effects and some rays

FIGURE 10 Model 2(Symmetrical model)

5 Conclusion

In this paper, we proposed a novel and efficient simplification algorithm for 3D point cloud data, whose shape is symmetrical. The experimental results have shown that within a certain degree of simplification, this algorithm achieves good simplification effects and high efficiency. The proposed algorithm is especially suited to ball and symmetrical models. To increase the adaptability to solve the problem whereby an unreasonable central point of the model negatively affects the simplification effect, the central point of the model will be set as a flexible point in the future research. In this way, the position of the central point can be adjusted according to the actual simplification model, thereby improving the simplification results and the efficiency.

References

[1] Wang Y, Li H, Ning X, Shi Z 2011 A new interpolation method in mesh reconstruction from 3D point cloud *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry* 235-42

[2] Nguyen H, Kim J, Lee Y, Ahmed N, Lee S 2013 Accurate and fast extraction of planar surface patches from 3D point cloud *Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication* **84**

[3] Shi B, Liang J, Liu Q 2011 Adaptive simplification of point cloud using k-means clustering *Computer-Aided Design* **43**(11) 910-22

[4] Song H, Feng H 2008 A global clustering approach to point cloud simplification with a specified date reduction ratio *Computer-Aided Design* **40**(3) 281-92

[5] Yu Z, Wong H, Hong P, Ma Q 2010 An adaptive simplification method for 3d point-based models *Computer-Aided Design* **42**(7) 598-612

[6] Miao Y, Pajarola R, Feng J 2009 Curvature-aware adaptive re-sampling for point-sampled geometry *Computer-Aided Design* **41**(6) 395-403

[7] Wang Y, Feng H, Delorme F, Engin S 2013 An adaptive normal estimation method for scanned point clouds with sharp features *Computer-Aided Design* **45**(11) 1333-48

[8] Zhu Y, Kang B, Li H, Shi F 2012 Improved algorithm for point cloud data simplification *Journal of Computer Applications* **32**(2) 521-3

[9] Song H, Feng H 2009 A progressive point cloud simplification algorithm with preserved sharp edge data *The International Journal of Advanced Manufacturing Technology* **45**(5-6) 583-92

[10] Su J, Srivastava A, Huffer F W 2013 Detection, classification and estimation of individual shapes in 2D and 3D point clouds *Computational Statistics and Data Analysis* **58** 227-41

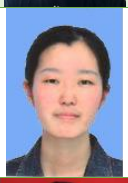
[11] Li H, Zhang X, Jaeger M, Constant T 2010 Segmentation of forest terrain laser scan data *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry* 47-54

[12] Chen Feizhou, Chen Zhiyang, Ding Zhan, et al. 2006 Filling Holes in Point Cloud with Radial Basis Function *Journal of Computer-Aided Design & Computer Graphics* **18**(9) 1414-9

Authors

Changsu Liao

University studies: undergraduate student at College of Information Engineering, Northwest A&F University.
Scientific interests: computer animation, computer graphics and geometric modelling.



Xiaojing Niu

University studies: graduate student at College of Information Engineering, Northwest A&F University.
Scientific interests: image processing, computer graphics and geometric modelling.

Meili Wang

Current position, grades: lecturer at College of Information Engineering, Northwest A&F University.
University studies: Ph.D. degree at the National Centre for Computer Animation, Bournemouth University.
Scientific interests: computer graphics, geometric modelling, image processing, visualization and virtual reality.

Dongjian He

Current position, grades: Professor at Mechanical and Electronic Engineering, Northwest A&F University.
Scientific interests: intelligent detecting and control, image analysis and recognition, agricultural information technology, digital plant and Virtual technology.

Virtual machine resource allocation algorithm in cloud environment

Lei Zheng[1, 2*]

[1]*School of Information Engineering, Shandong Youth University of Political Science, Jinan 250103, Shandong, China*

[2]*Key Laboratory of Information Security and Intelligent Control in Universities of Shandong, Jinan 250103, Shandong, China*

Received 1 July 2014, www.cmnt.lv

Abstract

To resolve the problem that virtual machine deployment reservation scheme waste a lot of resources and single-objective deployment algorithm is not comprehensive, a virtual machine resource allocation algorithm based on virtual machines group multi-objective genetic algorithm is proposed. The algorithm is divided into group coding and resources coding. Resources coding integrated coding according to the history resource need of virtual machines to physical machine and integrate number of physical machine and resource need of physical machine occupied by virtual machine through improved crossover and mutation operations. The experimental results show that the algorithm is effective to reduce the number of physical machine and resource utilization of physical machine, saving energy as much as possible.

Keywords: Cloud computing, Resource allocation, Virtualization, Energy-saving, Genetic algorithms

1 Introduction

Cloud computing is a type of new computing model, providing all kinds of serviced through Internet. Users can gain access to the cloud service anytime, anywhere and on any device. Virtualization technology plays a critical role in management of cloud resource and dynamic configuration, for all kinds of underlying hardware resources can be encapsulated by virtualization technology and provides services to users with virtual machines as the basic resource unit [1]. However, since cloud computing platform includes highly dynamic and heterogeneous resources, virtual machines has to adapt to the dynamic cloud computing environment [2]. The purpose of virtual machines deployment strategy is realizing ideal result by changing layout and placement of overall virtual machines to optimize the objective in meeting constraint condition. The problem of virtual machines placement proved to be *NP* problem. Thus, it is a research hotspot in current cloud computing filed how to conduct virtual machines placement effectively.

Currently, single-objective resource allocation and deployment method are usually adopted in the field of virtualized server technology application. With the goal of maximizing utility, the literature [3] utilizes *NUM* model in computer network. One or more physical machines resources, like bandwidth of network link, distributes one or more job by virtualization technology, reaching higher level of allocation of computing resources of physical machines and optimizing it by algorithm. With the goal of energy-saving, the literature [4] dynamically deploys virtual machines application by energy-aware heuristic algorithm. The literature [5] put forward an improved preferential cooperation descending method to solve the problem of node bin packing. The method merely involves node integration during peak load situation, without consideration of constraint that the goods may be incompatible with the box. The literature [6] suggests an adaptable management frame for virtual machines placement, and studies on the solution of genetic algorithm to the overall placement of virtual machines, effectively reduce the number of physical machines and migration times, but not considering the integration of physical machines resources by virtual machines in the solving process. Nowadays, the optimal method of most virtual machines placement is transforming multi-objective optimization problem to several single-objective optimization problems to be solved in stages. It rarely happens that multi-objective is optimized at the same time. In most time, only partial optimal solution rather than global optimal solution is gained.

For the issue of server integration and resource allocation in cloud computing, a virtual machines resource allocation algorithm based on virtual machines multi-objective genetic algorithm is proposed. With the aim of reduce the number of physical machines and resource allocation, a best solution is searched by genetic algorithm to saving energy as much as possible.

2 Resource allocation algorithm in cloud computing

The definition of bin packing problem [7] is that a set S of M in size and a set P of N in size given, how all

* ***Corresponding author*** e-mail: 1292815809@qq.com

elements of *S* are packed in elements of *P* with least elements of *P* used. *BPP* problem, a difficult *NP* problem, cannot be done by a known optimal algorithm in polynomial time. The problem of virtual machines deployment is actually bin packing problem. In cloud computing, how to reasonably deploy virtual machines to relevant nodes shall be considered, realizing optimal usage of resources while meeting service objectives of different applications. The virtual machines placement may be regarded as vector bin packing problem. The goods being packed are the virtual machine under operation, and the resources of virtual machine are the changeable size of goods. The box is the physical node, and the capacity of the box is the usage threshold of node resource. The number of types of resources is the number of dimensions of vector bin packing problem. Assuming that the number of physical nodes is *M* and the number of virtual machines is *N*, the solution space from the virtual machines to the physical nodes is M^N . It is a *NP* problem similar with bin packing problem that requires an approximate optimal solution.

2.1 DESCRIPTION OF ISSUES IN MULTI-OBJECTIVE VIRTUAL MACHINES DEPLOYMENT ALGORITHM

The resource asked by users to the cloud platform is equal to s virtual machine requiring specific resource, and the applications package of each user operates on their own virtual machines. It is an academic research hotspot how to save energy and utilize cloud computing resource as much as possible to deploy multi-objective virtual machines. Deploying multi-objective virtual machine problem is a multiple combination optimization problem, as well as multi-objective optimization problem. The available resource of each physical machine is multi-dimensional vector, with each dimension as one of all resources of physical machines, and the resource needed by each physical machine is also multi-dimensional vector. The objective is to allocate several virtual machines to several physical machines, maximizing each resource utilization rate of physical machines and minimizing the number of virtual machine immigration. The multi-objective deployment problem is described as follows:

Make N_{PM} as the physical machine set in cloud computing, N_{VM} as virtual machine set in cloud computing, *N* as the total number of virtual machines, N_R as the available allocated resources set in cloud computing and *K* as the total number of available allocated resources.

The objective: $max\sum_{m=1}^{M}\sum_{k=1}^{K}U_{m,k}$ and $min\sum_{m=1}^{M}P_m$.

$\forall n \in N_{VM}, \forall m \in N_{PM}$, among them, $U_{m,k}$ is usage rate of the *K* type of resource by physical machine *m*, P_m is the number of nodes of physical machines.

Constraint:

$$P_m \in \{0,1\} . \quad (1)$$

If $P_m = 1$, it means using new physical machine.

$$U_{m,k} < C_{m,k}, \quad (2)$$

$$\sum_n U_{n,k}^{\ m} < C_{m,k} . \quad (3)$$

Among them, $C_{m,k}$ means the threshold value of the *K* type of resource of *m* physical machine. $U_{n,k}^{\ m}$ is the usage rate of the *K* type of resource by the *n* virtual machine under *m* physical machine. The operation of each type of resource of each physical machine shall be less than the threshold value of each type of resource during the virtual machine deployment process. When there are several virtual machines under deployment in *m* physical machine, the total usage rate of resources of virtual machine under physical machine shall be less than the threshold value of each type of resource.

2.2 DESIGN AND REALIZATION OF MULTI-OBJECTIVE VIRTUAL MACHINES DEPLOYMENT ALGORITHM

For the huge cloud computing centre, combinational explosion may occur in combinatorial optimization. Genetic algorithm is one of the methods for solving combination problem now, since it can concurrently handle with all objectives and avoid priority ordering among objectives. Therefore, genetic algorithm is very suitable for solving multi-objective optimal issues [8]. A virtual machine resource allocation algorithm based on virtual machines multi-objective genetic algorithm is proposed, on the basis of multi-objective virtual machines deployment problem in cloud computing centre.

2.2.1 Coding

In the virtual machines deployment problem, there are three types of genetic coding methods: (1) the representation based on box; (2) the representation based on goods; (3) the representation based on group. Since the objective function of bin packaging problem relies on the goods group, the former two coding methods face single goods, with shortcoming of unclear grouping information. The shortcoming of the third coding method is relying on goods group, neglecting the difference of utilization of physical machine resources by each virtual machine in the crossover and mutation process. In the paper, combining with the coding method based on group

and goods, dynamic allocation genetic algorithm of cloud computing resources is proposed.

The coding based on goods in the paper mainly adopts the coding based on the resource need of virtual machine to the physical machine. The resource need of virtual machine to the physical machine taking CPU, disc and network as example, by *N* number of samplings of *i* virtual machine in a while *T*, calculates the number of operations of CPU, disc and network according to the sampling points. Then the number of operations is coded, and the number of operations of CPU, disc and I/O are showed as formulas (4), (5), and (6). In order to understand the change of demand of virtual machine for resources, the author of the paper adopts the energy efficiency models in the literature [9] for data sampling.

$$L_c^i(T)=\sum_{t=1}^{N}\left[C_f^i(t)\times C_u^i(t)\times C_c^i\times C_{fn}\right]. \quad (4)$$

Among them, $C_f^i(t)$ is the CPU frequency of *i* virtual machine at *t* time; $C_u^i(t)$ is the usage rate of CPU of *i* virtual machine at *t* time; C_c^i is the number of CPU cores of *i* virtual machine; C_{fn} is the number of calculation of floating point in each period.

$$L_d^i(T)=\sum_{t=1}^{N}\left[D_r^i(t)+D_w^i(t)\right]. \quad (5)$$

Among them, $D_r^i(t)$ is the data amount read from disc of *i* virtual machine per second at *t* time; and $D_w^i(t)$ *i*s the data amount written to disc of *i* virtual machine per second at *t* time.

$$L_n^i(T)=\frac{1}{2}\sum_{t=1}^{N}\left[N_r^i(t)+N_w^i(t)\right]. \quad (6)$$

Among them, $N_r^i(t)$ is the data amount received by network card of *i* virtual machine per second at *t* time; and $N_w^i(t)$ is the data amount sent by network card of *i* virtual machine per second at *t* time.

Then calculate the probability of the calculation amount of three types of resources needed by each virtual machine in the calculation amount of physical machine recourses and conduct normalization processing. The formula of probability of the calculation amount is as the following formula (7).

$$P_i=\frac{\mu_i}{\sum_{i=1}^{Z}\mu_i}. \quad (7)$$

Among them, μ_i is the calculation amount of CPU, disc and I/O of *i* virtual machine; and *Z* is the number of virtual machines in physical machines of *i* virtual machine. P_i may be the CPU calculation probability of CPU, disc and I/O of *i* virtual machine.

Then normalization processing is conducted according to the ratio of probability *H* of in the entire set of all types of resources in current physical machine. The value is in the range of [0, 10]. and the probability distribution is shown in figure 1.

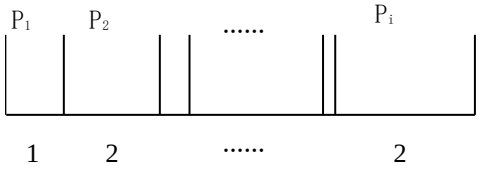


FIGURE 1 Probability distribution graph

Make c_i , m_i , n_i represent the coding of the resource need by CPU, disc and network, and make the proportionality distribution of all types of resources of *i* virtual machine after normalization processing as the resource coding of *i* virtual machine.

Constraint: $c_i\in[0-9], c_i\in N^*$, $m_i\in[0-9], m_i\in N^*$, $n_i\in[0-9], n_i\in N^*$.

Make T_c as the threshold value of CPU resource of physical machine, T_m as the threshold value of disc resource of physical machine and $T_{I/O}$ as threshold value of CPU resource of physical machine. In order to better illustrate the algorithm rose in the paper, the values of T_c, T_m and $T_{I/O}$ in the genetic operation process is 8. The threshold less than 10 is for reserving part of resource space for immigration of virtual machines.

The coding method is shown as figure 2. It mainly takes group as chromosome that has uneven length for the inconsistent number of genes in chromosome. There are three types of deployment of 9 virtual machines: the length of EBA and FCQ are the same but the types of deployment are different; both the length of chromosome of EBA and FCQ as well as the types of deployment are different.

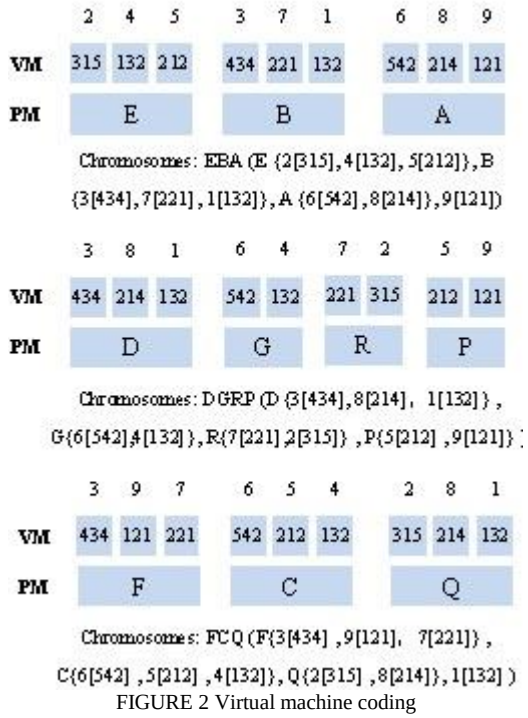


FIGURE 2 Virtual machine coding

2.2.2 Evaluation of fitness

Genetic algorithm evaluates the pros and cons of individuals according to fitness. Fitness function means the corresponding relevance between the whole subjects and their fitness. Evaluation of fitness conducts evaluation of each individual and prepare for the next genetic operation. Since the objective function adopts physical machines as less as possible to place virtual machines as more as possible, the fitness function is shown as the following formula (8), according to the objective function and constraint condition:

$$Fitness_j = \sum_{i=1}^{P_j} \sum_{k=1}^{K} U_{i,k}{}^{j} / P_j . \quad (8)$$

Among them, j means the number of father node, P_j as the total number of physical machines used by father node, and $U_{i,k}$ as the utilization rate of k type of resource of i virtual machine.

2.2.3 Crossover

The main function of the crossover process in genetic algorithm is letting the next generation inherit the excellent genes from the parents and have chance to produce more excellent generations. There are two parts of the crossover process: one is crossover process based on group coding aiming to minimize the number of physical machines, and the other one is crossover process based on resource coding aiming to maximizing the resource of physical machines.

Crossover process based on group coding is shown in figure 3 with steps as follows:

1. Randomly select two father nodes, cross part and cross dot.
2. Insert a selected virtual machine to father node to form new deployed physical machine setoff virtual machines.
3. Delete the repetitive virtual machines in the new physical machine set.

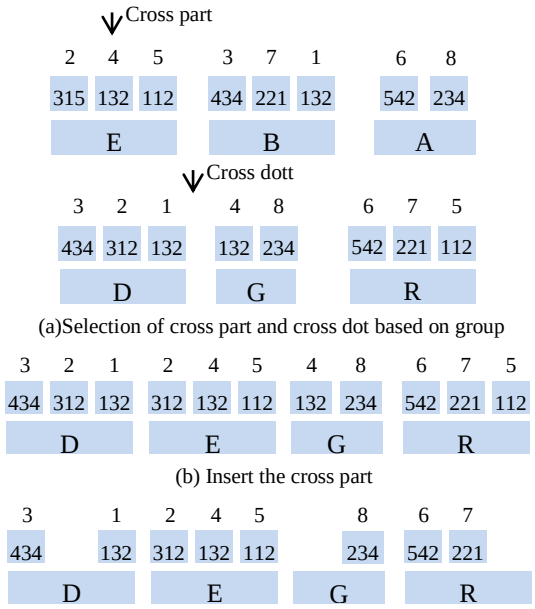


FIGURE 3 Coding cross process based on group

Based on the resource coding crossover process and group coding crossover process, the genetic operation of resource coding can fasten the convergence speed of the group coding genetic process, and integrate the resource of physical machine occupied by the virtual machines. In order to reserve the group crossover result, the resource coding crossover process will reserve the inserted virtual machine group of group coding, not conducting resource coding crossover to it.

Crossover process based on resource coding is shown in figure 4 with steps as follows:

1. Select the remaining virtual machine of the first father node and the second father node (excluding the cross part) of the group coding crossover result as the father nodes. Select the cross part and cross dots. Now the cross part is the virtual machine but not the virtual machine group.
2. Insert the selected virtual machine to virtual machine group.
3. Combine independent virtual machines and delete repetitive physical machine as well as physical machine without any virtual machine.
4. Combine the results of group coding crossover process and resource coding crossover process to get the crossover process result.

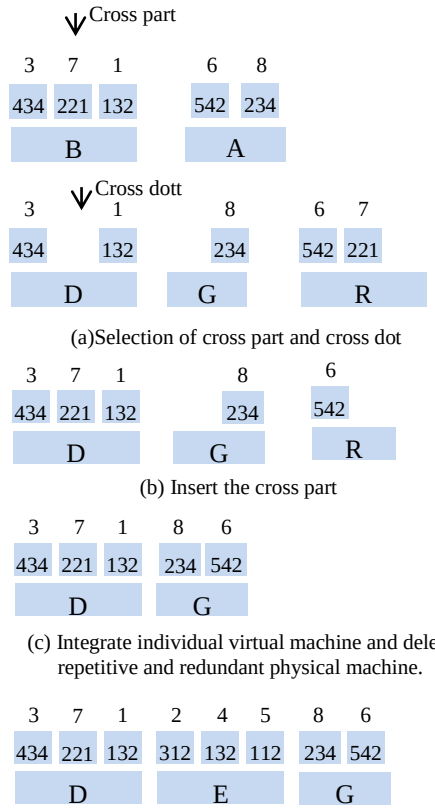


FIGURE 4 Cross process based on resource coding

3 Experimental analysis

In order to verify the proposed algorithm, the author conducts simulation experiment in CloudSim [10]. With the purpose of verifying the effectiveness and deployment scheme, we select the following two classical virtual machine deployment algorithm （Multi-object virtual machines resource allocation Algorithm, MOA） to compare with the multi-objective virtual machine resources distribution algorithm.

Best Fit Algorithm (BFA) means to select the physical machine that meets the resource need of virtual machine with least remaining resource during the virtual machine deployment process, making the physical machine least remaining resource. First Fit Algorithm (FFA) means to search physical machines in order during the virtual machine deployment process, letting virtual machine directly deployed in the physical machine that meets the resource need of virtual machine.

Experiment 1 Calculation of number of physical machines

Deploy 100 virtual machines in 50 physical machines using three types of algorithm independently, with the same nature of physical machines and virtual machine tasks excepting the deployment method and resource threshold value. Among them, the crossover proportionality and mutation proportionality of multi-objective virtual machine deployment algorithm is 0.7 and 0.5 respectively, and the genetic algebra is set as 10. There are load parameter and change of virtual machine resource need during the experiment, taking 10 minute as a time unit to record the change of number of physical machines in 10 time units by three types of algorithm. The experiment result is shown as figure 5:

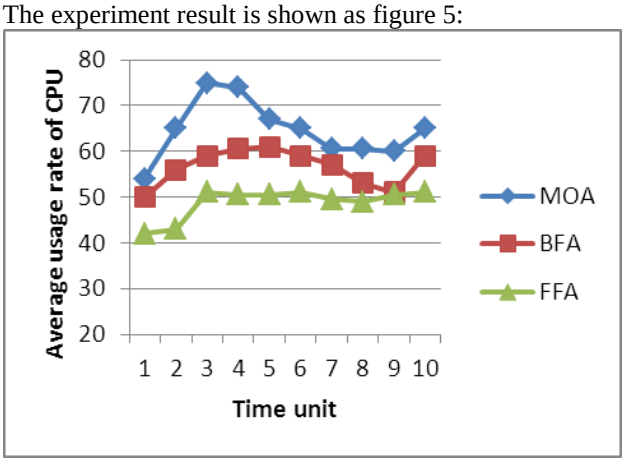


FIGURE 5 Comparisons of number of physical machines

The experiment shows that as time goes on, with the dynamic change of virtual machine's need of resource, the number of physical machines in MOA algorithm is less than that of BFA and FFA. It is because that in a dynamic process, MOA algorithm searches the least number of physical machines in generic operation that meets the constraint condition. It shows that MOA algorithm can effectively reduce the number of physical machines.

Experiment 2 Calculation of resource utilization rate

Calculate the average resource utilization rate of physical machines by three types of algorithm, taking 10 minute as a time unit to record the change of usage rate of CPU and inner storage of physical machines in 10 time units, and calculate the average resource utilization rate. The experiment results of average usage rate of CPU and inner storage by three types of algorithm are shown in figure 6 and 7.

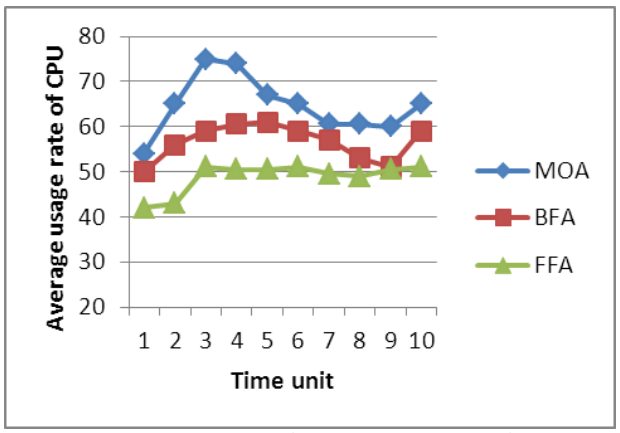


FIGURE 6 Comparisons of the average usage rate of CPU

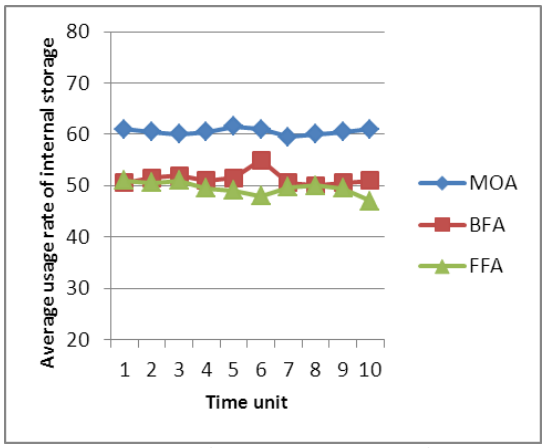


FIGURE 7 Comparisons of the average usage rate of internal storage

It shows from the compared experimental results that the average usage rate of CPU and internal storage of physical machines deployed by MOA algorithm is evidently higher than that of BFA and FFA algorithm. It is because MOA try to improve the resource usage rate as much as possible by using genetic algorithm to adjust virtual machine group during the deployment process, while BFA algorithm try to deploy physical machines as less as possible but not considering improving the resource usage rate, and there are randomness in FFA algorithm that does not consider the resource usage rate. It shows that MOA algorithm can improve the resource usage rate and save energy to some extent.

4 Conclusion

In the paper, after analysing the research situation of virtual machine deployment scheme in cloud computing, the author put forward improved genetic algorithm, conduct group coding and resource need coding of virtual machine, and improve the crossover and mutation operation to resolve the problem of energy waste in cloud computing. Under experimental condition, it shows that the algorithm can not only reduce the number of physical machines but also improve the resource utilization rate. In further research, the issue whether the performances among virtual machines are related will be introduced.

References

[1] Jinhua Hu, Jianhua Gu, Guofei Sun, et al. 2010 A scheduling strategy on load balancing of virtual machine resources in cloud computing environment *Third international symposium on parallel architectures, algorithms and programming* **89**(96) 18-20
[2] Cherkasova L, Gupta D, Vahdat A 2007 When virtual is harder than real: Resource allocation challenges in virtual machine based it environments *Technical Report HPL-2007-25*
[3] Truong H L, Dustdar S 2010 Composable cost estimation and monitoring for computational applications in cloud computing environments *Procedia computer science* **1**(1) 2175-84
[4] Bo Li, Jianxin Li, Jinpeng Huai, et al. 2009 EnaCloud: an energy-saving application live placement approach for cloud computing environments *IEEE international conference on cloud computing* **17**(24) 21-5
[5] Ajiro Y, Tanaka A 2007 Improving packing algorithms for server consolidation *Proceedings of the 33rd International Computer Measurement Group Conference* San Diego 399-406
[6] Qi G, Ji Q, Pan J Z, Du J 2011 Extending description logics with uncertainty reasoning in possibilistic logic *International journal of intelligent systems* **26** 353-81
[7] Aktas H, Cagman N 2007 Soft sets and soft groups *Information sciences* **177** 2726-35
[8] Sun Y L, Perrott R, Harmer T, Cunningham C, Wright P 2010 An SLA focused financial services infrastructure *Proceedings of the 1st International Conference on Cloud Computing Virtualization (CCV 2010)*, Singapore, 2010
[9] Rudolph S 2011 Foundations of description logics In: Polleres, A., D'Amato, C., Arenas, M., Handschuh, S., Kroner,P., Ossowski, S. and PatelSchneider, P.F., Eds. *Reasoning Web. Semantic Technologies for the Web of Data, Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg 76-136
[10] Calheiros R N, Ranjan R, De Rose C A F, et al. 2009 Cloud-Sim: A novel framework for modeling and simulation of cloud computing infrastructures and services *Parkville VIC*: The University of Melbourne Australia, Grid Computing and Distributed Systems Laboratory.

Authors

Lei Zheng, born on August 3, 1980, China

Current position, grades: researcher at Shandong Youth University of Political Science, China.
University studies: master's degree in Computer Software and Theory from Shandong Normal University, China in 2006.
Scientific interests: cloud computing and distributed computing.

Multi circle detection by using evidence accumulation

Hudai Fu*, Jingang Gao

School of Mechatronics Engineering, Changchun Institute of Technology, Changchun 130012, JiLin, China

Received 1 July 2014, www.cmnt.lv

Abstract

The traditional circle detection algorithm has complex computation, large memory space occupation and other deficiencies. It has low detecting efficiency and not suitable for the multi circle detection. So it proposes a multi circle detection method based on global search and evidence accumulation in the paper. The evidence accumulation and the weighted average are combined in the algorithm. The pseudo centre processed during the evidence accumulation are classified and analysed. Three kinds of pseudo centres are eliminated by class. Finally, the other circle parameters are calculated. It is proved that the algorithm has high precision, high efficiency, and low sensitivity to the defect of local information. Also the detecting time will not increase with the number of circles. The multi circle detection effect is obviously superior to the traditional randomized circle detection algorithm.

Keywords: Multi Circle Detection, Global Search, Evidence Accumulation

1 Introduction

Circle is one of the basic geometries in image analysis. The relevant information of the circles in image is required to rapidly detect in many fields, such as quickly detecting the location and size of round holes in the part, detecting round PCB printing plate in industry and so forth. Therefore, the circle detection has special significance in image analysis, pattern recognition and computer vision [1, 2].

The methods such as standard Hough transform (SHT) are commonly used to detect the circles [3]. It needs complex computation and large memory space, on account of using "one to many" mapping. Also it is not good at multi circle detecting. In recent years, the domestic and foreign scholars have done some researches in multi circle detection methods [4]. Xu presented randomized Hough transform (RHT), which took "many to one" mapping [5, 6]. It avoided huge amount of calculation comparing to SHT. But the sampling without targets could cause a large number of invalid samplings and invalid accumulations. So the performance of the algorithm was reduced. In addition, Chen proposed a randomized circle detection algorithm (RCD) not in RHT series [7]. The detecting speed was faster than RHT in the case of below medium noise ratio.

RCD detection algorithm and Hough transform have many problems in practical application. The probability of invalid sampling is great in multi circle detection. And the efficiency of the algorithm is low, because there are many square and square root operations in it. The test results have greater offset, because the threshold parameters have been set are different in both algorithms. The detecting speed will suffer great interference when the circle is incomplete, the image has large disturbances and so on. Therefore, a circle detection method based on global search is proposed. The evidence accumulation and the weighted average are combined in the algorithm. The pseudo centre processed during the evidence accumulation are classified and analysed. Three kinds of pseudo centres are eliminated by class. Finally, the other circle parameters are calculated.

2 Algorithm principle

2.1 GLOBAL SEARCHING CENTRE COORDINATE

Please see Figure 1. Suppose that the image size is M pixel× N pixel. After edge detection, there is

$$I(x,y)=\begin{cases}255, & if\\ 0, & else\end{cases}, \tag{1}$$

where $I(x,y)$ $(0\le x<M, 0\le y<N)$ is the grey value of row x the column y in the image.

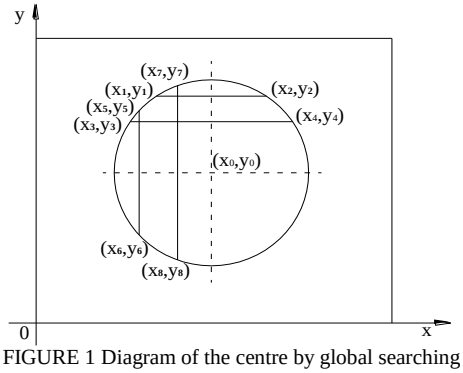


FIGURE 1 Diagram of the centre by global searching

Two arrays a_{2M} and b_{2N} are created. The position of each element in the array corresponds to the pixel coordinate of the centre. The size of the element stored in

* ***Corresponding author*** e-mail: fuhudai@126.com

the array is namely the evidence accumulation value of the corresponding pixel becoming the centre. Searching the coordinates of the centre consists of the following steps.

First, the pixels in the first row are searched. The horizontal ordinate average of each edge pixel and its nearest edge pixel is calculated successively. The value $a[2\times p_x]$ in the array a_{2M} will be added 1 after calculating one average value p_x each time. Then the pixels in each column will be processed as the above steps. The value $b[2\times p_y]$ in the array b_{2N} will be added 1 after calculating one average value p_y each time. See Equation (2).

$$\begin{cases} a[i]=a[i]+1 & if \quad i_1+i_2=i \\ b[j]=b[j]+1 & if \quad j_1+j_2=j \end{cases}. \tag{2}$$

In Equation (2), i_1, i_2 are the horizontal ordinates of the two adjacent pixels; j_1, j_2 are the vertical ordinates of the two adjacent pixels; and i, j is the position of the corresponding array element.

After completing global searching, the threshold T is set. The values in two arrays will be respectively compared with the threshold *T*. The position of the element which is greater than T in the array a_{2M} is twice of the centre abscissa value. And the position of the element which is greater than T in the array b_{2N} is twice of the centre ordinate value.

The evidence accumulation value in the array reflects the probability of each coordinate value taken as the centre coordinate, because of the symmetry of the circle. Since the accumulation value is gotten from calculating the average of the adjacent pixels, it has high degree of distinction, and the other kinds of edge interference can be effectively removed.

2.2 CHARACTERISTIC ANALYSIS OF PSEUDO CENTRE

For the multi circles detection, a variety of pseudo centres will be emerged after completing evidence accumulation. Because the digital image is in single pixel, the evidence accumulation values of the two adjacent coordinates in one circle are both greater than the set threshold, and they are not much difference. So the pseudo centre is produced. Please see Figure 2.

Pseudo centre can be divided into three categories:

1) Type 1 pseudo centre. Because the abscissa and ordinate values obtained from the evidence accumulation have many different combinations, as well as the adjacent coordinate probably can generate pseudo centre, one horizontal coordinate value can correspond multiple possible vertical coordinate values. The resulting multiple pseudo centres such as A C, E, F, G, H are shown in Figure 2.

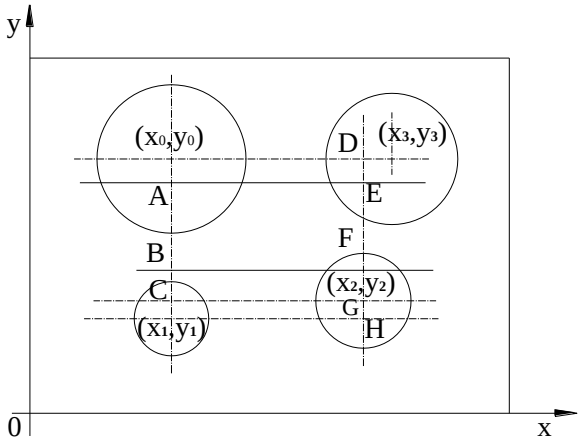


FIGURE 2 Diagram of pseudo centre

2) Type 2 pseudo centre. There will be a vertical coordinate value having larger evidence accumulation value between the two circles, if the centre horizontal coordinate values of the two circles are equal. This point is obtained from the evidence accumulation of the lower edge in the above circle and the upper edge in the below circle. Similarly, there will be a horizontal coordinate value having larger evidence accumulation value between the two circles, if the centre vertical coordinate values of the two circles are equal. Point B is this type of pseudo circle. Please see Figure 2.

3) Type 3 pseudo centre. The circles corresponding to (x_2, y_2) and (x_3, y_3) are shown in Figure 2. If the range of the two circles has larger overlap in the abscissa, the evidence accumulation values of (x_2, y_2) and (x_2, y_3) are both large, when doing evidence accumulation along the y axis for the edge points near x_2 and x_3. x_3 has the same situation. Also, the similar situation will also appear, if the range of the two circles has larger overlap in the ordinate. Point D is this type of pseudo circle. Please see Figure 2.

For type 1 pseudo centre, the vertical coordinate value corresponding to each horizontal coordinate can be found, only to do evidence accumulation for the edge points nearby every possible horizontal coordinate value along the vertical coordinate.

However, type 2 pseudo centre is not in the circle. The edge points nearby every possible horizontal coordinate value along the vertical coordinate are scanned. So the vertical coordinate of the edge point when doing evidence accumulation to the vertical coordinate of type 2 pseudo centre must vary greater, the further it is from the horizontal coordinate value. Type 2 pseudo centre can be eliminated according to this characteristic.

The vertical coordinate of type 3 pseudo centre (such as point D in Figure 2) is scanned. The difference of the edge points is great, when doing evidence accumulation to the vertical coordinate of point D on both sides of x_2. So, type 3 pseudo centre can be eliminated.

2.3 ELIMINATING PSEUDO CENTRE

According to the analysis in Section 2.2, all kinds of pseudo centre will be respectively eliminated.

1) Eliminating the adjacent centre and type 1 pseudo centre

The adjacent coordinate points have been found out by scanning the horizontal and vertical coordinates. The larger point of the evidence accumulation values is selected. If the difference between them is very small, the average of two points is taken as the centre coordinate. Setting thresholds t_1 and t_2, and the pixels of $2t_1+1$ columns from column (x_0-t_1) to column (x_0+t_1) are selected. The pixels per column are processed according to the method in section 2.1.

$$b[j]=b[j]+1, \tag{3}$$

if $j_1+j_2=j$ and $(x_0-t_1)\le i_0\le(x_0+t_1)$, where j_1 and j_2 are the vertical coordinate values of the nearest edge point in column i_0, and $b[j]$ is the evidence accumulation value for the vertical coordinate j. The vertical coordinates whose evidence accumulation value is greater than the threshold t_2 are the probable vertical coordinates $y_0^{'}, y_1^{'}\cdots$ corresponding to x_0. The vertical coordinate values of the edge points among the possible vertical coordinate values are recorded. Six pseudo centres including A, C, E, F, G, H can be eliminated, after x_0, x_1, x_2 have been processed in order.

2) Eliminating type 2 pseudo centre

The edge points carried out the evidence accumulation to the longitudinal coordinate $y_0^{'}$ are gathered in point group $C_{y_0^{'}}$. The maximum longitudinal coordinate point in column (x_0-t_1) or column (x_0+t_1) among the point group is chosen. If the ordinate of the point is larger than that of the maximum ordinate point in column x_0, it is the pseudo centre. The pseudo centre B can be eliminated, after x_0, x_1, x_2 have been processed in order.

3) Eliminating type 3 pseudo centre

In point group $C_{y_0^{'}}$, the longitudinal coordinate value which is greater than $y_0^{'}$ subtracts the one less than $y_0^{'}$ from column (x_0+1) to column (x_0+t_1), and the subtraction result is added up. Accordingly, the longitudinal coordinate value, which is greater than $y_0^{'}$ subtracts the one less than $y_0^{'}$ from column (x_0-t_1) to column (x_0-1), and the subtraction result is added up. Finally, the two calculated sums are subtracted.

$$\delta_0=\sum_{x_0+1}^{x_0+t_1}(y_i-y_{\ j})-\sum_{x_0-t_1}^{x_0-1}(y_k-y_{\ l}) \quad (y_i,y_{\ j},y_k,y_{\ l}\in C_{y_0^{'}}). \tag{4}$$

In equation (4) y_i is the ordinate value greater than $y_0^{'}$ from column (x_0+1) to column (x_0+t_1) in point group $C_{y_0^{'}}$, and y_j is the ordinate value less than $y_0^{'}$. y_k is the ordinate value greater than $y_0^{'}$ from column (x_0-t_1) to column (x_0-1) in point group $C_{y_0^{'}}$, and y_l is the ordinate value less than $y_0^{'}$. All possible ordinate values are calculated in turn by using equation (4). The corresponding differences having been gotten are $\delta_0,\delta_1,\cdots$. The threshold ε is set. The point whose difference δ is greater than the threshold ε will be eliminated. The pseudo centre D can be eliminated, after x_0, x_1, x_2 have been processed respectively.

So far, the pseudo centres corresponding to the abscissa x_0, x_1, and x_2 have been all eliminated. Similarly, the abscissa x_3 and the ordinate values $y_0,y_1\cdots$ can be processed as the above. The centre coordinates corresponding to it can be found. Finally, the evidence accumulation values of the true centre and its 8 adjacent points are taken as the weight. The coordinate values are carried out the weighted average, and the centre coordinates are gotten.

2.4 CALCULATING THE RADIUS

The images often have all kinds of noises in the actual industry detecting. In order to avoid noise interference, cluster analysis and the weighted average are used to minimize the error. The distance L from each edge pixel to the detected centre is calculated. The pixels close to the distance L are gathered as a class. Then the threshold ξ is set, the point group $C_{\max}$ gathered most points in the range of 2ξ is selected. The equation for calculating the weighted average of the radius is as follow. Please see Equation 5.

$$\sum_{L_n\in C_{\max}}\frac{P_n}{P_{sum}}L_n=L, \tag{5}$$

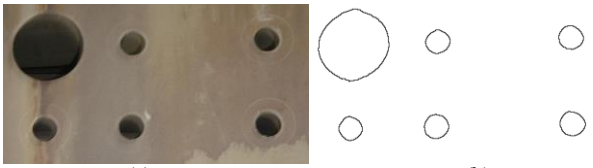
where L_n is the distance between the edge point and the centre in class $C_{\max}$. P_{sum} is the total number of points in class $C_{\max}$. P_n is the number of edge points whose distance is L_n in class $C_{\max}$. The value L calculated by equation (5) is the radius.

The clustering method can give full play to the effect of the evidence accumulation. The result can effectively avoid deviating from the correct value. The weighted average method used to calculate the radius can reduce

the inherent pixel error in digital image, and improve the reliability and accuracy of detection.

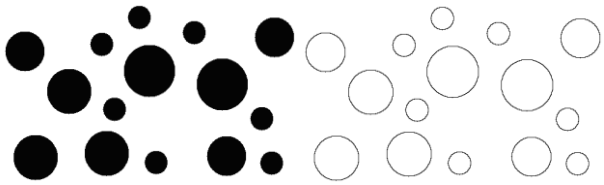
3 Experiment results and analysis

Please see Figure 3 (a) and Figure 4 (a). The sizes of the two experimental images are both 280pixel×166 pixel. First, an array *a*[560] and *b*[332] are created and initialized, and the parameter values $T=4$, $\varepsilon=2$, $\xi=0.5$, $t_1=4$ and $t_2=5$ are set. The proposed algorithm in the paper is used to separately detect the circles in two images. Please see Figure 3 (b) and Figure 4 (b).Then RCD algorithm is used to separately detect the circles in two images. The above two detections are carried out in the same hardware experiment environment.



(a) (b)

FIGURE 3(a) Automobile part image. (b) Proposed algorithm multi circle detecting result



(a) (b)

FIGURE 4(a) Experimental image. (b) Proposed algorithm multi circle detecting result

The comparisons of detected value and real value of Figure 3(a) are shown in Table 1.The error between the test results by the proposed algorithm and the real value is less than one pixel. Its accuracy is obviously superior to the RCD method. Please see Table 2. When the number of the circles increases, the executing time of RCD algorithm obviously increases. However, the executing time of the proposed algorithm is almost unchanged.

TABLE 1 Comparison of detected value and real value of Figure 3(a) (Unit: pixel)

Circle number (centre, radius)	**1**	**2**
Proposed algorithm	(41.8,117.9,33.7)	(38.9, 35.3,11.5)
RCD	(42.7,118.5,32.9)	(40.2, 35.8,12.7)
Real value	(42,118,34)	(39, 35,12)
Circle number (centre, radius)	**3**	**4**
Proposed algorithm	(118.4,37.3,12.1)	(119.7,121.2,11.7)
RCD	(120.3,36.4,10.8)	(120.5,122.6,12.9)
Real value	(119,37,12)	(119,122,12)
Circle number (centre, radius)	**5**	**6**
Proposed algorithm	(243.1,125.5,11.9)	(244.5,40.1,12.1)
RCD	(245.3,124.1,11.7)	(243.7,38.8,11.3)
Real value	(244,125,12)	(244,40,12)

TABLE 2 Comparison of executing time between the proposed algorithm and RCD algorithm (Unit: ms)

Images	Figure 3(a)	Figure 4(a)
Proposed algorithm	78	85
RCD	356	1054

4 Conclusion

The experiment and analysis showed that the proposed algorithm had better performance than the traditional circle detection algorithm. First, it avoids the shortcoming of largely invalid random sampling probability in RHT and RCD algorithms, because the proposed algorithm applied the global search to the image. Secondly, it used the simple averaging operation instead of the square and square root operations in the traditional algorithm, and increased the efficiency of the algorithm. Again, it adopted the evidence accumulation combining with the weighted average to calculate the parameters of the circle, and improved the stability and accuracy of the results. It is proved, that the proposed algorithm had fast detection speed, high precision, and great practical value after detecting multiple images.

References

[1] Blanco M, Penedo M G, Barreira N, et al. 2006 Localization and extraction of the optic disc using the fuzzy circular Hough transform *Artificial Intelligence and Soft Computing* 4029 712-21

[2] D'orazio T, Guaragnellab C, Leo M, et al. 2004 A new algorithm for ball recognition using circle Hough transform and neural classifier *Pattern Recognition* **37**(3) 393-408

[3] Huda Loannou D, Laine W, et al. 1999 Circle recognition through a 2D Hough transform and radius histogram *Image and Vision Computing* **17**(1) 15-26

[4] Yu Qiao, Ong S H 2004 Connectivity-based multiple circle fitting *Pattern Recognition* **37**(4) 755-65

[5] Xu L, Oja E 1993 Randomized Hough transform (RHT): basic mechanisms, algorithms, and computational complexities *CVGIP: Image Understanding* **57**(2) 131-54

[6] Lei Xu, rkki OJA, Pekka Kultanen 1990 A new curve detection method: randomized Hough transform RHT *Pattern Recognition Letters* **11**(5) 331-8

[7] Chen Tehchuan, Chung Kuoliang 2001 An efficient randomized algorithm for detecting circles *Computer Vision and Image Understanding* **83**(2) 172-91

Authors

Hudai Fu, born on November 25, 1977, China

Current position, grade: researcher at Changchun Institute of Technology, China.
University studies: M.E. degree in mechatronics engineering from Changchun University of Technology, China in 2004.
Scientific interests: mechatronics technology and intelligent detection.

Jingang Gao, born on September 9, 1976, China

Current position, grade: researcher at Changchun Institute of Technology, China.
University studies: M.E. degree in mechanical engineering from Huaqiao University, China in 2004.
Scientific interests: mechatronics technology and intelligent detection.

Applications of dynamic adaptive bee colony algorithm in multi-threshold image segmentation

Ye Chen*, Xiaoqun Qin

School of Information Science and Engineering, Hunan International Economics University, Changsha 410205, Hunan, China

Received 1 July 2014, www.cmnt.lv

Abstract

Artificial bee colony (ABC) is an evolutionary computation method, which is inspired from the specific collaborative social group behaviour among the individual bees in the colony and which is a heuristic optimization algorithm based on population search strategy. This paper has proposed a quick dynamic adaptive bee colony algorithm, which analyses the performances of the artificial bee colony algorithm and it designs a multi-threshold image segmentation method realizing a dynamic adaptive artificial bee colony (DAABC) with multi-threshold OTSU as the fitness function. The main characteristics of this method include: reducing the noise interferences in the multi-threshold image segmentation; effectively narrowing down the search range of the threshold; guaranteeing the quickness of the segmentation speed; determining the search range of the reconnaissance ants with adaptive dynamic control and accelerating the convergence speed of bee colony algorithm. The experimental results demonstrate that the method in this paper is better than the image segmentation method based on particle swarm optimization (PSO) and artificial fish swarm algorithm (AFSA).

Keywords: Artificial Bee Colony Algorithm, Multi-Threshold, Image Segmentation, OTSU

1 Introduction

ABC is a newly-emerging intelligent heuristic algorithm, which was proposed by Professor Karaboga in Erciyes University in 2005 and it comes from the research and simulation of honey-collecting behaviour of the bees. Compared with PSO, ant colony optimization (ACO) and differential evolution (DE), this method has simple operations, fewer parameters and strong robustness; therefore, it has attracted extensive attention of domestic and international scholars and it has successfully settled numerous practical problems. However, since ABC hasn't appeared until the recent years, it still has a short development history and there are still many problems to be solved, including the weak theoretical support, the critical times setting of the appearance of reconnaissance bee and the search behaviour of the honey-collecting bees.

The improvements of ABC mainly include the following three aspects: (1) the improvements of the search behaviour of the honey-collecting bees, including the suppression of bad food sources from becoming the neighbourhood bees and the accelerated movements the honey-collecting bees make on the excellent food sources; (2) the replacements of bad food sources. In the population evolution, replace the bad food sources in a narrow space in every circle to improve the quality; (3) the improvements of the appearance of the reconnaissance bees. In a new search, since the food sources which have been abandoned for too many retention times have been differentiated for the position in the population, adopt reverse learning strategy if it is global optimal; otherwise, cross them with better food sources[1].

In recent years, OTSU based on the maximum between-cluster variance has been successfully used in image dual threshold segmentation. This paper realizes multi-threshold image segmentation with OTSU. Additionally, it has used it in the artificial bee colony search algorithm and proposed a dynamic adaptive bee colony algorithm to analyse the value and status of bee colony algorithm in multi-threshold image segmentation.

2 The construction of the objective function in multi-threshold image segmentation

2.1 THE CONSTRUCTION OF OBJECTIVE FUNCTION

Firstly, assume that a grey-scale image can be shown with L grey scales ($L \in [1,256]$); the possibility for the pixel with i grey scales is $P_i(P_i = \frac{f_i}{N}, P_i > 0, \sum_{i=1}^{L} P_i = 1)$; f_i is the sum of the pixels with i grey scales and N is the total pixels of the image. If the threshold t segments the image l into the objective and the background, then the possibilities of the objective and the background are ω_0 and $\omega_1(t)$ separately and the mean value of the objective and the background are $\mu_0(t)$ and $\mu_1(t)$ and $\omega_0 + \omega_1 = 1$.

* ***Corresponding author*** e-mail: 153569367@qq.com

In the above formula, $\omega_0 = \sum_{i=0}^{t} P_i$, $\omega_1(t) = \sum_{i=t+1}^{L-1} P_i$, $\mu_0(t) = \sum_{i=0}^{t} iP_i / \omega_0$ and $\mu_1(t) = \sum_{i=t+1}^{L-1} iP_i / \omega_1$, $\omega_0 = \sum_{i=0}^{t} P_i$, $\omega_1(t) = \sum_{i=t+1}^{L-1} P_i$, $\mu_0(t) = \sum_{i=0}^{t} iP_i / \omega_0$, $\mu_1(t) = \sum_{i=t+1}^{L-1} iP_i / \omega_1$. Then the difference between the objective and the target is: $D(t) = \omega_0(t)\omega_1(t)(\mu_0(t) - \mu_1(t))^2$. Therefore, t with the biggest difference is the optimal threshold.

As mentioned above, assuming that the thresholds of the image to be segmented are m, then its difference is:

$$\begin{aligned} D(t_1, t_2, \ldots, t_m) &= \omega_0\omega_1(\mu_0 - \mu_1)^2 + \omega_0\omega_2(\mu_0 - \mu_2)^2 \\ &+ \cdots + \omega_0\omega_m(\mu_0 - \mu_m)^2 + \omega_1\omega_2(\mu_1 - \mu_2)^2 \\ &+ \omega_1\omega_3(\mu_1 - \mu_3)^2 + \cdots + \omega_{m-1}\omega_m(\mu_{m-1} - \mu_m)^2 \end{aligned} \quad . \quad (1)$$

In the formula, $\omega_{m-1} = \sum_{i=t_{m-1}+1}^{t_m} P_i$ and $\mu_{m-1} = \sum_{i=t_{m-1}+1}^{t_m} iP_i / \omega_{m-1}$. Assuming the segmentation threshold of the image is $(t_1^*, t_2^*, \ldots, t_m^*)$, then the optimal threshold of the image to be segmented is $(t_1^*, t_2^*, \ldots, t_m^*) = \underset{0 \le t_1 \le t_2 \ldots \le t_m}{Arg \max} D(t_1, t_2, \ldots, t_m)$ from the below formula according to DAABC [2, 3].

2.2 THE FEASIBILITY ANALYSIS OF THE OBJECTIVE FUNCTION

In the swarm intelligent optimization algorithm, the fitness function has a significant importance on the population optimization and it guides the population to forage or evolve towards a certain direction and finally reaches the optimal status.

This method sees the image threshold as the bee of the ABC and designs the fitness function of ABC through two-dimensional OTSU and it gets closer to the optimal threshold generation by generation through the collaboration and information share of the honey-collecting bees, the reconnaissance bees and the observing bees [4].

ABC searches the optimal nectar source through the collaboration and information share of the honey-collecting bees, the observing bees and the reconnaissance bees. The following analyses the working mechanism of the bee population by integrating the application environment of this paper. When ABC is used in the image segmentation, every nectar source position corresponds to a threshold and the honey volume of the nectar source corresponds to the fitness function value. The honey-collecting bee produces and compares a new position by searching the neighbourhood according to its memory. If the new position is superior to the optimal position in memory, replace it with the new position; otherwise, keep the original optimal position [5]. After all the honey-collecting bees finish searching, they share and nectar source information with the observing bees through "swing dance" and then the observing bees choose a nectar source position with the possibility related to the honey volume according to the information obtained and search and produce a new position by searching near the nectar source. After that, they compare with the previous position and always remember the optimal position. If a nectar source can't be improved through limited circulations, then abandon it and use the new position randomly producing by the reconnaissance bees [6].

In the bee colony search algorithm, the number of nectar sources can be seen as the number of thresholds of the image to be segmented, namely the number of variables of OTSU function. Every nectar source is deemed as the possible threshold of the image, namely the parameter $(t_1, t_2, \ldots, t_m)$ of OTSU function and when segmented 256-greyscale image, $(t_1, t_2, \ldots, t_m)$. Therefore, the objective function of the multi-threshold segmentation in (1) in this paper can be used as the fitness function of the bee colony search algorithm and the process to get a group of optimal parameters of multi-threshold OTSU function by using bees to find the optimal nectar source is the process to find the maximum value of Formula (1). In this way, it is feasible for the artificial bee colony search algorithm to be applied in the multi-threshold image segmentation [7, 8].

3 The analysis of segmentation object

Choose 4 commonly-used standard test images: Railway, Skyscraper, Bridge, Sunshine and woods, as the images to be segmented. The following are the grey-scale histograms of the 4 images. Analyse their histogram features.

FIGURE 1 Railway

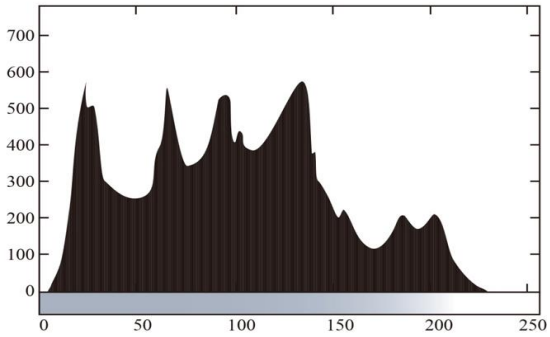


TABLE 1 The Histogram of Fig.1

FIGURE 2 Skyscraper

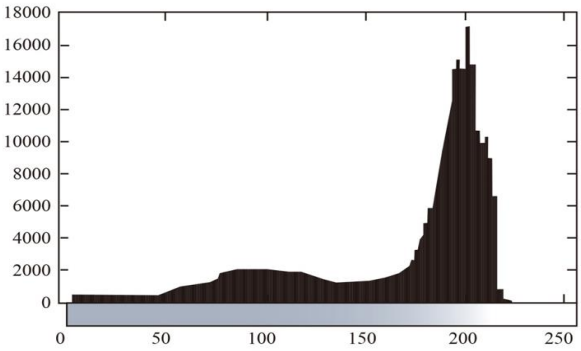


TABLE 2 The Histogram of Fig.2

FIGURE 3 Bridge

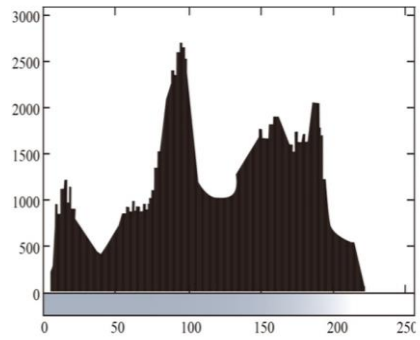


TABLE 3 The Histogram of Fig.3

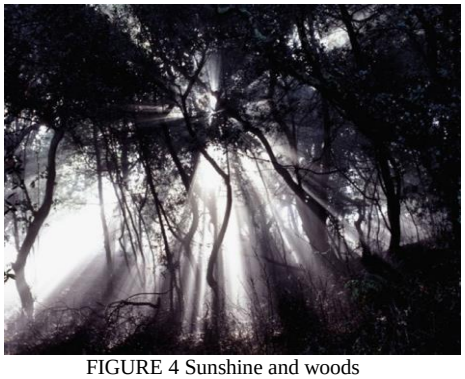


FIGURE 4 Sunshine and woods

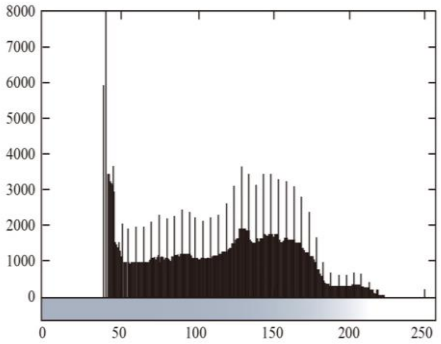


TABLE 4 The Histogram of Fig. 4

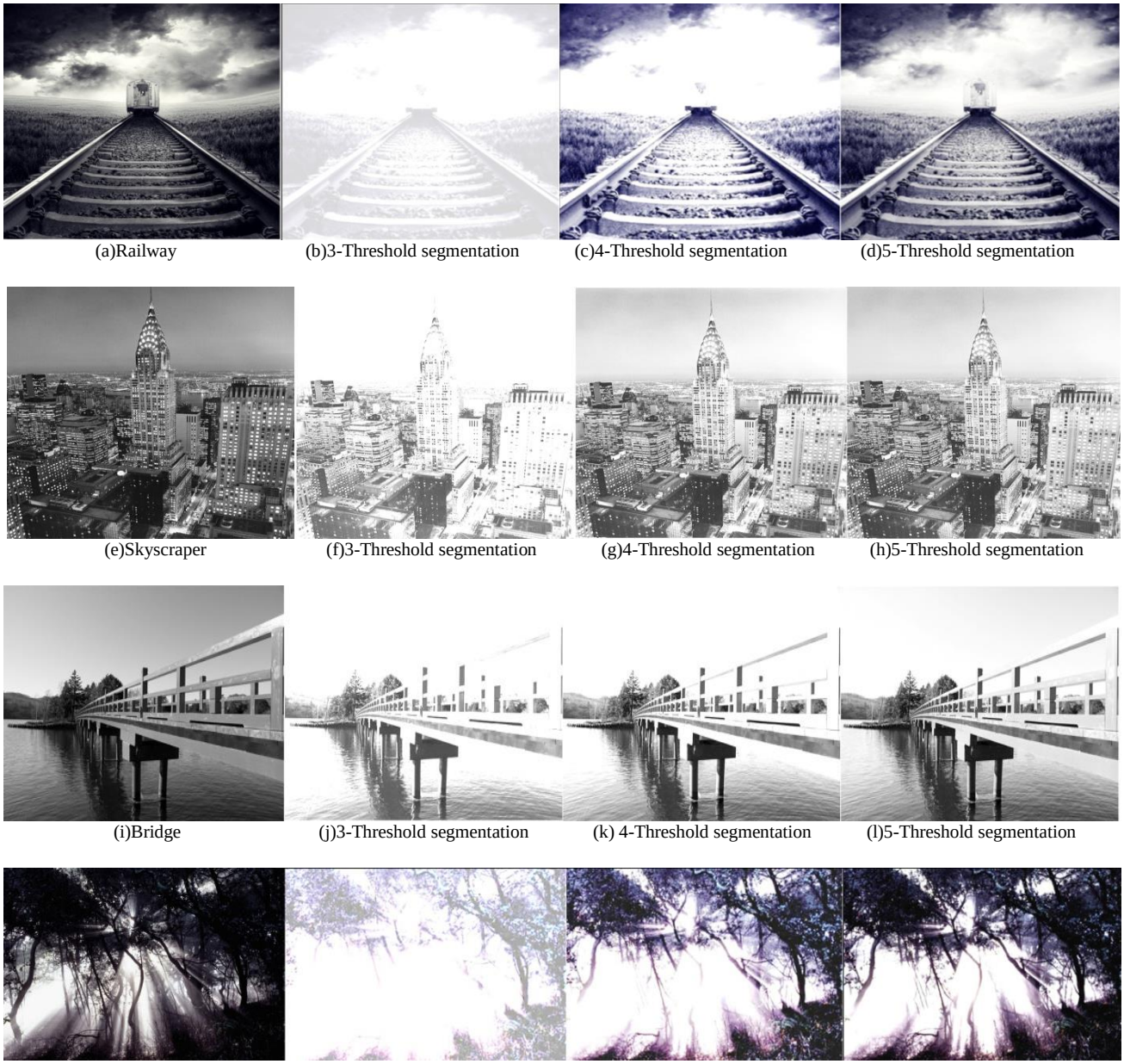
It can be seen that the histograms of the chosen standard test images from Table 1 to Table 4 have multiple peaks instead of only two peaks. The image histogram directly reflects the grey-scale distribution of the pixel points in the image; while single threshold can only segment the images with only two peaks in the histograms. Therefore, the previous single threshold segmentation method cannot precisely segment the images from Fig.1 to Fig.4 and only multi-threshold segmentation methods can be used to get ideal segmentation results.

4 Experimental analysis

In order to verify the effectiveness of DAABC in multi-threshold image segmentation, compare the performances of DAABC, PSO and basic AFSA in multi-threshold image segmentation with the above 4 standard test images as the images to be segmented and starting from the segmentation effects, segmentation speed and the quality of the optimal solutions.

4.1 THE COMPARISON OF SEGMENTATION RESULTS

With Formula 1 as the fitness function, obtain the 3-threshold, 4-threshold and 5-threshold of the above 4 standard test images through DAABC and get their segmentation images. As indicated in Fig.7, the left are the standard test images while the rights are their 3-threshold, 4-threshold and 5-threshold segmentation images successively.



(a)Railway (b)3-Threshold segmentation (c)4-Threshold segmentation (d)5-Threshold segmentation

(e)Skyscraper (f)3-Threshold segmentation (g)4-Threshold segmentation (h)5-Threshold segmentation

(i)Bridge (j)3-Threshold segmentation (k) 4-Threshold segmentation (l)5-Threshold segmentation

(m)Sunshine and woods (n)3-Threshold segmentation (o)4-Threshold segmentation (p)5-Threshold segmentation

FIGURE 5 The results of segmentation image by DAABC

As shown in Fig.5, with the increase of the segmentation thresholds, the segmentation image is closer to the original image with better quality by integrating human visual features and comparing them with the original images. It is certain that if the threshold number of the segmentation image is closer to the number of the bottoms in the grey-scale histogram of the original images, the segmentation image is closer to the original image.

4.2 THE COMPARISON OF SEGMENTATION SPEED

In the operating environment of MATLAB, compare the performances of DAABC, PSO and AFSA in multi-threshold segmentation. Set the parameters of the 3 algorithms as follows: iterations: 50; population scale: 30 and fitness function: OTSU.

TABLE 5 The comparison of computation time of 3 swarm intelligent optimization algorithms (Unit/S)

Fig.5	M	DAABC	PSO	AFSA
(a)	3	8.5718	19.5173	13.5714
	4	10.1306	19.4181	12.8185
	5	10.7585	21.3673	15.3728
(e)	3	9.6138	19.5286	13.3185
	4	10.1582	19.6196	12.8396
	5	11.35518	20.1467	15.8179
(i)	3	10.9327	17.2868	12.5261
	4	11.6384	20.3628	15.9383
	5	10.7386	18.9515	15.2693
(m)	3	9.6281	20.7295	12.5718
	4	11.5795	19.5383	13.4156
	5	10.6215	20.3572	16.8627

Run the procedures 50 times and the time the 3 algorithms need to find the optimal threshold is as indicated in Table 7. According to this table, the order of optimization speeds of the 3 algorithms is DAABC>AFSA>PSO with the same thresholds. This is because bee colony algorithm is an algorithm of positive feedback and it is the most important link for the bees to exchange information. The dancing area is the most important information exchange place in the honeycomb. The bee dance is called swing dance. The information of the food sources is shared with other bees through swing dance in the dancing area and it leads the bees to show the yield rate of the food sources through the swing dance time; therefore, following the bees can observe many dances and choose which food source to collect the honey according to the yield rate. The yield rate is proportional to the possibility to be chosen. Therefore, the possibility the bees are recruited to a certain food source is in proportion to the yield rate of the food source. This positive feedback process leads the entire system to evolve towards the optimal solution. Therefore, positive feedback is an important characteristic of bee colony algorithm and it promotes the algorithm evolution to proceed.

It is remarkable that when segment the same image with the same algorithm of different thresholds, the time increases with the increase of thresholds because with the increase of the thresholds, the dimensions of the problems increases, thus increasing the time complexity of the algorithm.

4.3 THE COMPARISON OF QUALITY OF OPTIMAL SOLUTION

Set the same parameters as 4.2 in this paper. Run the procedures for 50 times and the results the optimal fitness function value and the optimal threshold are indicated as Table 7.

The fitness function demonstrates the difference between the target and the background. The bigger, the better. From Table 8, it can be seen that the order of the solution quality of the 3 algorithms is DAABC>AFSA>PSO on the same image with the same thresholds. Its reason is that as a parallel algorithm in nature, bee colony algorithm can be seen as a distributed multi-agent system. It has more points in the problem space and it conducts independent solution search, which not only increases the reliability of the algorithm, but also makes the algorithm have strong global search capacity. Besides, it can not only expand the search range and increase population diversity, but it is also easier to jump out of the local optimal point.

TABLE 6 The comparison of solution results of 3 swarm intelligent optimization algorithms

Fig.5	m	Fitness function value			Optimal threshold		
		DAABC	PSO	AFSA	DAABC	PSO	AFSA
(a)	3	2451.8656	2360.3573	2448.8316	57,458,383	75,63,718	63,557,381
	4	2568.3738	2578.6932	2538.6831	48,52,159,185	32,71,852,391	49,51,651,783
	5	2782.6337	2685.4873	2786.8263	45,89,736,138,663	50,95,754,134,036	47,76,652,876,275
(e)	3	2784.6583	2563.6753	2165.5647	83,674,673	94,735,456	82,874,564
	4	2254.6737	2363.2341	2356.3421	71,116,352,241	58,56,116,352	70,352,342,241
	5	2342.8741	2534.6541	2251.8652	56,54,81,643,245	18,23,81,319,377	49,37,20,341,452
(i)	3	2653.7351	2852.8361	2851.2341	65,367,465	71,85,345	61,763,875
	4	2974.3561	2867.6238	2847.6748	48,67,478,357	45,478,367,478	40,46,478,487
	5	2836.4789	2747.4879	2746.4799	45,47,479,467,278	47,46,478,783,783	51,46,467,478,478
(m)	3	2146.4799	2046.4798	2146.4748	75,367,478	57,46,457	72,467,478
	4	2673.4748	2546.4783	2367.4782	67,478,457,578	67,47,371,478	63,567,478,387
	5	2498.4673	2367.4678	2378.4789	63,46,674,467,376	61,467,467,673,987	52,76,235,765,387

5 Conclusion

This paper has introduced the dynamic adaptive bee colony algorithm into the multi-threshold image segmentation; design multi-threshold fitness function with OTSU and guides the bee colony to quickly find the optimal segmentation threshold. By testing the standard images and comparing the optimization performance of the other two swarm intelligent algorithms, the experimental results show that the algorithm in this paper has faster optimization speed and high optimization quality compared with the other two swarm intelligent algorithms on the same images to be segmented with the same thresholds.

References

[1] Kuo R J, Huang Y D 2014 Automatic kernel clustering with bee colony optimization algorithm *Information Sciences* **283**(1) 107-22

[2] Tsai Hsing-Chih 2014 Integrating the artificial bee colony and bees algorithm to face constrained optimization problems *Information Sciences* **258**(10) 80-93

[3] Doğan Aydin, Serdar Özyön, Celal Yaşar, Tianjun Liao 2014 Artificial bee colony algorithm with dynamic population size to combined economic and emission dispatch problem *International Journal of Electrical Power & Energy Systems* **54** 144-53

[4] Fatma Latifoğlu 2013 A novel approach to speckle noise filtering based on Artificial Bee Colony algorithm: An ultrasound image application *Computer Methods and Programs in Biomedicine* **111**(3) 561-9

[5] Tien Jia-Ping, S Li Tzuu-Hseng 2012 Hybrid Taguchi-chaos of multilevel immune and the artificial bee colony algorithm for parameter identification of chaotic systems *Computers & Mathematics with Applications* **64**(5) 1108-19

[6] Fatih Tasgetiren M, Pan Quan-Ke 2013 A discrete artificial bee colony algorithm for the no-idle permutation flowshop scheduling problem with the total tardiness criterion *Applied Mathematical Modelling* **37**(10) 6758-79

[7] Gao Weifeng, Liu Sanyang, etc. 2012 A global best artificial bee colony algorithm for global optimization *Journal of Computational and Applied Mathematics* **236**(11) 2741-53

[8] Kalayci C B, Gupta S M 2013 Artificial bee colony algorithm for solving sequence-dependent disassembly line balancing problem *Expert Systems with Applications* **40**(18) 7231-41

Authors

Ye Chen, born on November 2, 1979, China

Current position, grades: researcher at Hunan International Economics University, China.
University studies: Master's degree in Computer Application technology from Hunan University, China in 2011.
Scientific interests: Data mining and image processing

Xiaoqun Qin, born on June 13, 1978, China

Current position, grades: researcher at Hunan International Economics University, China.
University studies: Master's degree in Computer Application technology from Central South University, China in 2007.
Scientific interests: Data mining and image processing.

Alignment-based approximate SPARQL querying on linked open data

Yu Liu[1, 2], Lei Chen[1*], Shihong Chen[1]

[1]*School of Computer Science and Technology, Wuhan University, Wuhan 430072, Hubei, China*

[2]*School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China*

Received 1 June 2014, www.cmnt.lv

Abstract

With the growth of Linked Open Data, more and more applications are developed to take full advantage of its massive data. However, all these applications face an inevitable problem - how to retrieve information from these datasets with different schemas, which results in that a query for a dataset may get none answer from other datasets. To solve this problem, ontology alignment has been adopted in some Linked Open Data querying systems. In this paper, we follow this idea and make further efforts to find more approximate answers by employing relations and probability values in the result of ontology alignment. The fundamental of our method is the similarity between entities, which is used to evaluate the similarity of rewritten query relative to original query. In order to facilitate user to query other dataset with original query, an algorithm for alignment-based approximate querying is proposed. In experiments, the SPARQL queries for DBpedia are rewritten on the basis of alignment result between DBpedia and YAGO. The results of experiments show that alignment-based approximate querying can not only retrieve approximate results, but also overcome the problem caused by imprecise result of ontology alignment, which is very common for most of alignment techniques.

Keywords: Linked Open Data, Approximate Querying, Ontology Alignment, SPARQL

1 Introduction

Since the W3C Linked Open Data (LOD) project launched in 2007, more and more datasets were published in the web of Data, which have covered a diversity of areas such as nature, geography, government, and so on. It offered a great incentive for researchers to develop various applications based on the LOD. The Traffic LarKC combined DBpeida, a central part of LOD, with an eventful wrapper and datasets of two Milano municipalities to implement an intelligent question answering system about the traffic [1]. On the basis of information theory, Meymandpour R. et al. took advantage of LOD in the domain of university to rank universities and compare the results with the international ranking system [2]. Obviously, the inevitable problem for those applications is how to retrieve information form the LOD efficiently.

There are several approaches to query the LOD and each of them has its own advantages and limitations [3]. Data warehouses and query federation, two traditional approaches for query answering over distributed data, can be employed to execute the complex and structured queries over LOD [4, 5]. The search engine for the LOD can retrieve the web by following RDF links and provide query interfaces for users to search information in the LOD, just as Sindice [6]. In contrast to above methods, traversal based query execution over LOD, a novel query execution paradigm, can intertwine query pattern matching over a continuously growing dataset with the traversal of links in order to discover data that might be relevant to answer the executed query [7]. However, no matter what approach the user adopts, the query patterns must be provided in advance, that means the user should be familiar with all schemas of the datasets he wants to query.

In fact, it is an impossible task for a user because the number of datasets in LOD increases continuously and each dataset always has a particular schema. For example, DBpedia [8] and YAGO [9] have their respective properties to express the relationships between the instances even if two properties have the same meaning, such as “geo:long” and “yago:hasLongitude” shown in Figure 1. Suppose a user want to query the information of longitude, latitude and population about the city Chengdu, the SPARQL query in formula (1)[1] can retrieve the results from DBpeida, but YAGO returns nothing because of the mismatches between query patterns and schemas of YAGO. Except for the above inconvenience, querying on a single dataset may lose the chance to get answer that exists in other datasets. In Figure 1, the property “dbo:capital” between “dbr:Sichuan” and “dbr:Chengdu” is not stated in DBpeida, so that the query pattern “dbr:Sichuan dbo:capital ?c” finds nothing in DBpedia. But, “yago:Sichuan yago:hasCapital ?c”, the query pattern with the same meaning, can make up for the

* ***Corresponding author*** e-mail: Chen_lei0605@sina.com

[1]All queries in this paper omit the definition of prefixes declared in Figure 1

discount in DBpeida by obtaining the answer from YAGO.

```
Select   ?lat     ?long    ?p
Where { dbr:Chengdu geo:lat ?lat.                    .        (1)
        dbr:Chengdu  geo:long ?long.
        dbr:Chengdu dbo:populationTotal ?p.}
```

In order to solve those problems, the approach that user can follow is ontology alignment. By combining the mechanism of query federation with the results of ontology alignment, an alignment based Linked Open Data querying system (ALOQUS) was proposed in [10]. It can map concepts in an upper ontology or domain specific ontology to concepts in other datasets, providing the capability to answer queries which cannot be answered by other state of the art system for LOD query processing. In this paper, we follow the idea of ALOQUA and make further efforts. Our method can find more approximate answers and overcome the problem caused by imprecise result of ontology alignment. The contribution of this paper is three-fold:

(1) We measure the similarity of rewritten query relative to the original query. The calculation of similarity depends on the result of ontology alignment, which include not only the relations between entities, but also the probability values of those relations.

(2) We propose an algorithm for alignment-based approximate querying, which can help user to query other dataset by using the original query and result of ontology alignment.

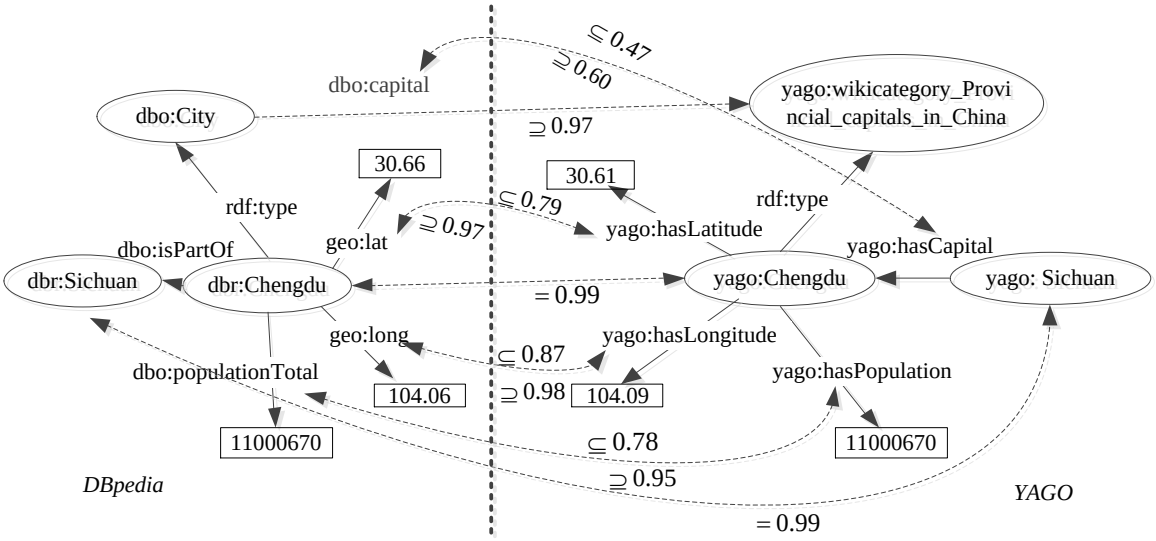
(3) We prove the validity of our approach through experiments on real-world datasets of LOD. The results show that alignment-based approximate querying can retrieve more answers.

The remainder of this paper is organized as follows. Section 2 discusses related work. In section 3, the SPARQL query rewriting is described with a formal way. How to calculate similarity between entities and queries on the basis of alignment result are explained in section 4. Section 5 proposes an algorithm for alignment-based approximate querying. The experiment results and their evaluations are presented in section 6. The conclusion of this paper is in section 7.

2 Related work

Hurtado et al. propose an RDF query relaxation method through RDF(s) entailment producing more general queries for retrieving potential relevant answers [11]. In order to ensure the desired cardinality and quality of answers, Huang et al. use the similarity between relaxed queries and original query to control the relaxation process [12, 13]. However, these works relax the original query based on a single schema, which is not suitable for various schemas of datasets in LOD. In [14, 15], Reddy et al. attempt to make use of ontologies available on the web of data to produce approximate answers, and integrate the approximate steps with query execution to improve the performance of query processing. Nevertheless, the similarity measure of relaxed SPARQL queries and the experiments in [14] are still based on the dataset with a single schema.

The purpose of ontology alignment is to find the sets of correspondences between entities belonging to the matched ontologies [16]. The result of ontology alignment is generated by hand or by ontology matchers and can be used for merging ontologies, linking datasets and transforming queries [17]. Just like ALOQUS introduced in [10], the best alignments in the result of BLOOMS mapping are used to transform the sub-queries, which are executed in the corresponding end-points later.



rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
dbo: <http://dbpedia.org/ontology/> dbr: <http://dbpedia.org/resource/>
yago: <http://yago-knowledge.org/resource/>

FIGURE 1 An example of ontology alignment between DBpedia and YAGO

Here, BLOOMS is an ontology matcher for schema alignment and can find the relations of including and equivalence between the classes belonging to the matched ontologies [18]. In general, relations found by an ontology matcher have some associated metadata. A frequently-used metadata element is confidence in the correspondence (typically in the [0, 1] range). The higher confidence means the higher likelihood that the relation holds [19]. Some ontology matchers use probability to describe the likelihood of relation between the entities, such as probability scores in PARIS (Probabilistic Alignment of Relations, Instances, and Schema) [20]. PARIS is a holistic ontology matcher--aligning not only instances but also properties and classes, which is a significant feature that can be employed to transform queries. Figure 1 presents some alignment results of PARIS about properties and instances between DBpeida and YAGO (e.g. "geo:lat" is a sub relation of "yago:hasLatitude" with the probability score 0.79 (Although values presented in this paper only reserve two decimal fractions, the type of values is double in the implementation procedure of alignment-based approximate SPARQL query)). In this paper, we will take the most advantage of PARIS to implement the approximate SPAQRL querying on LOD.

3 SPARQL query rewriting

With the purpose of retrieving the data stored in RDF format, W3C proposed SPARQL. The fundamental component of SPARQL query is triple pattern, which can be expressed as $(s, p, o) \in (I \cup V) \times (I \cup V) \times (I \cup V \cup L)$. Here, V is a set of variables, I is a set of IRIs (Internationalized URIs), and L is a set of literals. In general, a SPARQL query is built on some triple patterns with series operations of conjunction and disjunction. It is worth noting that other components of SPARQL query are not mentioned here. It is because that they have little influence on the SPARQL query rewriting.

Given Q is a SPARQL query and all IRIs in triple patterns of Q come from the dataset $D1$, Q can be denoted as Q^{D1}, which means the Q is designed for $D1$ and may retrieve some results from $D1$. Because of the diversity of datasets in LOD, the execution of Q^{D1} on another dataset $D2$ may return nothing, just as the example introduced above. By replacing IRIs in triple patterns of Q^{D1} with IRIs in $D2$, the SPARQL query rewriting can generate some new queries (denoted as Q^{D2}) that have the same or similar meaning with Q^{D1}. Then, users do not have to know the schema of D2 to query it with Q^{D2}. For example, the query for DBpedia in formula (1) can be transformed to the query in formula (2), so that user can achieve the desired answers from YAGO. Despite having the significant advantage, the SPARQL query rewriting faces a vital problem--how to choose IRIs in $D2$ to replace the IRIs in Q^{D1}. Apparently, the high similarity between them is preferred.

```
Select    ?lat    ?long    ?p
Where { yago:Chengdu yago:hasLatitude ?lat.
        yago:Chengdu  yago:hasLongitude ?long.
        yago:Chengdu yago:hasPopulation ?p.}
```
. (2)

4 Alignment-based similarity

4.1 SIMILARITY BETWEEN ENTITIES

To compute the similarity of IRIs from two different datasets, the results of ontology alignment can be employed. The result of ontology alignment between $D1$ and $D2$ is a finite set S, including a certain number of 4-tuples. Formally, a 4-tuple can be written as $<e1, e2, r, p>$. Here, $e1$ and $e2$ are entities in $D1$ and $D2$ respectively, and r is the relation between $e1$ and $e2$, such as equivalence ($\equiv$), including ($\supseteq$), included ($\subseteq$), and etc. The true probability value of the relation between $e1$ and $e2$ is p, which ranges from 0 to 1. For example, the relations between "geo:lat" and "yago:hasLatitude" in Figure 1 can be expressed as $<geo:lat, yago:hasLatitude, \supseteq, 0.97>$ and $<geo:lat, yago:hasLatitude, \subseteq, 0.79>$. Nevertheless, $<dbo:shipBeam, yago:hasLatitude, \subseteq, 0.78>$ also is a part of result when user adopts PARIS to align DBpeida and YAGO. At this time, the SPARQL query rewriting needs to decide which property is more similar with "yago:hasLatitude".

Taking the diversity of 4-tuple into account, our method computes the similarity between entities according to different conditions. First, the relationship between instances is always equivalence, so similarity between $i1$ and $i2$ is p when S includes $<i1, i2, \equiv, p>$. Here, $i1$ and $i2$ are instances in $D1$ and $D2$ respectively. In Figure 1, $<dbr:Chengdu, yago:Chengdu, \equiv, 0.99>$ implies that similarity between "dbr:Chengdu" and "yago:Chengdu" is 0.99, indicated as

$$Sim(dbr:Chengdu, yago:Chengdu) = 0.99. \quad (3)$$

Second, class or property in dataset is usually viewed as a set that means r in 4-tuple could be any relations for set, which make the computation of similarity for them more complex than instance. For definiteness and without loss of generality, only computing methods for similarity between classes are introduced here. Given $c1$ and $c2$ are classes in $D1$ and $D2$ respectively, the similarity between two classes can be defined as:

$$Sim(c1,c2) = \frac{|c1 \cap c2|}{|c1 \cup c2|}, \quad (4)$$

where $|c1 \cup c2|$ is the cardinality of union of $c1$ and $c2$. It is clear that similarity between classes is in the [0, 1]. Suppose that the relation set of classes includes equivalence ($\equiv$), including ($\supseteq$) and included ($\subseteq$). Then, the calculations of similarity based on S are divided into the following cases:

(1) $<c1,c2,\equiv,p> \in S$. In this case, similarity between *c1* and *c2* is *p*.

Proof: Given *U* includes all instances of classes in *D1* and *D2*, and *x* is an element in *U*, then

$$P(c1 \equiv c2) = P(x \in c1 \cap c2 \mid x \in c1 \cup c2) = \frac{P(x \in c1 \cap c2)}{P(x \in c1 \cup c2)} = \frac{\frac{|c1 \cap c2|}{|U|}}{\frac{|c1 \cup c2|}{|U|}} = \frac{|c1 \cap c2|}{|c1 \cup c2|} = Sim(c1,c2) \quad (5)$$

(2) $<c1, c2, \supseteq, p1> \in S$ and $<c1, c2, \subseteq, p2> \in S$. In this case, similarity between *c1* and *c2* is

$$Sim(c1,c2) = \frac{p1 \cdot p2}{p1 + p2 - p1 \cdot p2}. \quad (6)$$

Proof: Given *x* is an element in *U*, then

$$P(c1 \subseteq c2) = P(x \in c1 \cap c2 | x \in c1) = \frac{|c1 \cap c2|}{|c1|}, \quad (7)$$

$$P(c1 \supseteq c2) = P(x \in c1 \cap c2 | x \in c2) = \frac{|c1 \cap c2|}{|c2|}, \quad (8)$$

Because $P(c1 \subseteq c2) = p2$ and $P(c1 \supseteq c2) = p1$, then

$$|c1| = \frac{|c1 \cap c2|}{p2}, \quad (9)$$

$$|c2| = \frac{|c1 \cap c2|}{p1}, \quad (10)$$

Now, bring formula (9) and (10) into the definition of similarity, the conclusion can be achieved with the following inference.

$$Sim(c1,c2) = \frac{|c1 \cap c2|}{|c1 \cup c2|} = \frac{|c1 \cap c2|}{|c1| + |c2| - |c1 \cap c2|} = \frac{|c1 \cap c2|}{\frac{|c1 \cap c2|}{p2} + \frac{|c1 \cap c2|}{p1} - |c1 \cap c2|} = \frac{p1 \cdot p2}{p1 + p2 - p1 \cdot p2}. \quad (11)$$

For example, *<dbo:School, yago:wordnet_school_108276720, ⊇ ,0.64>* and *<dbo:School, yago:wordnet_school_108276720, ⊆ , 0.78>* are in the result of ontology alignment when PARIS is adopted to align DBpedia and YAGO. Hence, the similarity between "dbo:School" and "yago:wordnet_school_108276720" is 0.54.

(3) $<c1, c2, \supseteq, p1>$ *S* or $<c1, c2, \subseteq, p2> \in S$. Since only one relation between two classes found by ontology matcher, formula (6) cannot be used directly in this case. For optimal performance, ontology matchers always set some thresholds to reduce the amount of computations, such as PARIS assumes every value below to be zero [20]. It means the probability of undetected relation may be a value below the threshold. In addition, there are some relations between including probability and included probability. Table 1 presents some super classes of "dbo:School" in YAGO, which are found by PARIS. Obviously, more general super classes of original class *c* is prone to have bigger value of included probability and leads to smaller value of including probability – and vice versa. In order to approximately simulate the above-mentioned relations between included probability and including probability, the formula (12) is adopted here.

TABLE 1 Some super classes of "dbo:Hospital" in YAGO

DBpeida	Relation	YAGO	Probability
dbo:School	⊆	yago:wordnet_senior_high_school_108409617	0.47
dbo:School	⊆	yago:wordnet_secondary_school_108284481	0.61
dbo:School	⊆	yago:wordnet_school_108276720	0.78
dbo:School	⊆	yago:wordnet_institution_108053576	0.78
dbo:School	⊆	yago:wordnet_educational_institution_108276342	0.78
dbo:School	⊆	yago:yagoLegalActorGeo	0.79
dbo:School	⊆	yago:wordnet_entity_100001740	0.80

$$P(c \supseteq x) = \begin{cases} \frac{T \cdot (max(c,\subseteq) - P(c \subseteq x))}{max(c,\subseteq) - min(c,\subseteq)} & max(c,\subseteq) \neq min(c,\subseteq) \\ T \cdot (1 - P(c \subseteq x)) & max(c,\subseteq) = min(c,\subseteq) \end{cases}. \quad (12)$$

Here, *T* is threshold value of including relation, which is the upper limit for including relation, *max(c, ⊆)* is the maximum probability value in the superclass set of *c*, *min(c, ⊆)* is the minimum probability value in the superclass set of c. For example, *max(dbo:School, ⊆)* is 0.80 and *min(dbo:School, ⊆)* is 0.47, just as shown in Table 1. Once $P(c \supseteq x)$ gets calculated, the similarity between *c* and *x* can be easily achieved by bringing $P(c \supseteq x)$ and $P(c \subset x)$ into formula (6).

In fact, the computation of included probability based on including probability can adopt the same mechanism, just as formula (13) shows. In the same way, *T'* is threshold value of included relation; *max(c, ⊇)* is the maximum probability value in the subclass set of c; *min(c, ⊇)* is the minimum probability value in the subclass set of c.

$$P(c \subseteq x) = \begin{cases} \frac{T' \cdot (max(c,\supseteq) - P(c \supseteq x))}{max(c,\supseteq) - min(c,\supseteq)} & max(c,\supseteq) \neq min(c,\supseteq) \\ T' \cdot (1 - P(c \supseteq x)) & max(c,\supseteq) = min(c,\supseteq) \end{cases}. \quad (13)$$

Although formula (12) and formula (13) are exactly similar, there are some tricks, that should be mentioned. For instance, *T* and *T'* are two key parameters, which

reflect the tendency of users for precision or recall. If *T'* is bigger than *T*, classes in subclass of *c* is prone to achieve higher similarity, that means these classes can take precedence to be selected during the process of query rewriting. Then, the precision of query results may increase according to the mechanism introduced in section 5.

4.2 SIMILAR ENTITIES RETRIEVAL

Based on the definition of similarity between entities, an algorithm is given here to obtain the similar entities of target entity. The steps of the algorithm consist of initialization, retrieval of similar entities, sorting the set of similar entities, and so on. The details of the algorithm are described as follows:

Algorithm SER
Input: an entity in original dataset *e*, a result set of ontology alignment *resultSet*, threshold value of including relation *T*, threshold value of included relation *T'*.
Output: a set of entity with similarity *similarSet*
1: Initialize *similarSet*
2: find equivalent set of *e* from *resultSet*, denoted as *equalSet*
3: find superclass set of *e* from *resultSet*, denoted as *superSet*
4: find subclass set of *e* from *resultSet*, denoted as *subSet*
5: for each *x* in *equalSet* do
6: push <*x*, *p*> into *similarSet* according to formula (5); Here, *p* is probability value, with which *e* equals with *x*.
7: end for
8: for each *x* in *superSet*
9: if x is not in *similarSet*
10: if x is in *subSet*
11: calculate the similarity *s* between *x* and *e* according to formula (6)
12: else
13: calculate the similarity *s* between *x* and *e* according to T, formula (12) and formula (6)
14: end if
15: push <*x*, *s*> into *similarSet*
16: end if
17: end for
18: for each *x* in *subSet*
19: if *x* is not in *similarSet*
20: calculate the similarity *s* between *x* and *e* according to *T'*, formula (13) and formula (6)
21: push <*x*, *s*> into *similarSet*
22: end if
23: end for
24: sort *similarSet* in descending order by similarity
25: return *similarSet*

4.3 SIMILARITY BETWEEN QUERIES

In general, a SPARQL query is composed of several triple patterns, so that the similarity between original triple pattern and rewritten triple pattern should be defined first. Suppose that $q^{D1}(s^{D1}, p^{D1}, o^{D1})$ is a triple pattern for the dataset *D1* and $q^{D2}(s^{D2}, p^{D2}, o^{D2})$ is a rewritten triple pattern for the dataset *D2*, and then the similarity between q^{D1} and q^{D2} can be calculated by using formula (14).

$$Sim(q^{D1}, q^{D2}) = \frac{1}{3}(Sim(s^{D1}, s^{D2}) + Sim(p^{D1}, p^{D2}) + Sim(o^{D1}, o^{D2})) \quad (14)$$

Considering Figure 1 and Algorithm SER, *Sim(dbr:Chengdu, yago:Chengdu)* is 0.99, *Sim(geo:lat, yago:hasLatitude)* is 0.77, and then *Sim("dbr:Chengdu geo:lat ?c", "yago:Chengdu yago:hasLatitude ?c")* is 0.92.
Given an original query $Q^{D1}(q_1^{D1}, q_2^{D1}, \cdots, q_n^{D1})$ consists of n triple patterns and $Q^{D2}(q_1^{D2}, q_2^{D2}, \cdots, q_n^{D2})$ is the rewritten query, then the similarity between Q^{D1} and Q^{D2} is

$$Sim(Q^{D1}, Q^{D2}) = \prod_{i=1}^{n} w_i \cdot Sim(q_i^{D1}, q_i^{D2}), \quad (15)$$

where $w_i \in (0,1]$ is the weight of triple pattern q_i^{D1}, which stands for the importance of q_i^{D1} in Q^{D1}, When the value of each w_i is set to 1, the similarity between formula (1) and formula (2) is 0.78.

5 Alignment-based approximate querying

On the basis of result of ontology alignment, alignment-based approximate querying can be implemented by executing the following procedures. At first, the query parser locates the entities in Q^{D1} that should be replaced later, and then puts them into a container for subsequent processing. It is obvious that variables and constants in Q^{D1} do not need to be added. Except for variables and constants, some IRIs that are widely used in most datasets should be skipped as well, such as "rdf:type", "owl:sameAs", and so on. In the second step, similar entities of each entity in the container are obtained by calling the algorithm SER. Then, some rewritten queries for the dataset *D2* can be constructed by replacing original entities with similar entities. During the construction procedure, subject and object should be upside down when the property in dataset D1 is replaced by inverse property. For example, "dbo:creator" and "yago:created" have an inverse relationship. After sorting rewritten queries in descending order by similarity, the query engine executes these rewritten queries on dataset *D2* one by one and records the query answers until the termination criterions are satisfied. In the following algorithm, the concrete termination criterions are not

given, because user can set particular criterions that fit the requirement of specific application. For some applications, it could be a reasonable termination criterion when top k of the most similar queries is executed on dataset *D2*. The details of algorithm for alignment-based approximate querying are described below.

Algorithm AAQ
Input: an query Q^{D1} for the dataset *D1*, a result set of ontology alignment *resultSet*, threshold value of including relation *T*, threshold value of included relation *T'*.
Output: a result set with similarity *queryResult*

1: Initialize *queryResult*
2: Initialize *sContainer*. Here, *sContainer* is a container that includes entities in Q^{D1} and their similar entity in dataset *D2*
3: locate entities in Q^{D1} that should be replaced
4: put these entities into a container, denoted as *eContainer*
5: for each *e* in *eContainer*
6: Initialize *similarSet*
7: *similarSet = SER(e, resultSet, T, T')*
8: put <*e, similarSet* > into *sContainer*
9: end for
10: *rqSet* = RewrittenQuery(Q^{D1}, *sContainer);* Here, the function RewrittenQuery is to construct similar queries; *rqSet* is a set that includes all rewritten query for dataset *D2*
11: sort *rqSet* in descending order by similarity
12: for each *q* in *rqSet*
13: if(terminationSatisfied())
14: break
15: end if
16: *tmpResult = Engine.query(q, D2)*
17: *queryResult.add(tmpResult)*
18: end for
19: return *queryResult*

By analysing the algorithm AAQ, it can be easily concluded that the size of rewritten query set is closely related to two factors--one is the number of entities that should be replaced; the other is the number of similar entities of each entity, which depends on the algorithm SER. Hence, the size of queryResult can be counted with Formula (16).

$$|queryResult| = \prod_{e \in eContainer} |SER(e,resultSet,T,T')| . \quad (16)$$

6 Experiments

To testify the availability of alignment-based approximate SPARQL query, two key datasets in LOD (DBpedia and YAGO) are chosen as testing data sets in this paper. The versions of datasets are the same as the datasets used in [20]. The experiments show that the version of YAGO in [20] has no information about "rdf:type" facts of YAGO, therefore we download taxonomy data from the official website of YAGO and combine them with the above datasets. All data are imported into Virtuoso, which is an efficient RDF triple store and provides SPAQRL query language support. Just as explained in section 2, the results of ontology alignment found by PARIS are employed to evaluate the similarity between entities and rewrite queries. In addition, *T* and *T'* are set to 0.3 and 0.1 respectively because they conform to the results of PARIS.

In experiments, we construct five SPARQL queries based on the schema of DBpedia, which are listed in Table 2. The purpose of these queries is to retrieve all instances of specified type and some attributes about them. For instance, No.1 query is to find population size and council area of settlements in DBpedia; No.2 query is to find latitude and longitude of schools in DBpedia. Table 3 presents some experimental results. Here, these queries for DBpedia produce plenty of similar queries for YAGO, such as No.3 query having 13675 rewritten queries, similarity of which range from 0.51 to 0.43. The main reason why No.3 query derives so many rewritten queries is that 13670 subclasses and 7 super classes of "dbo:Organisation" are found in the results of ontology alignment. It is not necessary to execute all rewritten queries on YAGO because some queries with low similarity may return lots of inaccurate results. In this paper, only top 10 of rewritten queries are executed on YAGO, and the last column of Table 3 shows the number of result, which do not take duplicate results into account. We also submit the original queries to DBpeida and count the number of result set respectively, which are listed in the column of "Result Number of Original Query on DBpedia". By comparing result numbers of original queries and rewritten queries, it can be concluded that rewritten queries can improve the recall to some extent.
The similarity and the result number about top 10 of rewritten queries are listed in Table 4, in which two characteristics should be noticed. The first one is that few queries in the front of list of rewritten queries have distinctive similarity. In the rewritten queries of No.4 query, the similarities of the first 3 rewritten queries are noticeably greater than others. That is because the entities replacing "dbo:GolfPlayer" in these rewritten queries have both including and included relations with "dbo:GolfPlayer" in the result of ontology alignment. "yago:wordnet_golfer_110136959", "yago:wikicategory_American_golfers", "yago:wikicategory_PGA_Tour_golfers" are these entities. Obviously, the rewritten queries with these entities are more likely to achieve greater similarity according to the computational method of similarity introduced in section 4. Second, the rewritten query with the greatest similarity may return nothing despite obtaining many results in most cases. In Table 4, No.3 query is a typical example. By executing the algorithm SER on the result of PARIS, it is found that the

similarity between "dbo:Organisation" and "yago:yagoLegalActor" is greater than the similarity between "yago:wordnet_organization_108008335" and "dbo:Organisation", which results in "yago:yagoLegalActor" is used to replace "dbo:Organisation" in the top 1 of rewritten queries of No.3 query. However, for literal meaning and answer size "yago:wordnet_organization_108008335" is a better substitute than "yago:yagoLegalActor". The same phenomenon happens on No.5 query. The causation of this phenomenon is that the result of ontology alignment is not absolute precision. It means that approximate query is more suitable for the information retrieval based on ontology alignment--because the rewritten query with the "most" similar entity may miss chance to find the information meeting requirements.

TABLE 2 Some SPARQL queries based on the schema of DBpedia

No.	Queries fo DBpedia
1	SELECT ?s ?p ?c WHERE {?s rdf:type dbo:Settlement. ?s dbo:populationTotal ?p. ?s dbo:councilArea ?c.}
2	SELECT ?s ?lat ?long WHERE { ?s rdf:type dbo:School. ?s geo:lat ?lat. ?s geo:long ?long.}
3	SELECT ?o ?c ?m WHERE { ?o rdf:type dbo:Organisation. ?c dbo:creator ?o. ?o dbo:motto ?m. }
4	SELECT ?g ?a WHERE { ?g rdf:type dbo:GolfPlayer. ?g dbo:award ?a. }
5	SELECT ?a ?n ?d WHERE { ?a rdf:type dbo:Artist . ?a dbo:notableWork ?n. ?a dbo:deathDate ?d }

TABLE 3 Some features of original queries and their rewritten query

No.	Number of Rewritten Query	Max of Similarity	Min of Similarity	Result Number of Original Query on DBpedia	Result Number of Top 10 of Rewritten Queries on YAGO
1	7500	0.52	0.41	495	210
2	2506	0.75	0.58	9304	2973
3	13675	0.51	0.43	5	23
4	187	0.7	0.5	754	29
5	8345	0.53	0.42	1310	93

TABLE 4 Experimental details about top 10 of rewritten queries

No	1		2		3		4		5	
Top10RQ.	Similarity	|Result|	Similarity	|Result|	Similarity	|Result|	Similarity	|Result|	Similarity	|Result|
1	0.52	108	0.75	2837	0.51	0	0.70	8	0.53	0
2	0.51	23	0.73	116	0.50	23	0.61	0	0.53	0
3	0.50	31	0.70	1	0.46	0	0.60	4	0.49	1
4	0.49	45	0.63	23	0.46	0	0.54	4	0.45	67
5	0.45	0	0.63	13	0.46	0	0.54	23	0.45	6
6	0.45	0	0.63	1	0.46	0	0.54	0	0.45	0
7	0.45	1	0.63	0	0.46	0	0.54	0	0.45	6
8	0.45	0	0.63	0	0.46	0	0.54	0	0.45	2
9	0.45	2	0.63	5	0.46	0	0.54	0	0.45	1
10	0.45	0	0.63	11	0.46	0	0.54	0	0.45	10

7 Conclusion

Aiming at the problem that various schemas of datasets make it inconvenience to retrieve information from LOD, an approach of approximate SPARQL querying based on ontology alignment is proposed in this paper. On the basis of the formal result of ontology alignment, our approach quantitatively measures the similarity between entities in different conditions, and then the similarity between rewritten queries. Further, our approach can use these rewritten queries to obtain approximate answers from other dataset. The experiments show that alignment-based approximate SPARQL querying can not only retrieve approximate answers, but also overcome the problem caused by imprecise result of ontology alignment, which is very common for the techniques of ontology alignment. In the future work, we will improve alignment-based approximate querying from two aspects: one aspect is to increase the accuracy of ontology alignment between datasets in LOD; the other is to combine alignment-based approximate querying with other query execution paradigm, such as traversal based query execution over LOD [7].

Acknowledgments

This work was partly supported by the Science Guidance Project of Education Department of Hubei Province under Grant No. B2014085, Science Foundation for Yong Teachers of Wuhan University of Science and Technology under Grant No. 2012XZ015.

References

[1] Celino I, Dell'Aglio D, Della E, et al. 2011 Integrating Machine Learning in a Semantic Web Platform for Traffic Forecasting and Routing *Proc. of the 3rd ESWC Workshop on Inductive Reasoning and Machine Learning*

[2] Meymandpour R, Davis J G 2013 Ranking universities using linked open data *Proc. of the 6th Workshop on Linked Data on the Web*

[3] Hartig O, Langegger A 2010 A database perspective on consuming linked data on the web *Datenbank-Spektrum* **10**(2) 57-66

[4] Weiss C, Karras P, Bernstein A 2008 Hexastore: sextuple indexing for semantic web data management *Proc. of the 34th International Conference on Very Large Data Bases* **1**(1) 1008-19

[5] Quilitz B, Leser U 2008 Querying distributed RDF data sources with SPARQL *Springer Berlin Heidelberg* 524-38

[6] Oren E, Delbru R, Catasta M, et al. 2008 Sindice. com: a document-oriented lookup index for open linked data. International *Journal of Metadata, Semantics and Ontologies* **3**(1) 37-52

[7] Hartig O, Freytag J C 2012 Foundations of traversal based query execution over linked data *Proc. of the 23rd ACM conference on Hypertext and social media* 43-52

[8] Bizer C, Lehmann J, Kobilarov G, et al. 2009 DBpedia-A crystallization point for the Web of Data *Web Semantics: Science, Services and Agents on the World Wide Web* **7**(3) 154-65

[9] Hoffart J, Suchanek F M, Berberich K, et al. 2013 YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia *Journal of Artificial Intelligence* **194** 28-61

[10] Joshi A K, Jain P, Hitzler P, et al. 2012 Alignment-based querying of linked open data *Proc. of On the Move to Meaningful Internet Systems* 807-24

[11] Hurtado C A, Poulovassilis A, Wood P T 2006 A Relaxed Approach to RDF Querying *Proc. of the 5th International Semantic Web Conference* 314–28

[12] Hai Huang, Chengfei Liu, Xiaofang Zhou 2008 Computing Relaxed Answers on RDF Databases *Proc. of Workshop on Web Information Systems Engineering* 163-75

[13] Hai Huang, Chengfei Liu, Xiaofang Zhou Approximating query answering on RDF databases *World Wide Web* **15** 89–114

[14] Reddy B R K, Kumar P S 2010 Efficient approximate SPARQL querying of Web of Linked Data *Proc. of 6th international workshop on Uncertainty Reasoning for the Semantic Web* 37-48

[15] Reddy K B R, Kumar P S 2013 Efficient trust-based approximate sparql querying of the web of linked data *Proc. of 9th international workshop on Uncertainty Reasoning for the Semantic Web* 315-30

[16] Euzenat J, Shvaiko P 2007 Ontology matching *Springer Heidelberg*

[17] David J, Euzenat J, Scharffe F, et al. 2011 The alignment api 4.0. *Journal of Semantic Web* **2**(1) 3-10

[18] Jain P, Hitzler P, Sheth A P, et al. 2010 Ontology alignment for linked open data *Proc. of The 9th International Semantic Web Conference* 402-17

[19] Shvaiko P, Euzenat J 2013 Ontology matching: state of the art and future challenges *IEEE Transactions on Knowledge and Data Engineering* **25**(1) 158-76

[20] Suchanek F M, Abiteboul S, Senellart P 2011 Paris: Probabilistic alignment of relations, instances, and schema *Proc. of the 37th International Conference on Very Large Data Bases* **5**(3) 157-68

Authors

Yu Liu, born on January 20, 1980, China

Current position, grades: PhD candidate in School of Computer Science and Technology of Wuhan University.
University studies: master degree in School of Computer Science and Technology of Huazhong University of Science and Technology.
Scientific interests: semantic web and knowledge engineering.

Lei Chen, born on November 21, 1978, China

Current position, grades: instructor in School of Computer Science and Technology of Wuhan University.
University studies: master degree in National University of Singapore and doctoral degree in Huazhong University of Science and Technology.
Scientific interests: knowledge engineering and complex system.

Shihong Chen, born on May 12, 1949, China

Current position, grades: professor in School of Computer Science and Technology of Wuhan University. He also is the deputy director of the national engineering research center for multimedia software.
Scientific interests: software engineering and knowledge engineering.

A fractal de-noising algorithm based on least absolute deviation method

Hui Guo*, Jie He

School of Information and Electronic Engineering, Wuzhou University, Wuzhou 543002, Guangxi, China

Received 1 June 2014, www.cmnt.lv

Abstract

Against the shortcoming that the traditional method of fractal image compression coding has inferior decoding quality on the original image subject to salt-and-pepper noise interference, this paper raises a least absolute deviation (LAD) method to be applied in fractal image compression, which can replace the method of least square error in computing contrast and brightness adjustment value and solve the L1-norm recursive problem using weighted median. The experimental result indicates that the LAD method has a very good anti-noise effect on the outliers introduced by salt-and-pepper noise.

Keywords: Fractal Image Compression, Least Square Error Method, Least Absolute Deviation Method, Salt-and-Pepper Noise

1 Introduction

Put forward by Mandelbrot in 1975 [1], the fractal theory was first applied in image compression by Barnsley in 1988 [2]. It was not until Jacquin raised iterated function system (IFS) and local iterated function system (LIFS) [3] that the theoretical foundation of fractal image compression technique has been laid. Afterwards, this theory was used in image retrieval [4], digital watermarking [5], image inpainting [6] and image denoising [7], etc., respectively. The algorithm raised by Jacquin is based on the concept of local self-similarity of image, in which the local blocks can identify another similar block through three adjustment methods - contrast, brightness and reversal. The LIFS can automatically convert images into affine transform coefficients, which can achieve the aim of compression only by being stored up. Fractal image compression enjoys advantages like high compression ratio, irrelevance of resolution ratio, fast decoding and more, whereas the disadvantage is the overlong compressing hour as result of using the global searching method to seek for the optimal matching blocks. Therefore, in order to speed up the encoding and strengthen the image detail compensation, some scholars combine fractal with other algorithms. He Jia, zheng-kai liu [8] used low frequency coefficients in DCT as the matching feature, and made matching more rapid and accurate.

The LAD method was put forward by Boscovich in 1757, namely the L1-norm, whose aim is to minimize the absolute deviation through estimative parameters, though the weakness of non-differentiability left it unused for complicated computation. Many parallel methods, such as the Barrodale-Roberts [9] Algorithm, the Bartels-Conn-Sinclair Algorithm and the Maximum Likelihood Estimation Algorithm [10] did not rise until Edgeworth put forward resolving the problem of non-differentiability using weighted medians [11] in 1887 and Harris solved the LAD method using the simple concept of linear programming in 1950 [12].

In traditional fractal image compression algorithms, the least square error (LSE) method is adopted to determine brightness and contrast adjustment coefficients by contrasting the range blocks to the domain blocks. It can be known from the recursive principle that the LSE method is of no robustness, so the restored image has inferior quality upon being interfered by noise after undergoing compression. Replacing the traditional LSE method with the LAD method, we take the advantage of robustness with the LAD method to remove noise from the fractal images straightforward during the compression procedure without undergoing the pre-processing prior to noise-removal so that the quality of image is not affected by noise.

2 Theory evidence

2.1 LEAST SQUARE ERROR (LSE) METHOD

Fractal encoding performs compressed encoding by exploiting the characteristic of image's local self-similarity. Each block within an image may correspond to a big similar block. Taking the two groups of blocks in Figure 1 as an example, one group is the hat edge in correspondence to the hat in the mirror, the other being the similar but larger block in correspondence to the part of shoulder. The larger block approximates the smaller counterpart within the same group after being minimized and adjusted in contrast and brightness.

* ***Corresponding author*** e-mail: 13066724@qq.com

A 256-by-256 image can fall into non-overlapped n-by-n range blocks, the set of which is a range pool. Then the original image is divided into 2n-by-2n domain blocks that can be overlapped, the set of which is called a domain pool. The procedure of fractal image compression is to pinpoint the most similar domain block from the domain pool for each range block in the range pool. This procedure can be viewed as finding a domain block through convergent operations of affine transform.

FIGURE 1 Local Self-similarity of Fractal Image Compression

In the contrasting procedure, the square error is used to estimate the difference between the range block v and the domain block u after the sub-sampling. Smaller is indicative of higher similarity between the domain block u and the range block v, as shown in Equation (1).

$$E_k = \left\| (p_k u_k + q_k) - v \right\|^2, \; k=0,1,...,7, \tag{1}$$

where k denotes 8 directions of reversal; E_k, p_k, u_k and q_k represent the square error value, the contrast adjustment value, the block after the secondary sampling, and the brightness adjustment value, respectively, of the difference degree of the estimative block under the k^{th} reversal. $\hat{p}_k$ and $\hat{q}_k$ are the predicted values of p_k and q_k, respectively, which can derive from Equation(1)through partial differential. This method to compute p and q is just the LSE method, N in the following Equation being the size of the range block.

$$\hat{p}_k = \frac{N^2 < u_k, v > -(\sum_{j=0}^{N-1}\sum_{i=0}^{N-1} u_k(i,j)) \cdot (\sum_{j=0}^{N-1}\sum_{i=0}^{N-1} v(i,j))}{N^2 < u_k, u_k > -(\sum_{j=0}^{N-1}\sum_{i=0}^{N-1} u_k(i,j))^2}, \tag{2}$$

$$\hat{q}_k = \frac{1}{N^2}(\sum_{j=0}^{N-1}\sum_{i=0}^{N-1} v(i,j) - \hat{p}_k \sum_{j=0}^{N-1}\sum_{i=0}^{N-1} u_k(i,j)). \tag{3}$$

LAD Method

The LAD method is interpreted against the concept of linear regression, a model of which is assumed as Equation (4).

$$y_i = ax_i + b + \varepsilon_i, \; i=1.....L, \tag{4}$$

where x is the input value, y is the output value, and ε is the error term. When ε_i meets the following assumptions:

(a) $E(\varepsilon_i) = 0$;

(b) The variance number is constant;

(c) ε_i is a Gaussian distribution function whose formula of Gaussian distribution follows Equation(5), where the mean μ = 0 and the variance is σ;

(d) Each follows independent distribution.

When ε_i meets the above assumption, it is the optimal choice to solve Equation (4) using the LS method. When ε_i is assumed as a Laplace distribution function whose formula of Laplace distribution follows Equation (6) where the mean μ = 0 and the variance $\sigma = \frac{\sqrt{2}}{\lambda}$, and when ε_i meets the above assumptions (a), (b) and (d) with a Laplace distribution, it is the optimal choice to solve Equation (4) using the LAD method. Different methods are employed for the optimal solution according to different assumptions of ε_i.

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x)^2}{2\sigma^2}), -\infty < x < \infty, \tag{5}$$

$$f(x) = \frac{\lambda}{2} \exp(-\lambda|x|), -\infty < x < \infty, \tag{6}$$

The LAD method and the traditional LSE method are applied to Equation 4 respectively to produce Equation (7) and Equation (8).

$$E = \sum_{i=1}^{L} |y_i - (ax_i + b)| = \sum_{i=1}^{L} \mathrm{sgn}(e_i) \cdot e_i, \tag{7}$$

$$E = \sum_{i=1}^{L} (y_i - (ax_i + b))^2 = \sum_{i=1}^{L} e_i \cdot e_i. \tag{8}$$

In the above Equations, E is the sum of the parameters in group L, e_i is the error value of the parameters in the i[th] group. For the same group of parameters, e_i has identical value. Between both methods mentioned above, the difference lies in the fact that the LAD method takes the absolute value of e_i. When e_i is positive sgn (e_i) becomes 1, when e_i is negative sgn (e_i) becomes -1. While in the LSE method, the error value E will become greater after the value of e_i is squared. Under the circumstance where the LAD method is used, there is no significant difference with e_i between the interval [-1, 1] from the LSE method; but when outliers emerge to the data, e_i will be magnified at a square rate when the LSE method is employed, which leads the error E to become too great to make good judgment and estimation, so the value of e_i shall be kept from being magnified fast to reduce the effect of a few outliers on all the data, which is the robustness in inhibiting outliers, as shown in Figure 2.

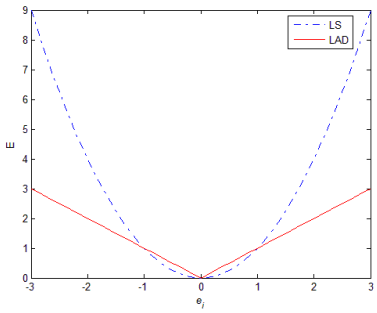


FIGURE 2 Relations between LAD, LSE and error E

2.2 MAXIMUM LIKELIHOOD ESTIMATES (MLE)

The LAD method cannot perform differential since absolute value exists, so the parameters with least absolute deviation value need to be determined through other methods. What is adopted in this paper is the maximum likelihood estimation algorithm [10], in which the weighted medians are operated. The weighted median starts from the viewpoint of geometric slope. Where the slope is greater than 0, the deviation value is minimum. It can be inferred as follows that the slopes after being sorted are on a progressive increase, turning from negative to positive as increasing to a certain value, which is the approximate minimum rising when the least deviation method is used. The proof goes as below:

$$F(a) := \sum_{i=1}^{L} |y_i - b - ax_i| = \sum_{i=1}^{L} |x| \left| \frac{y_i - b}{x_i} - a \right| = \sum_{i \in I\_(a)} |x_i| (a - \frac{y_i - b}{x_i}) + \sum_{i \in I_+(a)} |x_i| (\frac{y_i - b}{x_i} - a) \quad (9)$$

F(a) is the absolute value taken from Equation (4); *x* and *y* are the input and output, respectively, each holding the parameters in *L* Group, where:

$$I\_(a) := \left\{ i \in \{1.....L\} : \frac{y_i - b}{x_i} \le a \right\}$$

$$I_+(a) := \left\{ i \in \{1.....L\} : \frac{y_i - b}{x_i} > a \right\}$$

I_(a) is a set of parameters in *L* Group whose slopes are smaller than *a*; $I_+(a)$ is a set of parameters in *L* Group whose slopes are greater than *a*. Partial differential is applied to *a* in Equation (9) to get the slope as Equation (10). The minimum deviation value is determined by the LSE method when the slopes are 0. Since the LAD method is linear and continuous, it is taken that $F'(a) \ge 0$. While the parameter in correspondence to the slope turning from negative to positive reaches the minimum.

$$F'(a) := \sum_{i \in I\_(a)} |x_i| - \sum_{i \in I(a)_+} |x_i| \ge 0, \quad \sum_{i \in I\_(a)} |x_i| \ge \sum_{i \in I_+(a)} |x_i| \quad (10)$$

$$2 \sum_{i \in I\_(a)} |x_i| = \sum_{i \in I\_(a)} |x_i| + \sum_{i \in I\_(a)} |x_i| \ge \sum_{i \in I\_(a)} |x_i| + \sum_{i \in I_+(a)} |x_i| = \sum_{i=1}^{L} |x_i|$$

$$\sum_{i \in I\_(a)} |x_i| \ge \frac{1}{2} \sum_{i=1}^{L} |x_i| . \quad (11)$$

According to Equation (10), when 5 groups of parameters are input, the change of their slopes is shown as Figure3. The slopes are sorted as 1, 2, ..., 5 in an incremental order, whereas the slopes are $\frac{y_{[i]} - b}{x_{[i]}}$, and the corresponding parameters can be yielded. At Interval (3) $F'(a)$ is negative while becoming positive at Interval (4), so Interval (4) is the first group of data $x_{[4]}$ when $F'(a) \ge 0$. And Figure 3 indicates that the minimum should be among $x_{[3]}$. It can be known from Equation (11) that the input data are within the set of *I_(a)*. As the data $x_{[1]}$ input at the first time are accumulated to the data $x_{[i]}$ input at the ith time and when the accumulated sum is greater than one half of the sum of all input data, the input data at the ith time are the input ones of the first positive slope when the slope just turns from negative to positive. From Equation (11) it can be found the first datum where $F'(a) \ge 0$ is the one input at the ith time, so the minimum deviation value may occur when the datum is input at the (i - 1)th time.

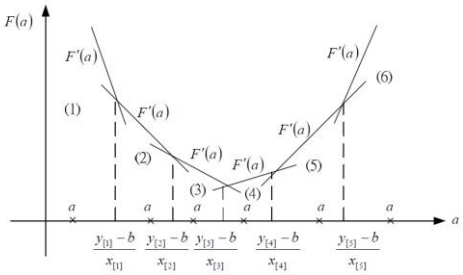


FIGURE 3 Conceptual Schematic Diagram of Weighted Median

Figure 3 reveals the minimum deviation value may occur among the data in the third group. It can be concluded from Equation (11) that, When the input data *L* group, the minimum deviation value may be among the data input at the (i - 1)th time. The weighted medians compute the slope of the data with the least deviation value. They are denoted by *MED*(•) as below:

$$a = MED(x_i \Diamond \frac{y_i - b}{x_i} |_{i=1}^{L}) . \quad (12)$$

The steps to compute the weighted medians are as follow:

Step I: Assume $w_i = \frac{y_i - b}{x_i}$, record the slope input each time.

Step II: Compute the threshold value $x' = (1/2)\sum_{i=1}^{L} x_i$.

Step III: Express all the slopes $w_1 ..... w_L$ as $w_{[1]} \le ..... \le w_{[L]}$ after sorting them; in correspondence to $w_{[i]}$ after being sorted, the relevant input data is $x_{[i]}$ as the weight.

Step IV: Apply summation to $x_{[i]}$ one after another; record the index value j that first meets $\sum_{i=1}^{j} x_{[i]} \ge x_0$, then the index value of the minimum is j - 1.

Step V: $w_{[j-1]}$ is the slope in correspondence to the minimum; the computation for the weighted medians is just done by letting $a = w_{[j-1]}$.

The above is introduction to the concept of and steps for weighted median, whereas the following provides illustration on how MLE figures out the slope and the gap by means of the weighted medians. In order to accelerate convergence after using the weighted medians, MLE transforms the space to finely adjust the slope, thereby to get more accurate slope, reduce the times of training, and get the slope with the least error and the corresponding gap through fast convergence.

The MLE calculation method 4 offers to pinpoint the parameters a and b in correspondence to the minimum E in Equation (7).

The steps are as follow:

Step I: Set the times of recursion as k = 0 at the very beginning; $b = b_k$ is yielded through calculation using the LSE method, as shown in Equation (13):

$$b_0 = \frac{\sum_{i=1}^{L}(x_i - \bar{x}) \times (\bar{y}x_i - \bar{x}y_i)}{\sum_{i=1}^{L}(x_i - \bar{x})^2}, \quad (13)$$

$$\bar{x} = (1/L) \times \sum_{i=1}^{L} x_i \ , \ \bar{y} = (1/L) \times \sum_{i=1}^{L} y_i$$

Step II: Compute the parameter a_0 using the weighted medians, expressed as Equation (14):

$$a_0 = MED(|x_i| \Diamond \frac{y_i - b_0}{x_i}) \Big|_{i=1}^{L}). \quad (14)$$

Determine the index value h for existence of least deviation value. The parameter a_0 can be yielded from Equation (14), accordingly update $a_{old} = a_0$.

Step III: Let $k = k + 1$ to convert the input spatial coordinate $z_i = x_i - x_h$.

Step IV: Compute $a'_{k-1} = a_{k-1}$, $b'_{k-1} = b_{k-1} + a_{k-1}x_h$ and the weighted median a'_k as Equation (15):

$$a'_k = MED(|z_i| \Diamond \frac{y_i - b'_k}{z_i}) \Big|_{i=1}^{L}). \quad (15)$$

Compute from Equation (15) the new index value m and the corresponding parameter a'_k.

Step V: Compute $a_k = a'_k$ and $b_k = b'_k - a'_k x_h$ by converting the original spatial coordinate.

Step VI: Let $h = m$. Stop where the variances of a_{old} and a_k are below the permitted value or where the time of recursion exceeds the set value; otherwise update $a_{old} = a_0$ and return to Step III. The above steps can be expressed by the flowchart as Figure 4.

3 Application of LAD in fractal image compression

The workflow of introduction of LAD in fractal image compression bears general similarity to traditional methods of fractal compression, the only difference being how to compute p and q and utilize the absolute deviation as an indicator of estimation, as shown in Figure 5.The main steps by which to replace LS with LAD are as follow:

Step I: Fall the original image into an 8-by-8 non-overlapped range block v and a 16-by-16 overlapped domain block u.

Step II: Apply secondary sampling to u and perform reversal at k.

Step III: The iteration time t = 0. Compute the initial $q_k = \hat{q}_k$

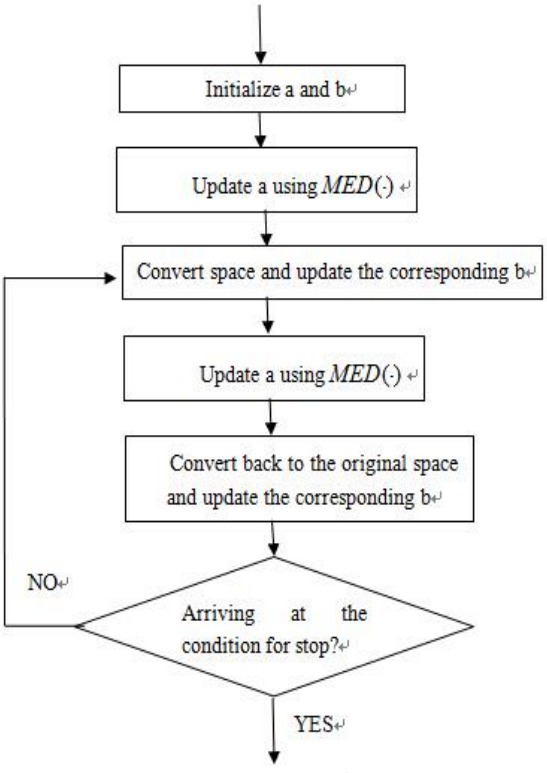


FIGURE 4 Flowchart of LAD

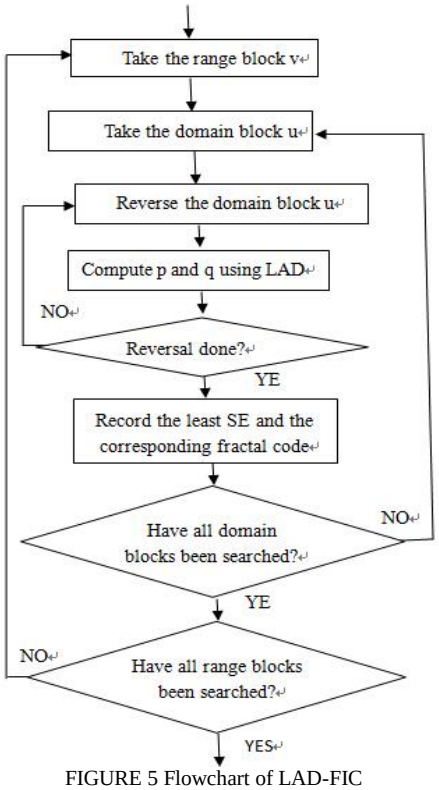


FIGURE 5 Flowchart of LAD-FIC

Step IV: Figure out pk using MED(•), expressed as Equation (16):

$$p_k = MED(|u_k(i,j)| \Diamond \frac{v(i,j)-q_k}{u_k(i,j)})\Big|_{i,j=0}^{N-1}), \qquad (16)$$

Where N is the size of the range block v. According to the above Equation, the least deviation value within the 8-by-8 $u_k(i, j)$ is $(h1, h2)$, whereby to update $p_{old} = p_k$.

Step V: Let $t = t + 1$. Convert the input space $z(i,j) = u_k(i,j) - u_k(h1,h2)$.

Step VI: $q_k = q_k + p_k u_k(h1,h2)$ and p_k after space conversion are shown as Equation (17):

$$p_k = MED(|z(i,j)| \Diamond \frac{y(i,j)-q_k}{z(i,j)}\Big|_{i,j=0}^{N-1}). \qquad (17)$$

According to the above Equation, the least deviation value within $z(i, j)$ is $(m1, m2)$.

Step VII: Convert to the original space, where $q_k = q_k + p_k u_k(h1,h2)$.

Step VIII: Let $h1 = m1$, $h2 = m2$. Stop where the variances of p_{old} and p_k are below the permitted value or where the time of iteration exceeds the set value; otherwise update $p_{old} = p_k$ and return to Step IV. LAD offers to compute p_k and q_k from Step IV through Step VIII.

Step IX: Return to Step II if none of the eight directions of reversal has been used.

Step X: Record the fractal code in correspondence to the least absolute deviation in Equation (15).

$\min(E_K = |(p_k u_k + q_k) - v|)$, $k=0,1,...7$

Step XI: Apply global search to the domain pool to pinpoint the domain block u in correspondence to each range block v. Thus the LAD-FIC computation is done.

4 Experimental result

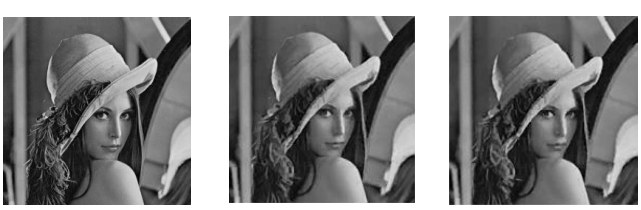
The tool for experiment in this paper is Visual C++ 6.0, the operating system is Microsoft Windows XP, the CPU is Intel Core i5 3450, and the internal memory is 4G. In this section, a comparison will be made between the traditional method of fractal compression and the LAD-introduced method of fractal image compression for multiple frames of images where different noises are included. The images in use are all 256-by-256, the size of range block being 8-by-8. The condition under which MLE stops is that the time of iteration reaches 5 or that the difference between the yielded p value and the previous p value is smaller than 0.05. PSNR, which is used to evaluate the quality of the decoded image, is defined as Equation (18), whereas MSE is defined as Equation (19).

$$PSNR(f, \hat{f}) := 10 \cdot \log_{10}\left(\frac{255^2}{MSE(f, \hat{f})}\right), \qquad (18)$$

$$MSE(f, \hat{f}) = \frac{1}{256^2} \cdot \sum_{i=1}^{256}\sum_{j=1}^{256}(f(i,j) - \hat{f}(i,j))^2, \qquad (19)$$

where f is the original image, $\hat{f}$ is the decoded image.

Initially, the Lena image in which no noise has been included offers as the image to test. The traditional fractal image compression (FIC) method and the fractal image compression method where LS gives way to LAD (LAD-FIC) are used separately to observe the effects of compression by both, as shown in Figure 6. Baboon is the initial image to decode. Through 9 times of iteration, Figure 6(b) is the image decoded by FIC, whose PSNR value is 28.86dB. Figure 6(c) is the image decoded by LAD-FIC, whose PSNR value is 28.27dB. The images decoded by both methods have almost the same PSNR, wherefrom both LAD-FIC and FIC methods can be verified to make the same effect in terms of compression.



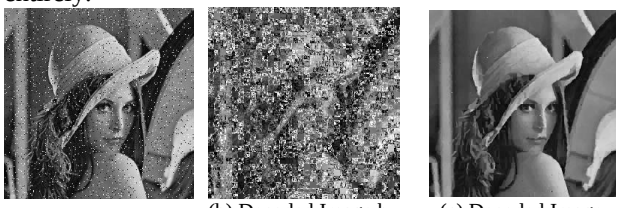
(a) Original Image Lena

(b) Decoded Image by FIC, PSNR = 28.86dB

(c) Decoded Image by LAD-FIC, PSNR = 28.27dB

FIGURE 6 Contrasted Effects of Lena Image Tested by FIC and LAD-FIC

Next, it is attempted to include 5% and 10% salt-and-pepper noises in the four images (Lena, Baboon, Pepper and F16). After being compressed by both methods, all the decoded images are based on Baboon as the initial image but Baboon itself, which is based on Lena as the initial image. The decoded images are produced after 9 times of iteration. Figure 7(a) is the image by including the 5% salt-and-pepper noise in Lena. Figure 7(b) and Figure 7(c) are the decoded images after FIC and LAD-FIC are used for compression. Obviously, it can be seen that Figure 7(b) has been vulnerable to noise in that its decoded image fails to restore Lena but presents only a few lines across big blocks which cannot clearly distinguish the original image. The difference between the PSNR values of Figure 7(c) and Figure 6(c) is only 1dB, wherefrom it can be seen that LAD-FIC can almost remove the effect of the 5% salt-and-pepper noise entirely.



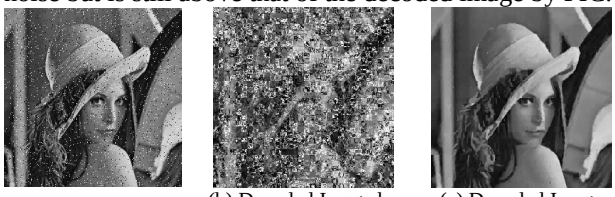
(a) 5% Salt-and-Pepper Noise

(b) Decoded Image by FIC, PSNR = 11.87dB

(c) Decoded Image by LAD-FIC, PSNR = 27.21dB

FIGURE 7 Lena Image for Testing 5% Salt-and-Pepper Noise

Next, 10% salt-and-pepper noise is then included in the test. Figure 8(a) is the image by including the 10% salt-and-pepper noise in Lena. Figure 8(b) is the decoded image by FIC whose PSNR value is 11.41dB, with only a bit more inferior effect than 7(b) though the quality of image has been so poor as to obscure the original image. Figure 8(c) is the decoded image by LAD-FIC whose PSNR value is 26.42dB, which falls below the counterpart of Figure 7(c) for the image's vulnerability to noise but is still above that of the decoded image by FIC.



(a) 10% Salt-and-Pepper

(b) Decoded Image by FIC, PSNR = 11.41dB

(c) Decoded Image by LAD-FIC, PSNR = 26.42dB

FIGURE 8 Lena Image for Testing 10% Salt-and-Pepper Noise

The test is conducted on the three other images by the same method. To illustrate it with Table 1, the parameter *S* in the table denotes the salt-and-pepper noise. The experiment demonstrates that the PSNR value of the decoded image produced through fractal compression on the image in which the salt-and-pepper noise has been included has decline to a considerable degree, whereas LAD-FIC remains not affected by this noise, therefore LAD-FIC has the effect of resisting the salt-and-pepper noise

TABLE 1 the PSNR (unit: dB) Values of Pepper, Baboon and F16 under 5%, 10% and 20% Salt-and-Pepper Noises

	Pepper		**Baboon**		**F16**	
	FIC	*LAD-FIC*	*FIC*	*LAD-FIC*	*FIC*	*LAD-FIC*
Original Image	29.78	29.31	20.03	19.82	25.17	24.43
S=5%	17.89	27.88	15.43	19.26	12.83	23.71
S=10%	17.96	27.21	17.56	20.12	5.53	23.23
S=20%	18.31	25.92	17.41	18.33	6.24	22.67

5 Conclusion

Traditional techniques of fractal image compression fail to discuss how to resist noise, so the quality of compression will be vulnerable to noise when the original image is interfered by noise. The LAD method of robustness used in substitution for the traditional LSE method endows the image with the capacity of resisting the salt-and-pepper noise, namely the LAD-FIC solution proposed in this paper can remove noise while compressing the image. This method makes extremely excellent effect on images vulnerable to the salt-and-pepper noise, yet there is no distinct effect for Gaussian noise and Laplace noise. Since LAD-FIC first uses FIC to estimate the rough contrast and brightness adjustment value prior to further adjustment to figure out the accurate contrast and brightness adjustment value, it is discovered through a comparison between both that the time cost for LAD-FIC is 18 times that for FIC. In the research direction in the future, it is hoped that the compressing time can be further reduced, or the contrast and brightness adjustment value can be changed to get their linear relationship, on the premise of guaranteeing its robustness, so as to achieve a higher quality of compression.

Acknowledgments

This paper is financially supported by Guangxi Natural Science Foundation Program (Grant No. 2013GXNSFBA019275, Grant No. 2013GXNSFBA019276) and Guangxi University of Science and Technology Research Program (Grant No.2013YB227, Grant No.2013YB228).

References

[1] Mandelbort B 1982 *The fractal geometry of nature* Freeman: San Francisco, California

[2] Barnsley M F 1992 *Fractals everywhere* Academic Press: New York

[3] Jacquin A E 1992 Image coding based on a fractal theory of iterated contractive image transformation *IEEE Transactions on Image Processing* **1**(1) 18-30

[4] Wang X Y A 2009 Aast fractal coding in application of image retrieval *Fractals* **17**(04) 441-50

[5] Tirkel A Z, Osbome C F, Hall T E 1998 Image and watermark registration *Signal Processing* **66**(03) 373-83
[6] Li Jin jiang, Zhang Caiming 2010 Image inpainting algorithm based on fractal theory *Acta Electronica Sinica* **38**(10) 2430-5
[7] Ghazel M, Freeman G H, Vrscay E R 2006 Fractal-Wavelet image denoising revisited *IEEE Transactions on Image Processing* **15**(09) 2669-75
[8] Hy Jia, Luır Zhengkai 2010 A fast fractal image coding method based on the DCT transform *Acta Electronica Sinica* **29**(6) 748-50
[9] Barrodale I, Roberts F D K 1973 An improved algorithm for discrete 1 L linear approximation *SIAM Journal on Numerical Analysis*, **10**(5) 839-48
[10] Li Y B, Arce G R 2004 A maximum likelihood approach to least absolute deviation regression *EURASIP Journal on Applied Signal Processing* **20**(04) 1762-9
[11] Edgeworth F Y 1887 A new method of reducing observations relating to several quantities *Philosophical Magazine* **24** 222-3
[12] Harris T E 1950 Regression using minimum absolute deviations *The American Statistician* **4**(1) 14-5

Authors

Hui Guo, born on September 28, 1981, China

Current position, grades: Associate Professor at Wuzhou University, China.
University studies: master degree in college of electronic engineering from Guangxi Normal University, China in 2008.
Scientific interests: image compression and image representation.

Jie He, born on February 13, 1982, China

Current position, grades: Associate Professor at Wuzhou University, China.
University studies: master degree in College of Computer Science & Information Technology from Guangxi Normal University, China in 2008.
Scientific interests: image compression and image representation.

Single threshold segmentation for noisy image based on fuzzy ant colony algorithm

Ye Chen[1*], Xiaoqun Qin[1], Xinmin Zhou[2]

[1]*School of Information Science and Engineering, Hunan International Economics University, Changsha 410205, Hunan, China*

[2]*School of Computer Science and Information Engineering, Hunan University of Commerce, Changsha 410205, Hunan, China*

Received 1 July 2014, www.cmnt.lv

Abstract

Firstly, this paper pre-processes the image to be segmented through grey-scale morphological method. Then, based on the in-depth analysis of basic ant colony algorithm, it explains the shortcomings of this algorithm; proposes the improved strategy of ant colony algorithm, namely fuzzy ant colony algorithm, which designs the fitness function of artificial ant colony algorithm with minimum cross entropy and applies the improved fuzzy ant colony algorithm in the spatial-domain noisy image single segmentation. Finally, starting from the segmentation results and convergence, it compares the performances of the improved ant colony algorithm and the basic ant colony algorithm, GA algorithm and AFS algorithm.

Keywords: Noisy Image, Ant Colony Algorithm, Histogram Feature, Threshold Segmentation, Cross Entropy

1 Introduction

This paper is aimed to study the swarm intelligent optimization algorithm based on ant foraging behaviour, namely the artificial ant colony algorithm. The early ant colony algorithm is mostly used in solving function optimization problems. For example, some scholars see the honey of the artificial ant colony algorithm as the possible solution of the function and achieve satisfactory results in optimizing function with ant colony algorithm. Other scholars consider the objective function in the engineering as virtual food and settle the optimization problems of two-dimensional functions with virtual ant colony algorithm. Later, starting from the optimization problems of high-dimensional functions, some scholars demonstrate the excellent performances of artificial ant colony algorithm by comparing particle swarm optimization and other algorithms before they use ant colony algorithm to solve the data clustering problems.

As a new swarm intelligent model, ant colony algorithm still encounters the following problems used in image segmentation. 1)The initialized population is randomly generated in the solution space without considering the characteristics of the image to be segmented; 2)In the threshold search of the basic ant colony algorithm, the individual movement in the population can be seen as a random process in a certain range; however, the actual individual movement has certain fuzziness[1].

Based on the above analysis, this paper improves the optimization efficiency of the intelligent model by integrating the histogram information of image to be segmented, the fuzzy theory and the artificial ant colony algorithm. On the basis of the improved ant colony algorithm, the design of noisy image segmentation is intended to analyse the characteristics of the image to be segmented and suppress the image noise through open operation and closed operation in the grey-scale morphology. Then design the fitness function of the artificial ant colony algorithm with the minimum cross entropy and narrow down the threshold search range of the ant colony by using the histogram information of the image in order to guarantee the image segmentation speed; consider the ant movement process as the fuzziness process and conduct fuzzy control with fuzzy theory. Finally, quickly find the optimal threshold with the improved artificial ant colony algorithm [2].

2 The proposed algorithm

With the advancement in networking and multimedia technologies enables the distribution and sharing of multimedia content widely. In the meantime, piracy becomes increasingly rampant as the customers can easily duplicate and redistribute the received multimedia content to a large audience.

2.1 THE IMAGE PRE-PROCESSING OF THE HISTOGRAM ANALYSIS BASED ON GREY-SCALE MORPHOLOGY

It is indicated in the documents that the grey-scale morphological pre-processing can effectively suppress the speckle noise in the noisy images. Therefore, in the noisy image segmentation, we first pre-process the image to be segmented with grey-scale morphology to reduce the noise in the image and then narrow down the

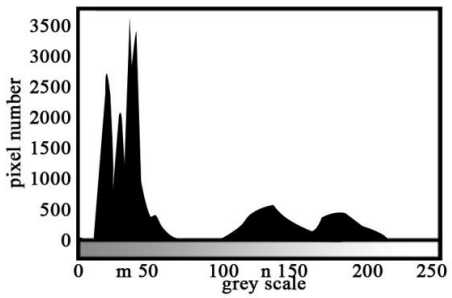
[*] ***Corresponding author*** e-mail: 153569367@qq.com

threshold distribution range according to the grey-scale histogram of the pre-processed image.

Take Fig.1 as example. Fig. 1(a) is the original noisy image and Fig.1(b) is the grey-scale histogram after pre-processing (a) with grey-scale morphology. Assume the possible threshold of the image is T, there are obvious differences in the grey scale between the segmentation target and the background, namely that there are two obvious twin peaks in the histogram of the pre-processing image. The threshold is located in the bottom of the twin peaks and its distribution range is between the twin peaks, namely $T \in [m,n]$ in Fig. 1(b).

The solving method of m and n is to calculate the average grey scale of the image in the histogram. Missing the grey scale among [0, c] with the biggest possibility and n is the grey scale among [c+1, 255] with the biggest possibility.

(a) Noisy image



(b) Histogram of noisy image

FIGURE 1 Histogram of noisy image

3 The fuzzy ant colony algorithm

3.1 THE PROBLEMS IN ANT COLONY ALGORITHM

There are still some problems when ant colony algorithm, a new swarm intelligent model is used in image segmentation, including:

1) The initialized population is generated in the solution space without considering the characteristics of the image to be segmented;

2) In the threshold search process of the basic ant colony algorithm, the individual movement in the colony can be seen as a random process in a certain range while the actual movement has certain fuzziness [3].

3.2 FUZZY CONTROL THEORY AND FUZZY MEMBERSHIP FUNCTION

Fuzzy control theory has been used increasingly extensive because no precise mathematical model of the controlled object is needed to be built when designing system. The existing Methods determine fuzzy membership function include:

1) Fuzzy statistical method: make a clear judgment whether a determined element belongs to a variable clear collection in the domain of discourse;

2) Exemplification: estimate the membership function of the fuzzy subset according to membership frequency of limited fuzzy subsets;

3) Expert experience method: determine the membership function according to the processing equation or the corresponding weight coefficient of the fuzzy information given by the actual expert experience;

4) Binary comparison ordering method: determine the general shape and order under certain characteristics through the paired comparisons of multiply objects so as to determine the membership function of the characteristics and these objects [4].

The reconnaissance ant in the basic ant colony algorithm generates a new position randomly in the search space to improve food source. Such new position is so uncertain that it can be seen as a fuzzy process.

In the below text, we determine the fuzzy membership function by using the exemplification in the fuzzy control so as to control the search range of the reconnaissance ant in the artificial ant colony algorithm. After repeated experimental observation, the ant colony has better optimization effects to fulfil this process according to Formula (1) and Formula (2) is the fuzzy control factor.

$$v_{ij} = x_{ij} + \varphi \times (\omega, \theta, t) \times x_{ij}, \quad (1)$$

$$\mu(\omega, \theta, t) = 1/[1 + \exp(-\omega t - \theta)]. \quad (2)$$

In the above formula, v_{ij} is a new possible solution generated by the ant; x_{ij} is the ant which needs improvements; i is the i^{th} ant; j is the number of dimension of the solution space; φ is a random number among [-1,1]; $\mu(\omega, \theta, t)$ is the fuzzy membership function and ω and θ are a set of values from the experiments. In this paper, $\omega = -0.2$, $\theta = 2$ and t is the current iterations [5, 6].

According to Formula (1), with the increase of iterations, the search range of the ant is self-adaptive to the changes. To be specific, in the initial search, the ant is far from the optimal solution and the search range is big, guaranteeing the diversity of the individuals in the colony ad in the later search, the ant is closed to the optimal solution and its search range is small, avoiding oversized step size and skipping the optimal solution [7].

4 The spatial-domain noisy image single threshold segmentation based on fuzzy colony algorithm

The general noisy image has some speckle noise due to other interferences. Therefore, effectively suppress the noise interference before segmenting the image and get the precise segmentation image. This paper pre-processes the image to be segmented with grey-scale morphology and reduce the noise interference. Then, it narrows down the search range of the threshold with the pre-processed image histogram and enhances the optimization speed. On this basis, it takes the two-dimensional cross entropy as the fitness function of the ant colony algorithm and finds the optimal segmentation threshold of the image through fuzzy ant colony algorithm.

In conclusion, the spatial-domain noisy image segmentation process based on fuzzy ant colony algorithm is classified as follows.

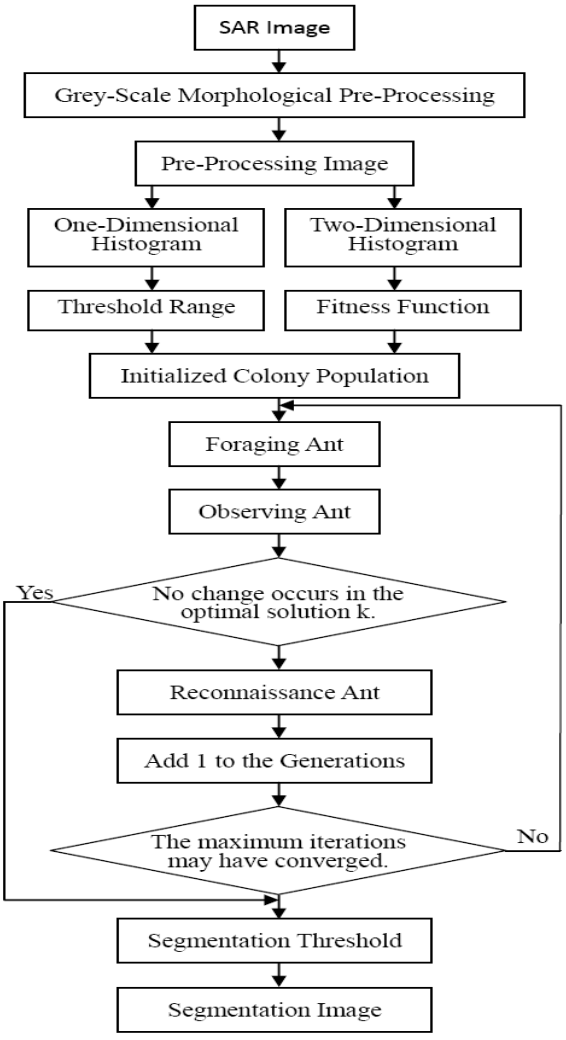


FIGURE 2 The spatial-domain noisy image segmentation based on fuzzy ant colony algorithm

The main steps are as follows:

(1) Pre-process the image to be segmented with appropriate grey-scale morphological factor and get the enhanced image;

(2) According to the histogram feature of the denoised image, narrow down the search range of the threshold to $[m,n]$, namely min and max in the artificial ant colony algorithm are m and n respectively;

(3) According to the grey scale and neighbourhood grey-scale information of the denoised image, get the grey-scale-neighbourhood-grey-scale histogram and calculate the minimum cross entropy in the image:

$$D(s,t)=\sum_{i=m}^{s}\sum_{j=m}^{t}\left[ijp_{ij}\ln\frac{ij}{\mu_1(s,t)}+\mu_1(s,t)p_{ij}\ln\frac{\mu_1(s,t)}{ij}\right]+\sum_{i=s+1}^{n}\sum_{j=t+1}^{n}\left[ijp_{ij}\ln\frac{ij}{\mu_2(s,t)}+\mu_2(s,t)p_{ij}\ln\frac{\mu_2(s,t)}{ij}\right] \quad (3)$$

In the formula, $P_1=\sum_{i=m}^{s}\sum_{j=m}^{t}p_{ij}$, $P_2=\sum_{i=s+1}^{n}\sum_{j=t+1}^{n}p_{ij}$, $\mu_1(s,t)=\frac{1}{P_1}\sum_{i=m}^{s}\sum_{j=m}^{t}ijp_{ij}$, $\mu_2(s,t)=\frac{1}{P_2}\sum_{i=s+1}^{n}\sum_{j=t+1}^{n}ijp_{ij}$, s is the grey-scale value; t is the average neighbourhood grey-scale value and p_{ij} is the grey-scale-neighbourhood-grey-scale joint probability;

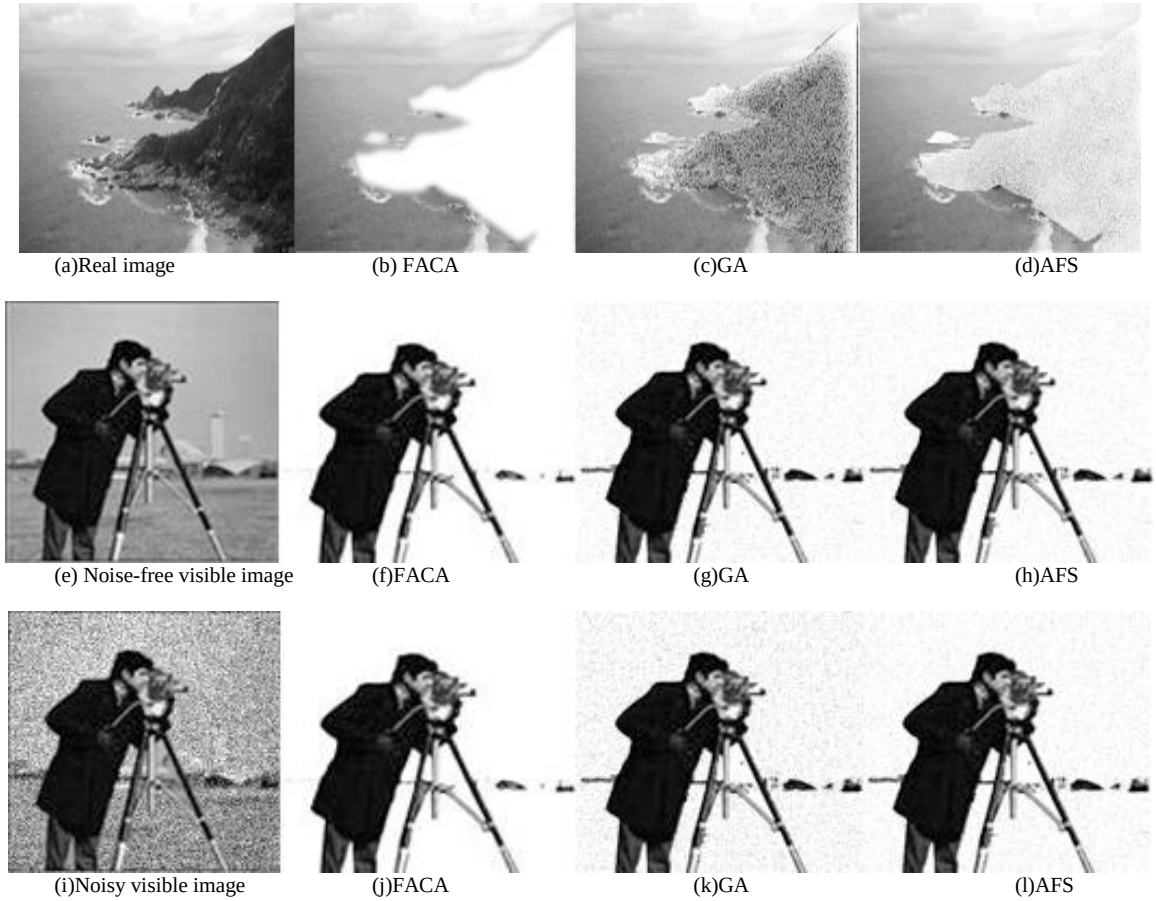
(4) Choose the reciprocal of Formula (3) as the fitness function of the artificial ant colony algorithm and then the corresponding (s,t) is the optimal threshold when the value of $\frac{1}{D(s,t)}$ is maximum;

(5) Segment the image according to the threshold (s,t) and the specific operation is as follows. If the sum of the grey scale of the pixel point (x,y) and the average neighbourhood grey scale is among $[0,s+t]$, then this point is the target pixel with a grey-scale value of 0; otherwise, this point is a background pixel with a grey-scale value of 1 [8].

5 Experimental results and analysis

5.1 THE COMPARISON OF SEGMENTATION RESULTS

The following are a group of image segmentation results by fuzzy ant colony algorithm, GA algorithm and AFS algorithm with visible image, noisy visible image and real image. Results are indicated as Fig. 3.



(a)Real image (b) FACA (c)GA (d)AFS

(e) Noise-free visible image (f)FACA (g)GA (h)AFS

(i)Noisy visible image (j)FACA (k)GA (l)AFS

FIGURE 3. The comparison of the image segmentation results by different optimization algorithms

From Fig.3, it is clear that no target regions has been split when segmenting the real noisy image; the noise has been split completely and there is a hole in the target when segmenting the noise-free visible image and the target has been split precisely when segmenting the noisy visible image since it is sensitive to the noise. Only part of target region has been split when segmenting the real noisy image; a hole appears in the target when segmenting the visible image and the noise and the target haven't been split well when segmenting the noisy visible image. The segmentation method based on fuzzy ant colony algorithm has strong noise immunity and get an intact and clear segmentation image target on the real noisy image with serious speckle noise, the visible image and the visible image with speckle noise.

5.2 THE COMPARISON OF ALGORITHM CONVERGENCE

In order to verify the quick optimization of fuzzy ant colony algorithm, the paper compares its convergence with the basic ant colony algorithm, the genetic algorithm and the artificial fish swarm algorithm with fixed experimental parameters. Make the individuals of the population in the fuzzy ant colony algorithm, the basic ant colony algorithm, GA algorithm and AFS algorithm as 20 and the population search times as 30. Compare the convergence of the fitness function curve when comparing the above algorithms in segmenting Fig. 3(a), as indicated in Fig. 4.

From Fig. 4, it can be seen that the fuzzy ant colony algorithm has converged in the 5th generation and it can search the optimal segmentation threshold; the basic artificial ant colony algorithm has converged in the 13th generation; the genetic algorithm hasn't converged in the 30th generation, so it can't find the optimal threshold and the artificial fish swarm algorithm has converged in the 19th generation, but it hasn't excellent stability. Therefore, the fuzzy ant colony algorithm can converge stably and quickly with small population and search times.

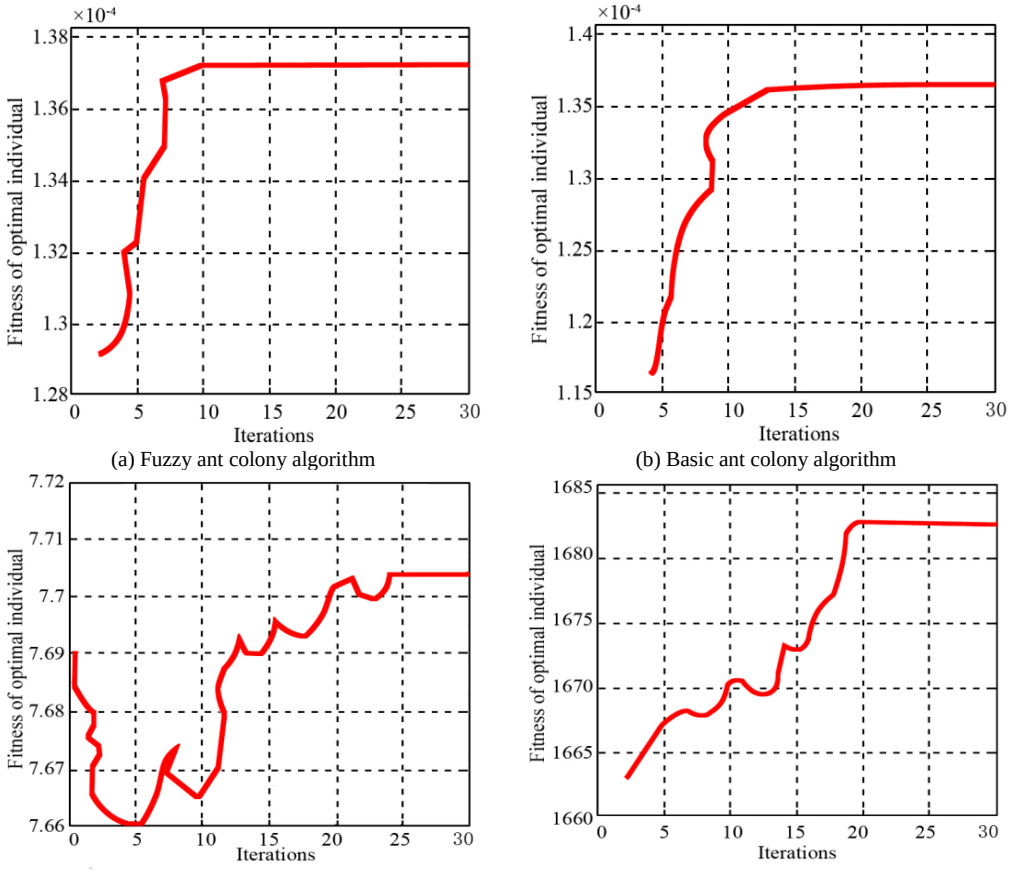


FIGURE 4 The convergence curve of the optimal

6 Conclusion

This paper has come up with a new spatial-domain noisy image segmentation method based on fuzzy ant colony algorithm. The main characteristic of this method is to make full use of the grey-scale features of the image to be segmented. It reduces the noise interference in the noise image segmentation through the grey-scale morphological pre-processing and effectively narrows down the threshold search range by pre-processing the histogram information of the image, which is good for the swarm intelligent algorithm to converge in a shorter time and guarantees the quickness of the segmentation speed. In improving the basic ant colony algorithm, it determines the search range of the reconnaissance ant via the fuzzy control dynamics and accelerates the convergence speed of the ant colony algorithm. The experimental results show that the method in this paper is better than the image segmentation methods based on the genetic algorithm and the artificial fish swarm algorithm in the noise immunity, precision, quickness and stability. The thing to note here is that this method is applicable for the image presenting twin-peak shape after grey-scale morphological processing and further exploration is needed to the images with mixed target and background noise.

Acknowledgments

This work was supported by Ministry of Education, Humanities and Social Sciences Research Projects (Grant No 12YJAZH216).

References

[1] Zhang B, Qi H, Ren Y T, Sun S C, Ruan L M 2014 Inverse transient radiation analysis in one-dimensional participating slab using improved Ant Colony Optimization algorithms *Journal of Quantitative Spectroscopy and Radiative Transfer* **133**(1) 351-63
[2] Mingyang Liu, Shufen Liu, Xiaoyan Wang, Ming Qu, Changhong Hu 2013 Knowledge-Domain Semantic Searching and Recommendation Based on Improved Ant Colony Algorithm *Journal of Bionic Engineering* **10**(4) 532-40
[3] Juan Rada-Vilela, Manuel Chica, Óscar Cordón, Sergio Damas 2013 A comparative study of Multi-Objective Ant Colony Optimization algorithms for the Time and Space Assembly Line Balancing Problem *Applied Soft Computing* **13**(11) 4370-82
[4] Zhe Yan, Hanming Gu, Chengguo Cai 2013 Automatic fault tracking based on ant colony algorithms *Computers & Geosciences* **51**(1) 269-81
[5] Salabat Khan, Abdul Rauf Baig, Waseem Shahzad 2014 A novel ant colony optimization based single path hierarchical classification algorithm for predicting gene ontology *Applied Soft Computing* **16**(3) 34-49

[6] Rana Forsati, Alireza Moayedikia, et al. 2014 Enriched ant colony optimization and its application in feature selection *Neuro computing* **142**(22) 354-71
[7] Martin Reed, Aliki Yiannakou, Roxanne Evering 2014 An ant colony algorithm for the multi-compartment vehicle routing problem *Applied Soft Computing* **15**(2) 169-76
[8] Wei-guo Zhang, Tian-yu Lu 2012 The Research of Genetic Ant Colony Algorithm and Its Application *Procedia Engineering* **37** 101-6

Authors

Ye Chen, born on November 2, 1979, China

Current position, grades: researcher at Hunan International Economics University, China.
University studies: the Master's degree in Computer Application technology from Hunan University, China in 2011.
Scientific interests: include Data mining and image processing.

Xiaoqun Qin, born on June 13, 1978, China

Current position, grades: researcher at Hunan International Economics University, China.
University studies: Master's degree in Computer Application technology from Central South University, China in 2007.
Scientific interests: Data mining and image processing.

XinMin Zhou, born on May 21, 1977, China

Current position, grades: researcher at Hunan University of Commerce, China.
University studies: Ph.D. degree in computer science and technology from Tongji University, China in 2010.
Scientific interests: text watermarking, information hiding and network security.

Face super-resolution algorithm based on SVM-improved learning

Jiali Tang[1, 2*], Chenrong Huang[3], Yijun Liu[2], Honghui Fan[2], Jianmin Zuo[1, 3]

[1] *School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, Jiangsu, China*

[2] *College of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, Jiangsu, China*

[3] *School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, Jiangsu, China*

Received 1 June 2014, www.cmnt.lv

Abstract

As many other inverse problems, human face image super-resolution is an ill-posed problem. The problem has been approached in the context of example-based superresolution learning. However, these methods need to run though all the sample set, which results in high calculation load and image degradation because of mis-matching. In this paper, we propose a new face image superresolution algorithm based on Support Vector Regression (SVR) pre-classified learning. A Principal Component Analysis (PCA) based pre-process is used to select a subset of samples. Then the best-matching sample images are trained to ensure the content relevance between the sample patch and the input low-resolution image. Further improvement involves a combination of classification and SVR-based techniques. Therefore, experiment results show that the proposed algorithm gets better reconstruction performance and faster program running speed.

Keywords: Face Super-resolution, Support Vector Machine (SVM), Principal Component Analysis (PCA), Example-based Algorithm

1 Introduction

The super-resolution reconstruction is the technology of using software algorithms to obtain higher resolution images without changing current imaging system hardware, which has been an active research in image processing and is widely applied in medical diagnosis, remote sensing and many other areas. Among them, face image has been specially focused in the field of image processing and pattern recognition. The traditional face super-resolution algorithms usually accord to the multiframe low-resolution (LR) images of sub-pixel displacement under same scenes to reconstruct high-resolution (HR) images. The single frame face image super-resolution algorithm based on training set reconstructs the high-frequency details of images by learning the relation between HR and LR images, which has become the mainstream of research area in recent years [1-5].

Human face super-resolution algorithm was first proposed by the Carnegie Mellon University's Simon Baker [6] etc., which builds a Gauss Pyramid with multi-resolution feature of HR images. By matching in different resolution spaces in the Pyramid and using the feature of images, the algorithm searches the corresponding HR patch from the side of matching the input image's feature. Framework of Bayesian inference is used to build the generating optimized model of HR images, then reconstructs the HR face images. This algorithm gets better reconstruction result than interpolation and traditional methods based on the Markov Network model. However, it is necessary to set up a huge sample database where we can do the traversal search, which leads to a heavy workload of calculating and the mismatching in search result will cause final quality decrease badly. Furthermore, the algorithm doing super-resolution reconstruction relies on its Gauss Pyramid model from the model of matched feature. It is limited for the improved room of subjective reconstructed quality when there is aliasing in sample feature extracting.

In this paper, we propose a face super-resolution algorithm based on SVM-improved learning. To solve the problems based on learning-based algorithm such as the heavy work of calculating and mismatching, we use a classifier to pre-classify the sample database and select the subset that is similar to the object's PCA feature, which reduces the mismatching rate of sample patch, shortens the program's running time and improves the image reconstruct performance.

2 SVM learning

SVM is on the base of Vapnik-Chervonenkis (VC) Dimension theory and Structural Risk Minimization (SRM) principle, to get better generalization ability according to limited sample details to find a best way to balance between the complexity of model and learning ability [7]. The main idea of SVM is to build a linear hyperplane as the decision boundary to make the margin

* ***Corresponding author*** e-mail: tangjl@jsut.edu.cn

between positive example and negative example maximum. Please see Figure 1.

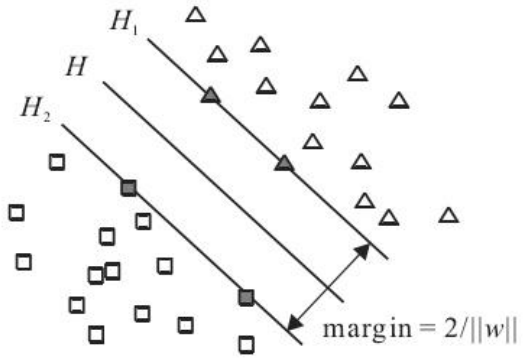


FIGURE 1 Schematic diagram of SVM learning.

More exactly to speak, SVM is a kind of generalized linear classifier based on SRM principle whose basic strategy is to confirm the empirical risk and to make confidence risk minimum.

The general type of classification hyperplane can be written as $wx+b=0$, after normalizated, which makes the linear separable sample set $(x_i, y_i), x_i \in R^d, y_i \in \{+1,-1\}, i=1,2,\cdots,n,$ meeting the following requirement:

$$y_i\left(wx_i+b\right)\geq 1, i=1,2,\cdots,n\ . \tag{1}$$

When margin is equal to $2/\|w\|$, so it is equal between maximizes margin and minimizes $\|w\|^2$. A hyperplane, which fits the condition (1) and makes $1/2\|w\|^2$ minimum is named optimal classification hyperplane, the training sample set is called support vectors, which makes the equal sign exist. Bring in the Lagrange function:

$$L=\frac{1}{2}\|w\|^2-\sum_{i=1}^{l}a_i y_i(w\cdot x_i+b)+\sum_{i-1}^{l}a_i\ . \tag{2}$$

Among in Equation (2), $a_i>0$ is the coefficient of Lagrange. Transform the question about optimal classification hyperplane into its dual problem by Lagrange optimization method, namely means becoming the question under the constraint of the inequality to search the best answering whose answer is unique solution. The optimal classification function is:

$$f(x)=\operatorname{sgn}\left\{\left(w^*x\right)+b^*\right\}=\operatorname{sgn}\left\{\sum_{i=1}^{n}\alpha_i^* y_i(x_i x)+b^*\right\}. \tag{3}$$

Among Equation (3), α_i^* is the corresponding Lagrange multiplier of each sample, we can prove only a few α_i^* is non-zero, then the corresponding sample is support vector. b^* is the classification threshold, which can acquire by any support vector or averaging from any couple of support vector.

For the sample set linearly non-separable case, we can put a slack variable $\delta_i\geq 0$, it becomes:

$$y_i\left(wx_i+b\right)\geq 1-\delta_i, i=1,2,\cdots,n\ . \tag{4}$$

And change the target into:

$$\min_{w,b,\delta}\varphi(w,x)=\frac{1}{2}w^T w+C\sum_{i=1}^{n}\delta_i\ . \tag{5}$$

To get the generalized optimal classification hyperplane by building a soft margin that considering in half way between minimum mistakes sub-sample and maximum margin, $C>0$ is a constant, which controls the level of punishment for the wrong sub-sample.

For the non-linear case, we can change it into a high-dimensional linear problem by non-linear variation. Generally speaking, this kind of non-linear variation is more complex, and hardly come true through the algorithm. We can use kernel function $K(x_i, x_j)$ who meets the Mercer's condition instead of original space's inner product, which will realize the linear classification after some non-linear variation, so that will avoid the concrete form of non-linear variation. The classification function is becoming:

$$f(x)=\operatorname{sgn}\left(\sum_{i=1}^{n}\alpha_i^* y_i K\left(x_i, x_j\right)+b^*\right). \tag{6}$$

The generator classification function is used to pre-classify the data in images training base, so as to judge the correlation between the training images and target images.

3 Face super-resolution reconstruction based on SVM-improved learning

Face super-resolution reconstruction means to estimate HR face by LR face, which is an application of imaging super-resolution technology in face area. As a kind of special image, due to the conspicuous structural features in face images, so the super-resolution algorithm adapt to common images can get a better effect in face reconstruction by combining these priori features. It is also widely applied in computer visual video surveillance, such as the field of video surveillance, usually the location of camera is quite far away from the face, so the captured face images are quite small. In order to better identify images, we first enlarge the face by super-resolution, then identify them, or doing the enlargement and identification at the same time.

Existing example-based learning face super-resolution reconstruction method uses the relation between the LR images and corresponding HR images in Markov network learning sample-base, then restoring the high-frequency details of input LR images through the relation we learned. This algorithm needs to build a huge face sample-base and do the traversal search in it. The HR sample can be the high-frequency details of the LR sample to restore the face. If and only if the searched face sample is similar to the input image sample and there is continuity in the content between corresponding HR

sample and other HR samples whose spatial position is contiguous, this HR sample can be the high-frequency details of the LR sample to restore the face. The searching process of the large-scale training sample base not only causes the time-consuming but also the mismatching of the search result will reduce the final quality.

According to the problems above, this paper proposes a face super-resolution algorithm based on SVM-improved learning. According to the PCA feature of human face, images are pre-classified by the algorithm. Before the matching search, SVM will pre-classify the sample base. So as to effectively extract and reconstruct the sub sample base, which has similar feature in target PCA, the algorithm establishes a SVM predictor for each class sample. By training samples, the algorithm trains the predictor and restores the prior knowledge in the form of predictiors' parameters. The experiment proves that it is an effective face super-resolution restoring method.

3.1 FACE FEATURE EXTRACTION

Face recognition technology has a wide application prospect in computer vision, guest identity authentication, multimedia data retrieval and etc.; face feature extraction is a necessary key step in process of face recognition. The face feature, which is related to application area, can be quickly and effectively extracted; becoming a key point whether realizes face automatic identification. The extraction process of human face features is a basis on a process that a transformation matrix maps human face image vector into feature vectors. During the mapping process, identification information is closely related with the application domain (such as different types of facial expression) can be extracted. And a large number of other irrelevant information (such as eyebrows shade, mouth size etc.) are discarded. In order to turn human face vector into human face feature vector with transformation matrix, it uses large quantities of training sample (known as the expression of face image type) to extract algorithm from human face feature. Then more training samples better the feature extraction.

PCA (Principal Component Analysis) is also called K-L transform (Karhunen-Loeve transformation), this method is to find a subset of main components which has the statistical distribution's data set at random, and corresponding base vectors satisfy the orthogonality [8]. The original data set is transformed into principal component space, which makes the cross-correlation of single data samples decrease to the lowest point. PCA is a classical method of feature extraction and data representation; it is widely applied in pattern recognition and computer vision, and becomes one of the most successful face recognition methods. Due to the dimension of face image sample is very high, if we directly process face images, the amount of calculation is large and the running time is long, so in recent years many face detection and face recognition algorithms use PCA method to reduce dimension first, then use other methods to process more.

A face image with $m \times n$ is rearranged as a column vector with $m * n$ dimension by PCA method, then all of the training images can get a group of column vectors after that change: $\{xi\}, x_i \in R^{m+n}, i = 1, 2, \ldots, N$, among that, N represented the number of the images in training sample set. Treat images as random column vectors, and training sample set is utilized to evaluate the average vector μ and covariance matrix S_T :

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \,, \tag{7}$$

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T = XX^T, X = (x_1, x_2, \cdots, x_n) \,. \tag{8}$$

Choose the projection matrix A corresponding to the first k and biggest eigenvalues of S_T . Equation (9) is utilized to reduce the dimension of original images:

$$y = (X - \mu) A_K \,. \tag{9}$$

After the eigenvectors of S_T are restored as image matrix, which is the standardized face. The basic idea of PCA is that approximating each face by the linear superposition of standardized face. And let these linear coefficients as face features, these features are used to classify. The way to classify face with the use of PCA technology is called Eigenfaces method.

3.2 PRE-CLASSIFICAITON OF SVM

If there are K kinds of images in the images training database, mark as $T = \{P1, P2, P3 \ldots Pk\}$, and have k semantic classifiers $\{C1, C2, C3 \cdots Ck\}$. For each SVM classifier, its training set $T = \{(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n)\}$, (x_1, y_1) is the labelled sample image which is given beforehand. $x_i \in R^2$ is the PCA feature of images; $y_i \in (-1, +1)$, $+1$ means images contain PCA semanteme, -1 means don't. Achieve images' semantic classifiers by using SVM to train these samples, and then these semantic classifiers are utilized to distinguish the left un-labeled images.

In this paper, using LibSVM 3.0 software package developed by Professor Zhiren Lin [9] of Taiwan University as SVM classify platform. LibSVM mainly includes *svmscale*, *svmtrain* and *svmpredict* three functions. Among that, *svmtrain* turn to train samples, according to input vectors and specified classification to generate suitable classifiers; when the way of classification is unknown, *svmpredict* is used to make input vectors map to the inner vector space of models based on specified classifiers, to find its classification principles; *Svmscale* is mainly for regularization of

numerical value, numerical value regularize input vectors to the boundary where LibSVM is fit for processing.

Due to the diversity of kernel function, one of the main points of designing SVM is that choose suited kernel function and corresponding parameters. According to the research by Vapnik and others, different kernel functions have little influence on the performance of SVM, but it is crucial for the performance of support vectors classifiers to choose penalty coefficient and corresponding paremeters of kernel function. In this paper, we use radial basis function, which is widely applied:

$$K(x, x_i) = \exp\left(-\gamma |x - x_i|^2\right). \quad (10)$$

Net check search is used to cross-validation C and γ. Try each basic couple (C, γ), and then choose the couple of parameters with the highest accuracy rate during the cross-validation. During the cross-validationn over and over again, determine two parameters, based on 1-a-1 multi-classification method to train image samples.

3.3 SVM PARAMETER OPTIMIZATION AND TESTING HR IMAGES GENERATION

In RBF-SVM, we need to determine two parameters which are planning factor C and Gaussian width σ. For C, if $C \to \infty$, means the rules of classification will meet all of constraint conditions, not only enhance training time and complexity of training, but also make classification have over-learning's situation, which reduce its generalization performance. So the data range of C should meet the situation of TAR and TRR, in order to keep the generalization performance of classifier, we need to take value as small as possible. For the Gaussian width σ, if without some ask for algorithm, we just set a group of probable values through the experience, and solving in turn under σ, then adjust according to the answer.

Do the SVM parameter optimization through net check search method. Net check search is reached through specially quantize and partly optimize method of exhaustion. In the method of net check search, firstly, confirm the data range of parameters, and then confirm the step of take value. Each parameters can start from the beginning of take value, take value to the end according to step, after enumerating all of parameter combination, then based on corresponding evaluation criterion to assess classification result, to find the optimum parameter.

When the classification is completed, the algorithm will establish groups of predictor corresponding with samples in each region, each category corresponding predictor to that class, i.e. if the samples divided into several K class, the groups of predictors in this region ought to contain the predictors of this K class. Multiclass predictor consists of a group of sub predictor; each sample of base classes corresponds to a linear sub predictor. A group of predictors operate the combined prediction to construct multiclass predictors. When predicting, firstly input data coding by the generator codebook in the process of sample classification, namely classification. Then according to the category, the system selects the corresponding sub predictors to predict. The training process of multiclass predictors is the process of training each single sub predictor. After the completion of training the predictors, the parameters will be stored for guiding the super-resolution restoration which inputting low resolution image. After the completion of training multiclass predictors, the system will use the multiclass predictors to reconstruct the HR face image. By using the bilinear interpolation to magnify HR face image size, the system inputs the LR face image, which is used as the estimation of initial image, for super-resolution restoration. When inputting low resolution block, according to the region which it belongs, the corresponding predictors operate high frequency prediction. The high frequency information blocks, which produced by those predictors, will be used to add up to the estimation of initial images in order to form the outcome of HR face image, namely the realization of face image super resolution restoration.

4 Simulation result and analysis

Text images come from a subset of international standard face image database, including 200 people's images from 7 angers and 1400 pictures in all, which includes location parameter of eyes, noses, mouths and etc. for each face image. Before the experiment, do the image intensity normalization and geometric normalization (80×80 pixels per image).

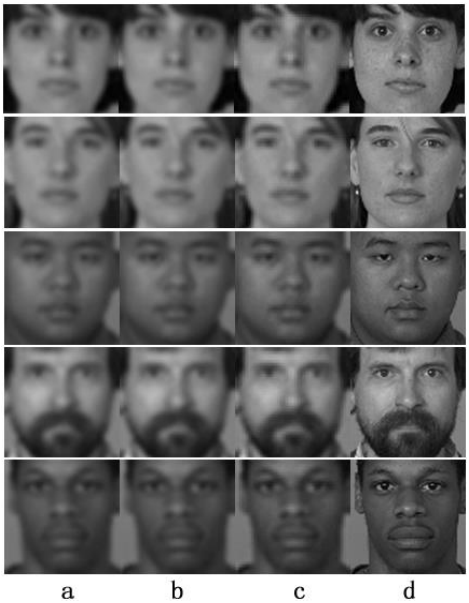


FIGURE 2 Reconstruction effects of three algorithms.
(**a** Freeman, **b** Reference [5], **c** Our method, **d** Original HD image)

Randomly choose 190 front view face images as training sample, PCA eigenvector is used as the input vector of support SVM, learning for the image categories,

pre-classify again based on output result. We choose the left 10 front face images in the database as text images, use a Gaussian blur template with the size 5×5 and variance $\sigma=0.85$ to degrade all HR images to the LR images with 40×40 pixels through Gaussian blur and down-sampling.

Do the contrast experiment apart with Freemam algorithm and document algorithm. We test the five face images' reconstructed effect (Figure 2).

As shown in these images, Freeman's algorithm [10] is the simplest, but the effect is poor. The algorithm of reference [5] is better in the central area of human face, but the other part especially the edge of face is fuzzy. Our approach is better than the first two algorithms in visual effect, but the edge of face and hair line still have some masks.

The value of PSNR is used to compare then text images' objective effect. Generally speaking, the higher value of PSNR is, the less distortion of image is. See Equation (11).

$$PSNR=10\times\lg\left[\frac{\left(2^n-1\right)^2}{MSE}\right]. \tag{11}$$

In Equation [11], *MSE* represents the error of mean square between original images and processing images. The result of quantizing PSNR is shown in Figure 3. We can reach the conclusion that the PSNR values in our approach basically higher than the other two algorithms. What's more, calculating apart the running time of image super-resolution reconstructed procedure on Freeman and our algorithm by *tic* and *toc* function in Matlab, the running speed of this paper enhanced obviously.

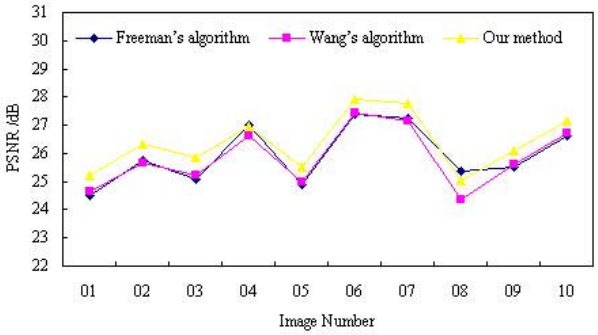


FIGURE 3 PSNR comparison of three algorithms

5 Conclusions

In this paper, SVM algorithm has been applied to example-based human face super-resolution reconstruction, which is a new attempt. SVM is used as the training classifier because its sample learning ability is very strong. We pre-search the sample database by defining PCA feature, find the sub-sample database whose feature is similar, then do the pixel-level exact matching search in the sub-sample set. The experiment shows that our approach can effectively reduce the mis-matching of image patches, save the running time and reach a higher quality in the reconstruction. In both subjective and objective way, restored image is better than the classical insert algorithm and the learning-based super-resolution method, and in the meantime the program running time is effectively improved.

Acknowledgments

The authors are indebted to their colleagues of School of Computer Engineering, Jiangsu University of Technology, for their suggestive discussion and zealous help. This study is supported by the National Natural Science Foundation of China (61302124), and the IUR Prospective Joint Project of Jiangsu Province (BY2014038-05), and also the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (13KJB520006).

References

[1] Gunturk B K, Batur A U 2003 Eigenface-domain Super-resolution for Face Recognition *IEEE Transactions on Image Processing* **12**(5) 597-606
[2] Huang H, He H T, Fan X 2010 Super-resolution of Human Face Image Using Canonical Correlation Analysis *Pattern Recognition* **43**(7) 2532-43
[3] Hu Y, Lam K M, Shen T Z 2011 A Novel Kernel-based Framework for Facial-image Hallucination *Image and Vision Computing* (29) 219-29
[4] Jia K, Gong S G 2008 Generalized Face Super-Resolution *IEEE Transactions on Image Processing* **17**(6) 873-86
[5] Wang X G, Tang X O 2005 Hallucinating Face by Eigentrans-formation *IEEE Transactions on Systems Man and Cybernetic* **35**(3) 425-34
[6] Baker S, Kanade T 2000 Hallucinating Faces *Proc. Conf. on Automatic Face and Gesture Recogniton* Grenoble, France 83-8
[7] Valentini G, Dietterlch T G 2004 Bias-variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods *Journal of Machine Learning Research* (5) 725-75
[8] Yang J, Zhang D 2004 Two-dimensional PCA: a new approach to appearance-based face representation and recogniton *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(1) 131-8
[9] Chang C C, Lin C J 2011 LIBSVM: A library for support vector machines *ACM Transactions on Intelligent Systems and Technology* **2**(3) 1-27
[10] Freeman W T, Jones T R**, Pasztor** E C 2002 Example-based super-resolution *IEEE Computer Graphics and Applications* **22(2) 56-65**

Authors

Jiali Tang, born on May 10, 1980, China

Current position, grades: associate professor at Jiangsu University of Technology, China.
University studies: Msc. Degree in Material Processing Engineering from Nanjing University of Aeronautics and Astronautics, China in 2005.
Scientific interests: As an IEEE member, his research interests include industrial control software development and image super-resolution reconstruction. He has published nearly twenty academic papers in related journals.

Yijun Liu, born on June 28, 1978, China

Current position, grades: works in School of Computer Engineering of the Jiangsu University of Technology.
University studies: bachelor's degree in Computer Science and Technology from Nanjing University in 2000, her Master's degree in Computer Science and Technology from Nanjing University in 2003.
Scientific interests: machine learning, data mining and intelligent information system.

Honghui Fan, born on October 24, 1980, China

Current position, grades: is a member of China Computer Federation. In 2011 joined Jiangsu University of Technology.
University studies: M.Sc. and Ph.D. from the Yamgata University of Japan in 2008 and 2011, respectively.
Scientific interests: ultrasound imaging, image restoration, digital image, and signal processing in biomedical engineering.

Jiamin Zuo, born on January 7, 1957, China

University studies: PhD supervisor in the School of Mechanical Engineering of Jiangsu University. **Current position, grades:** professor at Nanjing Institute of Technology, China.
Scientific interests: As a senior member of Chinese Mechanical Engineering Society (CMES), his research interests include electromechanical control and industrial automation.
Publications: over seventy academic papers in related journals.

Application of halcon in the image analysis of dry cutting gear meshing region

Jingang Gao, Shuang Zhang, Hua Wang*

College of Mechanical Science and Engineering, Changchun Institute of Technology, Changchun 130012, China.

Received 1 June 2014, www.cmnt.lv

Abstract

The master-slave bevel gear pair of some automobile rear axle is used to be as the research object, to obtain the image of the gear meshing contact region. Firstly, get the grey image of HSV space from RGB image to HSV space by the pre-processing algorithm; secondly, segment features of image accurately by the adaptive threshold segmentation algorithm; thirdly, fill the holes in the image of meshing region perfectly by proposed holes repair algorithm; fourthly, mark all connected domains of image with 4 -adjacent points domain labelling algorithm; finally, the Marking image area, width and the ratio of width and height features are selected to extract the image of the gear meshing region to obtain the geometry information of the gear meshing region. Research results show: contact centre E/L should be in 45% ~ 50%, and the length of the contact region B/L should be in 45% ~ 55%.

Keywords: HALCON, dry cutting gear, contact region, Image processing

1 Introduction

Dry machining is one of the development trends of metal cutting processing in the future. In recent years, especially the developed industrial country, gave great importance to the dry cutting. The quality of dry gear machining affected the performance of the gear transmission. To analysis the change of the meshing contact region before and after heat treatment, guarantee the system stability of the technological process for the dry cutting gear, the research group used Gleason Phoenix dry cutting machine to the master-slave bevel gear pair of some automobile rear axle, used as the study object.

The gear teeth surface contact region refers to the overlay of the contact lines on the gear meshing tooth's surface. The size, shape and position of the contact region not only reflect the manufacturing precision of a single gear, but also reflect the installation and transmission accuracy of the gear pair [3], which affect the bearing capacity, service life, quality and efficiency of transmission and noise etc. Therefore, scholars both at home and abroad, Litvin studied the dynamic performance of gear transmission for the teeth surface contact region [4], developed the theory of face gear transmission point contact. However, the reports are being rare by image processing technology, to analyse the size, shape and position of the tooth surface contact region.

In this paper, with HALCON image processing software, and through analysis the dry cutting gear meshing teeth surface image of the automobile rear axle, extract the grayscale image of HSV brightness space; use the adaptive threshold segmentation, holes filling algorithm and feature extraction algorithm to get the accurate geometry data of the gear meshing region; obtain the change of contact region before and after heat treatment; give the proportionality range of the required contact region for the final qualified products; and provide the basis of dry cutting gear manufacture.

2 Processing of the contact region image data for dry cutting gear before and after heat treatment

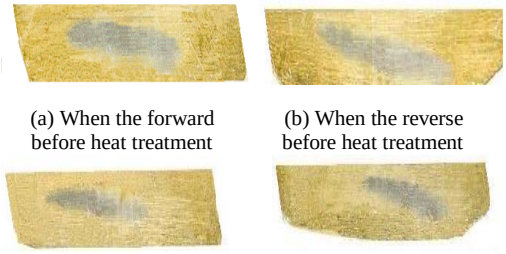
2.1 IMAGE ACQUISITION OF CONTACT REGION

During wet cutting manufacture, because the spiral angle change of big gear is smaller, change of the contact region in the tooth length direction is affected primarily by a small gear, namely, the forward surface turns towards the small end, and the reverse surface to the big end after heat treatment.

In order to get the meshing marks in the contact region of the master-slave bevel gear teeth surfaces, a thin layer of butter smears on the slave bevel gear tooth's surface, some resistance loads on the slave bevel gear to rotate the master gear in the positive and negative two directions, and then the imprinting is observed. Figure 2 shows the meshing contact surfaces image of the big gear tooth's surface before and after heat treatment twice, for the master-slave bevel gear pair of automobile rear axle. This system adopts high resolution digital camera (DH - HV1302UM) to collect images, produced by the Beijing DaHeng Company, a resolution of 1280 x 1024, and which is equipped with the image acquisition card, to

* ***Corresponding author*** e-mail: 394262133@qq.com

obtain images. The images of a tooth surface contact region for the gear are shown in figure 1.



(a) When the forward before heat treatment (b) When the reverse before heat treatment

(c) When the forward after heat treatment (d) When the reverse after heat treatment

FIGURE 1 The image of the teeth surface contact zone before and after heat treatment

Seen from Figure 1, there are the fixed contact regions on the tooth's surface for the different gear meshing. Through the information processing of those contact regions, the gear meshing situation can be obtained, thereby the gear machining and assembly process can be improved.

2.2 PROCESSING ALGORITHMS OF CONTACT REGION IMAGE

When measuring the gear contact region, based on the characteristics of the bevel gear meshing, some parameters are designed in this paper: L: the length of teeth, B: the length of contact region, E:, the distance from the centre of the contact region to the gear head, details is shown in figure 2.

FIGURE 2 The measured data of the contact region

2.2.1 Image transformation from RGB to HSV

Seen from the image of the gear meshing contact region, the boundary is fuzzy between the contact region and the contact region. It is difficult to identify the edge of the contact region accurately with the image segmentation algorithm. Even the owe or excessive segmentation problems may be lead [1]. The grey value of the contact region image and the background is close, when the colour image of the gear meshing contact region is transformed into the grey image. For this phenomenon, the grey value segmentation method is not good. According to the characteristics of the gear meshing contact region image, this article proposed algorithm based on HSV space V component to segment the gear meshing contact region image, and obtain the geometry information of the contact region.

The colour image segmentation algorithm based on HSV space in this paper, is aimed at the existing problem of RGB colour image segmentation, and combined with the three components in HSV colour space, which have nothing to do with the brightness. RGB colour space is converted to HSV space, there are:

$$\begin{cases} 0° & \max = \min \\ 60° \times \dfrac{g-b}{\max-\min} + 0° & \max = r, g \geq b \\ 60° \times \dfrac{g-b}{\max-\min} + 360° & \max = r, g < b \\ 60° \times \dfrac{b-r}{\max-\min} + 120° & \max = g \\ 60° \times \dfrac{r-g}{\max-\min} + 240° & \max = b \end{cases}, \quad (1)$$

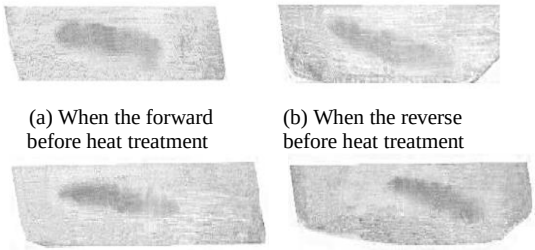
$$s = \begin{cases} 0 & \max = 0 \\ \dfrac{\max-\min}{\max} & otherwise \end{cases}, \quad (2)$$

$$v = \max . \quad (3)$$

The following is the programs from RGB to HSV, the transformed images are shown in figure 3:

```
min = min(R,G,B)
max = max(R,G,B)
V = max
if (max == min)
S = 0
H = 0
else
S = (max - min) / max
if (R == max)
H = ((G - B) / (max - min)) * 60
elif (G == max)
H = (2 + (B - R) / (max - min)) * 60
elif (B == max)
H = (4 + (R - G) / (max - min)) * 60
fi
fi
```

Range of values: H = [0;2pi], S = [0;1], V = [0;1].



(a) When the forward before heat treatment (b) When the reverse before heat treatment

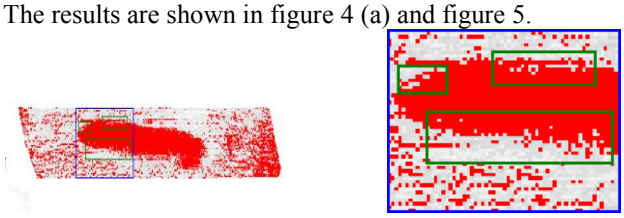
(c) When the forward after heat treatment (d) When the reverse after heat treatment

FIGURE 3 the bright image of HSV

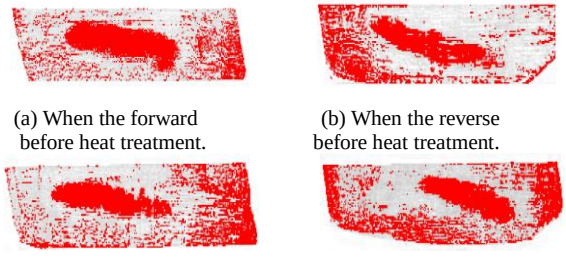
2.2.2 The adaptive threshold segmentation

In figure 3, the bright image of the gear meshing contact region is obtained after transformation, which is already a

pair of grey image. To extract the information of the contact region, the binarization processing of the image should be done, which is the key is to choose a suitable threshold. Such as Otus, the image segmentation algorithm can be used to get the better threshold with the least squares principle value [2]. In figure 3, the boundary is not obvious between the contact region and the contact region. When the change of grey value is the smallest, the effects of image segmentation are not ideal. Its main reason is that every pixel point in the image segmentation process all adopted the same threshold. So this paper introduced the adaptive threshold segmentation algorithm, its basic idea is that the image are divided into several sub images; for the neighbourhood window which is determined by the centre of each pixel itself, the average value of all the pixels in the window is used as a threshold value [3]. If the pixel of this point is less than the threshold, then this point is set to 0, otherwise it is set to 255.So the threshold of the whole image is not fixed, but adaptive. In this way, the part is segmented completely. To get ideal segmentation results, the key is the select the smoothing filter size and threshold [4-6]. The results are shown in figure 4 (a) and figure 5.



(a)The adaptive threshold image. (b)Image of the enlarged hole region.
FIGURE 4 The adaptive threshold segmentation image and local enlarged image



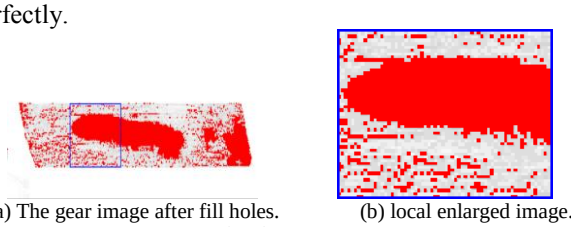
(a) When the forward before heat treatment. (b) When the reverse before heat treatment.
(c) When the forward after heat treatment. (d) When the reverse after heat treatment.
FIGURE 5 The image of the adaptive threshold segmentation

2.2.3 Feature repairing: holes filling in a region

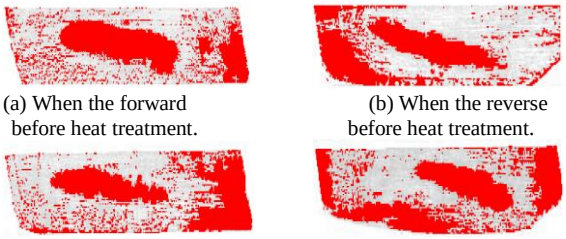
In region filling algorithm, in addition to the traditional scanning algorithm, Suzuki [7] proposed a method, which marks images and updates the array through the forward and backward scanning to decrease the scanning numbers of image greatly. Wu [8] during the linear scanning process, took into consideration the target pixel neighbourhood judging priority order, based on the different images which reduce the visits numbers neighbourhood and improve the scanning efficiency. He and others based on single linear sweep [9] and double linear sweep [10] of pixel neighbourhood characteristics, further development of this type of scan filling algorithm.

In region filling algorithm, in addition to the traditional scanning algorithm, Suzuki, [7] proposed a method, which marks images and updates the array through the forward and backward scan to decrease the scan numbers of the image greatly. Him and others, based on pixel neighbourhood features of single linear sweep [9] and double linear sweep [10], further developed the scanning filling algorithm.

In this paper, seen from the figure 4 (b), there region lots of holes and dispersed regions in the tooth's surface meshing region. These are adverse to extract the meshing region area information. So, firstly, holes need to be filled to maintain the integrity of the characters. The optimal neighbourhood region filling algorithm is adopted [11], and its calculation results are shown in figure 6 (a) and figure 7. Contrast the same region filling effect of figure 4 (b) and figure 6 (b), it can be seen that the algorithm not only ensure the accuracy of the gear meshing region edge image, but also fill holes in the meshing region image perfectly.



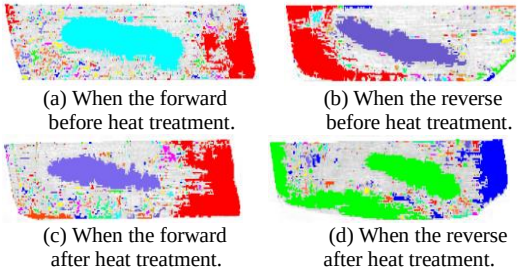
(a) The gear image after fill holes. (b) local enlarged image.
FIGURE 6. The gear image after fill holes and its local enlarged image.



(a) When the forward before heat treatment. (b) When the reverse before heat treatment.
(c) When the forward after heat treatment. (d) When the reverse after heat treatment.
FIGURE 7 The gear meshing region edge image after fill holes

2.2.4 Mark connects demean each other

In figure 7, after filling holes, the contact un-contact regions (in the middle of the image) of the gear meshing are separated clearly, but there are many noise spot images at the un-contact region of the image, which are not benefit to extract image information of the contact area image. To distinguish the contact and un-contact region, so all connected domains are marked in the image with different colours, as shown in figure 8.



(a) When the forward before heat treatment. (b) When the reverse before heat treatment.
(c) When the forward after heat treatment. (d) When the reverse after heat treatment.
FIGURE 8 the gear meshing region image of mark connects domain

In this paper, 4-adjacent point region labelling algorithm is adopted, the steps are as follows:

1) Scanning image line by line, when the unmarked pixel point P is found, whose pixels value is 255, this point is set to a new ID.

2) Checking 4 -adjacent points of P, assign adjacent points to the same ID with P, whose pixels value is 255.

3) According to the step (2), check all adjacent points of marking pixels, and assign the same ID.

4) The repeating steps (3) do not be stop until all interconnection pixel points are marked.

5) Return to step (1), find untagged pixel points again, if found, then repeat the previous step (2) - (4), otherwise, the marking process is over.

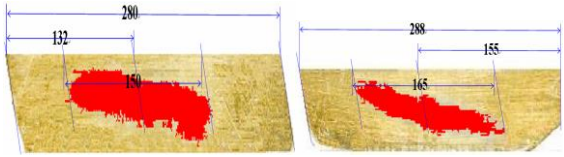
2.2.5 Feature selection and extraction

After the pixel connected domain, all obvious features (see figure 9) are connected into one region. Many connected regions are obtained. But only one is the desirable contact region image of the gear meshing. The image data of few larger regions in figure 9 and data analysis results are shown in table 1.

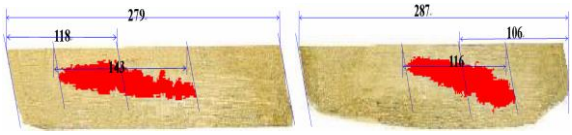
TABLE 1 Feature regional information and analysis

No.	Region	Width	Height	Width/ Height	Contact?
a	3046	150	41	3.7	Y
	1306	70	42	1.7	N
b	2376	165	36	4.6	Y
	2478	153	54	2.8	N
c	2060	143	27	5.3	Y
	2848	90	55	1.6	N
d	1837	116	36	3.2	Y
	2329	193	53	3.6	N

$$\begin{cases} 1500 < area < 3500 \\ 100 < width < 180 \\ 3.0 < \dfrac{width}{height} < 6 \end{cases} . \quad (4)$$



(a) When the forward before heat treatment. (b) When the reverse before heat treatment.



(c) When the forward after heat treatment. (d) When the reverse after heat treatment.

FIGURE 9 The final meshing region image and the analysis of Geometry data

3 The analysis and conclusions of experimental data

3.1 DATA ANALYSIS

Through the image processing of the gear tooth's surface before and after heat treatment, analysis results of geometry data information can be obtained and shown in table 2.

TABLE 2 Geometric data information analysis results of the gear meshing region image

	forward		reverse	
Design B/L	30%		30%	
	before the heat treatment	after the heat treatment	before the heat treatment	after the heat treatment
B/L	53.6%	51.3%	57.3%	40.4%
E/L	47.1%	42.3%	53.8%	36.9%

Data analysis shows:

1) The length of the actual contact region is longer than theoretical contact region;

2) After the heat treatment, the contact region is changed to the more narrow and shorter; as shown in figure 10, the forward surface is shorten about 3mm (7 pixels), and the reverse side is shorten the 20mm (47 pixels);

3) The forward and reverse surfaces are both moving towards the smaller end, where moving distance of the reverse surface is longer; as shown in figure 10, the forward surface is moving about 6mm (14 pixels), and the reverse side is moving the 21mm (49 pixels);

Through the above experiments, considering the change of the contact region in the meshing process, and the shape and position of the contact region of the final product, the general requirements of ensuring the contact region shall be as follows:

1) The forward surface: the contact region is above in the middle position of the tooth length and tooth root:

Contact centre E/L = 45% ~ 50%.

Length of contact region B/L is, generally in 45%~ 55%.

2) The reverse side: the contact region is slightly smaller than the end tip, and length slightly shorter:

Contact centre E/L = 40% ~ 45%.

Length of contact region B/L is about in 40%~ 45%.

3.2 CONCLUSIONS

In this paper, the meshing region image of the dry cutting gear for the automobile rear axle is studied by image processing technology. When the boundary is found to be the more vague between the tooth surface contact and un-contact regions, the preprocess algorithm is proposed to obtain the grey image in HSV space brightness; but the change of image grey difference is very small, the effect of the ordinary image segmentation is poor. In this paper, the adaptive threshold segmentation is used to segment features of the image; after image segmentation, there are lots of holes in the image, which are not benefit to extract

the area information of the meshing region. Holes must be filled. So the holes repairing algorithm are put forward to fill the holes in the meshing region image; all connected domains of image are marked with 4-adjacent points domain labelling algorithm; the marking image area, width and the ratio of width and height features are determined by the geometry information of every the connected domain; the final accurate image of the gear meshing region is obtained by three features.

According to the geometry data of the gear meshing region, the contact region change of dry cutting gear is obtained in the meshing process before and after heat treatment, and the required proportion range of the contact region shape and position is given to guarantee the qualified products, to provide the basis for dry cutting gear manufacture.

Acknowledgments

This work was supported by Jilin Province Development and Reform Commission of China: 120090023.

References

[1] Lucchese L, Mitra S K 2001 Color image segmentation: a state-of-the-art survey *Proceedings of the Indian National Science Academy* **67**(3) 207-21
[2] Mizushima A, Lr Ren-fu 2013 An image segmentation method for apple sorting and grading using support vector machine and Otsu' S method *Computers and Electronics in Agriculture* **94**(6) 29-37
[3] Manisha S, Vandana C 2012 Objective evaluation parameters of image segmentation algorithms *International Journal of Engineering and Advanced Technology* **2**(2) 84-7
[4] Ivan G D, Fernando D 2013 A region-centered topic model for object discovery and category-based image segrnentation *Pattern Recognition* **46**(9) 2437-49
[5] Chen Q, Sun Q S, Xia D S 2013 Serial slice image segmentation of digital human based on adaptive geometric active contour tracking *Computers in Biology and Medicine* **43**(6) 635-48
[6] Navon E, Miller O, Averbuch A 2013 Color image segmentation based on adaptive local thresholds *Image and Vision Computing* **23**(6) 69-85
[7] Suzuki K, Horiba I, Sugie N 2013 Linear-time connected component labeling based on sequential local operations *Computer Vision and Image Understanding* **89**(1) 1-23
[8] Wu K, Otoo E, Shoshani A 2005 Optimizing connected component labeling algorithms *In Proceedings of SPIE Medical Imaging Conference 2005* San Diego 558-70
[9] He L, Chao Y, Suzuki K 2010 An efficient first-scan method for label-equivalence-based labeling algorithms *Pattern Recognition Letters* **31**(1) 28-35
[10] He L, Chao Y Y, Suzuki K 2009 Fast connected-component labeling *Pattern Recognition* **42**(9) 1977-87
[11] Du Jianjun, Guo Xinyu 2013 A region filling algorithm based on optimal neighborhood relativity *Chinese Journal of Stereology and Image Analysis* 18(1) 972-8

Authors

Jingang Gao, born on September 9, 1976, China

Current position, grades: associate professor at Changchun Institute of Technology, China.
University studies: master degree in mechanical engineering from Huaqiao University, China in 2004.
Scientific interests: products quality inspection and control in the manufacture.
Publications: 9 academic papers, of which 6 papers was indexed by EI.

Shuang Zhang, 10. 11. 1979, China

Current position, grades: lecturer at Changchun Institute of Technology, China.
University studies: master degree in mechatronic engineering from Changchun University of Technology, China in 2006.
Scientific interests: products quality inspection and control in the manufacture.
Publications: 12 academic papers, of which 4 papers was indexed by EI.

Hua Wang, 8. 11. 1963, China

Current position, grades: professor at Changchun Institute of Technology, and a senior expert in Jilin province, China.
University studies: Ph.D. degree in mechanical Engineering from Jilin University, China in 2009.
Scientific interests: products quality inspection and control in the manufacture.
Publications: 22 academic papers, of which 10 papers was indexed by EI.

Editing method of virtual human motion path based on motion cycle step-length

Wenju Wang[1*], Liujie Sun[1], Zhang Xuan[2]

[1]*University of Shanghai for Science and Technology, Shanghai 200093, China*

[2]*Shanghai Conservatory of Music, Shanghai 200031, China*

Received 1 June 2014, www.cmnt.lv

Abstract

An editing method of virtual human motion path based on movement cycle step-length is presented: it uses the Cardinal spline interpolation to extract the original path and equably sets the path editing control points; alters the positions of path editing control points to generate a new path and adjusts the position and orientation of the original motion sequence frame's root joint nodes on the new path; automatically judges any path curve segment that generates the problems of footstep disharmony after editing; for every incorrect curve segment, deletes the frames of general motion cycle step-length between the two path editing control points that contain incorrect curve segment based on the definition of motion cycle step-length and adjusts the spacing and moving direction of root nodes in every frame on the paths between the two control points; finally does constraint re-establish to all floor frames in the new path. The experimental results show that: the motion path editing method is simple and easy to operate. The generated virtual human motion is natural and smooth. It can effectively eliminate the footstep sliding phenomenon to improve the reusability of motion capture data.

Keywords: Motion path editing, Cardinal spline interpolation, Motion cycle step, Constraint reestablishment, Footskate cleanup

1 Introduction

Virtual human animation is to use a computer program or software to simulate the human motion. In the game and entertainment, simulation training and simulation tests based on virtual reality, such as virtual scene display, sophisticated skill training, virtual manufacturing and ergonomics analysis, it plays a more and more important role [1, 2, 3]. The motion capture technology means to use the data capture device to record some real human motion, and then map the recorded motion data to the virtual human in computer, to drive the virtual human moving [4, 5]. Because it has accurately recorded a wealth of detailed information and can efficiently draw realistic and delicate human motion, the technology has become the current study hotspot of virtual human animation. But it can only generate a particular human motion prior captured and the acquired data can only be applied in certain circumstances, which is not adaptable, of poor flexibility and with low reusability [1]. In the actual animation designs, human-body motion is rich and varied. If re-capturing the motions for every slightly differences of design actions, it will bring high costs. So, editing and addressing the existing captured motions to fit design requirement, to improve the reusability of captured data to reduce production costs is significant. The motion editing is an effective way of editing that changes other actions of virtual human as little as possible, and edits motion paths to fit animation design requirements of different trajectories under different situations.

But in the motion path editing process, it often generates the footstep slip phenomenon. Despite adopting many effective measures and methods, it is not well resolved. For example, Michael Gleicher had proposed a path transformation method. He used the arc length parameter to map the virtual human location to the new motion path. Then, he used the method of geometric constraints to define the double-foot position or relative position. But the resulted effect was not realistic [5]. In 2002, Michael Gleicher proposed a way to automatically recognize previous foot location constraints and manually edit and generate new foot position by users. This process did not introduce artifacts, but increased the labour input editors [6]. Noah Lock used an improved deformation technology approximating to rigid body to modify the motion paths, and time bending technology to allow users to concentrate on meaningful space path editing. Meanwhile, it automatically improved the motions, to regularly maintain a biomechanical relationship between the path shape and speed. But time bending technology has strict rules on speed, which makes its applicable scope limited [7].

Lu Xaojun used the joint interpolation method to generate each frame's joint parameter from the original captured data. It got the animations of different motion paths by interpolating values to the positions of motion locus of human body basis points. The process contained no foot position constraint and was prone to appear foot

[*] ***Corresponding author*** e-mail: wangwenju666@163.com

sliding phenomenon, making the generated animation on poor quality [8]. Liu Shuang proposed an improved motion path editing method: evaluating the foot position of original motion data for each frame to use as the initial constraints. Then, it automatically generated new constraint position of supporting point according to the motion direction changes. Following this, it would use the real-time inverse kinematics algorithm according to frame editing to solve constraint conditions [9]. In the motion path editing process, Luo Zhixiang established the appropriate objective function by defining a set of space-time constraint conditions and used inverse kinematic and numerical optimization method to solve the motion posture meeting the constraint conditions [10]. Chen Zhihua introduced the motion auto simplification technology into the conventional space-time optimization method and introduced the physical constraints into the conventional path transformation method, to guarantee the rapid convergence of the optimization process and the physical authenticity of the resulting motion [11]. However, in the above-described algorithm, it added the physical constraints containing velocity and acceleration. They were both non-linear functions. This would greatly increase the complexity of time-space optimization solution, and the overall solving speed is slower.

SAFONOV A [12] and some others put forward a search model based on the motion diagram that means to connect different motions by the way of diagram. Thus, it changes the motion path editing problem to the problem of finding optimal path according to the motion diagram. However, the effect of splicing and fusing the motion fragments is not satisfying and the mode needs to spend more time on fusing motions.

To solve these problems, here proposes a motion path editing method based on motion cycle step-length. It can effectively eliminate the footstep sliding phenomenon to improve the reusability of motion capture data.

The section 2 describes related work of the motion path editing algorithm; the section 3 describes some preliminaries as the theoretical basis of the algorithm; the section 4 gives a virtual human motion path editing algorithm based on motion cycle step-length; the section 5 is experimental results and analysis of our algorithm; the section 6 gives the conclusion.

2 Preliminary knowledge

2.1 FORMAT OF MOTION CAPTURE DATA

The file format of motion capture data mainly contains BVH, HTR, ASF / AMC. (In which, the BVH format developed by the Biovision is more common. This paper will take data on this file format as the input object of motion path editing.)

Analysing these human motion capture data, the content is mainly divided into two parts [13]: 1. data being used to describe the structure and the initial position of human skeleton; 2. recording a set of specific data for each frame including displacement of each root node in every frame under the global coordinate system and motion data information between adjacent joints as Euler angle and so on.

The hierarchical skeleton model is usually used to describe the human skeletal structure of the virtual human (as in Figure 1). The skeletal system is represented as a set of interdependent joint chains in the hierarchical skeleton model, and can be described as a tree structure (as in Figure 2). The root node of the tree is Root. Other nodes correspond to each joint of the human body model. The whole human motion can be seen as being composed by translation and rotation, that it is achieved by the translation of the root node and the rotation of the rest nodes relative to their parent nods on the tree. Wherein, the translation of the root node determines the motion trajectory of human, and the rotation determines the motion direction of human. Except the root node, the rotation of the rest nodes is done in the local coordinate systems which take their parent nodes as coordinate origin. So, in each frame of the motion sequence, the position of each joint node p_{n+1} in human model can be solved in the world coordinate system by the following formula:

$$p_{n+1}^{world} = \prod_{i=Root}^{p_n} T_i R_i p_{n+1}^{local} \tag{1}$$

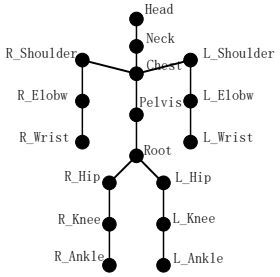


FIGURE 1 Structure diagram of human skeleton

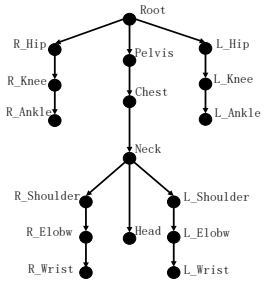


FIGURE 2 Hierarchical skeleton model of human

In the equation (1), p_{n+1}^{world} is the world coordinate of joint node p_{n+1}, p_{n+1}^{local} is the coordinate of joint node under the local coordinate system using p_{n+1} as origin and $T_i, R_i (i = Root, p_1, ...p_n)$ is the offset and rotation matrix of the child node relative to the parent node.

3 A motion path editing algorithm based on motion cycle step-length

In the motion path editing, it often appears a footstep slip phenomenon because of the original path changing. There are two main reasons for this phenomenon: 1. the location and motion direction of the virtual human are not timely adjusting. The total frame amount does not change, but the path length becoming relatively short after editing, that is, the relative distance between root nodes of motion sequence frames becoming smaller (what we called frame aggregation is too high). Their relative movement rate increases. But the step exercise frequency and step-length remain unchanged. These result the phenomenon of "out of step". 2. in the original motion sequence, a supporting point (i.e. a foot) is at the same position in space on two or more continuous moments. However, after motion path editing, the position of the supporting point has changed on the corresponding moments. If still using the original motion frame sequence to draw, it will produce footstep slip and appear motion disharmony.

To solve the footstep slip phenomenon, the new motion path editing method is proposed from these two areas. The algorithm steps are described as follows:

(1) Here takes global offset of each frame's root joint node in accordance with the Cardinal spline interpolation to extract the original motion paths and save, and calculates the slope of the spline interpolation curve of each frame's root node as a motion orientation.

(2) Evenly setting the path editing control points on the original motion path and taking two adjacent path editing control points as the starting terminal and ending terminal of a curve segment, by this it divides the original motion path curve into a number of curve segments. For each curve segment, along with the direction of human motion and with each path editing control point as starting point, it will find starting frames of each general motion cycle step-length in turn.

(3) It changes the location coordinates of the path editing control point, then rebuilds a new motion path by Cardinal interpolation, and adjusts the position and motion orientation of each frame on the new path.

(4) On the newborn motion path, here determines the curve segments with footstep disharmony phenomenon to construct the corresponding curve segment collection. If the curve segment collection is empty, namely no curve segment with footstep problem, then it directly turns to step (8).

(5) Here selects a curve segment with the phenomenon of footstep disharmony from curve segment collection. And it judges whether there is a general motion cycle step-length within the curve segment, if no, reading the next curve segment, Otherwise, starting to count the number of the to-be-deleted general motion cycle step-length on the curve segment. It assumes the curve segment on the original motion path which uses two adjacent control points as terminals AB, and the curve segment after path editing as AB'. The number of to-be-deleted general motion cycle step-length is calculated by equation (2), where λ is the general motion cycle step-length.

$$N' = \left\lceil \frac{\left| AB - AB' \right|}{\lambda} \right\rceil \tag{2}$$

(6) According to the number of to-be-deleted general motion cycle step-length in the curve segment, it will delete the frames of general motion cycle step-length between the frame sequences where the two control points are.

(7) It finds the nearest two editing path control points in the newborn path. The curve segment taking the two control points as terminals should contain the curve segment with the phenomenon of footstep disharmony. Between the two control points, it adjusts the frames that has deleted the general motion cycle step length. Checking if all the curve segments with the phenomenon of footstep disharmony have been treated in the curve segment collection, if still existing untreated curve segment, it turns to step (5), otherwise to step (8).

(8) It constrainedly reestablishes all the foot landing frames in the newborn path.

3.1 PATH EXTRACTION METHOD BASED ON CARDINAL SPLINE

To edit the motion path of virtual human, it has to extract its motion path from a section of known motion capture data. The new motion path extraction method based on Cardinal spline interpolation is as follows:

(1) On the equal time interval frame in the original motion sequence, it takes the global offset of root joint node as the position of Cardinal spline curve control point.

(2) For the translation trajectory of human root node, it carries out Cardinal spline interpolation to its horizontal axis component and longitudinal axis component (in the BVH files, because the world coordinate system is defined as the right-handed coordinate system and takes the positive direction of y axis direction as the positive direction of vertical axis, it takes the x axis component, z axis component as horizontal axis component, longitudinal axis component) conducting curve fitting to generate the original motion path.

In the process of extracting the original motion path, here selects the Cardinal spline interpolation because the slope of control points can be obtained from the coordinates of two adjacent control points and easy to implement. It assumes that $p(u)$ is the parameter cubic function between the two control points p_k and p_{k+1}. The four control points between p_{k-1} to p_{k+1} are used to

establish the boundary conditions of Cardinal spline segment [14], see equation (3):

$$\begin{aligned} p(0) &= p_k \\ p(1) &= p_{k+1} \\ p'(0) &= \frac{1}{2}(1-t)(p_{k+1} - p_{k-1}) \cdot \\ p'(1) &= \frac{1}{2}(1-t)(p_{k+2} - p_k) \end{aligned} \tag{3}$$

Then, the curve passing the two control points p_{k-1}, p_k can be noted as equation (4).

$$\begin{cases} p(u) = (-su^3 + 2su^2 - su)\,p_{k-1} + [(2-s)u^3 \\ \quad + (s-3)u^2 + 1]p_k + [(s-2)u^3 \\ \quad + (3-2s)u^2 + su)\,p_{k+1} + (su^3 - su^2)\,p_{k+2} \cdot \\ s = (1-t)/2 \end{cases} \tag{4}$$

According to equation (4), similarly the curve segment can be calculated passing p_k, p_{k+1} and the one passing p_{k+1}, p_{k+2} . Analysing this equation can get a certain interpolation curve of the curve segment that requires the coordinate values of four control points $p_{k-1}, p_k, p_{k+1}, p_{k+2}$ to obtain, so the original path is generated by combining groups of interpolation curves taking 4 control points as a unit, as in equation (4). In this formula, $u \in [0,1]$ and there is a variable t to control the tangent tensor of the being calculated spline, where t values zero.

3.2 SET OF PATH EDITING CONTROL POINT AND DEFINITION OF MOTION CYCLE STEPLENGTH

The step of evenly setting editing path control points on the original motion path is:

1. The number of user-defined path editing control points is N, $N \geq 4$, using the starting point and end point of the original motion path as the first and the last path editing control points, and it determines the time interval $\Delta t = frame\,time$ between frames of the original path (reading from BVH file) .

2. It calculates to determine the frame n_i where the i -th path editing control point located, in order to achieve an even set of the path editing control points. Wherein, $n_i = (i-1) \times \left\lfloor \frac{n}{N-1} \right\rfloor, i = 2 \cdots N-1$, n is the total frame number of the original motion sequence.

In the method above, it is about to remove part of the frames, to reduce the relative velocity between the root nodes of the frames. But if optionally removing a certain frame, it may cause incoherence of the motion action. We find that in most of the motion capture data there exists walking, running, small stride jumping and other actions. And actually, these actions are simple cyclical motion, such as walking that the action that left foot moves a step and then the right foot moves another step which can be regarded as a motion cycle. In order to keep the drawn motion action still smooth and continuous, after removing several sequence frames, we can take the motion cycle as a unit for deletion treatment to reduce the problem of motion actions excessively concentrating.

It defines the motion cycle step-length as follows: Any frame in which the left foot begins to lift is recognized as the beginning of a motion cycle. Through the left foot landing, and the right foot lifting and landing, the frame in which the right foot begins to land is defined as the final of the motion cycle. The spacing (in the plane of x, z) between the root nodes of the beginning frame and the final frame in the motion cycle is called the motion cycle step-length λ , as in Figure 3. Based on the definition of the cycle step-length we can also define the half-cycle step-length: Between the first frame any foot begins to lift and the first frame this foot lands and the other foot prepares to lift, the spacing (in the plane of x, z) between their root points is called the half-step-length of motion cycle, noted as $\lambda/2$.

Of course, it should be noticed that in a motion cycle step it may exist two cases as following. In Figure 3, if at the landing points p_{i1}, p_{i2} of the left and right foot, respectively comparing with the first frame the left or right foot lifting in the half-step-length containing these points, the human motion direction does not change to be a linear type or changes a small angle (≤ 60 degrees), it's called the general motion cycle step-length; If at the landing points p_{i1}, p_{i2} of the left and right foot, the human motion direction changes greatly(> 60 degrees), it's called special motion cycle step-length. To keep the important details of motion after editing, we will take the general motion cycle step-length as deletion objects.

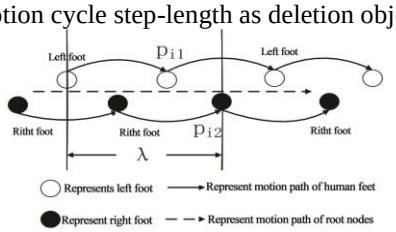


FIGURE 3 Motion cycle steplength

3.3 EDITING OF MOTION PATH AND FRAME ADJUSTMENT

In this paper, the path editing is realized by changing the position coordinates of the horizontal axis component and longitudinal axis component of path editing control points. In that, it does not change the vertical axis component to avoid artificially causing the phenomenon of feet suspending, feet into the ground and other issues. Therefore, it can only generate new motion path by carrying out Cardinal interpolation on the adjusted path

editing control point coordinates in the plane of horizontal axis and longitudinal axis.

It still needs to determine the position and motion orientation of each frame on the new path. The location treating process is as follow: the known number of original motion sequence frames is n, the number of path editing control points is N, and then it can calculate the incremental step length $step = (N-1)/n$ of u. Due to equation (4), in the curve segments $p_{k-1}p_k$, $p_k p_{k+1}$ and $p_{k+1}p_{k+2}$, every time u adding $step$ getting a point, separately calculating the horizontal axis and longitudinal axis coordinate components of the points, that is the location information of a certain frame root node in the new motion obtained by editing motion path.

About the adjustment of the human motion direction, it is known that in the motion path the motion direction of human at a certain moment is related to the slope of its position in the curve segment at that time. To this end, we will use the slope of each frame's root node position in the curve segment as the orientation of virtual human at this moment. So after editing the motion path, the position of the root node at time t also needs to be adjusted by changing of curve slope.

3.4 JUDGEMENT OF PATH CURVE SEGMENT ARISING FOOTSTEP DISHARMONY PHENOMENON

In the process of motion path editing, if the path change results in the moving velocity increasing of the root node (i.e. the curve segment length of the edited human motion path which passes the two control points is shorter than the length before editing) and the number of footsteps remains unchanged, it will cause footstep disharmony phenomenon, as in Figure 4. A, B are two adjacent edit path control points. L is the curve segment length of original motion path which passes these two points. L1 is the curve segment length of edited motion path which passes these two points. If $L_1 < L$, and because the number of frames between A, B is a fixed value, the moving velocity increasing of root node will surely cause the footstep disharmony phenomenon.

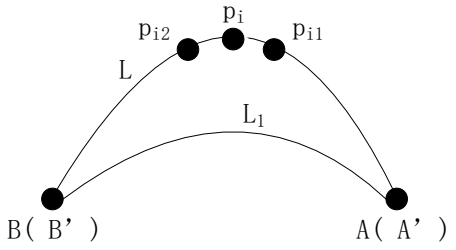


FIGURE 4 Variation diagrams of motion path before and after editing

Method of judging the specific curve segment arising footstep disharmony phenomenon in the motion path curve:

(1) It finds the root node collection $\{p_1, p_2, \cdots\cdots, p_n\}$ in the original path with the slope $k = 0$, and selects root node p_i from the collection in turn.

(2) It selects two frames which are $\pm m (5 \le m \le 10)$ frame away from the frame where p_i is. It calculates the slope k_{i1}, k_{i2} of root nodes p_{i1}, p_{i2} of these two frames in the original motion path curve and the slope k'_{i1}, k'_{i2} in the edited motion path curve. If one of k'_{i1}, k'_{i2} is positive and the other is negative and $k_{i1} \ne k'_{i1}, k_{i2} \ne k'_{i2}$, it executes (3), else it goes to (5).

(3) In the original path, that takes p_i as center to determine the two nearest control points A, B in the curve segment which passes the three points p_i, p_{i1}, p_{i2}.

(4) It calculates the curve segment lengths AB, AB' on the original and the edited path between the two control points A, B and the difference of the two lengths. If $AB - AB' > 0$, the curve segment in the edited path will arise footstep slip.

(5) Returning to (1), until finishing traversing the root node collection with the slope $k = 0$, the judging method ends.

3.5 ADJUSTMENT OF FRAMES AFTER REMOVING CYCLE STEPLENGTHS

In the newborn path, the specific method of adjusting two adjacent editing path control nodes after removing general cycle step-length can be illustrated by the following example. It assumes A, B are two adjacent editing path control points in the curve segment, which has been determined existing footstep slip phenomenon and noted as A', B' in the edited path. E, F is starting frame and ending frame of the N' to-be-deleted motion cycle step-lengths. $C(C'), D(D')$ are the editing path control points nearest to $A(A'), B(B')$, as in Figure 5.

(1) It separately calculates the number n_{EF} of to-be-deleted frames in curve segment EF. In the curve segment $C'E$ there contains $n_{C'E}$ frames. In the curve segment FD' there contains $n_{FD'}$ frames.

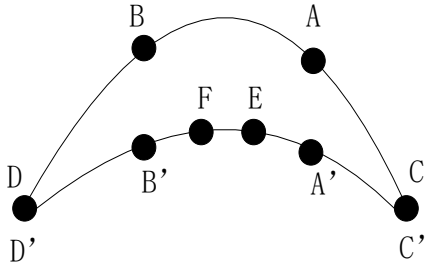


FIGURE 5 Removing Step-length

(2) It calculates the time T_{EF} that the frame sequence contained in curve segment EF spends for broadcasting, the time $T_{C'E}$ that the frame sequence contained in curve segment $C'E$ spends for broadcasting, and the time $T_{FD'}$ that the frame sequence contained in curve segment FD' spends for broadcasting.

(3) After removing the cycle step length in EF, it adjusts the interval length of the frames in $C'E$, FD' by equation (5) (6) to make it smooth and continuous.

$$j_i = \left\lfloor i \times \frac{T_{C'E} + 1/2\, T_{EF}}{n_{C'E} \times \Delta t} \right\rfloor \quad i = 1,2,3 \cdots\cdots n_{C'E}, \tag{5}$$

$$l_k = \left\lfloor (k+1) \times \frac{T_{FD'} + 1/2\, T_{EF}}{n_{FD'} \times \Delta t} \right\rfloor + j_{n_{C'E}} \quad . \tag{6}$$

$$k = 0,1,2,\cdots\cdots n_{FD'}$$

The equation (5) represents: after removing cycle step length, it determines the frame number j_i how far the to-be-adjusted frame, which is i-th frame apart from C' in the curve segment $C'E$, will be apart from C' in the newborn motion path after editing. The equation (6) represents: after removing cycle step-length, it determines the frame number j_i how far the to-be-adjusted frame, which is the k-th frame apart from F in the curve segment FD', will be apart from C' in the newborn motion path after editing.

(4) It takes the position information and motion direction information of the root node on frame j_i or l_k, to be the position information and motion direction information of the root node on the i-th or k-th frame after removing general cycle step-length, and takes equation (1) to calculate the position information of each joint node in the frame.

3.6 CONSTRAINT REESTABLISHMENT OF THE FLOOR FRAME

In the original motion sequence, the positions of the end effecters (feet) at two or more continuous moments are at the same landing position in space. However, after treated by path editing, it will make the human feet in different positions at these few moments appearing the footstep slip phenomenon. In order to ensure the quality of motion actions unchanged before and after motion path editing, in the newborn path that has deleted the cycle step-length, and it needs to constrainedly re-establish the foot landing frames (floor frame) inside the half-cycle step-length in the original path.

The judgment of human foot landing is distance, to check if the vertical distance between foot and ground is close to zero. The determining formula is equation (7).

$$\left|\vec{d} \cdot \vec{n}\right| \le H \ . \tag{7}$$

In equation (7), $\vec{d}$ represents distance vector from end effecter (foot) to the ground, $\vec{n}$ represents vertical unit vector and H represents the set threshold of landing detection. After landing detection, it will get the serial numbers of those frames at the foot landing moment. This allows the frame of the original motion sequence to classify, taking half-cycle step-length as unit to establish a collection S_L of frames of left foot landing moments and collection S_R of frames of right foot landing moments as landing constraint frames.

Setting the position of foot in the first landing constraint frame inside every motion half-cycle step-length on the new path as the position of foot in all the landing constraint frames inside this half-cycle step-length, it uses per-frame Inverse kinematics (inverse kinematics edited according to frame) [15] to calculate the rotation changes of other joints nodes to meet the position requirement.

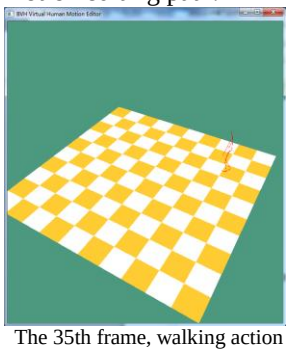
4 Experimental results and analysis

In this paper, it verifies the proposed virtual human motion path editing method based on the motion cycle step-length by experiment. Experimental environment: Operating system: Microsoft Windows XP; CPU: IntelQuadQ9550 2.8GHz; Memory: 3GB; Program compiler environment: Microsoft Visual Studio 2008. The Human body motion capture data is collected by the OptiTrack whole body motion capture system with 12 cameras by the frame rate of 60 frames/Sec, and is stored by BVH file format.

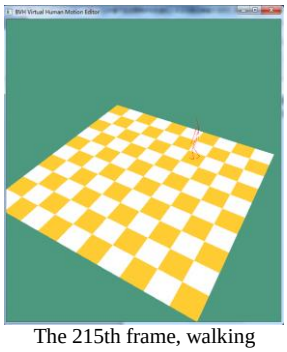
In reality, the human motion is often complex but not single cycle motion as walking, running, etc.. To illustrate the generality and the effectiveness of the proposed method, we take a motion capture sequence (as in Figure 6) with 740 frames mixed with walking and jumping by the right leg for verification. In 1- 660 frames, it records the walking action starting with left foot, In 661-710 frames, it records the jumping action by the right leg, and in 711-740 frames, and it records the action of left foot walking a step after right foot landing. In this hybrid action sequence, walking is general cycle action.

During the process of building original motion path, the root nodes of each frame are as control points of Cardinal spline curve. For the motion capture data is the BVH file format, it does Cardinal spline interpolation to the positions of root nodes in the plane xoz to obtain the original motion path, indicated by the solid line in Figure 7. Before editing the path, it selects 4 control points to be path editing control points from the original motion path which is a Cardinal spline interpolation curve. According

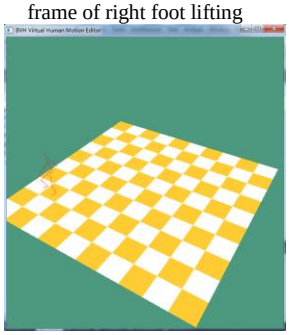
to the setting method of path control point in this paper: two control points are the starting terminal C (root node position of frame 1) and ending terminal D (root node position, frame 740) of Cardinal spline interpolation curve, while the other two control points A, B are the root node positions of 247-th and 494-th frames. Now, it edits the path by changing the positions of path editing control points: in the x axial direction, the point C moves 35 units down, the point A moves 140 units down, and point B moves 100 units down; in the z axial direction, point C moves 5 units to the left. The new motion path curve is generated by carrying out Cardinal spline interpolation to these four control points, indicated by dotted line in Figure 7. As shown in Figure 7, the two motion path curves are smooth, and the transition between action sequences of right leg jumping and walking frames is natural, indicating the effectiveness of the Cardinal spline interpolation for virtual human motion editing path.



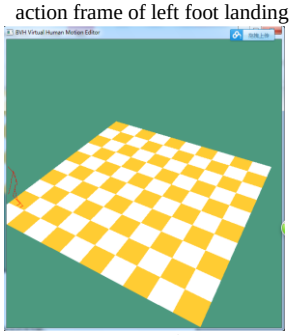
The 35th frame, walking action frame of right foot lifting



The 215th frame, walking action frame of left foot landing



The 685th frame, jumping action by the right foot



The 730th frame, walking action

FIGURE 6 Frame sequence of original motion

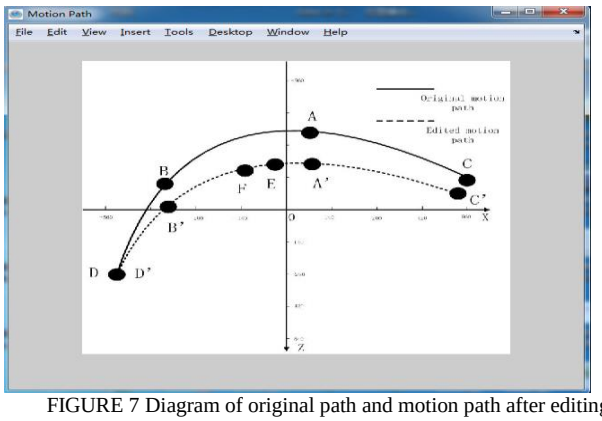
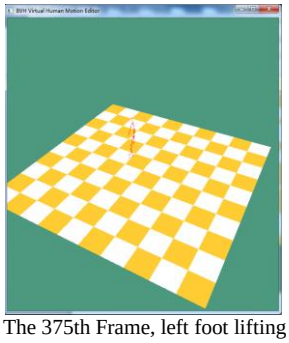
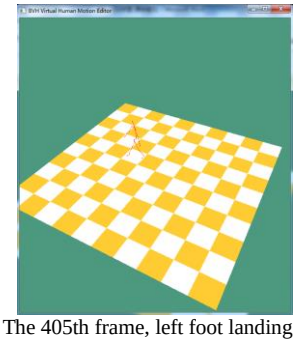


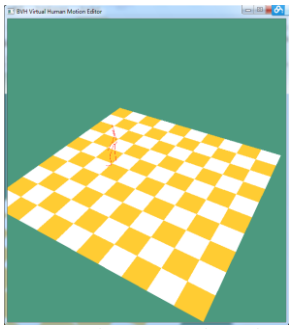
FIGURE 7 Diagram of original path and motion path after editing



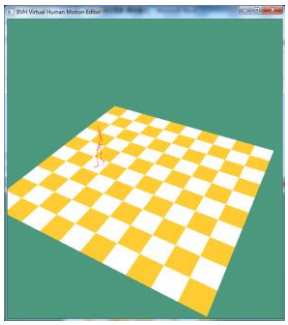
The 375th Frame, left foot lifting



The 405th frame, left foot landing



The 420frame, right foot lift



The 450 frame, right foot landing and left foot lifting

FIGURE 8 Frame sequence of footstep disharmony phenomenon

That is shown in Figure 8 is a motion frame sequence generated after editing the path. In which, it is found that there happens significant footstep disharmony phenomenon from the 375-th frame to the 450-th frame and human action is not natural and smooth. The reason is because after adjusting the editing path control points neither root nodes position or root nodes orientation of each frame are adjusted, and the shortening of curve length between the two points causes the frame aggregation to be too high. So, firstly, after adjusting the editing path control points, the root nodes position and orientation of each frame must be adjusted. Secondly, the problem of frame aggregation being too high should be solved by cutting the general motion cycle step-length. According to the judging method of path curves existing footstep disharmony phenomenon mentioned in section4.4, it can automatically determine the footstep disharmony phenomenon occurs in the curve segment with the path edit control points A, B as endpoints. It's determined by: the length of the curve segment with the endpoints A, B is 531cm, and after path editing its length is 510cm, the curve length becoming shorter; while the slope of the 280-th frame is 0, and the slopes of the 275-th frame and the 285-th frame which are 5 frames away from that point at its both sides are one positive and one negative. The 280-th frame is just in the curve segment with the two paths editing control points as starting and ending. This illustrates that the proposed method in this paper can achieve the automatic determination of path curve segment existing footstep disharmony phenomenon.

In the path curve segment AB, there exists general cycle step-length, and the starting frame and ending frame of

the first general cycle step-length are the 300-th frame and 360-th frame in which the root nodes identified by points E, F are. According to equation (2), it can get the difference between the lengths of the curve segment with the two path control points as terminals before and after the path editing. The ratio that the difference is divided by the general motion cycle step-length λ (here sets λ to be 100 cm, and it can be adjusted according to the different motion human subjects) is 0.21. So, it should delete a cycle step-length that is the motion frame between the two points E, F . Then, according to equation (5) and (6) to adjust the position and orientation of the frames between the curve segments EC, FD, the result is shown in Figure 9. It can be seen from this that although this method deletes the general motion cycle step-length, it does not affect the animation effects of right leg jumping, and this hybrid motion remains consistent and smooth.

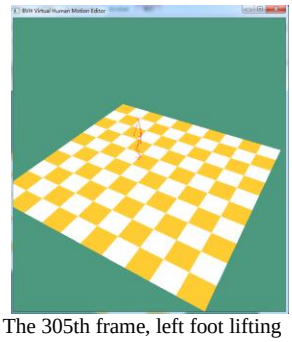
In addition, in Figure 9, we find no footstep slip phenomenon occurs after path editing. This is because it uses the equation (7) to find each floor frame, and using inverse kinematics to constrainedly re-establish, to make the foot in the frames of these types at the status of landing before and after path editing.

Of course, the algorithm can also be applied to other virtual human motion path editing which contains cycle motion such as running, jumping, walking and other complex motion. It can effectively remove footstep slip phenomenon to obtain satisfying human motion animation effect.

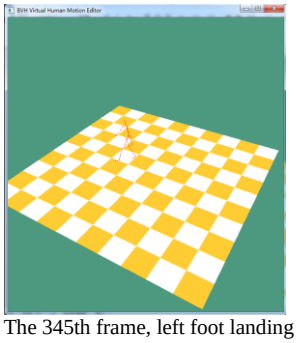
5 Conclusion

To improve the reusability of motion capture data, here presents an editing method of motion path: the Cardinal spline interpolation is used to extract the original path and set the editing path control points; it alters the positions of path editing control points to generate a new path, and adjusts the original motion sequence frame on the new path; it aims at the human foot disharmony phenomenon caused by the shortening of the motion path curve segment after editing path, based on motion cycle step-length, proposing the way of deleting the frame of general motion cycle frames and adjusting the root nodes of every frame which contains that curve segment and between the two nearest editing path control points in the motion path to remove it; finally it does constraint reestablishment to all floor frames with cycle step-length, to reduce footstep slip phenomenon. The experiment results show that: the motion path editing method is simple, with less calculation, and easy to operate. The generated virtual human motion is natural and smooth. It can effectively eliminate the footstep sliding phenomenon.

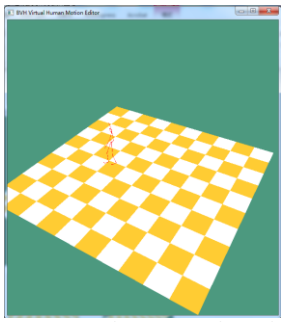
However, this method does not consider that in the motion process, with the dramatic changes of path curvature, the centre of gravity of virtual human will be shifting to cause the human body tilt phenomenon; and also does not consider the dynamic approach of human motion path when it meets obstacles. These two areas are the place that in the future we should work hard in.



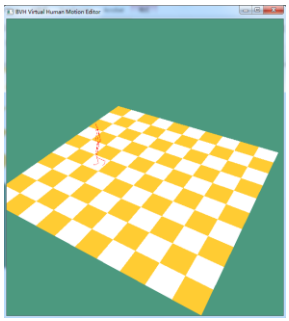
The 305th frame, left foot lifting



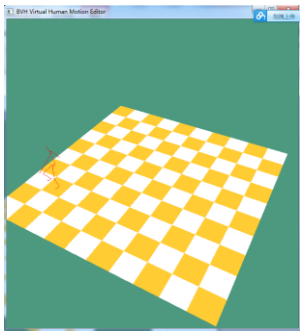
The 345th frame, left foot landing



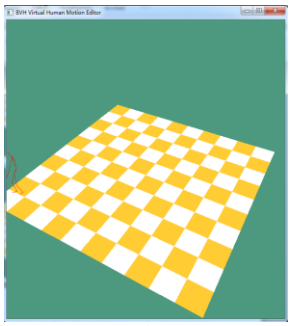
The 360th frame, right foot lifting



The 390th frame, right foot landing and left foot lifting



The 625th frame, jumping action by the right foot



The 670th frame, jumping action by the right foo

FIGURE 9 Frame sequence of human motion after editing path

Acknowledgments

This work was supported in part by a grant from the Starting Project of Doctor of University of Shanghai for Science and Technology (No.1D-13-309-005), Funding Scheme for Training Young teachers in Shanghi Colleges in 2013(No.51-14-309-101), Innovation Program of Shanghai Municipal Education Commission (No.13ZZ111), the bidding project of Shanghai research institute of publishing and media (No.SAYB1408) funded by the National Higher Vocational and Technical Colleges construction project of the Shanghai Publishing and Printing College.

References

[1] Jung S, Bajcsy R 2006 A Framework for Constructing Real-time Immersive Environments for Training Physical Activities *Journal of Multimedia* **1**(7) 9-17

[2] Boon-Seng C, C Lap-Pui, Y Kim-Hui 2009 Progressive Transmission of Motion Capture Data for Scalable Virtual Character Animation *Proc IEEE Int Symp Circuits Syst* Taipei, Taiwan,1461-4

[3] Gleicher M 1999Animation from Observation: Motion Capture and Motion Editing *Computer Graphics* **33**(4) 51-4

[4] Kovar L, Schreiner J, Gleicher M 2002 Footskate Cleanup for Motion Capture Editing *Proceedings of Computer Animation* TX, United states 97-104

[5] Lockwood N, Singh K 2011 Biomechanically-inspired Motion Path Editing *Proceedings - SCA 2011: ACM SIGGRAPH / Eurographics Symposium on Computer Animation* Vancouver, Canada, 267-76

[6] Lu X, Li Y, He H 2005 Reusability of Virtual Human's Walking Motion Capture Data based on Path Transformation *Journal of System Simulation* **17**(SUPPL.) 170-3

[7] Liu S, Sun J 2006 Improved Motion Path Editing *Journal of Computer Applications* **26**(12) 2680-2

[8] Luo Z, et al. 2002 Space-time Constraints based Motion Editing and Motion Retargeting *Journal of Computer-Aided Design and Computer Graphics* **14**(12) 1146-51

[9] Chen Z, et al. 2006 Editing Human Motion Path *Journal of Computer Aided Design & Computer Graphics* **18**(5) 651-5

[10] Safonova J K, Hodgins 2007 Construction and Optimal Search of Interpolated Motion Graphs *ACM Transactions on Graphics* **26**(3) 106-11

[11] Meredith M, Maddock S 2001 *Motion Capture File Formats Explained. Department of Computer Science* Technical Report CS-01-11, University of Sheffield, 1-36

[12] Grigorieff R D 2013 On Cardinal Spline Interpolation *Computational Methods in Applied Mathematics* **13**(1) 39-54

[13] Gleicher M 2001 Comparing Constraint-based Motion Editing Methods *Graphical Models* **63**(2) 107-34

Authors

Wenju Wang, born on March 27, 1979, China

Current position, grades: lecturer at University of Shanghai for Science and Technology and research fellow of Shanghai research institute of publishing and media, China.
University studies: PHD degree in computer application technology from Tongji University, China, in 2012.
Scientific interests: virtual reality, computer animation and computer graphics.

Liujie Sun, born on August 15, 1965, China

Current position, grades: professor at University of Shanghai for Science and Technology, China.
University studies: Ph.D. degree in Optical engineering from University of Shanghai for Science and Technology, China in 2008.
Scientific interests: information security and image processing.

Zhang Xuan, born on April 28, 1982, China

Current position, grades: lecturer in the Musicology Department, Shanghai Conservatory of Music.
University studies: Zhang Xuan received her Master's Degree in music composition from National Academy of Chinese Theatre Arts, China, in 2007.
Scientific interests: Chinese traditional music and digital music.

Early fault warning mechanism based-on association rules in server clusters management system

Yumei Ning*, Zhenguo Ding, Ping Zeng

School of computer science and technology, Xidian University, Xi'an 710071, Shaanxi, China

Received 1 June 2014, www.cmnt.lv

Abstract

With the rapid development of Internet and business internal information technology, the problem of server cluster fault early warning becomes important in server clusters management. To solve the problem, a server cluster fault early warning system (SCFEWS) based on event association rules is presented. This system is mainly based on the fault event association tree. First obtain event relationship and association rules. Construct early warning events associated tree with system logs and association rules, and store it in binary tree linked list. Then by using the warning event filtering algorithm, redundancy fault early warning events are filtered out and only source event for early-warning notification are kept. Experiment shows the proposed algorithm can effectively improve the accuracy of fault location for server clusters management.

Keywords: Server Cluster, Early Warning, Fault Association Tree, Warning Event Filtering Algorithm

1 Introduction

With the rapid development of Internet and business internal information technology, enterprise server clusters are gradually used to store a big data processing to meet growing business needs. The server clusters size and the number of servers upgrade and expand constantly with the development of enterprise business. But an increasing number of server components makes server clusters management difficulty and improves the probability of server clusters fault at the same time, which bring tremendous challenge to the entire server clusters manageability. So it has great significance to improve the cluster service availability about how to guarantee the stability of the server clusters system and how to find and promptly remove the hidden dangers in cluster system before fault occur.

Now various faults in server clusters are usually detected and located manually lacking server cluster fault warning management mechanism. So it is the priority problem monitoring real-time status for server cluster resources and giving corresponding warning to the administrator. With faults warning system, administrator can monitor server clusters comprehensively, detect server clusters anomalies timely, locate the fault timely and accurately, take action to avoid or reduce the risk and prepare for effective response [1]. Nowadays, although the major IT giants have their own server clusters management system, it does not open to all. And there is no universal solution for server cluster faults warning.

Early Warning Systems play a highly critical role in monitoring, prediction and reaction on upcoming disasters [2]. Early Warning Systems are found in many areas: natural catastrophic detection [3, 4] of tsunami, flood and earthquakes; detection of sudden and significant economic changes [5]; engineering and scientific area [1, 6], etc. In spite of variations between monitored domains, their main goal is similar: to reduce economic losses and mitigate the number of deaths from disasters by delivering information which allows people and organizations to prepare for emerging disasters [1]. So Early Warning Systems can also be used in server cluster management. Upcoming faults can be monitored and predicted by using Early Warning Systems.

This paper presents a server cluster management system based on the idea of Early Warning Systems. In a Server Clusters Fault Early Warning System (SCFEWS) put forwarded in this paper, fault warning module not only monitor real-time status for server cluster resources but also send early-warning information to the administrator before a fault occurs. It guarantees the reliable and stable running and the processing big data of the server clusters system.

In this paper, SCFEWS based on the warning event tree is used to filter out triggered fault early warning events and retain only source event for early warning notification. And the destination is avoiding the redundant warning events.

The remainder of this paper is organized as follows: Section 2 describes the formulation of the fault early-warning problem in server cluster system. Section 3 gives a precise problem statement and details our algorithm of the SCFEWS. Section 4 is the experimental evaluation in a few respects: its performance and the validity. And the Section 5 concludes the paper with a summary of our findings.

* ***Corresponding author*** e-mail: ssss@dddd.com

2 The proposed algorithm

This paper presents a server clusters fault early warning system in network resource management, which relies on client agents [7, 8] and provides early warnings against equipment failure such as CPU utilization, memory utilization, disk space, etc.

2.1 EARLY WARNING SYSTEM

The basic idea behind early warning [1] is that the earlier and more accurately we are able to predict potential risks, the more likely we will be able to manage and mitigate disasters.

An effective early warning system needs an effective communication system. Early warning communication systems are made of two main components:

(1) Communication infrastructure hardware that must be reliable and robust.

(2) Appropriate and effective interactions among the main actors of the early warning process.

Communication infrastructure for SCFEWS uses intelligent mechanisms [9] for communication and scatter of fault alerts through agents on which the server clusters can be remotely monitored and managed. SCFEWS use association rules tree for fault prediction. With them SCFEWS can better guarantee the server clusters running stable and reliable.

2.2 EXPERT SYSTEM

Expert system is intelligent computing Machine application system that has a lot of specialized knowledge and experience. Expert System contains the following modules [10].

(1) The knowledge base, where the knowledge of experts is represented using rules, frames, semantic networks, and first-order logic based methods, etc.

(2) The working memory module that stores the input data and the information generated by the processing of rules.

(3) The inference engine where the processing of the rules and the reasoning of the Expert system take place.

(4) The user interface module facilitates the interaction between the user and the Expert system.

(5) The knowledge acquisition facility provides the user with appropriate useful tools during knowledge acquisition procedures.

(6) The explanation module allows the Expert system to present its reasoning regarding its conclusions.

The main components involved in Expert system are the knowledge base and the domain expert.

2.3 EVENT CORRELATION

Event Correlation is an important fault localization strategy. The basic idea is to correlate multiple events to a single concept to filter out unnecessary events and unrelated events, reduce the amount of information presented to the network administrator, and increase the semantic content of the information represented by the association process.

2.4 ALGORITHM DESCRIPTION

The thought of the SCFEWS is showed below. Please see Figure 1. Initial stage, there is small amount of the managed server clusters monitored by management centre. With the expansion of the managed server cluster scale, the probability of server cluster faults is increasing accordingly. Therefore, corresponding early warning notification should be sending out before a fault occurs. It is important finding usable system early-warning information from all monitoring information. Thus, with the warning notification system administrator can take early preventive measures rather than remedial measures before a server fault occurs. And that is the processing of SCFEWS. Thus, the system robustness and the scalability to manage server cluster fault are improved.

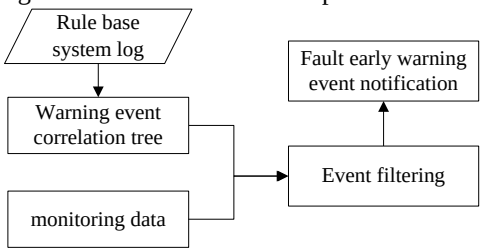


FIGURE 1 Structure chart of the SCFEWS

There are four mainly components in SCFEWS: the monitoring database, the warning event tree, event filtering module, and fault early-warning event notification module.

The monitoring data is comprised of server hardware such as temperature, fan, voltage, and power etc., system resources such as CPU utilization, memory utilization, disk space, and network speed etc., service resources such as processes, system services and application services etc. and network connectivity etc.

The warning event correlation tree is used to store the relationships between the various fault phenomena and is dynamically maintained for there is a certain correlation between various fault phenomena. The relationships data come mainly from the association rules for warning event and system logs.

Event filtering module: in server clusters, the generation of fault is often not a single fault phenomenon, but multiple fault phenomena occurring at the same time, that is to say cascading fault phenomenon caused by single fault. There is a certain correlation between fault phenomena, so event filtering module need to filter out triggered fault early-warning events and try to keep only source event for early-warning notification.

Fault early-warning event notification module sends the early-warning notification to system administrator for rapid positioning faults. And early preventive measures

are taken to prevent fault risk caused by server clusters before a server fault occurs.

3 Problem solution

The Server Clusters Fault Early Warning System (SCFEWS) aims to filter fault early-warning events according to generated warning event tree and try to keep only source of information for early-warning event for notification. The system consists of a server-side proxy equipment, communication modules and early warning systems management console.

The follow is the specific approach. Real-time monitoring various state of system equipment via proxy server clusters, storing all kinds of the collected information, judging its status data according to the threshold value. When monitoring event is judged to be early warning event, it will be encapsulated into object-oriented fault early warning event and delivered to early warning filtering module. Warning filtering module will filter out triggered fault early warning events according to generated warning event tree, and only retain source early warning event for notification of early warning event.

The emphasis of the SCFEWS is early warning filtering algorithm and the warning event tree with which early warning events is handled with.

3.1 ALGORITHM DESCRIPTION

Usually there are three main methods for early warning judging. They are threshold judging, Status Judgment and shazam judging. This paper uses threshold judging and shazam judging. Early warning judging [1] model is mainly responsible for processing the data obtained from the condition monitoring model. It judges the data according to the threshold value and quick change. When monitoring event is judged to be early warning event, the detailed event information will be encapsulated into fault event and delivered to memory fault events waiting for subsequent processing by early warning filtering module. Adopt object-oriented method [11] and construct quaternion form $E=(U,S,O,D,T)$ for storing early warning event-related information to descript the fault event in this paper. And symbol definition is as follows:

$U(U\text{ei})$: Identification of early warning events means such events is monitoring by the corresponding event listener for subsequent processing.

S=(Source) : Source of early warning events.

$O=(Object)$: Object of fault event. That is the fault device type.

$D=(D\text{escription})$: Fault description in detail, including fault identification, fault monitor and fault description, etc.

$T=(T\text{ime})$: Time for early warning events.

Early warning judging process is as follows:

(1) Get effectively monitor data from status monitoring model.

(2) Select the appropriate judging method based on monitoring data type.

(3) If the monitor data exceeds a threshold value or if it meets the fast changing limit, the monitor data is defined as early warning information and stored, else go to steps 1.

3.2 EARLY WARNING ASSOCIATION RULES TREE

When a fault event occurs in server clusters, it may cause other fault events closely related. There may be a relationship between them, such as time correlation, physical correlation [12], etc. Fault event sequence [13] is the combination of some fault events according to certain rules in a certain time interval.

Fault events associated tree is constructed based on the Fault event sequence in the rule base.

3.2.1 Get association rules

System administrators formulate relationship between warning events before server cluster fault early warning system is running, using algorithms inference log data and obtain relationship between warning events.

Association rules reveal the unknown dependencies between data items. According to the mining associated relationship, information from a data object can be used to infer the information about another data object [12, 13]. Describe as following.

Set $I=\{i_1,i_2,...,i_n\}$ is a project collection. X and Y represent item sets. X is the precursory item event and Y is the subsequent item event. Then Association rules is defined an implication relationship of the following forms: $X\rightarrow Y, where X\subset I, Y\subset I, and X\bigcap Y=\varnothing$.

3.2.2 Construct early warning events associated tree

Read the warning events association rules in rule base, according binary tree structure to construct and store the structure of early warning events associated tree. When management control starts, binary tree linked list is to be constructed and the warning events associated tree is to be generated according to the algorithm. Once relationship between early warning events in rule base change, the warning events associated tree need to be re-generated.

The node of early warning events is mainly composed of four components.

String *id* represents early warning event.

EventNode *child* represents the child node of early warning event.

EventNode *brother* represents the brother node of early warning event.

Variable *occur* represents whether the early warning event occurred. See Definition (1).

$$occur=\begin{Bmatrix}0,\text{event does not occur}\\1,\text{event occur}\end{Bmatrix}. \quad (1)$$

Binary Tree linked list for the early warning event is structured as follows in Figure 2.

```
public class WarningNode{
    String id;
    WarningNode child;
    WarningNode brother;
    int occur;
}
```

FIGURE 2 Warning node class

3.2.3 The algorithm of warning event correlation tree

Construct the algorithm of warning event correlation tree depending on the warning events association rules in rule base. Define *endFlag* to be the tag of scanning rule database. See Definition (2).

$$endflag=\begin{Bmatrix}0,\text{there exist unprocessed event}\\1,\text{scan end}\end{Bmatrix}. \quad (2)$$

The algorithm process is as follows:

(1) Generate Root node, set *endFlag* value to 1, and began to build children brothers linked list.

(2) Scan the rule base, add early warning event node whose precursor event is empty and subsequent events is not empty to the second layer of early warning events associated tree, and then set *occur* value to 1.

(3) Set *endFlag* value to 1, scan the rule base.

(4) If there is the early warning event node whose precursor event is not empty and *occur* value is 0, add the node to the sub-linked list of precursory early warning events node and set *occur* value to 1. If the node does not exist, the *endFlag* value will to be set 0 waiting for next scan processing.

(5) If *endFlag* value is 1, set all the warning events *occur* value to 0 within the linked list, scan end and the algorithm end. Go to steps (3) if *endFlag* value is 0.

3.3 EARLY WARNING EVENT FILTERING ALGORITHM

The early warning filtering algorithm need to depend on early warning association rules tree, and it mainly adopts the thoughts of expert systems. The aim of the algorithm is obtaining only source of information for early warning event for notification. Relying on early warning association rules tree, the algorithm filters all generated early warning events stored in filtering database, and eliminates redundant warning events.

The flow sheet for early warning filtering algorithm is shown in Figure 3.

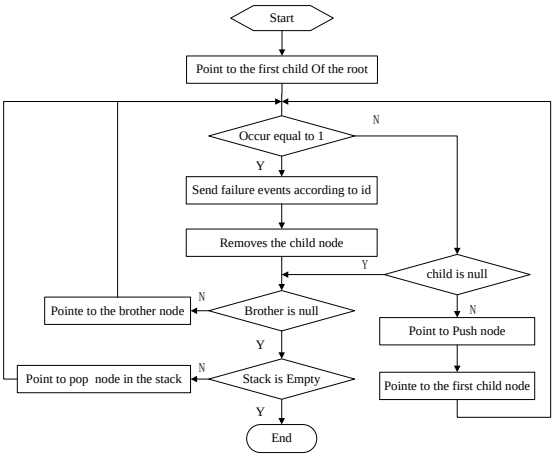


FIGURE 3 Flow sheet for filtering algorithm

The process of the early warning filtering algorithm description:

(1) Determine whether the event data in rule base changes.

(2) Call the early warning association rules tree algorithm and re-construct the warning event correlation tree, if the change exists. Call the warning event correlation tree constructed above, if the change does not exist.

(3) Traverse the warning event correlation tree, mark the warning event already stored in storage. Set variable *happen* value to 1.

(4) Traverse the warning event correlation tree using depth-first traverse principles. If there is a warning event whose precursor event does not occur or is empty, deliver it to early warning notification module. At the same time, if it has sub-event, remove all of its sub-events.

This paper uses a timer to trigger the early warning filtration module. The module will automatically carry out once filtering operation at regular intervals, and the administrator can set the interval. Finally the early warning event filtered will to be send to the warning notification module and there to be dispatched to administrator. Please see Figure 4.

```
LinkedList<WarnRule> events = new LinkedList<WarnRule>();
for(WarnNode w : events)
    markOccor(w);
Iterator<WarnEventRule> iterator = warnEventList.iterator();
filterEvent(iterator);
public void filterEvent(iterator<WarnRule> iterator){
    while(iterator.hasNext()){
        WarnRule occorRule=iterator.brother();
        if(occorRule.occor == true)){
            getWarnNodeSender().SendNode(changeWarnNode(occorRule));
            occorRule.child=null;
        }else  if(occorRule.child!=null){
            iterator=occorRule.child.iterator();
            filterEvent(iterator);
        }
    }
}
```

FIGURE 4 The early warning filtering algorithm

The early warning filtering algorithm can not only ensure the manageability of SCFEWS, but also ensure redundancy warning information not to be sent out.

4 Results

In order to verify the reliability of the Server Clusters Fault Early Warning System (SCFEWS), we programmed and simulated in the Network experimental Environment as shown in Figure 5. Management console was mainly consisted of management server and database server. Managed devices included five sets of managed servers, four managed PC and two mobile terminals. They were Lenovo IdeaCentre, AMOI N820 and IBM X5 respectively. The testing and analysis is only for server cluster fault warning system functional modules.

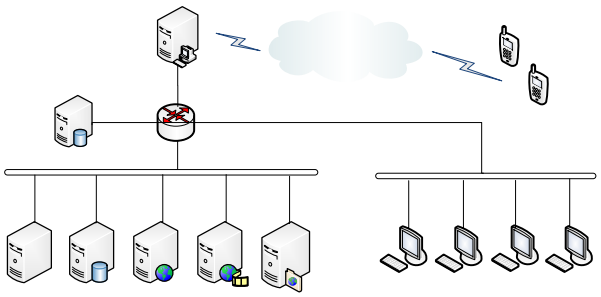


FIGURE 5 Network experimental Environment

For the convenience of testing the rationality and effectiveness of the model algorithm, the warning event correlation rules for the experimental system is part of preset in this paper as shown in Table 1. Program algorithm is omitted here.

The model algorithm is analysed from the performance, the warning filtration effect and the network utilization in the network experimental environment above.

TABLE 1 Correlation rules definition

No	Early warning UEI	Early warning description	precursory early warning	precursory early warning description
1	AgentDown	Client Agent shut down	DeviceDown	Server shut down
2	GetDataFail	Fault to getting data	AgentDown	Client Agent shut down
3	ServiceClosed	Host server is shut down	DeviceDown	Server shut down
4	DeviceUnreachable	Host network is unreachable	DeviceDown	Server shut down
5	ApacheDown	Apache Application stops	DeviceDown	Server shut down

Process is as follows: After the Device Agent of Managed device is launched, its various types of information is collected and incorporated into the monitoring database. If the monitoring data exceeds the defined threshold, then the event data is encapsulated into early warning and early warning is transmitted to the early warning filter module for pre-process based on the warning event correlation rules. And the early warning notifications being filtered are sent to the management client.

Experimental results show that the system is running continuously seven days in monitoring, during which the system produced a total of 44 early warning events. There are only 25 early warning events being warning information after filtering module filters the 44 early warning events, which is basically close to the original number of early warning events. Thus early warning filtration achieves a desired effect and the effectiveness of the warning filtering algorithm is proved in this paper. Please see Figure 6.

The following is the SCFEWS performance experimental.

Server is usually used to provide services. Because the proxy client software runs on the server equipment, and communicates with the management console, there must be low resource occupancy not to impact on performance of the server. So the CPU usage of proxy client and procedures network occupancy rate is experimented.

Server device can obtain CPU usage via top and other procedures when the proxy client software is installed on the managed device and turned on in the daemon. By setting the management console monitor polling cycle and sampling the CPU utilization, the CPU utilization of proxy client server is measured. Please see Figure 7.

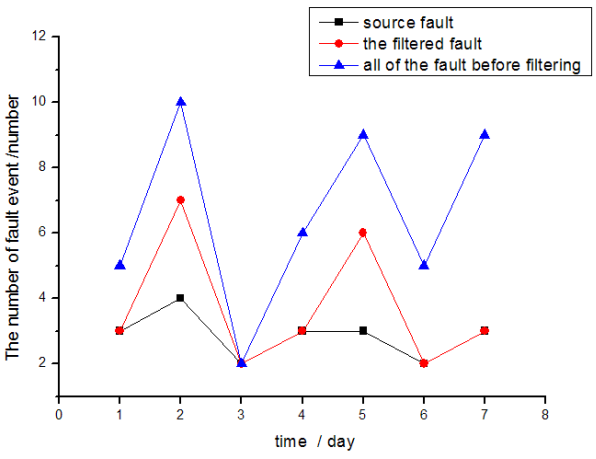


FIGURE 6 Comparison of early warning event

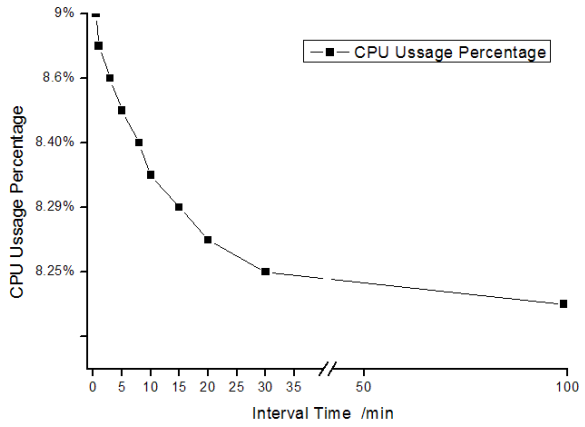


FIGURE 7 CPU utilization of server proxy client

It can be seen from the above figure that the monitoring polling cycle time is inversely proportional to the CPU utilization. Once polling cycle time shortens, CPU occupancy rate goes up. When proxy software is not running, CPU occupancy rate maintains at about 8.3%. When the proxy software is opened and the polling cycle is set to be thirty minutes, at this time, although management client does not send the command, the CPU utilization increases slightly for activity agent software monitoring ports and the activity agent process taking some of the resources. Along with shortened polling cycle, proxy software is more and more frequent to get early warning surveillance system data, so the value is also gradually increasing in CPU utilization. When polling cycle is set to half a minute, the CPU utilization maintains at about 9.0%. Thus the agent software on server equipment occupies very little CPU resources, and it will not have big impact on the server computing performance.

Server device can obtain the network rate via iftop and other procedures when the proxy client software is installed on the server device and turned on in the daemon.

Sampling the network speed during the agent software opening and the management end different monitoring polling cycle stetted, server agent can be analysed terminal network rate which is showed in Figure 8.

It can be seen from the above figure that the monitoring polling cycle time is inversely proportional to the network traffic. Once polling cycle time shortens, CPU occupancy rate goes up gradually. When proxy software is not running, network traffic maintains at about 0.28KB/S. When the proxy software is opened and the polling cycle is set to be thirty minutes, the network traffic decreases slightly around 0.31KB/S. Along with shortened polling cycle, network traffic is also gradually increasing. When polling cycle is set to half a minute, the network traffic maintains at about 0.8KB/S. Thus, the software routine has slight impact on the server computing performance when interval time is set a certain value.

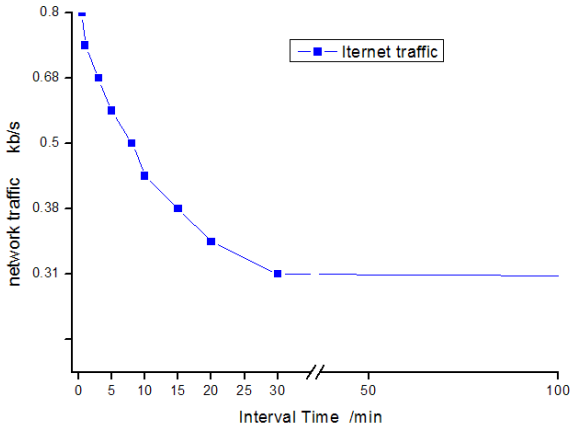


FIGURE 8 The network traffic of server proxy client

5 Conclusion

It can be seen from the above experimental data that the warning model algorithm presented in this paper has certain advantages in fault advance warning. The Server Clusters Fault Early Warning System (SCFEWS) designed in this paper integrating hardware status monitoring, system resource monitoring, early warning service monitoring, device control and device resource management has achieved the early warning and management of server clusters, also has not a major impact on the server computing performance. It greatly facilitates the operation and maintenance of the server clusters management. The algorithm has high value for reference to large-scale server clusters management.

The SCFEWS designed in this paper need to deal with system fault human involvement after the early warning events occur. Failover automatically and automatic downgrade functionality will to be introduced into early warning in the next research, trying early warning automated processing and restore.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (No.JY10000917003).

References

[1] Grasso V 2012 *Early Warning Systems: State-of-art Analysis and Future Directions* UNEP
[2] López V F, Medina S L, Juan F de P 2012 Taranis: Neural Networks and Intelligent Agents in the Early Warning against Foods *Expert Systems with Applications* **39** 10031-7
[3] Hong Y, Adler R F 2007 Towards an Early-warning System for Global Landslides Triggered by Rainfall and Earthquake *International journal of remote sensing* **28**(16) 3713-9
[4] Bartosz Balisa, Marek Kasztelnikb, Marian Bubaka 2011 The UrbanFlood Common Information Space for Early Warning Systems *Proc. Conf. on Computational Science* Japan 96-105

[5] Ng G S, Quek C, Jiang H 2008 FCMAC-EWS: A Bank Failure Early Warning System based on a Novel Localized Pattern Learning and Semantically Associative Fuzzy Neural Network *Expert Systems with Applications* **34**(2) 989-1003
[6] Nasonov D, Nikolay B 2014 Hybrid Scheduling Algorithm in Early Warning Systems *Proc. Conf. on Computational Science* Nevada, USA 1677–87
[7] Corchado J M, Tapia D I, Bajo J 2012 A Multi-agent Architecture for Distributed Services and Applications *International Journal of Innovative Computing, Information and Control* **8**(4) 2453-76
[8] Andreas P, Martin F 2004 Secure Network Management within an Open Source Mobile Agent Framework *Journal of Network and Systems Management* **12**(1) 9-31
[9] Dokas I M, Karras D A, Panagiotakopoulos D C 2009 Fault Tree Analysis and Fuzzy Expert Systems: Early Warning and Emergency Response of Landfill Operations *Environmental Modelling and Software* **24**(1) 8–25
[10] Turban E 1997 *Decision Support and Expert Systems: Management Support Systems* Prentice-Hall
[11] Rajkumar Buyya 2000 PARMON: a Portable and Scalable Monitoring System for Clusters *Software: Practice and Experience* **30**(7) 723-39
[12] Liu G, Mok A K, Yang E J 1999 Composite Events for Network Event Correlation *Proc. Conf. on Integrated Network Management of the Sixth IFIP/IEEE* Boston, MA, USA 247-60
[13] Xu J P, Zeng Z Q 2011 Applying Optimal Control Model to Dynamic Equipment Allocation Problem: Case Study of Concrete-faced Rock Fill Dam Construction Project *Journal of Construction Engineering and Management* **137**(7) 536-50

Authors

Yumei Ning, born in December, 1978, China

Current position, grades: working toward the PhD degree in the School of Computer Science and Technology at the Xidian University of Xi'an.
University studies: master degree in information science from the Xidian University, Xi'an, China, in 2006.
Scientific interests: computer network management technology and computer image processing.

Zhenguo Ding, born in October, 1959, China

Current position, grades: Professor in the School of Computer Science and Technology at the Xidian University, Xi'an, China. And he is the education technology academic leaders in Xidian University.
University studies: PhD degree in the School of Computer Science and Technology from the Xidian University, Xi'an, China, in 2005.
Scientific interests: range from computer network technology to remote education technology. He has published more than 40 technical papers in these areas.

Ping Zeng, born in February, 1956, China

Current position, grades: Professor and PhD supervisor in the School of Computer Science and Technology at the Xidian University, Xi'an, China. He is a senior member of Chinese Institute of Electronics.
Scientific interests: computer graphics, computer image processing, color management technology and computer peripherals technology.
Publications: more than 70 technical papers in these areas.

A new dynamic regulation UIPO model of groundwater based on cloud computing and hydroinformatics

Wensheng Wang*, Huifeng Xue, Feng Zhang

School of automation, Northwestern Polytechnical University, Xi'an 710072, China

Received 3 March 2014, www.cmnt.lv

Abstract

In order to solve the problem of excessive mining of the coal resources in Yulin mine area caused enormous damage to groundwater and in turn threatening the regional sustainable development, proposed a new dynamic regulation UIPO model of groundwater, which combined hydroinformatics, cloud computing and multi-source data fusion algorithm, and the mine hydrogeological spatial database, the visualization of 3D geological model and groundwater dynamic evolution model are created. Simulation results show that the UIPO complete with these models and with groundwater - ecological environment - economy system model all connected data analysis for decision support system and with complete hydrogeological and spatial process method by means of big data, can simulate the temporal and spatial variations of groundwater resources, forecast future impact on Yulin mine area groundwater for sake of large-scale exploitation.

Keywords: hydroinformatics, cloud computing, groundwater, dynamic regulation, big data

1 Introduction

It is a very complex system engineering to dynamic regulation of groundwater resources [1], and integrated studies it involving cross between different disciplines. To effectively achieve the dynamic regulation of groundwater resources must first create the refinement of the dynamic control model. After 20 years in Yulin in the mining of groundwater resources monitoring period, the accumulated number of dollars in TB groundwater data assets and still maintain a geometric growth. These data include mining high-resolution remote sensing, InSAR, sensors and other monitoring underground mining "monitoring wells" information and "cadastral" information, including a diverse [2], heterogeneous, multi-scale spatial information. Faced with such a large scale, diversity and rapid proliferation of "big data", the model has not been available for large data deduction, making a simple mechanism based on the traditional assumption of certainty or aquatic ecosystem dynamics model approach has been limited.

So how to improve the "big data" management and service levels, to achieve its "crossover reuse" and "holographic visible" will be an important issue to realize mining groundwater resources dynamically regulated faces. Pressing need to design an application for large data natural sciences, social sciences and engineering disciplines fusion of dynamic regulation of cross-model integrated studies methods, to be able to effectively analyse "all data" relationship between groundwater mining "big data" and the ecological environment and socio-economic systems, so that show great value in use, can simulate mining temporal and spatial variation of groundwater resources, projections of future large-scale mining impact on groundwater, forming a technical methods and theoretical system and coupled ecological, socio-economic, groundwater "big data" decision support system dynamic regulation of groundwater resources.

This paper proposed and studied by means of information theory and method of water including cloud computing, networking, and big data analysis and other next-generation IT and the combination of three-dimensional GIS to solve the dynamic regulation of groundwater mining elm God unified model construction problems. Hydroinformatics methods have become a new trend in groundwater monitoring and numerical simulation of the ecological environment development. Foreign researchers have tried in the waters of rivers, groundwater, rainwater, water dynamics, water ecological environment to calculate [3], groundwater flow numerical simulation [4] variation of water environment [5], etc. into water information science, exploration of cross discipline between information technologies combined with the common water science. Now that the hydroinformatics is related to hydrological data management, analysis, water resources protection, investigation and assessment of water resources, it is very effective in hydrological models [6]. So, this article intends to use the hydroinformatics multidisciplinary and comprehensive features to create a dynamic regulatory model for groundwater resources in water-based informatics, combined with groundwater dynamics theory, create the basic framework for the evolution model of groundwater resources under conditions of coal mining activities associated with water cycle changes in the region, and then put forward have the versatility and

***Corresponding author** e-mail: wwsh2007@163.com

systematic UIPO (Unified-Input-Process-Output) conceptual model, it is a further foundation for the other models.

2 UIPO regulatory model based on hydroinformatics

This paper first carried on the dynamic monitoring method thorough analysis on groundwater hydrology spatial data, then the systematic research on the UIPO model, UIPO model framework, as shown in Figure 1, finally proposes the effective implementation method and experimental verification. The UIPO model mainly includes:

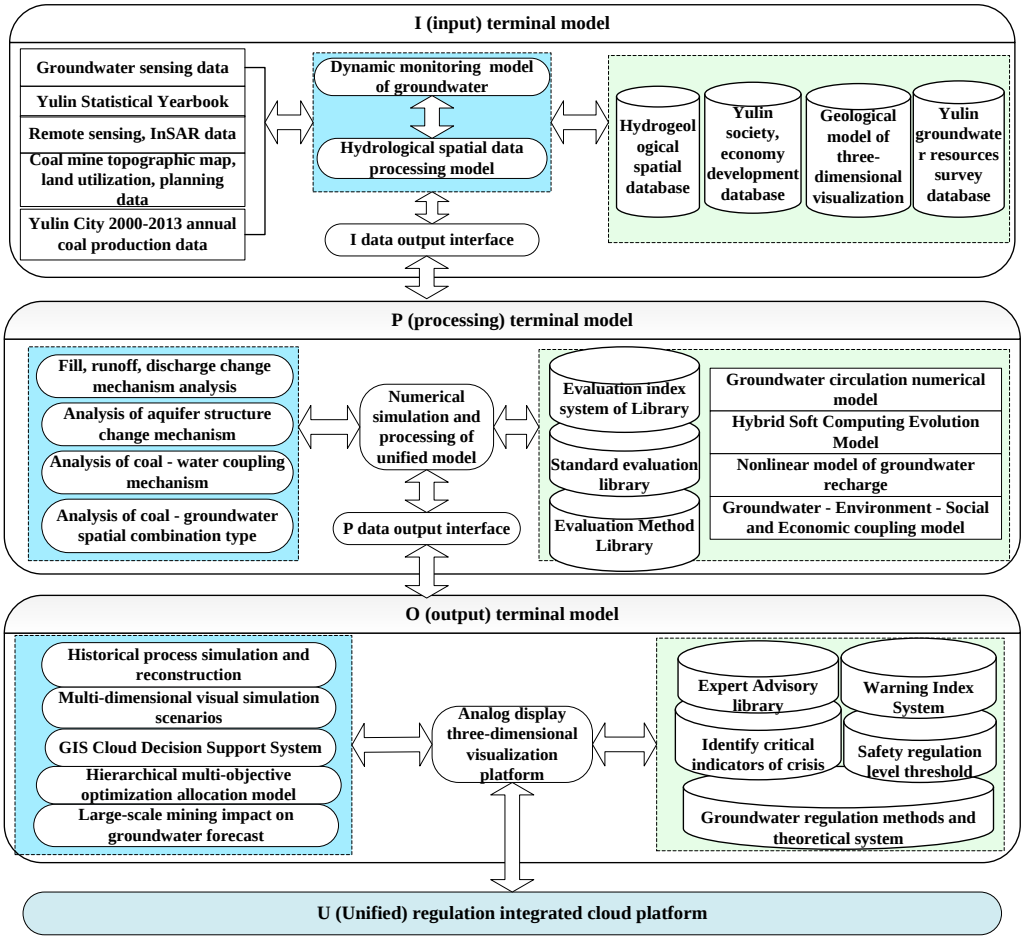


FIGURE 1 IPO model framework

2.1 CREATING OF A THREE-DIMENSIONAL GEOLOGICAL MODEL AND VISUALIZATION OF LARGE DATA PROCESSING HYDROLOGICAL MODEL SPACE

Effective realization of groundwater resources of refinement dynamic control problem was to obtain a large number of dynamic control data, In this paper, using of remote sensing, networking and a series of probing, testing techniques, to obtain the area aquifer hydrogeological parameters, create hydrogeological spatial database mining and visualization geological model, Build dynamic monitoring of groundwater resources and groundwater hydrology model space large data processing model, formation of I terminal model in UIPO. The main contents include:

1) The hydrological spatial data semantic modelling. Due to the diversity and heterogeneity of hydrological data sources a large space, research-based ontology semantic technology hydrological data for unified expression of a large space, to achieve an integrated processing, service publishing and storage technology for hydrologic data provide a large space data representation model.

2) Multi-source remote sensing data fusion model. Using of Yulin City TM images and MODIS images and Yulin 20 years of mining groundwater resources monitoring data and field data, based on pluralistic, multi-temporal, multi-scale, multi-source remote sensing image mining integration of heterogeneous data processing and evaluation techniques, The two information extraction and mining groundwater monitoring wells measured data from remote sensing

image analysis, generate the groundwater of multi-source remote sensing data fusion and monitoring model.

3) Dynamic monitoring model of groundwater. Groundwater resources in network research on perception, model structure, deployment framework, in order to realize the dynamic monitoring of groundwater resources, access to the aquifer hydrogeological parameter data fusion method, establishment of monitoring data based on D-S evidence theory.

4) A large space hydrological data processing methods. In order to meet the long-term dynamic monitoring of groundwater resources in the mining area, a hybrid cloud method of dynamic organization of information and structured, semi-structured and unstructured data storage model, Using NoSQL, MapReduce and machine learning techniques, structure support large data processing centre hydrological spatial data infrastructure, be able to multi-source, heterogeneous, lots of space for large hydrological data for dynamic organizations, the establishment of hydrogeological spatial database mining and visualization of geological models.

2.2 CREATING OF GROUNDWATER RESOURCES UNDER COAL MINING CONDITIONS ASSOCIATED WITH THE MINE WATER CYCLE CHANGES IN A DYNAMIC EVOLUTION MODEL

According to the key scientific problems of groundwater resource dynamic evolution system of mine hydrogeological characteristics, groundwater recharge, runoff and discharge characteristic, construct the index system, evaluation criteria and evaluation methods of evaluation, the formation of P terminal model in UIPO. The main research contents include:

1) Analysis of groundwater hydrodynamic field characteristics and evolutionary process. Using of historical data, related data model and I model output terminals nearly 20 years Yushen mining groundwater "recharge, runoff, discharge," the temporal variation, revealing the dynamic evolution of groundwater recharge. Dividing seams - a combination of the type of ground space, study the impact of coal mining impermeable layer stability, building evaluation index system, evaluation criteria and evaluation methods.

2) High Performance hybrid soft computing model. Study on groundwater dynamic process "recharge, runoff and discharge" model of temporal and spatial variation characteristics and mechanism, Create mining groundwater circulation numerical model and the dynamic evolution of high-performance mixed soft computing model to simulate groundwater mining cycle.

3) Groundwater recharges nonlinear model. Study on mining area containing variation of aquifer structure, groundwater recharge revealed qualitative and quantitative variation characteristics and formation evolution mechanism and time and space, a nonlinear model for mining area groundwater recharge aquifer structure change.

2.3 BUILD DYNAMIC CONTROL MODEL OF GROUNDWATER RESOURCES

Through the study of groundwater regulation methods and theoretical system, to build a unified regulation of groundwater comprehensive integration platform, achieve reasonable regulation of groundwater resources and the purpose of sustainable development and utilization. To form of O terminal model in the UIPO and U (Unified) control platform. The main research contents include:

1) Building of underground water resources regulation platform. Using of multi-disciplinary, multi-technology comprehensive integration and big data, cloud computing and other next-generation IT to create a "dynamic regulation of groundwater mining Yulin comprehensive integration cloud platform" (hereinafter referred to as U-control platform), realization of dynamic monitoring of groundwater and unified regulation. To create a unified mine three-dimensional groundwater flow model and mining groundwater flow - land subsidence coupling model on platform U-control, revealing the intrinsic link groundwater resources and water level changes between coal mining and land subsidence mechanism.

2) Groundwater regulation methods and theories. Based on the U control platform, using the P model output and the evaluation index system, evaluation criteria and evaluation method, constructs the groundwater of mining area ecological environment and social economy system is a complex coupling model of water resource optimal allocation model of multi-objective hierarchical; using GIS cloud data integration, statistical, spatial analysis, systems evaluation mining groundwater resources based on created three-dimensional visualization of the geological model and hydrological space large data processing model, building analysis based on all the data big data associated decision support systems. Affected by coal mining on water resources evaluation system, revealing the impact of coal mining on water resources and characteristics of the way to establish an expert advisory library. Based on the results of a variety of typical coal mining geological conditions were analysed to establish critical identity crisis groundwater mining index system, early warning indicators and safety regulation of groundwater level threshold.

3 Groundwater regulation of cloud computing architecture platform

In this paper, using of cloud computing solutions to build a data centre, including the design of a public and private cloud hybrid cloud groundwater cloud computing platform, as shown in Figure 2, Corresponding groundwater cloud storage can be divided into the

corresponding public cloud storage and private cloud storage. Public cloud built on the Internet, and to learn building a private cloud within the Intranet, both through a bridge (Bridge) before they can communicate with each other, the daily works of the two are physically isolated.

3.1 PRIVATE CLOUD PLATFORM OF GROUNDWATER REGULATION

Private cloud combines the realization of cloud computing technology and methods to focus on building the Intranet; you can realize the integration and sharing of groundwater data. With the inner scale geographic, geologic map superposition, fusion, and further can be related to the ongoing construction of potentiality assessment of mineral resources, reserves by using survey data integration and sharing.

Groundwater regulation of private cloud within the LAN using server clusters, storage devices, networking equipment and other infrastructure, Using virtualization software background memory, memory, CPU division, loading a different operating system [7], in which the deployment of geological database, application server software platform. But we can also use an application-level virtualization technology, virtual application table level, so that users can seem like using a local machine using a remote operating system, applications, and so on. Groundwater regulation database is a large logical database, is more centralized, distributed database cluster, while the actual database stored in the background groundwater regulation of cloud storage, the formation of Earth Science Data Centre Data Warehouse based on groundwater regulation. Groundwater regulation database data centre computing centres through groundwater regulation of WebServices provide all kinds of groundwater regulation outward data services supported by the SOA Manager GIS server as a spatial data services and computing services, and constitutes a cluster of servers. Because Web Services returned to the user is the result of converting it later, it effectively hides the pluralistic, heterogeneous raw data, to the integration and sharing purposes.

3.2 PUBLIC CLOUD PLATFORMS OF GROUNDWATER REGULATION

Groundwater regulating public cloud is constructed in Internet, will be able to open basic geographic, geologic map publishing network support for SOA GIS server, released to the public resource called API. Groundwater can be achieved in the regulation of public cloud services and the Azure service resources, shared resources (such as Google, ESRI spatial services, etc.) Internet converge on the interaction of groundwater can provide cloud computing services in the public Internet, and background data services, computing services, etc. can be flexibly increased or decreased according to demand, but can also be integrated with other systems.

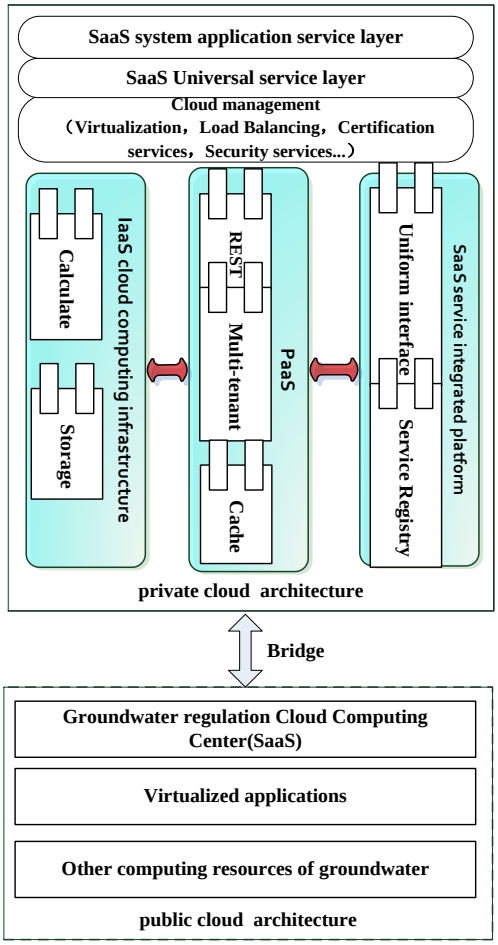


FIGURE 2 Load balancing test result

3.3 DESIGN AND IMPLEMENTATION OF VIRTUALIZED APPLICATIONS

In order for the existing hardware, software resources (genuine software can effectively solve the problem) were fully utilized, rational allocation of resources with virtualization technology, cloud computing applications using international recognized VMWare, Citrix, IBM and other virtualization products, building groundwater regulation of software and hardware resources virtualized application platform. Set one server into multiple virtual servers, we can also create a number on a single server or desktop level to achieve application-level virtualization application examples (the operating system can be Windows, Linux, Unix and other operating systems), managed by the virtual instance is ready to create a new instance of on-demand, upon request, and may also install supporting software. To meet people while online application, each person using different virtual instance, users can experience the benefits of virtualization vital brought. In a Windows operating system desktop virtual application instance, we can run ArcGIS, Oracle, Office and other software applications, and even the development of application software; Installing ArcGIS Server, Java EE applications on Linux and other

operating systems. An example in ArcMap version management software in virtualized applications, install all virtual instances share a license manager [9], log a user license is available subtrahend one, until exhausted. Data and applications can be separated or stored centrally. Backup mechanism virtual instance is optional, virtual instance can be unified updates and upgrades, updates and upgrades can also be customized, but all users do not need, all done on the server side, the user can at any location simply by browsing connected to the server can be achieved as local as processing data. Multiple servers reasonable allocation of hardware resources, which will be a server as a central node server virtualization, virtualization management platform to establish, manage multiple situated LAN, WAN physical servers or virtual servers [8]; We can use virtual machines, virtual desktops run in two modes. Virtual machine model is not installed on the client virtual machine, but according to the user's application, automatically creates a virtual machine installed on the server according to various software applications, databases, software development and other virtual machine template for applicants and assign the appropriate hardware resources (disk space, memory, network card, bandwidth, etc.).

3.4 DESIGN AND IMPLEMENTATION OF GROUNDWATER REGULATION OF BIG DATA CLOUD STORAGE PLATFORM

Use of Yulin City TB level storage device (the future will use PB grade, EB level or ZB level) to build groundwater regulation cloud storage, and expandable distributed storage nodes as the provinces. Using IBM, HP and other multiple servers form data storage pool, cloud storage device is a large physical storage device through optical fibre connection; Data storage uses zoning, fast file method, data access optimization in the interview. Taking into account the future internal and external services also needs to be divided into public storage cloud storage cloud, private storage cloud. Which is the corresponding public storage cloud computing external public cloud services, public cloud analysis, and private cloud computing is on a private cloud inside the private network, private cloud analysis. The groundwater regulation of big data cloud storage platform is as shown in Figure 3.

Data stored in the data warehouse using the form geosciences organizations, and is divided into content database, file databases, relational databases and object types, geological maps, gravity data, geochemical data, isotope geology of the work, geological work arrangements, as preserve the original format mining belt, administrative divisions and other data in a file database, so easy future use. These data systems data checking, cleaning, conversion, and unified loaded into Oracle spatial database, spatial data using the Oracle Spatial option to achieve. Fast indexing by using spatial index, and the objects in the database are saved in different tables paces, and a table space unified stored in the cloud storage device. Data storage uses parallel file system (Parallel File System) or concurrent file system, distributed file system, data management is the table. A key feature relative to the general system is the stand-alone file can store the same block of a file stored in the group of nodes, while allowing access to a number of computing nodes concurrently, without confusion and error, each node can use the own local file system, the establishment of efficient synchronization between each other and co-processing capabilities.

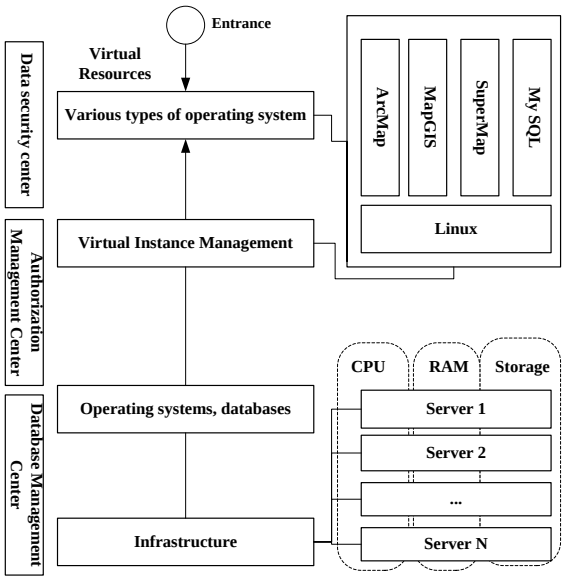


FIGURE 3 Groundwater regulation of big data cloud storage platform

3.5 MULTI-SOURCE DATA FUSION ALGORITHM FOR GROUNDWATER REGULATION BASED ON KALMAN FILTER

Mining groundwater regulation data involves the mineral resource potential assessment, field verification of mining rights, the extent of the work of geological data, geological work deploy data, geological maps and other data, the data is produced in different periods, data formats, and so have different spatial reference, the need for its data conversion, normalization processing operations. The original data on groundwater regulation database to keep a unified cloud storage device [10]. Combined with their standard-depth study of the actual construction of the library content of these heterogeneous databases, these heterogeneous data systematically cleaned to a unified coordinate system, unified administrative map, unified base map, unified geological bottom photo shows the basis for the conversion of spatial data. These databases are related to similar data dictionary table's entity name, domain, semantic standardization, to perform various special interactions, query analysis association.

Information fusion technology is proposed in the 1970s, which consists of a variety of information sources such as sensors, databases, knowledge bases and

humanity itself to obtain the relevant information and filtering [12], correlation and integration, thus forming an expressed frameworkб which is suitable to get information about decisions, explanation of the message to reach the target system, sensor management and system control. Information fusion is integrated multi-source information processing, with the nature of complexity. Signal processing and estimation theory indispensable for it's laid the theoretical foundation, which due to the high computational efficiency of the Kalman filter algorithm mature, etc., have been widely used.

Kalman filter based on minimum mean square error criterion for the best estimate, the basic idea is the use of state-space model of signal and noise, using the estimated value of the previous time and the current observation time updates on the status of the estimated variable, find the current the estimated value of the time [11, 13]. Assume that the discrete model for linear system:

$$x_k = F_k x_{k-1} + w_{k-1}, \tag{1}$$

$$z_k = H_k x_k + v_k, \tag{2}$$

where $v_k \sim N(0, R_k)$, $w_k \sim N(0, Q_k)$ are independent of each process, and $x_0 \sim N(x_0, P_0)$ independently in the initial state. Then for any loss function, the following Kalman filter equations:

The time update equations are given:

$$x_{t+1|t} = F_{t+1} x_{t|t}, \tag{3}$$

$$P_{t+1|t} = F_t P_{t|t} F_t^T + Q_{t+1}. \tag{4}$$

The measurement update equations are given:

$$K_{t+1} = P_{t+1|t} H_{t+1}^T (H_{t+1} P_{t+1|t} H_{t+1}^T + R_{t+1})^{-1}, \tag{5}$$

$$x_{t+1|t+1} = x_{t+1|t} + K_{t+1}(z_{k+1} - H_{k+1} x_{t+1|t+1}), \tag{6}$$

$$P_{t+1|t+1} = (1 - K_{t+1} H_{t+1}) P_{t+1|t}. \tag{7}$$

Kalman filter algorithm provides an effective way to filter out noise, is widely used to control, track, spacecraft orbit correction and other fields, has made many achievements. But the Kalman filter can only deal with the linear object, and the modernization of industrial production process the relationships among the variables are often very complex, using the linear model to reflect the relation between variables.

The Extended Kalman Filter (EKF) provides an online linearization method, namely the linearization of the nominal trajectory, and then uses Kalman filtering formula, which can solve the non-linear problem object to a certain extent. Set k time obtained system filtering system value is $x_{t-1|t-1}$ and $x_{t|t-1}$. The state equation is discrete systems in the vicinity of the $\hat{x}_{t-1|t-1}$ by Taylor series expansion, the linear model of the top two can have the original nonlinear system [8], it is given

$$\begin{cases} x_t = F_t x_{t-1} + U_{t-1} + \Gamma_{t-1} W_{t-1} \\ z_t = H_t x_t + Y_t + V_t \end{cases}, \tag{8}$$

where

$$F_t = \frac{\partial f(x_{t-1}, t-1)}{\partial x_{t-1}} | x_{t-1} = x_{t-1|t-1},$$

$$H_t = \frac{\partial h(x_{t-1}, t-1)}{\partial x_t} | x_t = x_{t|t-1},$$

$$U_{t-1} = f(x_{t-1|t-1}, t-1) - \frac{\partial f(x_{t-1}, t-1)}{\partial x_{t-1}} | x_{t-1} = x_{t-1|t-1},$$

$$Y_t = h(x_{t|t-1}, t) - \frac{\partial h(x_{t-1}, t-1)}{\partial x_t} | x_t = x_{t|t-1}.$$

By Kalman filter, EKF, UKF and other methods as well as other information fusion technology for large data processing, you can filter out the noise and easier to obtain the corresponding relationships between variables to improve the accuracy of analytical results.

4 Experiment and analysis

In order to test Yulin city groundwater monitoring data standardization process, we simulate multiple concurrent processing requests, validate the multitasking parallel speedup ratio and load balancing effect. Software and hardware environment test system is as follows. Hardware: X86-64 architecture server specific CPU (8 an 8 nuclear, frequency 2.6GHz), DDR3 32GB memory, SCSI hard disk (1TB, 10000rpm), Gigabit ethernet. Software: CentOS 5.2, Python 2.5.4, PostgreSQL 9.

1) Concurrent performance speedup test. At the start of single processing node A and dual processing nodes A, B, simulation 200 concurrent processing request, test and record the operation of each node respectively 1, 2, 3, 4, 5, 6, 7, 8 parallel processing example is complete when all tasks required processing time, finally the corresponding calculation speedup.

2) Load balancing test. Start the two processing node, A, B, each instance running 4 parallel processing, were recorded 20, 50, 200 systems, 260 processing tasks, each task number A, B node.

According to the concurrent performance speedup test scheme, processing and recording processing time, rendering acceleration as shown in Figure 4.

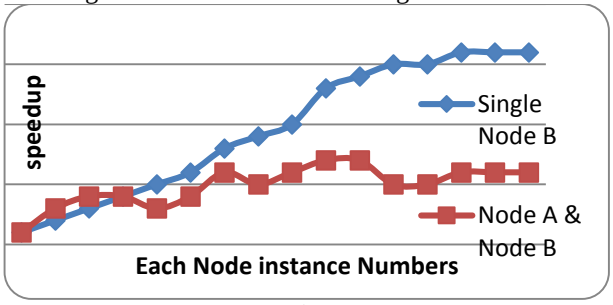


FIGURE 4 Concurrent performance speedup test result

From an overall perspective, both single node or double points, speedup ratio of parallel processing task instances with the growth in the number of approximate linear growth, with better performance of parallel processing system. With the parallel increase in the number of instances of processing tasks, accelerating gradually deviated from the theoretical speedup ratio, mainly due to an increase in instances of parallel processing with the internal system affect the communication time consuming and central control node and processing nodes for parallel processing speedup also gradually increased.

Testing according to the load balancing scheme, each node records the amount of processing to complete the task, the task of load balancing based on the drawing result of the number shown in Figure 5.

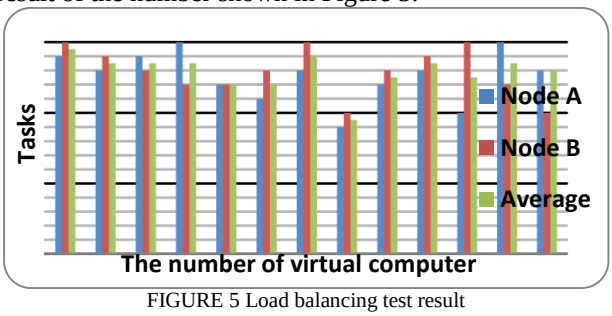


FIGURE 5 Load balancing test result

The test is repeated 15 times, the number of tasks handled similar to the average of each node, the total number of tasks as the load increases, the deviation gradually decreases, load balancing stability increase. Two node load quantity deviation, mainly because of the surplus is less than the number of instances of processing task concurrent processing, unable to realize the equalization processing, further processing resources or capability of instance share there are fluctuations and deviation.

5 Conclusions

This article has conducted the system research to the mine hydro geological spatial database and visualization of 3D geological modelling method, and created the data processing framework. From the perspective of water informatics starting to construct dynamic monitoring of groundwater resources groundwater model and a complete text of a large space method of data processing, thereby establishing groundwater regulation of technical methods and theoretical system of environmental, economic constraints, combined with groundwater regulation comprehensive integration platform implementation, and ultimately to create a unified UIPO models. In the model implementation, we have proposed a comprehensive cloud-based computing platform integration framework. Practice shows that this design is very suitable for implementing UIPO model.

Acknowledgments

This work is partially supported by Funding Project for Department of Education of Shaanxi Province of China (#14JK1864), Natural Science and Technology Project Plan in Yulin University of China (#14YK39), Thanks for the help.

References

[1] Quiroga, V M, Popescu, I, Solomatine D P, Bociort L 2013 Cloud and cluster computing in uncertainty analysis of integrated flood models *Journal of Hydroinformatics* **15**(1) 55-70
[2] Park Y, Hwa C K, Kang J H, Lee S W, Кшь J H 2014 Developing a flow control strategy to reduce nutrient load in a reclaimed multi-reservoir system using a 2D hydrodynamic and water quality model *Science of the Total Environment* 466-467(2014) 871-80
[3] Yu F, Lu W, Li P, Xin X, Li J 2012 Dynamic optimal control for groundwater optimization management with covariates *Journal of Hydroinformatics* **14**(2) 386-94
[4] Booth C J 1986 Strata-movement concepts and the hydrogeological impact of underground coal mining *Groundwater* **24**(4) 507-15
[5] McKinney D C, Cai X 2002 Linking GIS and water resources management models: an object-oriented method *Environment Modeling and Software* (17) 413-25
[6] Awan U K, Tischbein B, Martius C 2013 Combining hydrological modeling and GIS approaches to determine the spatial distribution of groundwater recharge in an arid irrigation scheme *Irrigation Science* **31**(4) 793-806
[7] Zhang F, Xue H-F 2013 Big data cleaning algorithms in cloud computing *International Journal of Online Engineering* **9**(3) 77-81
[8] Zhang F, Xue H-F 2012 Cloud manufacturing-based enterprise platform architecture and implementation *3rd International Conference on Digital Manufacturing and Automation* 190-191 60-3
[9] Younge A J, Laszewski G, Wang L, Lopez-Alarcon S, Carithers W 2010 Efficient resource management for Cloud computing environments *International Green Computing Conference* 357-64
[10] Menache I, Ozdaglar A, Shimkin N 2011 Socially optimal pricing of cloud computing resources *Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools* 16-25
[11] Kalman R E 1960 A new approach to linear filtering and prediction probles *Journal of basic engineering* 82 35-45
[12] Dhiman G, Marchetti G, Rosing T 2010 vGreen: A System for Energy-Efficient Management of Virtual Machines *Transaction on Design Automation of Electronic System* **16**(1) 1-27
[13] Ma S 2012 A review on cloud computing development *Journal of Networks* **7**(2) 305-10

Authors

Wen-sheng Wang, born on June 28, 1969, LiaoNing, China

Current position, grades: PhD in Northwestern Polytechnical University.
University studies: PhD in Systems Engineering from Northwestern Polytechnical University from 2008.
Scientific interest: hydroinformatics, cloud computing.

Hui-feng Xue, born on June 16, 1964, Shanxi Yuncheng, China

Current position, grades: professor in Northwestern Polytechnical University.
University studies: PhD degree in Water resource economics from Xi'an Polytechnic in 1995.
Scientific interest: the modelling of complex systems, simulation and performance evaluation, management, systems engineering, energy and environmental systems engineering, computer control, intelligent control, network control.

Feng Zhang, born on June 26, 1980, Shannxi Yulin, China

Current position, grades: PhD in Northwestern Polytechnical University.
University studies: MS degree in Computer science from Xidian University in 2009.
Scientific interest: cloud integrated manufacturing technology, the modelling of complex systems, the Internet of things applications.

Composition of web applications in clouds environment

Zuowen Wang*

School of Computer, Wuhan Polytechnic, Wuhan, 430074, China

Received 1 March 2014, www.cmnt.lv

Abstract

In recent years, for the advances of Cloud Computing technologies, cloud applications have been popularity for their rich set of features. The advantages of cloud applications include that users can utilize them in a low cost, threshold, and risk way; these applications can be quickly deployed on the clouds without duplication of work such that developers can focus on enhancing their QoS to improve core competitiveness. Therefore, their practical use on business with promising values can be expected. As such, cloud applications are recognized as a trend for the next generation of business applications, and hence how to migrate these on-premise applications to the clouds becomes a desired field in the literature. For this need, we present an ontology-based method for the composition process that specifically addresses the cloud features and the composition of on-premise applications into the clouds. In particular, for enabling the selection of desired clouds, the method imposes semantic ontologies on the specifications of the candidate clouds from which the desired ones can be effectively selected. For illustration, the method is applied to the composition of a CSS application to its cloud version.

Keywords: cloud computing, cloud application, semantic ontology, service composition

1 Introduction

Web service has already become an important paradigm for developing web applications. Growing number of web services raise the issue of efficiently locating the desired web services. Many approaches have been proposed with respect to the way in which services are described. Semantic communities provide ontology languages for web services such as OWL-s and WSMO. Services are organized in ontology. To locate desired services depends on semantic match of services. For the advances of Cloud Computing technologies in recent years, cloud applications have been popularity for their rich set of features. The advantages of cloud applications include that users can utilize them in a low cost-, threshold-, and risk-way; these applications can be quickly deployed on the clouds without duplication of work such that developers can focus on enhancing their QoS to improve core competitiveness. Therefore, their practical use on business with promising values can be expected. As such, cloud applications are recognized as a trend for the next generation of business applications.

In terms of the architecture for on-premise applications (e.g., web information systems), client-server patterns were most commonly used in the past decades; almost all kinds of existing on-premise applications were constructed using this style of architectures. However, as stated above, cloud applications have been recognized in recent years as applications to the clouds for taking advantage of cloud applications becomes a desired field in the literature. For this need, some discussions about the composition of on-premise applications into the clouds have been presented in [1-7]. In general, these statements have clarified some important issues about the composition and then proposed respective tips for addressing such issues. Thus, there are already many ideas about the cloud characteristics for composition; for instance, how on-premise applications are smoothly migrated to either of the three service models – SaaS, PaaS, and IaaS – supported in various cloud environments.

For the composition process, nonetheless, any methods that take into considerations of the architecture and characteristics of both on-premise applications and clouds to provide guidance on their composition are still missing. In addition, for the most critical issues in the composition process: the identification of candidate clouds and then the selection of desired ones from these candidates, few discussions can be found for addressing them by explicit formal approaches. Such methods and any formal approaches for addressing these critical issues, in our opinion, should not be negligible since a well-guided process and suitable formal issue approaches are most important for directing the composition of the many on-premise applications in a systematic and managed manner.

In this paper, we therefore present a method for directing the composition process. The method starts at the identification of the architecture and profile of the on-premise application to be migrated, through the discussion of the requirements for clouds and also the identification of the candidate clouds from which desired ones are selected with service models – SaaS or PaaS or IaaS – satisfying the cloud requirements, and finally ends

* ***Corresponding author*** e-mail: zuowenwangwh@163.com

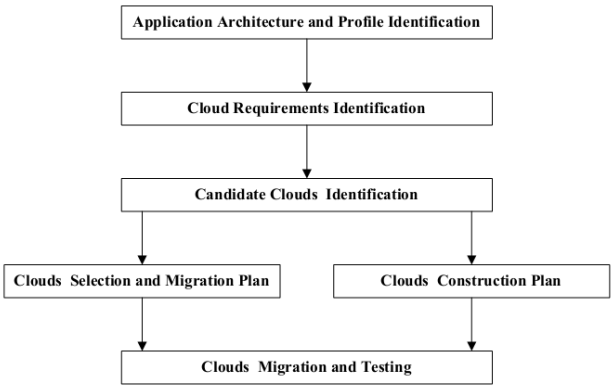


FIGURE 1 The Composition Method

at the deployment of the application on these selected clouds where a deployment and test plan is specified for conducting the deployment and tests. In particular, for enabling the selection of desired clouds, the method imposes semantic ontologies on the specifications of the candidate clouds such that desired clouds can be effectively selected by matching the desired cloud requirements with the provided services of these candidate clouds. Further, for addressing the exceptional situation that no candidate clouds can be found for smooth compositions, some discussions are also presented for conducting how on-premise applications may achieve their cloud-based versions via proper virtual mechanisms where either of the three service models – SaaS, PaaS, and IaaS – can be virtually supported in these mechanisms. For illustration, the method is applied to the composition of a Customer Support System (CSS) [8, 9] application to its cloud version that particularly emphasizes on both of collecting customer information (i.e., knowledge about/from customers) for enterprises and reversely delivering services information from enterprises to benefit customers.

This paper is organized as follows. Section 2 presents the composition method that encompasses three processes: an application-description process, a cloud-identification process, and an application- deployment process. The construction of semantic ontologies for enabling the selection of desired clouds is then presented in Section 3. For illustration, the method is applied in Section 4 to the composition of a CSS application to its cloud version. Finally, Section 5 has the conclusions and future work.

2 The composition method

As shown in Figure 1, the composition method has the following six steps:

1. Application Architecture and Profile Identification that determines (1) the architecture of the on-premise application where imposed components and their dependencies to support functional/non-functional purposes are specifically addressed; and (2) the profile of the on-premise application that includes the usage data about its executions (e.g., CPU, memory, storage, I/O, and network usage data) as well as the action data about its users (e.g., the number of active users, request rates, transaction rates, and request/transaction latencies).

2. Cloud Requirements Identification that clarifies the cloud requirements for the on-premise application based on its architecture and profile, including (1) for its components, the requirements for their deployment on the prospective services in clouds such as virtual machines, data storages, and a/synchronous message channels; and (2) for its usage and user actions, the requirements for their QoS in clouds such as customized user interfaces and access modes, performance, reliability, security, and scalability.

3. Candidate Clouds Identification that identifies the candidate clouds whose services (i.e., service models – SaaS or PaaS or IaaS – provided in clouds) satisfy the cloud requirements identified above. It is noticed that for the components of the on-premise application that have deployment requirements on clouds, there may necessarily have various clouds that collaboratively provide services to satisfy these requirements. Also, in particular, semantic ontologies are imposed on the specifications of these candidate clouds such that the services of these clouds are specified in an integrative and consistent manner.

4. Clouds Selection and Composition Plan that determines from the identified candidates which clouds will be selected for the composition of the on-premise application. As in above, since these candidate clouds are specified in integrative ontologies, the selection of desired ones from them can be effectively achieved by matching the desired cloud requirements with their provided services. After then, the plan about the activities and relevant artifacts involved in the composition into selected clouds will be specified. In general, the activities include (1) deploying the application components on the services in clouds; (2) deploying the interaction mechanisms among application components on the inter/intra-cloud interaction solutions over/in clouds; and (3) refactoring/restructuring deployed components for satisfying the usage and user actions requirements such as customized user interfaces and access modes, performance, reliability, security, and scalability.

5. Clouds Construction Plan that identifies and schedules the alternatives for the situation that no candidate clouds can be found for smooth compositions as stated above. In such a situation, the on-premise application may be considered to achieve its cloud-based version via virtual mechanisms where either of the three service models – SaaS, PaaS, and IaaS – can be virtually supported in these mechanisms.

6. Clouds Composition and Testing that realizes the composition of the on-premise application into selected or constructed clouds in accordance with the composition or construction plan identified above. As in usual, testing of the composition proceeds in accordance with the activities involved in the composition process. where the first step addresses an application-description process, the

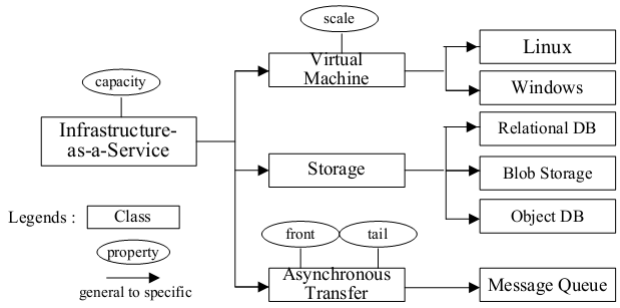


FIGURE 2 The hierarchy of classes for Infrastructure- as -a -Service

middle two encompass a cloud- identification process, and the last three cover an application-deployment process.

2.1 THE COMPOSITION METHOD

In this step, two descriptions about the on-premise application are addressed: (1) the architecture of the on-premise application; and (2) the profile of the on-premise application that includes the usage data about its executions as well as the action data about its users. For the architecture of the on-premise application, it specifically addresses desired architectural components with imposed behaviours and collaborations to support functional/non-functional purposes. As an example, the architecture of a Customer Support System (CSS) [8, 9] that emphasizes via the participant community [10-12] on both of collecting customer knowledge [13-16] for enterprises as well as reversely delivering services information to benefit customers: It is a 4-layer architecture of collaborative components where Customers interact with Enterprises via three intermediaries: Community, Customer Knowledge Agent, and Task Service Provider. It emphasizes on Community to help Customers share information about their desired tasks (e.g., buy or rent services from Enterprises). It emphasizes on collecting customer knowledge by Customer Knowledge Agent to help Enterprises catch their needs (e.g., provide services satisfying their desired tasks). It focuses on delivering services information by Task Service Provider from Enterprises to help Customers make recognition and comparisons.

With the architecture of the on-premise application, it is then good time to capture the profile of the on-premise application that can help size the application before it is migrated to the clouds. In general, the application data should be collected for at least 10 to 14 days to allow figuring out any variances in daily or weekly usage patterns. There are two kinds of data about the application: (1) the usage data about its executions (e.g., CPU, memory, storage, I/O, and network usage data); and (2) the action data about its users (e.g., the number of active users, request rates, transaction rates, and request/transaction latencies). With such application data, it is feasible to have an initial estimate of the cloud resources for the application to be migrated.

2.2 CLOUD REQUIREMENTS IDENTIFICATION

The second step is to identify cloud requirements for the on-premise application based on its architecture and profile, including (1) for its components, the requirements for their deployments on the prospective services in clouds such as virtual machines, data storages, and a/synchronous message channels; and (2) for its usage and user actions, the requirements for their QoS in clouds such as customized user interfaces and access modes, performance, reliability, security, and scalability. For the Customer Support System (CSS) as an example, its five components may require respective deployments on various cloud environments to support the functional/non-functional purposes via deployed clouds. Further, for its purposes of collecting customer knowledge for enterprises and delivering services information to benefit customers, it may require such QoS from these deployed clouds as customized user interfaces and access modes, performance, reliability, security, and scalability.

2.3 CANDIDATE CLOUDS IDENTIFICATION

The third step is to identify the candidate clouds whose services (i.e., service models – SaaS or PaaS or IaaS – provided in clouds) satisfy the cloud requirements identified above. For this, therefore, it is good to consider all of the cloud environments available on-line whose service models may satisfy the cloud requirements identified above. The following describes the possible service models in clouds:

1. Software-as-a-Service (Saas): In this model, the cloud provides application services, which may replace those provided by the on-premise application. With such SaaS services, many QoS features need to be evaluated for determining their replacement with the on-premise application as below.

(1) Their Service-Level-Agreements (SLAs) for availability, scalability, security, and performance; note that the evaluation for specific SLAs such as availability, scalability, and performance can be achieved by assessing the profile of the on-premise application.

(2) The compatibility of the application services with those offered by the SaaS.

(3) The portability of the application data into the SaaS for being accessed by the SaaS services.

(4) The portability of the access control by the application users into the SaaS for the access control by the SaaS users.

(5) The portability of the application interoperability with other services into the SaaS for the interoperable operations by SaaS services.

2. Platform-as-a-Service (Paas): In this model, the cloud provides platform services on which the on-premise application based on certain platforms such as JEE and MS.NET may be deployed under some QoS features as below.

(1) Their SLAs for availability, scalability, security, performance, and configuration (e.g., platform versions, APIs); note that as in SaaS, the evaluation for such SLAs as availability, scalability, and performance can be achieved by assessing the usage data about the application executions and the action data about the application users.

(2) The deployment of the application components and their interaction mechanisms on the PaaS for supporting the functional/non-functional purposes.

(3) The portability of the application services into the PaaS for the access by the deployed application users.

(4) The portability of the application data into the PaaS for being accessed by the deployed application.

(5) The portability of the access control on platforms (e.g., virtual servers) by the application users into the PaaS for the access control on clouds (e.g., virtual machines) by the deployed application users.

(6) The portability of the application interoperability with other services into the PaaS for the interoperable operations by the deployed application.

(7) The portability of the application management into the PaaS for monitoring and managing the deployed application.

3. Infrastructure-as-a-Service (Iaas): In this model, the cloud provides infrastructure services such servers, storages, and networks that the on-premise application and its deployed platforms may use under some QoS features as below.

(1) Their SLAs for availability, scalability, security, performance, and configuration; note that as in PaaS, the evaluation for such SLAs as availability, scalability, and performance can be achieved by assessing the profile of the on-premise application.

(2) The portability of the application services into the IaaS for the access by the deployed application users.

(3) The portability of the application data into the IaaS for being stored in the IaaS storages.

(4) The portability of the access control on infrastructure services (e.g., physical servers) by the application users into the IaaS for the access control on clouds (e.g., physical machines) by the deployed application users.

(5) The portability of the application interoperability with other services into the IaaS for the interoperable operations by the deployed application.

As a result, after considering all possible cloud environments that provide either of the above three service models, it is expected to identify some of them whose service models may satisfy the cloud requirements and then become the candidates to be selected for the realization of the composition. Since, as mentioned above, for enabling the selection of desired clouds from these candidates, semantic ontologies are imposed herein for the specifications of these candidate clouds such that the services of these clouds can be specified in an integrative and consistent manner. For the CSS as an example, Figure 2 shows the possible candidate clouds with service models that may satisfy the cloud requirements for the CSS.

2.4 CLOUDS SELECTION

The fourth step is to determine from the candidates identified above which clouds are selected for the composition of the on-premise application. As mentioned above, since these candidate clouds are specified in integrative ontologies, the selection of desired ones from them can be effectively determined by matching the desired cloud requirements with their provided services: (1) cloud requirements can be easily specified in terms of the structures and vocabularies of these ontologies, and (2) the matching of these requirements with the provided services of these candidate clouds can be easily undertaken. After determining the selection of targeted clouds, the plan about the activities and relevant artifacts involved in the composition into these selected clouds will then be specified. In general, the activities include (1) deploying the application components on the services in respective clouds; (2) deploying the interaction mechanisms among application components on the inter/intra-cloud interaction solutions over/in respective clouds; and (3) refactoring/restructuring any deployed components for satisfying the usage and user actions requirements such as customized user interfaces and access modes, performance, reliability, security, and scalability.

2.5 CLOUDS CONSTRUCTION

The fifth step is to identify and schedule the alternatives for the situation that no candidate clouds can be found at step 3 for smooth compositions. In such a situation, some alternatives may be considered; for instance, the on-premise application may be tailored to achieve a cloud-based version via specific virtual mechanisms [17-27] where either of the three service models – SaaS, PaaS, and IaaS – can be virtually provided through the support of these mechanisms.

The last step is to realize the composition of the on-premise application into selected or constructed clouds in accordance with the composition or construction plan identified above. As in usual, testing of the composition proceeds in accordance with the activities involved in the composition process.

3 Conclusion

In this paper, we present an ontology-based method for directing the composition of on-premise applications into selected clouds. The method takes into considerations of the architecture and characteristics of both on-premise applications and clouds to provide guidance on their composition. It therefore starts at the identification of the architecture and profile of the on-premise application to be migrated, through the discussion of the requirements

for clouds, the identification of the provided services of the available clouds, and the selection of the candidate clouds whose service models – SaaS or PaaS or IaaS – satisfy the cloud requirements, and finally ends at the deployment of the application into selected clouds where a deployment and test plan is specified for conducting the deployment and tests. In particular, for enabling the selection of desired clouds, the method imposes semantic ontologies on the specifications of the candidate clouds from which the desired ones can be effectively selected. For illustration, the method is applied to the composition of a CSS application to its cloud version that particularly emphasizes on both of collecting customer knowledge for enterprises and reversely delivering services information from enterprises to benefit customers. Since cloud applications have been recognized in recent years as a trend for the next generation of business applications, how to migrate the many existing on-premise applications to the clouds for taking advantage of cloud applications has thus become a desired field in the literature. However, current discussions about this need mainly focus on some important issues about the composition and then present respective tips for addressing such issues. For the composition process, any methods that take into considerations of the architecture and characteristics of both on-premise applications and clouds to provide guidance on their composition are still missing. Further, for the most critical issues in the composition process: the identification of candidate clouds and then the selection of desired ones from these candidates, few discussions can be found for addressing them by explicit formal approaches. Such methods and any formal approaches for addressing these critical issues, in our opinion, should not be negligible since a well-guided process and suitable formal issue approaches are most important for directing the composition of the many on-premise applications in a systematic and managed manner. The method and corresponding semantic ontologies presented herein provides an effort on this need.

As our future work, we will continue to explore the real composition of existing on-premise applications to the clouds where specific PaaS and IaaS offerings such as Google GAE and Amazon EC2 will be selected as the deployed or used platforms. As one may conceive, while migrating these applications, experiences about the composition can be collected correspondingly for validating the usefulness and effectiveness of the method. In fact, with our systematic and managed steps for gradually identifying and specifying application/cloud features and then conducting the deployment of the applications on the clouds, quality of these migrated applications on clouds can be expected.

References

[1] Banerjee U 2010 *Five Examples of Composition to Cloud* May 2010 http://setandbma.wordpress.com/ 2010/05/28/5-examples-of-composition-to-cloud/

[2] Blaisdell R 2011 *How to Plan Your Composition to the Cloud* Nov. 2011 http://www.rickscloud.com/how-to-plan-your-composition-to-the-cloud/

[3] Carr N, 2011 *Best Practices for Achieving Composition to a Cloud Model* March 2011 http://blogs.dlt.com/practices-achieving-composition-cloud-model/

[4] Cisco Systems 2010 *Planning the Composition of Enterprise Applications to the Cloud* http://www.cisco.com/en/US/services/ps2961/ps10364/ps10370/ps11104/Composition_of_Enterprise_Apps_to_Cloud_White_Paper.pdf

[5] Huey G, Wegner W 2010 *Tips for Migrating Your Applications to the Cloud* August 2010 http://msdn.microsoft.com/ en-us/magazine/ff872379.aspx

[6] Mallya S 2010 *Migrate Your Application to Cloud: Practical Top 10 Checklist* April 2010, http://www.prudentcloud.com/cloud-computing-technology/composition-to-cloud-top-10-checklist-24042010/

[7] Momentum S I 2013 *Cloud Composition* http://cloudcomposition.com/

[8] Lin J 2009 An Object-Oriented Development Method for Consumer Support Systems *International Journal of Software Engineering and Knowledge Engineering* **19**(7) 933-60

[9] Orman L Consumer Support Systems *Communications of the ACM* **50**(4) 49-54

[10] Armstrong A, Hagel J 1996 *The Real Value of Online Communities* Harvard Business Review 134-41

[11] Hoadley C, Kilner P Using technology to transform communities of practice into knowledge-building communities *ACM SIGGROUP Bulletin* **25**(1) 31-40

[12] Krieger B, Müller P Making internet communities work: reflections on an unusual business model *ACM SIGMIS Database* **34**(2) 50-9

[13] Bueren A, et al. 2004 Customer Knowledge Management – Improving Performance of Customer Relationship Management with Knowledge Management *Proc. of 37th International Conference on System Sciences, IEEE Computer Society* 1-10

[14] Davenport T, Klahr P 1998 Managing Customer Knowledge *California Management Review* **40**(3) 195-208

[15] Garcia-Murillo M, Annabi H 2002 Customer Knowledge Management *Journal of the Operational Research Society* **53** 875-84

[16] Mobasher B, Cooley R, Srivastava J 2002 Five Styles of Customer Knowledge Management, and How Smart Companies Use Them To Create Value *European Management Journal* **20**(5) 459-69

[17] Sunrun SunrunVas http://ventem.com.tw/DM11.aspx

[18] Bruijn J, Fensel D, Keller U, Lara R 2005 Using The Web Service Modeling Ontology to Enable Semantic E-Business *CACM* **48**(12) 43-7

[19] Zhang Y, Huang H, Yang D, Zhang H, Chao H, Huang Y 2009 Bring QoS to P2P-based Sematic Service Discovery for the Universal Network *Journal of Ubiquitous Computing* **13**(7) 471-7

[20] Noy N, McGuinness D 2001 *Ontology Development 101: A Guide to Creating Your First Ontology* Stanford Knowledge System Lab Technical Report KSL-01-05, March 2001

[21] Martin D, et al. 2004 *OWL Web Ontology Language for Services (OWL-S)* W3C Member Submission, Nov. 2004: www.w3.org/Submission/OWL-S/

[22] METEOR-S: Semantic Web Services and Processes: http://lsdis.cs.uga.edu/projects/meteor-s/.

[23] Gong H, et al. 2008 Research on the Building and Reasoning of Travel Ontology *Proc. of International Symposium on Intelligent Information Technology Application Workshops, IEEE Computer Society* 94-7.

[24] *Google App Engine (GAE)* https://appengine.google.com/

[25] *Windows Azure* http://www.windowsazure.com/zh-tw/

[26] *Google Compute Engine (GCE)* https://cloud.google.com/products/compute-engine

[27] *Amazon EC2* http://aws.amazon.com/ec2/

Authors

Zuowen Wang, born in October, 1966, Hubei Province, China

Current position, grades: Associate Professor
University studies: Computer science and technology
Scientific interest: Computer software, Software engineering
Publications: 20 papers published in the international or national journals
Experience: He is an associate professor in School of Computer, Wuhan Polytechnic, China. His research interests include computer software and software engineering.

OSTU image segmentation algorithm of fruit fly optimization algorithm

Chunyan Han*

Department of Computer Science, Sichuan University for Nationalities, Sichuan, 626001, China

Received 1 May 2014, www.cmnt.lv

Abstract

Traditional OSTU algorithm has the disadvantages of a large amount of calculation and low calculating speed. Based on the combination of Fruit Fly Optimization Algorithm and OSTU algorithm, an image segmentation algorithm is created from Fruit Fly Optimization Algorithm to improve OSTU, stressing the basic principles and calculation procedures of this revised algorithm. In order to verify the validity of this algorithm, the work compared the quality of image segmentation, segmentation speed and algorithm stability of 4 sets of standard test images. The simulation results show that the segmentation speed of revised OSTU algorithm is much faster than that of traditional OSTU algorithm when Fruit Fly Optimization Algorithm is applied to improve OSTU algorithm. Meanwhile, the quality of image segmentation is also more stable under the same condition of time limit.

Keywords: OSTU algorithm, image segmentation, Fruit Fly Optimization Algorithm, grey value, stability

1 Introduction

Image segmentation is the key technology of computer image visual processing. It has received extensive attention from researchers since it was proposed. So far, there are nearly one thousand kinds of image segmentation algorithm [1-3]. However, these algorithms are poor in versatility and are only applicable to some specific image segmentation issues rather than all images segmentations.

Based on the characteristics of large amount of calculation and low calculation speed of traditional OSTU algorithm, the work combined Fruit Fly Optimization Algorithm and OSTU algorithm and proposed an improved OSTU image segmentation algorithm by Fruit Fly Optimization Algorithm. Meanwhile, the work also compared the improved algorithm and OSTU, showing that both the segmentation quality and speed are superior to that of the traditional OSTU

2 OSTU method (also known as OSTU algorithm)

OSTU method is to set threshold value according to the sequence of grey value, and then classify images into 2 categories. Then, obtain the interclass variance through calculating the pixel and average grey-value of each category. In the end, set the threshold value as the final segmentation threshold when OTSU reaches its maximum. The calculation method of segmentation image is as follows [4-5]:

Classify the pixel of one digital image $f(x,y)$ into C_0 and C_1 by threshold T according to its grey levels. That is: $C_0 = \{f_1(x,y) \mid f_{\min} \le f(x,y) \le T\}$, $C_1 = \{f_2(x,y) \mid f_{\max} \ge f(x,y) > T\}$.

Here, $f_{\min}$ and $f_{\max}$ are respectively the minimum and maximum of the grey level of image $f(x,y)$. If N_i is set as the pixel when grey value is i ($f_{\min} \le i \le f_{\max}$), the total pixel of image $f(x,y)$ is $N = \sum N_i$. Therefore, the probability of the occurrence of each grey level is $P(i) = N_i / N$ and the total probability of the occurrence of C_0 is [6-7]: $P_0 = \sum_{i=f_{\min}}^{T} P(i)$.

Its mean value is: $\mu_0 = \sum_{i=T+1}^{T} iP(i) \Big/ P_0$.

The total probability of the occurrence of C_1 is: $P_1 = \sum_{i=T+1}^{f_{\max}} P(i)$.

Its mean value is: $\mu_1 = \sum_{i=T+1}^{f_{\max}} iP(i) \Big/ P_1$.

The mean value of image $f(x,y)$ is: $\mu = \sum_{i=f_{\min}}^{f_{\max}} iP(i) = \sum_{i=f_{\min}}^{T} iP(i) + \sum_{i=T+1}^{f_{\max}} iP(i) = P_0\mu_0 + P_1\mu_1$.

Define the interclass variance of the two categories is $\sigma^2(T)$: $\sigma^2 = P_0(\mu - \mu_0)^2 + P_1(\mu - \mu_1)^2$.

OSTU takes the interclass variances of the two categories as the determine evidence of selecting threshold value. It suggests that the best threshold value T^* should be the one when interclass variance $\sigma^2(T)$

* ***Corresponding author*** e-mail: han@scun.edu.cn

gets its maximum. That is: $T^* = \left\{ T^* \mid \sigma^2\left(T^*\right) \geq \sigma^2\left(T\right), \forall T \in \left[f_{\min}, f_{\max} \right] \right\}$.

3 Fruit fly optimization algorithm

3.1 OVERVIEW OF FRUIT FLY OPTIMIZATION ALGORITHM

Pan Wenchao [13], a young teacher in Taiwan, proposed Fruit Fly Optimization Algorithm (FOA), which is a new evolutionary calculation method. Fruit fly has the superiority on the senses of smell and visual, shown as Figure 1. Fruit fly can find food through searching the food sources in air, keenly locate food and its companions through its exquisite visual and finally fly to the food. Therefore, this method suggests that fruit fly usually determines an approximate position of food through its sense of smell and then locates the exact position of food through its sense of visual. This new method is for global optimization deduced from the action of fruit flies searching for food.

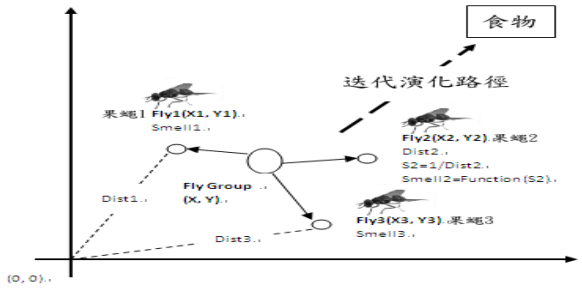


FIGURE 1 Fruit Flies Iteratively Search for Food

3.2 PROCEDURES OF FRUIT FLY OPTIMIZATION ALGORITHM

Fruit Fly Optimization Algorithm contains 7 steps. The specific procedure is as follows:

(1) As shown in Figure 1, initialize the location of fruit fly group and set the initialization results as $InitX\_axis$ s and $InitY\_axis$.

(2) When searching direction RV_x and RV_y are set, the stochastic searching distance of individual fruit fly can be obtained through the following formula:

$$Xi = Init\ X\_axis + RV_x,\ Yi = Init\ Y\_axis + RV_y. \quad (1)$$

(3) Estimate the distance $Disti$ between the current location of individual fruit fly and its original location because the location of food is unknown. Then, calculate the determination value of taste concentration S_i. Determination value of taste concentration equals the reciprocal of distance.

$$Disti = \sqrt{Xi^2 + Yi^2},\ Si = 1/Disti. \quad (2)$$

(4) Apply the determination value of taste concentration into determination function of taste concentration to calculate the taste concentration of this fruit fly at current location.

$$Smelli = Function(Si). \quad (3)$$

(5) The best taste concentration of fruit fly group can be calculated through the following formula:

$$[bestSmell\ bestIndex] = \max(Smelli). \quad (4)$$

(6) Keep the best taste concentration value of fruit fly group as well as its corresponding X coordinate and Y coordinate. At this time, fruit fly group can locate the food source according to their sense of visual and then fly to the food position.

$$\begin{aligned} &Smellbest = bestSmell, \\ &X\_axis = X(bestIndex), \\ &Y\_axis = Y(bestIndex). \end{aligned} \quad (5)$$

(7) Enter the iterative optimization step. Repeat iterative step (2) to (5) and determine whether the taste concentration is better than that of the previous iteration at the same time. If it is, then perform step (6).

4 USTO image segmentation of fruit fly optimization algorithm

4.1 DESCRIPTION OF THE ALGORITHM

When applying initialization, certain amount of fruit flies will be created at random. Calculate corresponding variance of the grey level of each fruit fly and find the maximum of variance among fruit flies. Then, update the location of fruit flies according to fruit flies optimization algorithm principles and procedures. When getting the set algebra, take the corresponding grey value of the maximum variance as the optimal threshold value to achieve image segmentation [8-9].

Specific procedure of the algorithm is as follows:

Step 1: Initializing the location of fruit flies and set the iteration times and the size of fruit fly group at the same time;

Step 2: Calculating the corresponding variance of the grey level of each fruit fly and compare the historical optimal value of individual and that of fruit fly group. Keep the location of current value and update the historical optimal value if the variance is superior to the historical optimal value of individual or fruit fly group; whereas, keep the current optical value;

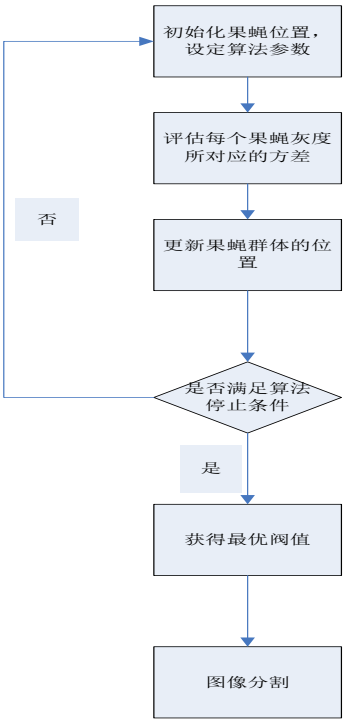
Step 3: Updating and changing the location of fruit fly according to the fitness value of fruit flies;

Step 4: If the iteration times $Iteration < Maxgen$, searching optimization completes; whereas, return to Step 2;

Step 5: Obtaining optimal threshold value and applying it to achieve image segmentation.

4.2 ALGORITHM FLOW CHART

The flow chart of Fruit Fly Optimization Algorithm improving OSTU algorithm is shown as Figure 2:



Initiate the location of fruit lies and set the parameter of the algorithm

Evaluate corresponding variancefor each grey value of fruit flies

Update the location of fruit fly group

Whether it meets the conditions to stop the calculation

Yes

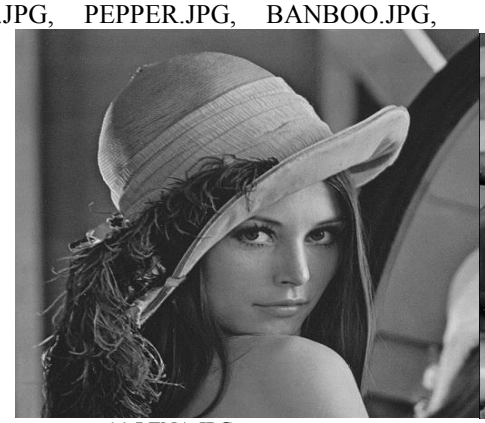
Obtain the optimal threshold value

Segment the images

FIGURE 2 OSTU Segmentation Method of Fruit Fly Optimization Algorithm

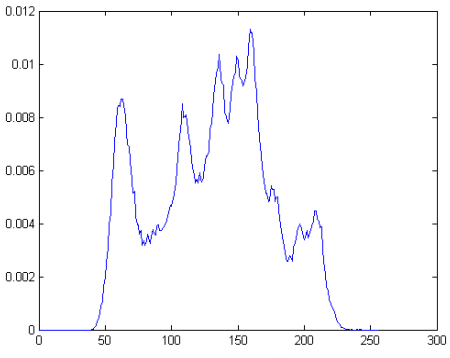
5 Algorithm test

Based on four standard images as test images [10-12], namely LENA.JPG, PEPPER.JPG, BANBOO.JPG, HUNTER.JPG, the work compared the segmentation effect of OSTU algorithm and FOAOSTU algorithm. The test images are shown as Figure 3:

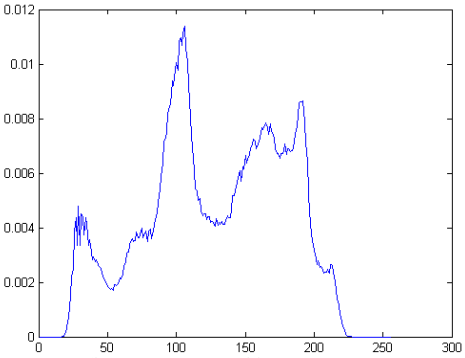


(a) LENA.JPG

(b) PEPPER.JPG

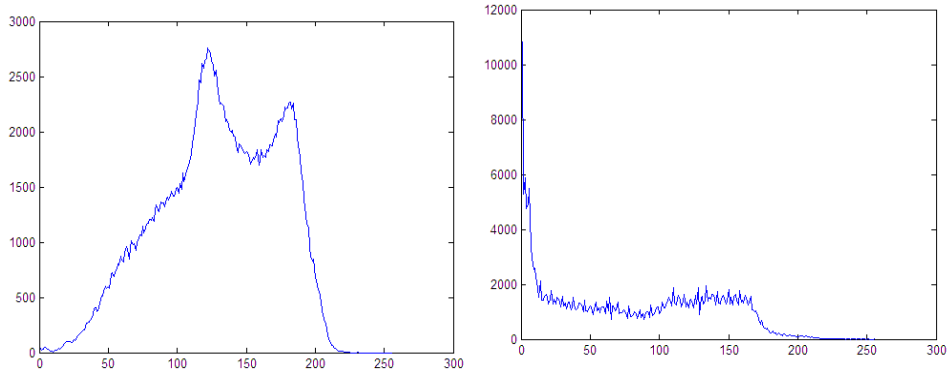


(a') LENA.JPG histogram



(b') PEPPER.JPG histogram

(c) BANBOO.JPG (d) HUNTER.JPG



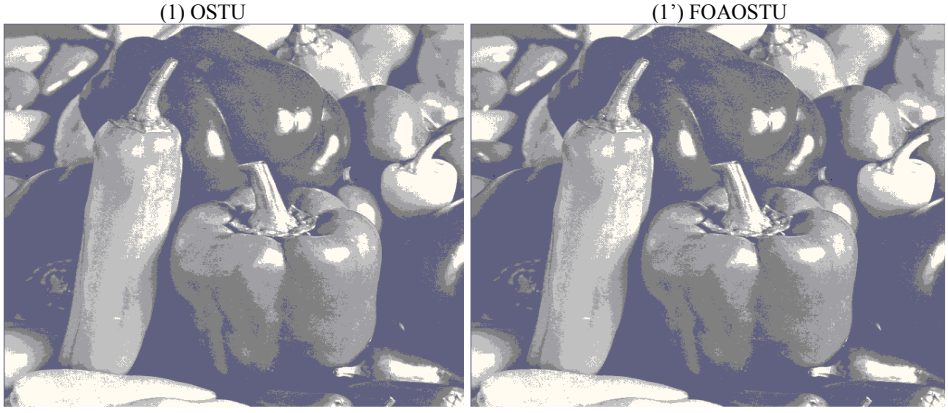
(c') BANBOO.JPG histogram (d') HUNTER.JPG histogram

FIGURE 3 Test Images and Corresponding Histogram

5.1 COMPARISON OF SEGMENTATION QUALITY

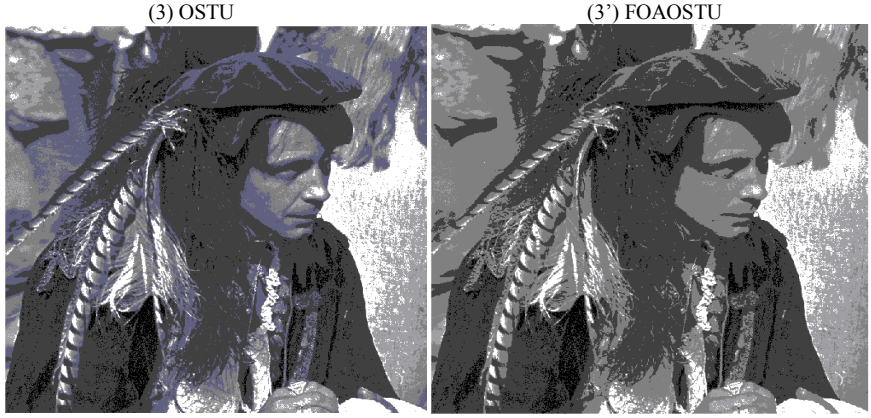
Figure 3 shows test images and the corresponding histograms. The segmentation comparison results of OSTU algorithm and improved OSTU algorithm by Fruit Fly Optimization are shown as Figure 4:

(1) OSTU (1') FOAOSTU



(2) OSTU (2') FOAOSTU

(3) OSTU (3') FOAOSTU



(4) OSTU (4') FOAOSTU

FIGURE 4 Comparison pictures of segmentation results of OSTU and FOAOSTU

Figure 4 shows that the segmentation of FOAOSTU algorithm is superior to that of OSTU algorithm, which ensures the segmentation speed and effect when achieving the edge information of images.

5.2 COMPARISON OF STABILITY

In order to compare the stability of OSTU algorithm and improved OSTU algorithm, the work applied formula (6) to evaluate the astringency [14-15] of the algorithm results:

$$std = \sqrt{\sum_{i=1}^{n} (\sigma - \tilde{\sigma}) / n} \,. \tag{6}$$

In the above formula, n refers to the times (set as 100 in the work) of repetitive computation of the algorithm; σ means the optimal solution obtained from every computation of the algorithm, and $\tilde{\sigma}$ refers to the average value of σ. Therefore, formula (6) is considered as the standard deviation of σ, which is labelled as std. Under the same condition, the bigger std is, the more unstable the algorithm is; whereas, it means the algorithm is more stable.

TABLE 1 Comparison results of the standard deviation of OSTU algorithm and FOAOSTU algorithm

Images	**Standard Deviation**	
	FOAOSTU	*OSTU*
LENNA	0.0912	6.8173
PEPPER	0.4738	3.3477
BANBOO	0.3015	5.3859
HUNTER	0.0968	3.9608

Table 1 shows that the stability of FOAOSTU algorithm is superior to that of OSTU algorithm, but its standard deviation is much smaller than that of OSTU algorithm.

5.3 COMPARISON OF SEGMENTATION SPEED

Table 2 shows that the segmentation speed of FOAOSTU algorithm is superior to that of OSTU algorithm, but its segmentation time is shorter than that of the OSTU algorithm.

TABLE 2 Comparison results of the segmentation speed of OSTU algorithm and FOAOSTU algorithm

Images	Segmentation Speed (Unit: Second)	
	FOAOSTU	*OSTU*
LENNA	0.32	0.75
PEPPER	0.38	0.67
BANBOO	0.43	0.82
HUNTER	0.56	0.79

6 Conclusions

According to the disadvantages of large amount of calculation and low calculation speed of traditional OSTU algorithm, the work applied fruit fly optimization algorithm to improve it and then segment images after obtaining optimal threshold value through this algorithm. The work carried out comparisons on image segmentation quality, segmentation speed and algorithm stability for 4 sets of standard test images. The simulation results show that when fruit fly optimization algorithm is applied to improve OSTU algorithm, and the segmentation speed of the improved algorithm is much faster than that of traditional OSTU algorithm. Meanwhile, the segmentation quality is also more stable under the same condition of time limit.

References

[1] Senthil Arumugam M, Rao M V C, Tan A W C 2009 A novel and effective particle swarm optimization like algorithm with extrapolation technique *Applied Soft Computing* **9**(1) 308-20

[2] Erwie Zahara, Yi-Tung Kao 2009 Hybrid Nelder-Mead Simplex search and particle swarm optimization for constrained engineering design problems *Expert Systems with Applications* **36**(2) 3880-6

[3] Krohling R A, dos Santos Coelho L 2006 Coevolutionary Particle Swarm Optimization Using Gaussian Distribution for Solving Constrained Optimization Problems *IEEE Trans. on Systems, Man, and Cybernetics* **36**(6) 1407-16

[4] Abido M A 2009 Multiobjective particle swarm optimization for environmental/economic dispatch problem *Electric Power Systems Research* **79**(7) 1105-13

[5] Junjie Yang, Jianzhong Zhou, Li Liu, Yinghai Li 2008 A novel strategy of pareto-optimal solution searching in multi-objective particle swarm optimization *Com. & Math. With Applications* (in press)

[6] Ababneh J I, Bataineh M H 2008 Linear phase FIR filter design using particle swarm optimization and genetic algorithms *Digital Signal Processing* **18**(4) 657-68

[7] Luitel B, Venayagamoorthy G K 2008 Differential evolution particle swarm optimization for digital filter design *CEC* 3954-61

[8] Boo Junyou 2007 Stock Price forecasting using PSO-trained neural networks *IEEE Congress on Evolutionary Computation* 2879-85

[9] Cheng-Jian Lin, Yong-Cheng Liu 2009 Image backlight compensation using neuron-fuzzy networks with immune particle swarm optimization *Expert Systems with Applications* **36**(3) 5212-20

[10] Chao-Ming Huang, Fu-Lu Wang 2006 An RBF Network with OLS and EPSO Algorithms for Real-Time Power Dispatch *IEEE Trans. on Power Systems* **22**(1) 96-104

[11] Jun Ying Chen, Zheng Qin, Ji Jia 2008 A PSO-based subtractive clustering technique for designing RBF neural networks *IEEE world Congress on Computational Intelligence* 2047-52

[12] He Wanjun, Qiang Wenyi, Chai Qingxuan 2007 Fuzzy Controller Design on the Basis of Particle Swarm Optimization *Control and Decision* **22**(5) 585-8

[13] Wen-Tsao Pan 2012 A new fruit fly optimization algorithm: Taking the financial distressmodel as an example *Knowledge-Based Systems* **2012**(26) 69-74

[14] Rigau J, Feixas M, Sbert M, et al. 2004 Medical image segmentation based on mutual information maximization *In: Proc. Of MICCAI 2004 (Saint-Malo), Belin*: Springer 135-42

[15] Kim J, Fisher III J W, Yezzi A, et al. 2005 A nonparametric statistical method for image segmentation using information theory and curve evolution *IEEE Trans. Image Process* **14**(10) 1486-502

Authors

Chunyan Han, born in March, 1971, Mianyang, Sichuan Province, China

Current position, grades: Associate Professor, Master
University studies: Computer science and technology
Scientific interest: Computer Applications and Algorithm and Design
Publications: 20 papers published in the international or national journals
Experience: She is an associate professor in Department of Computer Science, Sichuan University of Nationalities, China. Her research interests include computer applications and algorithm and design.

Colour recognition system based on TCS3200D

Lili Jing*, Yang Nie, Lifang Zhao

Jining Normal University, Inner Mongolia, 012000, China

Received 1 March 2014, www.cmnt.lv

Abstract

The colours of objects can be converted into pulses with different frequencies by TCS3200D. Colour recognition system was designed through the combination of Single Chip Microcomputer (SCM) and PC on the basis of the colour vision principle of TCS3200D. Fully utilizing TCS3200D, the system integrates the advantages of photodiodes and converter of light intensity to frequency to simplify circuit. The effects of factors, such as illuminant, orientation and the surfaces of objects, on the measurement were eliminated by the white balance adjustment. Additionally, colour vision errors were effectively reduced by the correction of the measured data with ANFIS and subtractive clustering.

Keywords: TCS3200D, colour recognition, ANFIS

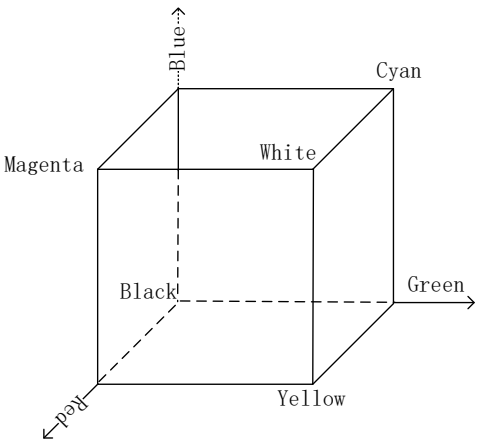
1 Introduction

As an important application, colour recognition has been widely used in various industrial detections. A separate photodiode is covered with the corrected red, green and blue filters in the typical colour sensor at present. Then, colour signal is recognized by the corresponding processing of the output signal. Some sensors combine two steps together to output analog signal. However, front channel processing, such as amplification, filtering and A/D conversion, should be conducted before computer processing in order to perform recognition. Consequently, the complexity of the circuit increases with a large recognition error, affecting the results [1]. Colour recognition is realized by TCS3200D made by TAOS (Texas Advanced Optoelectronic Solutions) Company. Square wave with a 50% duty cycle is directly produced by the conversion of light intensity to frequency of TCS3200D. Meanwhile, the output using the digital interface brings many new, excellent properties to TCS3200D.

2 Colour recognition principle of TCS3200D

2.1 RGB MODEL

The surface of object absorbs part coloured components in white light (sunlight) shining on it and reflects the other coloured parts to the eyes. Any colour can be obtained by the mixture of three primaries (red, green and blue) in different ratios. Three axes represent R, G and B in the RGB model of colour space (Figure 1). Origin corresponds with black (0, 0, 0), and the vertex farthest away from the origin white (255, 255, 255). The line connecting the origin with the farthest vertices represents grayscale distribution from black to white. The other six vertices of the cube represent red, yellow, green, cyan, blue and magenta, respectively. Each colour has a corresponding RGB value [2].



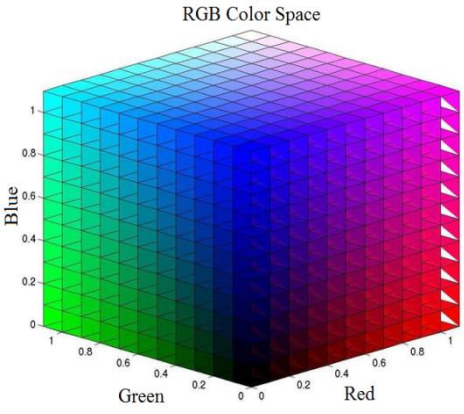


FIGURE 1 RGB model

***Corresponding author** e-mail: 15847418543@163.com

2.2 COLOUR RECOGNITION PROCESS OF TCS3200D

The colour of the tested object can be obtained with known RGB values according to the RGB model of colour space. TCS3002D only allows some particular colour to access but prevents the other colours after a colour filter is chosen. For example, only red incident light can access if red filter is chosen, while blue and green will be prevented. Then, the red light intensity can be measured. Similarly, the intensity of the blue or green light can be obtained if the other filters are chosen. RGB values are quantized by the linear conversion of light intensity to frequency signal. Then, the colour of the light projected to TCS3200D can be analysed.

The quantization process of RGB values is as follows. First, the RGB intensity value of reference light () was measured. Then, the intensity value $\left(P_{or}, P_{og}, P_{ob}\right)$ of light reflected by object was obtained under standard light. The ratio of two values was the reflection/transmission property of the object - the actual colour of the object (Equation 1).

$$K_r = P_{or} / P_{sr}, K_g = P_{og} / P_{sg}, K_b = P_{ob} / P_{sb}. \quad (1)$$

The standard coordinate of colour is among 0 to 255 in the RGB coordinate, so the standard RGB values were obtained by the result multiplied by 255.

2.3 WHITE BALANCE ADJUSTMENT

It should be noted that different light reflected by the object has different light intensities. Moreover, intensity components reflected by the object are different for non-standard white light (three unequal RGB values). Theoretically, white is composed by the same amount of red, green and blue. However, the amount of three primaries actually is not entirely equivalent in white. Equation (1) can be used to eliminate the effect [3]. The ratios are all 1 when an object is white. Therefore, the ratios multiplied by 255 were (255, 255, 255) —RGB values.

On the other hand, TCS3200D has different sensitivities to three primary colours, resulting in three unequal RGB values. The light intensity values $\left(P_{or}, P_{og}, P_{ob}\right)$ of three primary colours are not equal with the same induction time when TCS3200D is under white light. Therefore, white balance adjustment should be carried out before the test, adjusting the TCS3200D induction time of three primaries. For instance, induction time should be decreased with a higher sensitivity of red; otherwise the time should be increased. The induction time of three primary colours is determined as follows. First, the filters of three colours are sequentially gated, and then the output pulses of TCS3200D are counted. The time of pulses getting through each channel is calculated when the count is 255. The time is corresponding to the time standard used by each TCS3200D filter in actual tests. The numbers of pulses measured during the time are corresponding light intensity values - P_{or}, P_{og}, P_{ob} values.

3 TCS3200D

TCS3200D is the programmable converter proposed by TAOS Company of colourama to frequency. It integrates configurable silicon photodiode and a current-to-frequency converter on a single CMOS circuit. Moreover, three filters (red, green and blue (RGB)) are integrated on a single chip. Different combinations of pins S2 and S3 are controlled by photodiode, thus selecting different types of filters. Square waves with different frequencies (50% duty cycle) are outputted after current-to-frequency conversion with incident light projecting onto TCS3200D, corresponding to different colours and light intensities. TCS3200D can be directly connected to microprocessors or other logic circuits because the output signal with digital quantity can drives standard TTL or CMOS logic input. In addition, A/D conversion circuit is no longer needed because digital output can realize the conversion precision of more than 10 bit in each colour channel. Thus, the circuit becomes simpler. The pin and the function of TCS3200D are shown in Figure 2 [4].

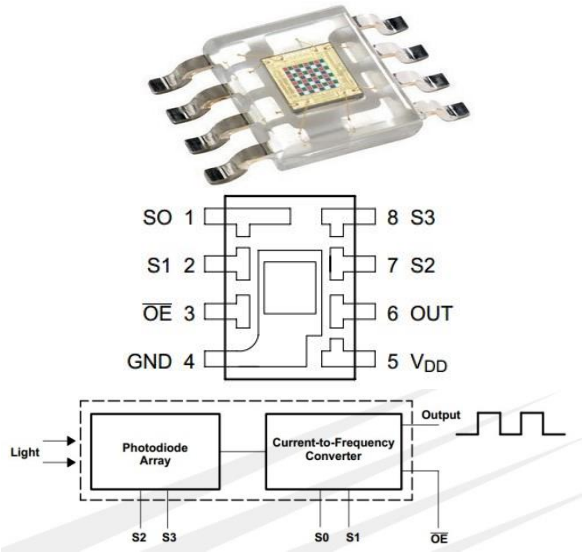


FIGURE 2 Pin and function of TCS3200D

TCS3200D (Figure 2) uses the 8-pin SOIC surface-mount package and integrates 64 photodiodes on a single chip. The photodiodes are classified into four types—16 with red filters, 16 with green filters, 16 with blue filters and the rest without filters (all optical information can access). The unevenness of the incident radiation is maximized through the stagger of these photodiodes in the chip, thus increasing the accuracy of colour recognition.

On the other side, the 16 multiple-connected photodiodes with the same colour are distributed evenly throughout the array to eliminate location bias among

colours. The typical range of output frequency is 2~500KHz. Users can choose the output scale factor (100%, 20% or 2%) or power down model *via* two programmable pins. The sensor output is applicable to different measurement ranges through the scale factor, improving its adaptability. For example, a small scaling value is selected to make the output frequency of TCS3200D match counter when low-frequency counter is used. The available combinations of S0, S1, S2 and S3 are shown in Table 1.

TABLE 1 Options of combinations

S0	S1	Output frequency scaling	S2	S3	Filter type
L	L	Power down	L	L	Red
L	H	2%	L	H	Blue
H	L	20%	H	L	Clear
H	H	100%	H	H	Green

4 Practical applications of TCS3200D

The hardware structure in this system adopted SCM application system and modular design. The principle block consists of LED module, colour recognition module, a microcontroller module and host computer module, as shown in Figure 3. Furthermore, display module, data storage module, operation module and serial communication module were included in the host computer module. The LED module improves the accuracy of colour recognition. The uniform arrangement along circle of four high-brightness, white LED makes the cross-section intensity of emitted beam evenly distributed, thus obtaining better recognition effect.

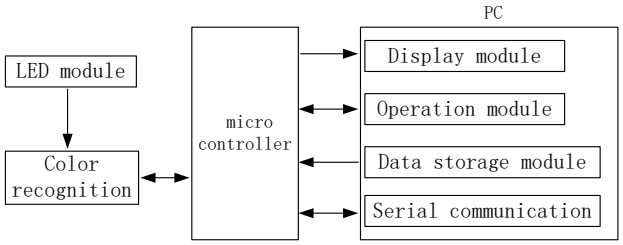


FIGURE 3 Connection between TCS3200D and Microcontroller

Microcontroller was connected to TCS3200D through SCM STC89C51 (Figure 4). First, TCS3200D driver determined whether white balance was need, and then the TCS3200D filters were gated successively. Subsequently, the timing was performed on the basis of the time standard in white balance adjustment. Meanwhile, the number of TCS3200D output pulses was counted. Finally, the results were transmitted to the PC through the serial port.

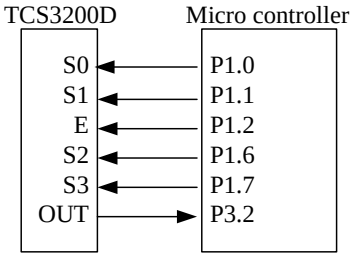


FIGURE 4 Connection between TCS3200D and Microcontroller

5 Colour vision optimization with ANFIS and subtractive clustering

The colour sensor and hardware circuit design can cause some systematic deviations of system measurement, which cannot be described with existing physical knowledge. There is an unknown nonlinear relationship between the output value and the actual value of sensor. With strong nonlinear approximation ability, ANFIS (Adaptive-Network-based Fuzzy Inference System) is entirely feasible to the non-linear calibration of sensor. Therefore, the calibration model based on ANFIS and subtractive clustering was adopted for the correction of the values measured by colour sensor.

First, samples were obtained by calibration experiment. Then, ANFIS network structure was determined by subtractive clustering. Subsequently, the system was trained with hybrid learning algorithm. Finally, the trained system was applied to the characteristic correction of sensor. The correction process is shown in Figure 5 and Figure 6 [5].

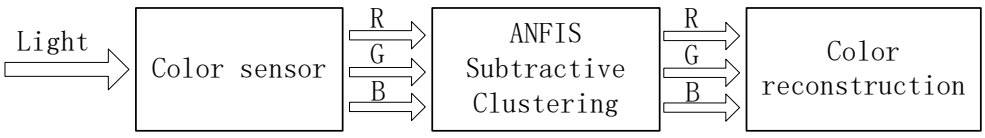


FIGURE 5 Optimization of colour sensor's output

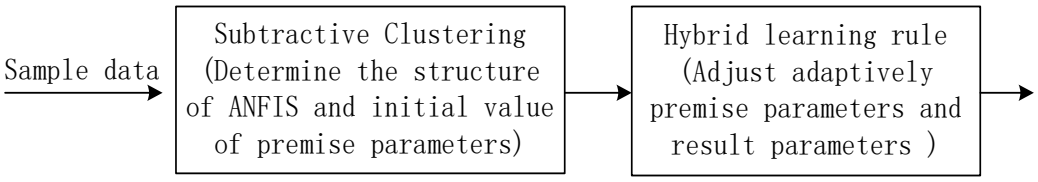


FIGURE 6 Process of designing ANFIS

Each output corresponds to a three-input and single-output ANFIS system (Figure 5). It should be noted that calibration points cannot be arbitrarily selected for training samples, while sample data should be selected on the basis of certain law. Thus, the data are representative so that correction model ANFIS is finally obtained with higher accuracy and universal applicability. The more the selected samples are, the better the correction effect of the model on results will be. Varying among 0-255, RGB selected the values of five specific points—0, 64, 128, 192 and 255. There are a total of 125 sample data, used to build correction model ANFIS [6].

ANFIS was used to compensate the sensor output as follows:

1) The related training and test samples were obtained.

2) The fuzzy rules of ANFIS and the initial values of parameters were determined using subtractive clustering according to training samples.

3) ANFIS training was conducted, and its performance was judged using test samples. If the performance was not good, then the evaluation threshold of clustering centre should be modified in the clustering algorithm. Thus, the number of ANFIS fuzzy rules was changed until ideal correction results were obtained.

4) The sensor output of actual colour was corrected using ANFIS trained well.

The data sent by the SCM through serial were processed using VISUAL C++ programming at the PC-terminal. The colour vision results, optimized using ANFIS and subtractive clustering, are shown in Figure 7. As can be seen from Figure 7, the RGB values after recognition are very close to those of recognized colour. Therefore, the effect of colour vision is excellent.

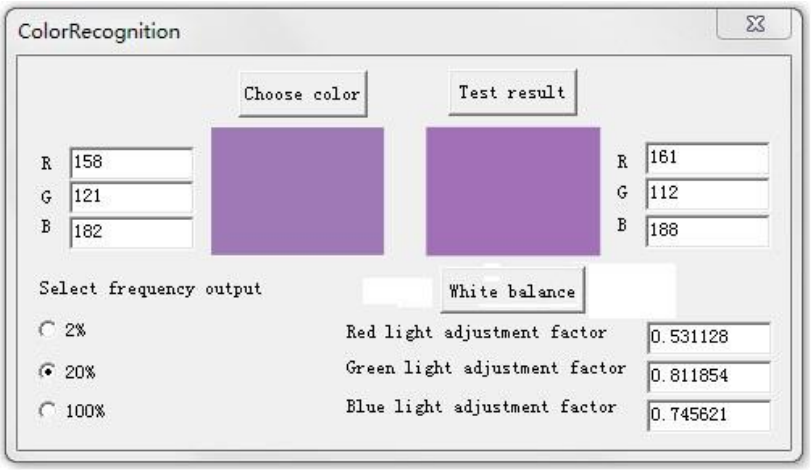


FIGURE 7 TCS3200D colour vision results

In addition, the following questions should be noted in the TCS3200D application:

1) Colour recognition is susceptible to interference from the external environment, e.g., stray light, etc., will affect the accuracy of colour recognition. Therefore, LED lighting system, the object to be tested and colour sensor were placed in the shading light cylinder in the design of system structure. Then, the effect of light from the outside world can be avoided during colour measurement to reduce measurement errors.

2) White balance adjustment is necessary for the first use of TCS3200D, reboot, replacement of lighting source, etc.

6 Conclusions

Light, observation orientation and surface of measured object and other factors have important effects on measurement results, which is the difficulty in colour measurement. The colour vision principle of TCS3200D was analysed, and its characteristics, such as high integration and digital output interface, were fully used. The output data of sensor was collected through SCM STC89C51, avoiding complex circuit processing. LED light effectively eliminates the effect of light instability on measurement. Client was designed through VC++ at the PC-terminal. Moreover, ANFIS and subtractive clustering were used to optimize the results of colour vision, thus greatly reducing measurement error. TCS3200D can be widely used in various industries where the measurement, analysis and recognition of light colour components are required.

References

[1] Oestreich J M, Toliey W K, Rich D A. 1995 The development of a colour sensor system to measure mineral compositions *Minerals Engineering* **8**(1) 31-3
[2] Leon K, Mery Domingo, Pedreschi Franco, etc. 2006 Colour measurement in Lab units from RGB digital images *Food Research International* **39**(10) 1084-91
[3] Kurioka 2002 Influence of light source and illurninance on Bonham type subjective colours *SPIE* 4421 426-29
[4] Texas Advanced Optoelectronic Solution Inc. TCS3200D Programmable colour light-to-frequency converter [EB/OL]
[5] Pollakova R 2005 Test methods for evaluation of colour fastness and comparison of the results *Vlana a Testile* **12**(l) 37-8
[6] Ying H 2008 *IEEE Transactions on System Man and Cybernetics* **28**(4) 515-20

Authors

Lili Jing, born in January, 1981, Inner Mongolia, China

Current position, grades: a lecturer of Jining Normal University, China.
Scientific interest: control engineering and computer engineering.
Publications: 10 papers.

Yang Nie, born in February, 1980, Inner Mongolia, China

Current position, grades: a lecturer of Jining Normal University, China.
Scientific interest: digital signal processing and control engineering.
Publications: 10 papers.

Lifang Zhao, born in December, 1980, Inner Mongolia, China

Current position, grades: a lecturer of Jining Normal University, China.
Scientific interest: electronic communications and control engineering.
Publications: 10 papers.

Distributed ontology based information retrieval using semantic web

Chun Zhang*

Shantou Radio and TV University, Shantou, Guangdong, 515041, China

Received 1 March 2014, www.cmnt.lv

Abstract

In recent years, user demand for integrated searches over several independently operating semantic web systems have been increasing rapidly. Integrated semantic searches enable more meaningful results to be generated because information with similar meanings in diverse areas and domains is likely to be used for inference. However, it is not easy to integrate physically independent, distributed, and heterogeneous database systems to provide a single, integrated semantic web system to end-users. In this paper, we propose a novel system that integrates heterogeneous semantic web systems based on schema mapping. The user can generate only one SPARQL query using the integrated schema without the necessity of checking the schemas of individual systems each time thereby reducing additional costs to generate queries for individual systems. Furthermore, the user is not required to collect individual query results manually after performing a query and additional costs for establishing systems can be reduced because no change in existing system structures is required. If currently established systems are expanded by adding the schema structures of other ontology systems, the cost to establish another integrated retrieval system can be saved. To evaluate the effectiveness of our approach, we have implemented a prototype that integrates two national information retrieval systems.

Keywords: integrated information retrieval, ontology, schema mapping, semantic web

1 Introduction

As the semantic search has been positioned as a killer service, many conventional information retrieval systems have been transformed into semantic web systems. In the early days, individual semantic web systems were built based on their own requirements and operated independently. As a result, users of a particular semantic web system were presented with the information managed only by the system. More recently, user demands for integrated searches over several independently operating semantic web systems have been increasing rapidly. This has occurred because integrated semantic searches enable more meaningful results to be generated, as information having similar meanings in diverse areas and domains is likely to be used for inference. However, it is not an easy task to integrate physically independent, distributed, and heterogeneous database systems to provide a single, integrated semantic web system to end-users. For physical integration, existing legacy data from participating systems must be transformed according to the integrated schema and whenever new data are accumulated in the participating systems, the transformation process must be repeated.

The Semantic Web [10-13] is a collaborative movement led by the international standards body, the World Wide Web Consortium (W3C). The standard promotes common data formats on the World Wide Web. By encouraging the inclusion of semantic content in web pages, the Semantic Web aims at converting the current web dominated by unstructured and semi-structured documents into a "web of data". The semantic web technology can be used in the pattern recognition layer. For example, with the help of the specific domain ontologies, the objects and relations of videos can be detected accurately.

A semantic link network (SLN) [14-17] is a relational network consisting of the following main parts: a set of semantic nodes, a set of semantic links between the nodes, and a semantic space. Semantic nodes can be anything. The semantic link between nodes is regulated by the attributes of nodes or generated by interactions between nodes. The semantic space includes a classification hierarchy of concepts and a set of rules for reasoning and inferring semantic links, for influence nodes and links, for networking, and for evolving the network. The semantic link network can be used in the video resources layer [18-20]. For example, with the help of the semantic link network model, the videos can be organized with their semantic relations.

In this paper, we propose a novel system that integrates heterogeneous semantic web systems based on schema mapping. The proposed system works by first creating integrated schema that includes all the attributes of the ontology schemas of participating semantic web systems (e.g., local schema). In the process, it maintains schema-mapping information that indicates which attribute of the local schema corresponds to that of the integrated schema. Second, user queries are generated against the integrated schema. Third, for query execution, the system re-

***Corresponding author** e-mail: cronychun@vip.qq.com

generates actual queries from the original user query in such a way that the attributes of the integrated schema are replaced with the corresponding attributes of the local ontology schema of the individual semantic web systems using the schema-mapping information. The core concept of the above-mentioned process is depicted in Figure 1.

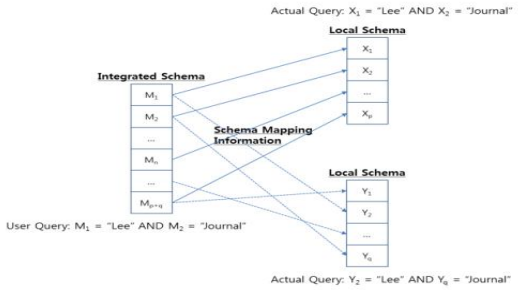


FIGURE 1 The schema mapping and query regeneration concept

To evaluate the effectiveness of our approach, we have implemented a prototype that integrates two national information retrieval systems, the National Discovery of Science Leader (NDSL) and the National Science and Technology Information Service (NTIS). The NDSL is an online database retrieval system that began service in 1962; it retains more than one hundred million pieces of information regarding research papers and reports, patents, standards, factual information, etc. The NTIS was developed in 2008 to enhance R&D investment efficiency by sharing and jointly utilizing information related to national R&D projects and science technology that are managed separately by individual government departments. Approximately 3.5 million pieces of data have been accumulated in the NTIS including national R&D and evaluation committee members, project information, result information, and research equipment information. Since the NDSL maintains only those pieces of information contained in journals, it does not retain information on what research projects‟ outcome the papers contained in the journals are. On the other hand, the NTIS does contain information about papers and reports for the outcomes of R&D projects carried out since 2008. Therefore, if the NDSL and the NTIS were linked to perform integrated retrieval, the value and the reliability of information from the two systems could be enhanced to be complementary.

The organization of this paper is as follows. Section 2 presents related work and Section 3 describes the proposed system architecture in detail. Section 4 presents classification of integrated queries and methods for sub-query generation. Section 5 presents experimental results. Finally, Section 6 provides a summary and proposals for future work.

2 Related works

Similar to ours, the ISENS [1] system provides a function to integrate and retrieve different real-world data sources having different ontologies. This system is less useful because queries in this system cannot be answered independently using a single system; instead, integrated queries can be made only to systems composed of mutually complementary data. In [1], mapping information for ontology schemas was not gathered in one place, but instead was made only for two ontologies with fields that can be mapped. Therefore, the existence of mapping information cannot be known without accessing the source system, which means that the mapping information can be accessed only through navigation. The largest difference between the present paper and [1] is that, instead of creating queries appropriate for individual local sites using mapping information, performing the queries, and integrating the results, in the case of [1]. The same query is performed using mapping information, the next system is accessed using the KAON2 reasoner to collect information, and results are presented but each database system cannot be accessed without using the system's source description and the source selection algorithm.

The DARQ system in [3] is also intended to perform integrated retrieval for distributed systems. However, this study is quite different from the study set forth in the present paper. In the DARQ system, heterogeneous data should be accessed using wrappers and the service description describes the kinds of data and access patterns that can be used for individual sites (endpoint) using sets of predicates. The DARQ system is different from the system in the present paper in that it focuses on query optimization using statistical information for integrated retrieval.

The SECO system in [2] enables efficient collection of any RDF files existing on the Web and provides interfaces in the form of HTML so that users can easily identify integrated data. This system is composed of a collector, a wrapper, a transformer, a user interface, a remote query interface, and data storage. The data storage is composed of multiple different sets of RDF data. Among them, MetaModel has Metadata information collected from files. SourceModel stores original RDF triples collected from files existing on the web and triples created here are purified through the transformer and stored in the TargetModel thereafter. The TargetModel enables access to user interface for creating HTML and remote query interfaces for query processing. The MetaModel, the UsageModel, and the TargetModel are described as ontologies and the SourceModel is composed of diverse schemas without any particular form.

3 The proposed architecture

Figure 2 depicts the overall architecture of the proposed system. The primary components of the system are the Schema Manager, the Query Manager, and the Result Manager. In what follows, we present detailed information about each component.

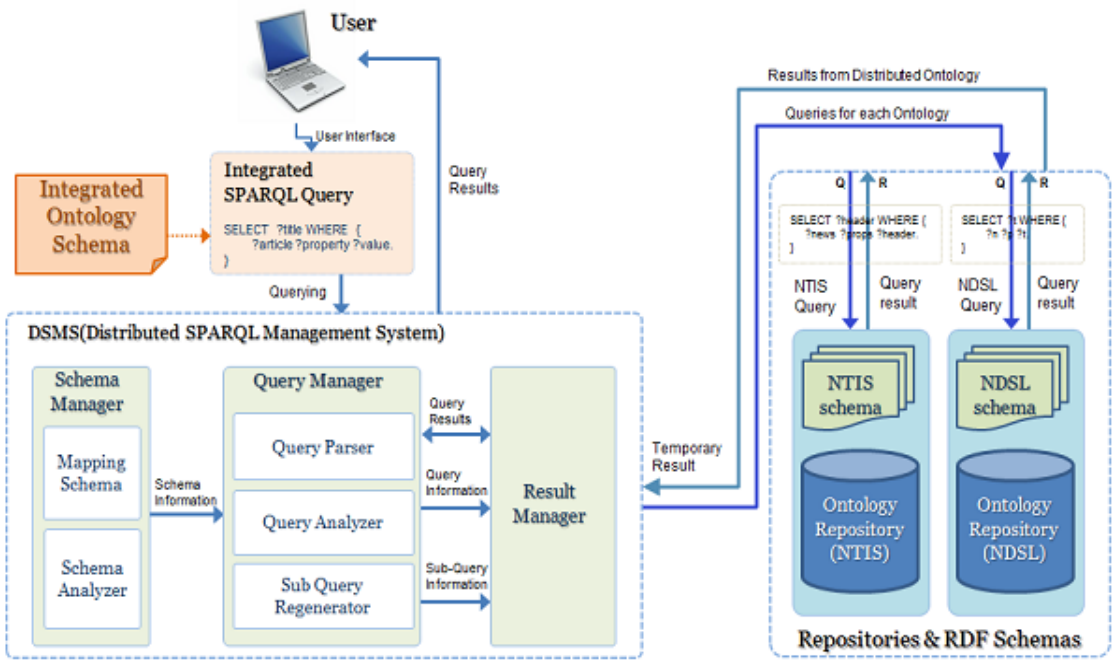


FIGURE 2 The system architecture

The NDSL and the NTIS define their own ontology schemas. Using the individual schemas, the administrator connects classes and attributes having the same meaning but different names with each other to create integrated schema that can link and retrieve two ontologies together. Therefore, the Schema Manager must maintain the Resource Description Framework (RDF) schema created by the administrator, as well as RDF schemas for individual ontologies of the NDSL and the NTIS. Both the NDSL and the NTIS have information on papers and the authors of the papers in common but they represent the information in different ways. For example, the People class in the NDSL schema and the Person class in the NTIS schema both indicate a human, but they are expressed differently. In contrast, both the NDSL and the NTIS schemas have a Journal class to represent a journal. Among the attributes defined in the classes, korName, hasName, write, and hasFirstAuthor are the attributes with the same meanings in both schemas. Therefore, the primary role of the Schema Manager is to maintain information about how individual classes and attributes defined in one schema are mapped to those defined in another schema. Note that classes and attributes in the integrated schema contain the level information, which is the core information needed by the sub-query generator. For example, classes such as Paper and Project are assigned to level 1, while hasFirstAuthor and korName attributes are assigned to level 2 and level 3, respectively. The level is related to the linkage of the class or the attribute. For instance, the hasFirstAuthor attribute has the identifier of the author but it does not have the name of the author. To acquire the author name, another triple condition such as ID :korName authorName must be executed. Therefore, the levels are constructed using the concept of linkage. How the level information is utilized will be described in the next section. The Schema Analyser analyses the schema information in the single SPARQL query submitted by the users. Then, it compares it with the RDF schemas of individual ontologies. Next, it delivers the information of the corresponding ontologies to the Query Manager.

The user creates queries based on integrated schema regardless of whether the ontologies managed by different information retrieval systems exist. The user query is first analysed using individual ontology schemas and the mapping schema. Subsequently, sub-queries are generated that are suitable for the corresponding ontologies. The Query Manager consists of the query parser, the query analyser, and the sub-query generator. When the query is verified by the query parser, the type of the query is determined through the query analyser. Using the integrated schema created by the administrator, the query parser checks the validity of the class and the attributes in the query. Then, using the schemas of individual ontologies, the query analyser determines whether the user query can be transformed into sub-queries for individual ontologies. Finally, the sub-query generator re-generates sub-queries. The sub-query generator and the query type will be dealt with in detail in the next section.

The Result Manager manipulates the intermediate results obtained by performing sub-queries in individual ontology for further processing or preparation for the final results. For example, depending on the query type, the Results Manager uses the results from one sub-query as filter information for another sub-query or it combines the results of sub-queries for the final query result.

4 Query generation

The query types determined by the query analyser are automatically generated into four types depending on the content of the information that the integrated query will retrieve. Type-1 queries represent those queries using the schema that exist only in one ontology. The integrate query includes the „Project" class, which exists only in the NTIS schema. Therefore, there is no corresponding sub-query for the NDSL. Figure 3 shows the control and data flows for processing the Type queries.

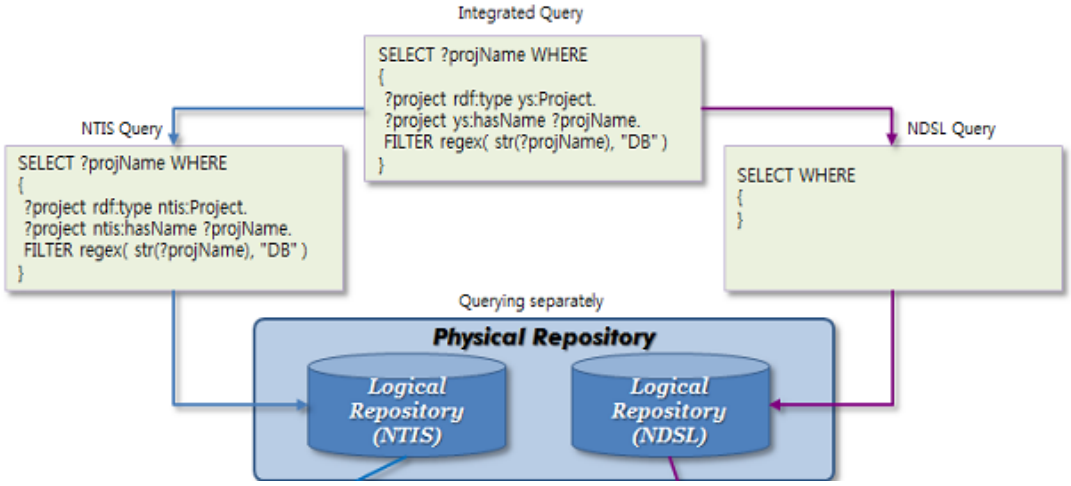


FIGURE 3 Control and data flows for processing Type-1 queries

Type-2 queries include classes and attributes that exist in every ontology. When retrieved using classes and attributes such as Paper and Author that the NDSL and the NTIS have in common, the query result is in the form of a combination of individual sub-query results. Figure 4 shows control and data flows for processing Type-2 queries. In Type-3 and Type-4 queries, some of the classes and attributes included in the queries exist in every ontologies, but the other classes and attributes are defined only in one ontology. For instance, the „Author" attribute exists both in the NDSL and in the NTIS schemas, whereas the Project attributes exists only in the NTIS schema. When queries are received of these types, the sub-query generator first separates the commonly existing classes and attributes from those that exist individually. For those classes and attributes that exist in only one ontology, they will be eliminated from the sub-queries for the ontologies that do not support them.

The example for a Type-3 query is "papers written by those who participated in the project for establishment of a driving safety DB and development of operating technology".

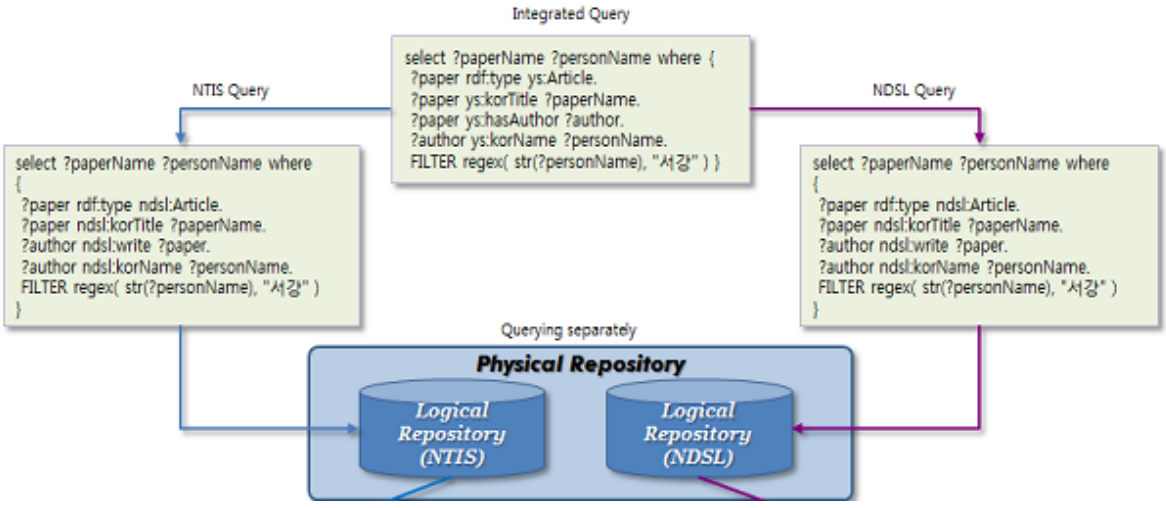


FIGURE 4 Control and data flows for processing Type-2 queries

The Figure 4 Control and data flows for processing Type-2 queries example for a Type-4 query is "projects in which the author of the paper „Diesel Engine participated". Project and participateIn in the integrated query do not

appear in the sub-query for the NDSL because the relevant classes and attributes do not exist in the NDSL schema. In the sub-query for the NTIS, the class Article is transformed into Paper and the attribute korTitle is transformed into hasName.

Type-3 and Type-4 queries have similar basic preconditions but differ from internal processing. For Type-3 queries, intermediate query results are first obtained by using classes and attributes that exist exclusively in one ontology. Then the commonly existing classes and attributes are applied to the intermediate query results. On the other hand, Type-4 queries apply query results obtained using common classes and attributes to queries for exclusive classes and attributes in a certain ontology.

To process Type-3 queries, given the two ontologies A and B, a query that includes the schema for A must be executed first. A query for B is also generated up to the WHERE clause excluding the FILTER clause. The FILTER clause to be included in the query for B will be generated by analyzing the intermediate query results from A and the queries for A and B (The query for A is complete, whereas the query for B is incomplete). First, all attributes that exist exclusively in the schema for A are extracted. Then, they are compared with the schema for B and the intersection of the two in the lowest level. Finally, the intersection and query results obtained from A are compared to generate the FILTER clause and complete the query for B. For instance, a Project exists in the NTIS schema and its level is one. The „Project" has participant for level two. The participant has a participantName, which is level three. The participantName is matched with authorName of Author of Paper in the NDSL schema. The FILTER clause is completed through these traces. For Type-4 queries, the FILTER clause is generated similarly to Type-3 queries. However, it is different from Type-3 queries in that the query results for individual ontologies extracted using classes and attributes that commonly exist in A and B. The results are then used to generate a FILTER clause for classes and attributes that exist exclusively in A or B.

5 Experiment and results

For each query type, we generated ten queries automatically and measured the performance of the proposed system in terms of the number of retrieved tuples and processing time. NDSL query results are not indicated because the integrate queries use only classes and attributes that exist exclusively in the NTIS schema. Sub-queries are generated for individual ontologies and the results are brought and integrated. In this case, the sum of the two query results differ slightly because overlapping values exist in the two query results, and they are treated as a single value.

NTIS query results are the same as the integrated query results. This indicates that the results obtained from the NDSL did not greatly affect obtaining the results from the NTIS. In other words, in the case of projects in which the author of a certain paper has participated", the author of the particular paper in the NDSL had hardly participated in NTIS projects or he participated in some cases but the cases were removed as overlapping results. This can be considered an issue of the scope and diversity of the data existent in each ontology. Queries for individual ontologies were not processed in parallel with each other because there are cases where the result values of queries for one ontology are included in the conditional clauses of queries for the other ontology. Although the results of individual queries were brought from ontologies, the integration of the results and the removal of overlapping results were made by the Result Manager using the memory of the relevant system. On reviewing individual experimental results, it can be seen that the time spent for integration is not something about which to be greatly considered.

6 Conclusions

In this paper, we proposed a novel system that will enable storing data from two different ontology systems in one physical system without converting the data to conduct integrated retrieval. Based on the individual schemas of the separated systems, the administrator can connect the schemas that have the same meanings but different forms of expression with each other between the two systems to generate integrated schemas. The user can then generate integrated SPARQL queries utilizing the schemas generated by the administrator to perform queries without recognizing the existence of individual ontologies. The user can generate only one SPARQL query using the integrated schema without the necessity of checking the schemas of the individual systems every time thereby reducing additional costs to generate queries for individual systems. Furthermore, the user is not required to collect individual query results manually after performing a query and additional costs for establishing systems can be reduced because no change in existing system structures is required. If currently established systems are expanded by adding the schema structures of other ontology systems, the cost to establish another integrated retrieval system can be saved. Although the complexity of applications will increase in this case, it should be a trivia l problem compared to the cost to integrate several millions or several dozen millions of triples manually.

In the future, the performance of result integration algorithms should be improved by adding more triple data and more query types should be added. If the system is complemented by establishing multiple ontologies in completely distributed network environments instead of a single physical storage, a more reliable system can be implemented.

References

[1] Abir Q, Dimitre A, Jeff H 2009 ISENS: A system for information integration, exp loration, and querying of multi-ontology data sources *Proceedings of the 2009 IEEE International Conference on Semantic Computing* 330-5

[2] Andreas H 2004 An integration site for semantic web metadata," proceedings of world wide web conference 1 229-34

[3] Bastian Q, Ulf L 2008 Querying Distributed RDF Data Sources with SPARQL

[4] Dimitre A, Dimitrov H J, Abir Q, Nanbor W 2006 Information integration via an end-to-end distributed semantic web system *Lecture Notes in Computer Science* 4273 764-77

[5] Mayfield J, Finin T 2003 Information retrieval on the Semantic Web: Integrating inference and retrieval *SIGIR 2003 Semantic Web Work shop*

[6] http://jena.apache.org/documentation/query/.

[7] http://linkeddata.org/.

[8] http://www.w3.org/TR/rd f-mt.

[9] http://www.w3.org/TR/rd f-sparql-query/.

[10] Liu Q, Zhang, Ni L M 2010 *IEEE Transactions on Parallel and Distributed Systems* **21**(3) 405-16

[11] Menzel M, Ranjan R, Wang L, Khan S, Chen J 2014 CloudGenius: a hybrid decision support method for automating the migration of web application clusters to public clouds *IEEE Transactions on Computers, in press*

[12] Hao F, Min G, Chen J, Wang F, Lin M, Luo C, Yang L T 2014 An optimized computational model for task-oriented multi-community-cloud social collaboration *IEEE Transactions on Services Computing in press*

[13] Qi L, Dou W, Chen J 2014 Weighted principal component analysis-based service selection method for multimedia services in cloud computing *Computing, Springer in press*

[14] Wang L, Tao J, et al. 2013 G-Hadoop: MapReduce across distributed data centers for data-intensive computing *Future Generation Computer Systems* **29**(3) 739-50

[15] Xu Z, et al. 2014 Knowle: a semantic link network based system for organizing large scale online news events *Future Generation Computer Systems*, 10.1016/j.future.2014.04.002

[16] Xu Z, Luo X, Zhang S, Wei X, Mei L, Hu C 2013 Mining temporal explicit and implicit semantic relations between entities using web search engines *Future Generation Computer Systems* DOI:10.1016/J.future.2013.9.027

[17] Liu Y, Ni L M, Hu C 2012 *IEEE Journal on Selected Areas in Communications* **30**(9) 1780-8

[18] Luo X, Xu Z, Yu J, Chen X 2011 *IEEE transactions on automation science and engineering* **8**(3) 482-94

[19] Hu C, Xu Z, et al. 2014 Semantic link network based model for organizing multimedia big data. *IEEE Transactions on Emerging Topics in Computing* 10.1109/TETC.2014.2316525.

[20] Liu Y, Zhu Y, Ni L M, Xue G 2011 *IEEE Transactions on Parallel and Distributed Systems* **22**(12) 2100-7

Author

Chun Zhang, born in May, 1970, Shantou, Guangdong Province, China

Current position, grades: a lecturer of Shantou Radio and TV University, China.
Scientific interest: data structure and management information systems
Publications: 11 papers.

Research on cost-sensitive ensemble classification for mining imbalanced massive data streams

Yuwen Huang[1, 2*]

[1]*Department of Computer and Information Engineering, Heze University, Heze 274015, Shandong, China*

[2]*Key Laboratory of computer Information Processing, Heze University, Heze 274015, Shandong, China*

Abstract

The existing classifiers for massive data streams do not consider the imbalance distribution and cost factors, so this paper proposes the approach of the cost-sensitive ensemble classification for imbalanced massive data streams (CECIDS). Firstly, this paper gives the construction method for cost-sensitive ensemble SVM Classification, which is integrated by the classifiers with oversampling, sub-sampling and reconstituted sample space. Secondly, we propose a classifier method BL_KNNModel, which is based on KNNModel algorithm for imbalanced massive data streams. BL_KNNModel can detect the concept drift streams by using the variable windows size, which has lower time complexity. At last, the cost-sensitive ensemble classifier for imbalanced massive data streams is given, which has the virtue of high classification and lower time complexity. In addition, the cost-sensitive ensemble SVM algorithm is used to handle the confused instances. The experiments using both synthetic and real datasets show that compared to the other classification algorithms for imbalanced data streams, CECIDS has higher evaluating indicator and more excellence integrated learning curve.

Keywords: Imbalanced data streams, Ensemble classification, Cost-sensitive SVM

1 Introduction

In recent years, with the development of computer techniques, more and more massive data Streams are produced in wireless sensor networks, network traffic monitoring, credit card fraud detection, earthquake monitoring and weather forecast. The traditional data mining methods are designed on static data, and massive data streams are real-time and dynamic variability, so the traditional data mining techniques can not effectively handle massive data and the research for massive data streams is the hot spot in mining data. The current model for massive data streams can be classified into two categories, one is the single-model structure, and the other is multi-model. A model structure is used to classify the data flow in single-model structure, and the common classification models are as follows. Abdulsalam proposed a method to use incremental learning for data stream classification, which can accelerate self-renewal by the incremental processing model [1]. Hashemi designed a complete data stream processing classification method, which is called the flexible decision tree model [2]. Tang proposed SVM model for unbalanced data stream classification [3]. Kuncheva proposed a data stream classification variable model with sliding window, which can automatically enlarge and reduce the size of the window when the concept drifts occur [4]. Liang extended the very fast decision tree algorithm for handling no labels data [5]. The integrated classification models for massive data streams use multiple identical and different classification models with the hybrid architecture to build strong classifiers. Amin was inspired by thoughts of the integrated learning, and used multiple single-layer neural networks to build an integrated learning model [6]. Avci proposed a mixed method based on the genetic algorithm and support vector machine model [7]. Bu built the hybrid learning model to use Bayesian networks and probabilistic neural network combination by the context of human motional data stream [8]. Chen proposed a mixed element based on Meta-evolution rules respond classification model [9]. Chi proposed an integrated learning ideological classification model based the hyper spectral remote sensing data stream [10]. Dembczynski built integrated learning classification model for the multi-standard problem to use the monotonic constraint method [11]. Hashemi built an integrated multi-decision tree classification model by a hybrid approach [12]. Huang used the ant colony optimization algorithm and support vector machine model to build an integrated classification model [13]. Adler used the bootstrap method by extracting different training set to build an integrated learning model [14]. Baumgartner and Serpen proposed a heuristic classification hybrid model [15]. The distributions of massive data streams in practical applications in are unbalanced, and the numbers of certain classes are less than other classes, which are called imbalanced massive data stream. There are little researches on imbalanced data stream classification. Gao proposed an ensemble classification model by oversampling for imbalanced data streams [16]. Adel proposed cost-sensitive neural network for imbalanced

* ***Corresponding author*** e-mail: 57065873@qq.com

data streams [17]. Liu gave algorithm of reusing data for classifying skewed data stream [18]. GAO proposed an approach for classifying data with imbalanced class distributions and concept drifts [19]. By improving KNNModel algorithm and constructing cost-sensitive ensemble SVM Classification, this paper proposes the approach of the cost-sensitive ensemble classification for imbalanced massive data streams.

2 Cost-sensitive ensemble SVM classification

2.1 COST-SENSITIVE SVM CLASSIFICATION

The goal of traditional SVM minimizes the objective function, and the linear inseparable is as follows.

$$R(w,\xi)=\frac{1}{2}\|w\|^2+C\left(\left(\sum_{i=1}^{n}\xi_i\right)\right), \tag{1}$$

$$s.t\ \ y_i\left(x_i\times w+b\right)\geq 1-\xi_i,\quad \xi_i\geq 0, i=1,...,n.$$

Misclassification is asymmetric, and cost-sensitive SVM is expressed as follows.

$$R(w,\xi)=\frac{1}{2}\|w\|^2+C\left(\left(\sum_{i=1}^{n}\cos t_i\xi_i\right)\right), \tag{2}$$

$$s.t\ \ y_i\left(x_i\times w+b\right)\geq 1-\xi_i,\quad \xi_i\geq 0, i=1,...,n.$$

$\cos t_i$ is misclassification of different sample, and $\sum_{i=1}^{n}\cos t_i\xi_i$ is experience cost. C is relaxation factor, and it controls the balance of the structure and experience cost.

The optimal function for cost-sensitive SVM classification in the case of the linear separable is as follows.

$$f(x)=\mathrm{sgn}\left\{\sum_{i=1}^{n}\alpha_i^* y_i\left(x_i\cdot x\right)+b^*\right\}. \tag{3}$$

The above formula is turned into dual form by the Lagrange equation, and the optimal function for cost-sensitive SVM classification with kernel function is as follows in the case of the linear inseparable.

$$f(x)=\mathrm{sgn}\left\{\sum_{i=1}^{n}\alpha_i^* y_i K\left(x_i\cdot x\right)+b^*\right\}. \tag{4}$$

2.2 OVERSAMPLING COST-SENSITIVE SVM CLASSIFICATION

$\cos t_i$ is misclassification cost of class i, and $\cos t_{ij}\left(i,j\in 1,2,...,C\right)$ is cost of classifying i to j. $\cos t_{ii}=0\left(i,j\in 1,2,...,C\right)$. $\cos t_i=\sum_{j=1}^{C}\cos t_{ij}$.

Oversampling method can change the distribution for the training set, and copy the samples with high cost, so the distribution of each class is proportional to its cost.

The Oversampling samples number is calculated as follows.

$$\lambda=\arg\min_{i\in C}\frac{\cos t_i}{N_i},\ i=1,2,...,C, \tag{5}$$

$$N_k^*=\left\lfloor\frac{\cos t_k}{\cos t_\lambda}N_\lambda\right\rfloor.$$

N_k is the number of class k, and N_k^*-N_k is the added samples number. The added samples are replaced randomly, and the description of oversampling Cost-sensitive SVM Classification is as follows.

Input: Training sample set $S=S_1\cup S_2\cup...\cup S_C$.

Output: Oversampling Cost-sensitive SVM Classification OCS_SVM

Step 1: $S^*=S$ is new data set.

Step 2: for (i=0;i<=C;i++)

{

$$\lambda=\arg\min_{i\in C}\frac{\cos t_i}{N_i},\ i=1,2,...,C.$$

$$N_k^*=\left\lfloor\frac{\cos t_k}{\cos t_\lambda}N_\lambda\right\rfloor;$$

If ($N_k^*\succ N_k$) N_k^*-N_k samples are randomly placed into s^* from S_k;

}

Step 3: Use data set S^* to train the cost-sensitive SVM.

2.3 SUB-SAMPLING COST-SENSITIVE SVM CLASSIFICATION

Sub-sampling method changes the distribution of the train sets, and leads that the samples number is proportional to its cost. Sub-sampling method is different with oversampling, and it decreases samples with low cost.

The sub-sampling samples number is calculated as follows.

$$\lambda=\arg\max_{i\in C}\frac{\cos t_i}{N_i},\ i=1,2,...,C.$$

$$N_k^*=\left\lfloor\frac{\cos t_k}{\cos t_\lambda}N_\lambda\right\rfloor.$$

N_k is the number of class k, if $N_k^*\prec N_k$, sample numbers N_k^*-N_k in class k are decreased. The description of sub-sampling Cost-sensitive SVM Classification is as follow.

Input: Training samples set $S=S_1\cup S_2\cup...\cup S_C$.

Output: Sub-sampling Cost-sensitive SVM Classification UCS_SVM

Step 1: $S^*=S$ is new data set.

Step 2: for (i=0; i<= C ;i++)

{

S^*-S_k Samples are placed into s^*

$$\lambda=\arg\max_{i\in C}\frac{\cos t_i}{N_i},\ i=1,2,...,C.$$

$$N_k^*=\left\lfloor\frac{\cos t_k}{\cos t_\lambda}N_\lambda\right\rfloor;$$

if ($N_k^*\prec N_k$) $\left\lfloor\frac{N_k^*}{2}\right\rfloor$ samples are randomly placed into s^* from S_k ;

else S_k samples are randomly placed into s^*.

If (the samples number with class k is more than N_k^*)

Delete randomly samples.

}

Step 3: Use data set s^* to train the cost-sensitive SVM.

2.4 COST-SENSITIVE SVM CLASSIFICATION OF RECONSTITUTED SAMPLE SPACE

The sample set $X=\{(x_1,y_1),(x_2,y_2)...(x_i,y_i)...(x_n,y_n)\}$, $x_i\in R^\ell$, $y_i\in\{C_1,C_2,...C_m\}$. $p(j\mid x)$ is the probability that sample x belongs to class j, The minimum conditional risk of classifying sample x into class i is as follows by Bayes decision theory.

$$\mathrm{R}(\mathrm{i}\mid x)=\sum_j P(j\mid x)C(i,j).$$

The reconstituted training samples $\bar{y}=\arg\min_{i\in m}\{R(i\mid x)\}$.

$$\cos t_{M*\mathrm{M}}=\begin{bmatrix}\cos t(a_1,\mathrm{a}_1) & \cos t(a_1,a_2) & \cos t(a_1,a_3) & ..... & \cos t(a_1,a_M)\\ \cos t(a_2,a_1) & \cos t(a_2,a_2) & \cos t(a_2,a_3) & ..... & \cos t(a_2,a_M)\\ \cos t(a_3,a_1) & \cos t(a_3,a_2) & \cos t(a_3,a_3) & ..... & \cos t(a_3,a_M)\\ ..... & ..... & ..... & ..... & .....\\ \cos t(a_M,a_1) & \cos t(a_M,a_2) & \cos t(a_M,a_3) & ..... & \cos t(a_M,a_M)\end{bmatrix}$$

M is class number, and $\cos t(a_\mathrm{i},a_\mathrm{j})$ is cost for forecasting a_i into a_j, $\cos t(a_i,a_i)=0$.

Test sample $X=\{x_1,x_2,x_3,.....,x_n\}$, $x_i=\{x_{i1},x_{i2},x_{i3},.....,x_{id}\}$, the basic classifier f_i consumes the misclassification cost as follows.

Input: cost-sensitive SVM, training sample set $X=\{(x_1,y_1),(x_2,y_2)...(x_i,y_i)...(x_n,y_n)\}$,. $y_i\in\{C_1,C_2,...C_m\}$.

Output: Cost-sensitive SVM Classification of reconstituted sample space RS_SVM

Step1: for(i=1;i<=n; i++)

for(j=1;j<=m; j++)

Calculate $P(C_j\mid x_i)$ by Cost-sensitive SVM Classification.

Step 2: for (k=1; k<=n;k++)

{

$$\mathrm{R}(i\mid x)=\sum_{i,j\in m}P(j\mid x)C(i,j)$$

$$\bar{y}=\arg\min_{i\in m}\{R(i\mid x_k)\}$$

}

Step 3: Reconstruct training set X into $\bar{X}$.

Step 4: Reconstitute Cost-sensitive classifier RS_SVM by training set $\bar{X}$ and Cost-sensitive SVM Classification.

2.5 COST-SENSITIVE WEIGHT INTEGRATION OF CLASSIFIER

The traditional integrated learning algorithms for massive data streams prefer the classification accuracy and overall performance, and set the accuracy as weights of the base classifier. For imbalanced massive data, we must not only consider the overall performance, but also the classification accuracy of minority class, so this paper proposes the cost-sensitive weight integration. Cost matrix is as follows.

$$\mathrm{mis\_cost}_{f_i}=\sum_{i=1}^{n}\cos t(a_i,a_j). \quad (6)$$

The test for sample is also paid, and test cost of sample x_i is as follows.

$$test\_\cos t_{f_i}=\sum_{i=1}^{n}\sum_{j=1}^{d}test(x_{ij})$$

This paper integrates OCS_SVM, RS_SVM and UCS_SVM. The weight of three cost-sensitive SVM is as follows.

$$w_{f_i} = \sum_{j=1}^{n} accurcy(x_i) \Big/ \left(\text{mis\_cost}_{f_i} + test\_cost_{f_i}\right), i = 1,2,3. \quad (7)$$

Use OCS_SVM, RS_SVM and UCS_SVM in current training samples, and structure the base classifiers $\left(f_1^1, f_2^1, ..., f_k^1\right)$, $\left(f_{K+1}^2, f_{k+2}^2, ..., f_{2*k}^2\right)$, $\left(f_{2k+1}^3, f_{3k+2}^3, ..., f_{3*k}^3\right)$. Select k classifiers with higher weight for all classifiers to form $\left(f_1^t, f_2^t, ..., f_k^t\right), t \in 1,2,3$.

$$f(x) = \sum_{i=1}^{k}\left(w_{fi} * f_i^t(x)\right). \quad (8)$$

$x \in R^n$, $f_i^t(x)$ is predicted value of sample x. $f(x)$ is classification result for test sample.

3 Improved KNNModel

Guo proposed KNNModel algorithm, which extends outward by each training centre, and build the multiple models cluster for training set, so that the models cover maximum similar instances without any heterogeneous instances. KNNModel algorithm has higher classification accuracy when the samples are covered by the model clusters, but if the classified samples are boundaries, the classification accuracy is affected. The time complexity of KNNModel is high, and is not suitable for rapid classification of data stream. This paper improves KNNModeld algorithm for the concept drift detection of imbalanced massive data streams.

The training set $X = \{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i), ..., (x_n, y_n)\}$, $x_i = \langle x_{i1}, x_{i2}, ..., x_{id}\rangle$, $y_i \in \{1,2,...,K\}$. Define the class cluster $unba\_class_i^k = \left(type_i^k, center_i^k, class_i^k, Tolerance_i^k, Radius_i^k, Num_i^k\right)$, $unba\_class_i^k$ is the model cluster i with class k. If class k is the minority class, $type_i^k = 1$. Otherwise $type_i = 0$. $Center_i^k = \frac{1}{Num_i^k}\sum_{y_i=i} x_i$. $class_i^k = k$ is the class of model cluster, and $Tolerance_i^k$ is tolerability. If the class is minority, $Tolerance_i^k = 0$. In other word, the model clusters of minority class are not different samples. $Radius_i^k$ is radius, and Num_i^k is sample number of model cluster.

Algorithm: Training process of BL_KNNModel

Input: Training data set X, the majority class $mc \in \{1,2,...,K_1\}$, the minority class $lc \in \{1,2,...,K_2\}$, the tolerability α for the majority class.

Output: Model cluster of each class.

Begin

For (k=1; k<=K_1;k++)

{

Use k-means algorithm to structure three cluster centre $center_1^k$, $center_2^k$, $center_3^k$ for class k.

$unba\_class_i^k = \left(type_i^k, center_i^k, class_i^k, Tolerance_i^k, Radius_i^k, Num_i^k\right), i = 1,2,3.$

}

For (k=1;k<=K_2;k++)

{

Label all samples into uncovered in class k.

Each uncovered sample is centre, and expand it the area, which covers the same class point without the other point. Each model cluster is $unba\_class_i^k = \left(type_i^k, center_i^k, class_i^k, Tolerance_i^k, Radius_i^k, Num_i^k\right), i = 1,2...$

Order degradedly the model clusters according to the samples number.

If the model cluster $unba\_class_i^k$ is covered by $\left\{unba\_class_0^k, unba\_class_1^k, ..., unba\_class_{i-1}^k\right\}$, delete it.

If the number that the model cluster covers is not more than three, delete it.

}

Structure cost-sensitive ensemble Classification.

End.

Algorithm 2: Classification process of BL_KNNMode

Input: Data streams DS, Cost-sensitive ensemble Classification, the model clusters of minority and majority class.

Output: Classification of data streams DS.

Begin

Calculate the distance form each test sample to each model centre.

If (The test sample is covered by the model clusters with the same classifier)

Label the test sample as the covered model cluster.

else if (The test sample is covered by different classifiers)

Label the classifier, which is the nearest form the test sample with the model centre.

else use cost-sensitive ensemble Classification.

BL_KNNModel runs K times k-Means cluster for structuring the model, and time complexity of k-Means cluster is $O(n_k * Num)$, and Num is cluster number, so the total complexity is $O(\max(k_1, k_2) * n_k * Num)$. $\max(k_1, k_2)$ and Num are free-running with n_k, $Num \ll n_k$, $\max(k_1, k_2) \ll n_k$, so the time complexity of BL_KNNModel is $O(n)$, and it is suitable for massive data streams classification.

4 Modules for concept drift detection

Concept drift is the characteristics of the data streams, and the overall data distribution changes over time, so that the accuracy of classification models is lower. According to concept drift problem, how to design to data streams classification model has become a hot research. At present, there are ways of instance selection, weighted instance and integration learning. This paper improves BL_KNNModel algorithm, if the density of minority class and majority class change, the concept drift occurs and the classification models should been updated. The existing concept drift detection ways use the fixed length window, but how to set the window length is difficult. The window is too long, and it is not sensitive to concept drift. On the other hand, the window is too short, and it cannot contain the enough new instances that update the model, so this paper proposes the up-and-down window. BL_KNNModel detects the massive data stream and finds the instances that are covered by any model cluster, if the uncovered instance number adds up to the fixed value $window\_length$, all current instances constitute the up-and-down windows. Detect the probability density of each model cluster, and if the differential density with previous windows exceeds the fixed threshold, the concept drift occurs.

Algorithm 3: Concept drift detection algorithm.

Input: Massive data streams DS, the probability density for each previous model cluster, the fixed $window\_length$ for the uncovered instance number, the probability density threshold τ_1 of minority class, the probability density threshold τ_2 of majority class.

Output: Occurrence of concept drift

Begin

Step 1: Run repeatedly BL_KNNModel for data streams DS, until the uncovered instance number adds up to the fixed value $window\_length$.

Step 2: Calculate the probability density of each current model cluster.

Step 3: Compare the probability density of the same model cluster in current and previous windows, if the difference exceeds the fixed threshold, the concept drift occurs.

End.

5 Cost-sensitive ensemble classifier for imbalanced data streams

The descriptions of Cost-sensitive Ensemble Classification for Imbalanced Massive Data Streams are as follows.

Input: Imbalanced Massive Data Streams $S=\{S_1,S_2,...,S_j,...\}$, the BL_KNNModel model clusters and cost-sensitive weight integration of classifier, the previous windows, the uncovered instance number $window\_length$

Output: Classification of $S_1,S_2,...,S_j,...$

Begin

i=0; l=0;

while (i<= $window\_length$)

{

Use algorithm BL_KNNModel to classify an sample S_i.

l++;

If (S_i is uncovered by any model)

i++;

}

Use algorithm 3 to test whether $\{S_1^{'},S_2^{'},...,S_l^{'}\}$ occurs concept drift.

If (Concept drift occurs)

Use BL_KNNModel to reconstruct the model that occurs concept drift, and the cost-sensitive weight integration of classifiers are also rebuilt.

Use rebuilt BL_KNNModel to classify $\{S_1^{'},S_2^{'},...,S_l^{'}\}$.

6 Simulation experiment

6.1 DATA SET

This paper selects the synthetic and real datasets. In view of rotating hyper plane, the synthetic datasets with concept drift are composed by the modified MOA Task Launcher. The sample $X=\{x_1,x_2,...,x_d,\}$. $\sum_{i=1}^{d}a_ix_i=a_0$, $a_0=r\sum_{i=1}^{d}a_i$. a_i is weight number, $r\left(r\neq 1/2\right)$ is gradient of slope of imbalanced datasets. This paper sets a series of parameters for concept drift in data streams. The parameter t is the change size of weight number a_i, The parameter $s_i\in\{-1,1\}(1\le i\le k)$ is the change direction. When a new sample is produced, a_i is adjusted by $s\bullet t$. The dimensionality for data streams is 50, the attribute varies with time, and the imbalance ratio is 0.05.

Use KDDCUP'99 as the real datasets. This paper chooses randomly Normal, Neptune and smurf as the majority, and they are negative and share 99.95%. The connection type of Land, Rootkit, Warezmaster are selected as the minority, and they are positive and share 0.5%. Sample randomly KDDCUP as the second experiment data streams which is named as Stream_KDD.

6.2 EVALUATION METRICS FOR IMBALANCED DATA STREAMS

In general, the overall accuracy is often used to measure the performance for classification model. However, the classification accuracy of the negative samples in unbalanced data streams is greater influence than the positive. Therefore, the classification accuracy is not suit for unbalanced data streams. *TP* is the positive samples number that are correctly classified and *TN* is negative. *FP* is the positive samples number that are improperly classified and *FN* is negative. This paper gives the performance evaluation as follows.

$$recall = \frac{TP}{TP + FN}, \tag{9}$$

$$Precision(\Pr) = \frac{TP}{TP + FP}, \tag{10}$$

$$G - Mean = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}}, \tag{11}$$

$$F - measure(Fm) = \frac{(1+\beta)^2 \times recall \times precision}{\beta^2 \times recall + precision}, \tag{12}$$

$$TPR = \frac{TP}{TP + FN}, \tag{13}$$

$$TNR = \frac{TN}{TN + FP}, \tag{14}$$

$$Mean\ Squared\ Error\ (MSE) = \frac{1}{|T|} \sum_{x_i \in T} \left(f(x_i) - p(+ \mid x_i) \right)^2. \tag{15}$$

T is test sample datasets, and $f(x_i)$ is classification result of sample x_i. $p(+ \mid x_i)$ is real posterior probability of sample x_i.

6.3 RESULT OF EXPERIMENTS

For test the efficiency of CECIDS, the test environment is Pentium D3.5GHz CPU, 8GM RAM windows 7+64bit, and select Weka+JDK6.0 as the development environment. The parameters of all algorithms refer to the papers [17, 19], and the results of synthetic datasets with concept drift are as follows. CECIDS that this paper proposes compares with RDFCSDS [16], SS [18], and OACLNIS [17], which are the other classification algorithms for imbalanced data streams.

TABLE 1 Results of synthetic datasets with concept drift

Algorithm	TPR	TNR	G-mean	Pr	Fm	MSE
SS	0.0067	0.9987	0.0923	0.9032	0.0107	0.0215
RDFCSDS	0.0246	0.9966	0.1235	0.9223	0.0345	0.0427
OACLNIS	0.0642	0.9952	0.2445	0.9465	0.1854	0.0564
CECIDS	0.0723	0.9941	0.3384	0.9668	0.2365	0.0736

The results of Stream_KDD datasets are as follows.

TABLE 2 Results of Stream_KDD datasets

Algorithm	TPR	TNR	G-mean	Pr	Fm	MSE
SS	0.0008	0.9967	0.0848	0.9254	0.0144	0.3564
RDFCSDS	0.0045	0.9943	0.1437	0.9365	0.0383	0.5324
OACLNIS	0.0093	0.9925	0.1865	0.9523	0.1624	0.6384
CECIDS	0.0102	0.9906	0.2873	0.9786	0.2413	0.8414

Form the table 1 and 2, for the evaluating indicator of TPR, G-mean, Precision, F-measure and MSE, CECIDS is higher than the other algorithms for imbalanced massive data streams, and the evaluating indicator TNR in CECIDS is lower, so CECIDS algorithm is more excellent than the other.

The classification accuracy is not suitable for imbalanced massive data streams, and this paper selects G-mean and F-measure as integrated learning curve for imbalanced massive data streams. The integrated learning curve for imbalanced massive data streams in the synthetic datasets is as follows in figure 1 and figure 2.

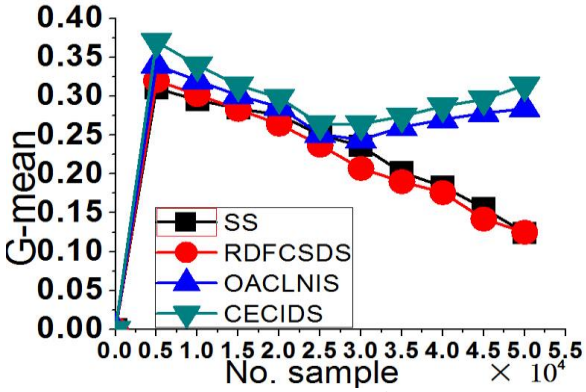


FIGURE 1 G-mean integrated learning curve for the synthetic datasets

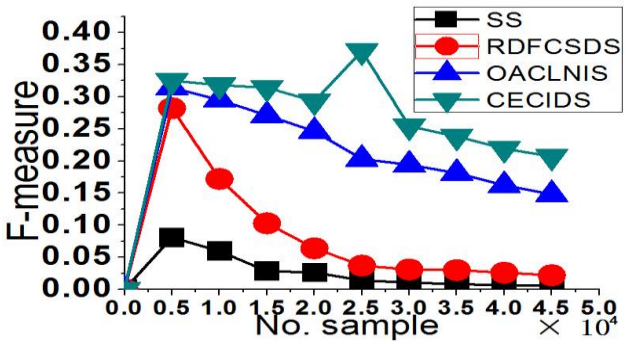


FIGURE 2 F-measure integrated learning curve for the synthetic datasets

The integrated learning curve for imbalanced massive data streams in the Stream_KDD datasets is as follows in figure 3 and figure 4.

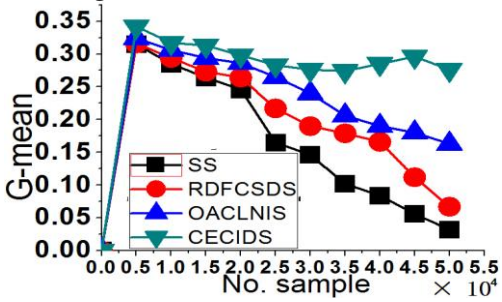


FIGURE 3 G-mean integrated learning curve for Stream_KDD datasets

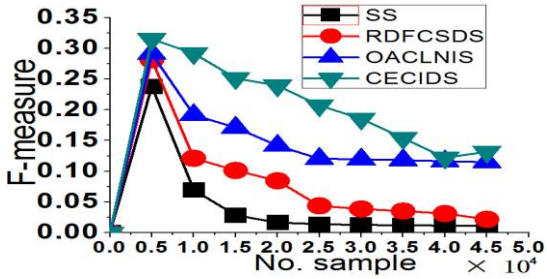


FIGURE 4 F-measure integrated learning curve for Stream_KDD datasets

From figure 1 to 4, when the data streams increase gradually, G-mean and F-measure integrated learning curve of CECIDS for synthetic and Stream_KDD datasets are more effective than the other classifiers for imbalanced data streams. Almost all the experiments validate that CECIDS that this paper proposes are more excellence than the other classification algorithms for imbalanced data streams.

7 Conclusions

The classification of concept drifting data streams has been a hot research, and this paper proposes the approach of the cost-sensitive ensemble classification for imbalanced massive data streams. A new classifier method BL_KNNModel based on KNNModel algorithm is given, which can detect the concept drift streams by using the variable windows size. By constructing the different model cluster for every class, if the concepts drift occurs on the model cluster, it only needs to rebuild the corresponding classification model. The experiments on the synthetic and real datasets prove the effectiveness of the algorithms that this paper proposes. The next work focuses on trying to adopt a new measure to calculate the distance between two samples, and distinguishes effectively the noise and concept drift.

Acknowledgments

This work is supported by the science and technology projects of the Shandong province universities (No. J14LN65, No. J13LN53). This paper comes from the research fund of Heze University (No. XY12KJ03).

References

[1] Aboalsamh H L 2009 A novel incremental approach for stream data mining *AEJ-Alexandria Engineering Journal* **48**(4) 419-26
[2] Hashemi S, Yang Y 2009 Flexible decision tree for data stream classification in the presence of concept change, noise and missing values *Data Mining and Knowledge Discovery* **19**(1) 95-131
[3] Tang Y, Zhang Y Q, Chawla N V 2009 SVMS modeling for highly imbalanced classification *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **39**(1) 281-8
[4] Kuncheva L I, Zliobaite I 2009 On the window size for classification in changing environments *Intelligent Data Analysis* **13**(6) 861-72
[5] Liang C, Zhang Y, Shi P, et al. 2012 Learning very fast decision tree from uncertain data streams with positive and unlabeled samples *Information Sciences* **213**(5) 50-67
[6] Amin M F, Islam M M, Murase K 2009 Ensemble of single-layered complex-valued neural networks for classification tasks Neurocomputing **72**(10-12) 2227-34
[7] Avci E 2009 Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm-support vector machines: HGASVM *Expert Systems with Applications* **36**(2) 1391-402
[8] Bu N, Okamoto M, Tsuji T 2009 A Hybrid Motion Classification Approach for EMG-Based Human-Robot Interfaces Using Bayesian and Neural Networks *IEEE Transactions on Robotics* **25**(3) 502-11
[9] Chen T C, Tsao H L 2009 Using a hybrid meta-evolutionary rule mining approach as a classification response model *Expert Systems with Applications* **36**(2) 1999-2007
[10] Chi M M, Kun Q, Benediktsson J A, et al. 2009 Ensemble Classification Algorithm for Hyperspectral Remote Sensing Data *IEEE Geoscience and Remote Sensing Letters* **6**(4) 762-6
[11] Dembczynski K., Kotlowski W, Slowinski R 2009 Learning Rule Ensembles for Ordinal Classification with Monotonicity Constraints *Fundamenta Informalicae* **94**(2) 163-78
[12] Hashemi S, Yang Y, Mirzamomen Z, et al. 2009 Adapted One-versus-all decision trees for data stream classification *IEEE Transactions on Knowledge and Data Engineering* **21**(5) 624-37
[13] Huang C L 2009 ACO based hybrid classification system with feature subset selection and model parameters optimization *Neurocomputing* **73**(1-3) 438-48
[14] Adler W, Brenning A, Potapov S, et al. 2011 Ensemble classification of paired data *Computational Statistics & Data Analysis* **55**(5) 1933-41
[15] Baumgartner D, Serpen G 20122 A design heuristic for hybrid classification ensembles in machine learning *Intelligent Data Analysis* **16**(2) 233-46
[16] Gao J, Ding B L, Fan W, Han J W, Philip S Y 2008 Classifying data streams with skewed class distributions and concept drifts *IEEE Internet Computing* **12**(6) 37-49
[17] Adel G, Reza M, Hadi S Y 2013 Ensemble of online neural networks for non-stationary and imbalanced data streams *Neurocomputing* **122** 535-44
[18] Liu P, Wang Y, Cai L J, Zhang L B 2020 Classifying skewed data streams based on reusing data *In Computer Application and System Modeling (ICCASM) of 2010 International Conference* **4** V4-90-V4-93
[19] Guo G D, Li N, Chen L F 2014 Concept Drift Detection for Data Streams Based on Mixture Model *Journal of Computer Research and Development* **51**(4) 731-42

Author

Yuwen Huang, born in 1978 at Shanxian

Current position, grades: lecturer at the Department of Computer and Information Engineering, Heze University.
University studies: Master of Engineering in Computer Science from the "Guangxi Normal university" in 2009.
Scientific interest: the data-mining, intelligence Calculation.

Accurate self-localization of mobile robots based on vision sensors

Xiaoming Dong*

Department of Information and Technology Anqing Teachers college, 128 Linghu South Road, Anqing, Anhui province, China

Received 6 July 2014, www.cmnt.lv

Abstract

Robot localization is a challenging problem in indoor environment since no GPS information is available. In this paper, an algorithm was proposed for accurate localization, which designed a delicate way to extract the feature points at first, then the position of the robot was determined using the relation of the features in different images, finally, the Kalman filer was designed to decrease the error caused by robot's moving. Experiments showed the accuracy and robustness of this algorithm.

Keywords: mobile robot, multi-cues fusion, vision sensor, Kalman filer, localization

1 Introduction

Self-localization of Mobile robots is one of the most important tasks in the various applications. GPS is a commonly used technique in outer environments. However, GPS cannot be used in some special conditions such as indoor environments [1-4]. What's more, GPS cannot achieve the high accuracy required in some applications of mobile robots. Vision sensors based on cameras have long been used to implement the localization of mobile robots, which is a promising technique that improves the capability of a mobile robot to estimate its position and the speed of motion. Extensive researches have been done to make use of vision sensors in practice, monocular image sequences or stereo image sequences are used in several algorithms [5-9]; however, because of the inaccurate feature matching, there are still many difficulties to estimate the robots position effectively. The aim of this paper is to design a good method to achieve the task of localization in indoor environments.

The paper is organized as follows. In section 2, we designed a delicate way to detect the position of the distinguished feature points. Then the outlier rejection is discussed, section 3 discussed the way we used to estimate the position of robots, section 4, to decrease the error further in the time that robots move, Kalman filter is designed. In Section 4, the algorithm is verified in various experiments. Finally, the conclusions are given in section 5.

2 Feature extractions and outlier rejection

The primary idea in vision sensors is to select the distinguished features to describe the environments. How to gain the accurate point features are the most important problem in robot localization [10-12]. In our proposed algorithm, considering the real environment characters, we designed a hierarchical framework to obtain the accurate feature points. The framework mainly contains two parts, which are the global search and local search. In global search, we extract feature points as many as possible, while in local search, we only choose the feature points that have a good effect in pose estimation.

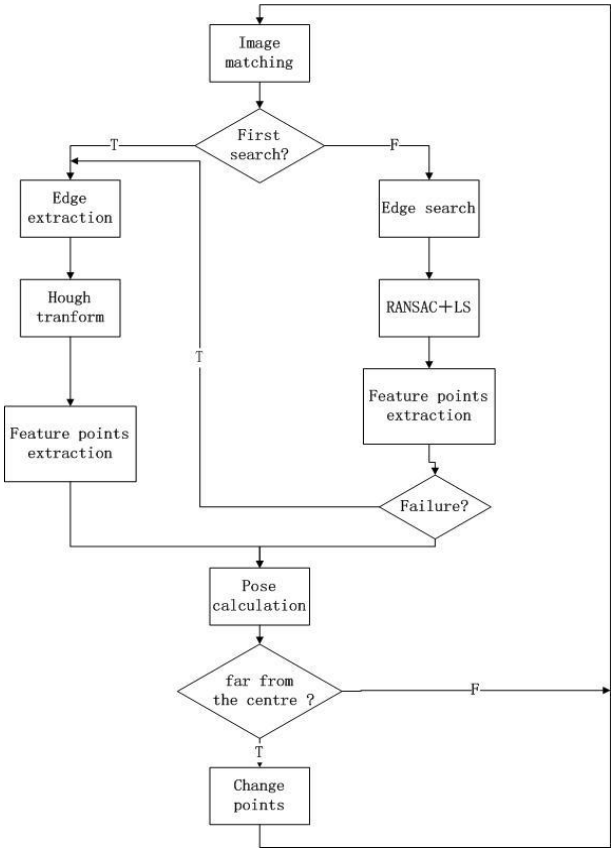


FIGURE 1 The flow graph of the feature extraction

* ***Corresponding author*** e-mail: dongxiaoming008@126.com

Figure 1 shows the detailed process, which can be described as follow:

1) When the robot is the first time to run, the edge points is to be extracted in the global scope.

2) The parameter of the lines is obtained by using Hough Transform.

3) The optimization is implemented by using least square.

4) The cross of the lines is chosen as the feature points.

Considering the speed of the calculation and the changes in the continuous images are little, so we firstly extract the lines in the neighboring area of the last image.

When the robot is not the first time to run, and the Hough transform plus the least square are implemented similarly. If it had a great change in environments, which contributes to the results that the local search loses effectiveness, the robot turns to the global search.

Integrating the two policies to extract the point features, our algorithm can get the distinguished points to pose estimation and acceptable speed. Figure 2 shows the points that robot has obtained from ceiling.

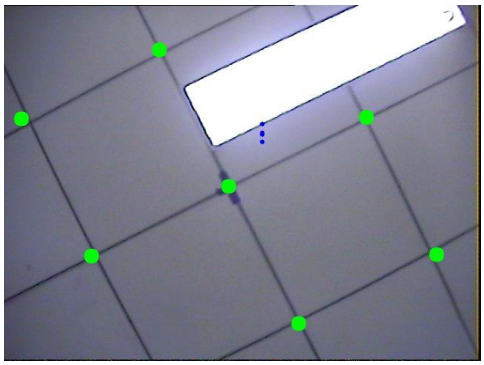


FIGURE 2 the feature points detected

For the accuracy of the point features, the method designs two steps for outlier's rejection. Firstly, an initial estimation of the homography matrix H is achieved by the RANSAC algorithm, secondly, the unmatched point features are picked out according to the homography matrix H. the process is described as follows in calculating the homography matrix based on the correspondent features: it is assumed that the correspondent features are $P_i(x_i, y_i)(i = 1,...,5)$, and then we can compute $\mathbf{H}$ as follows. P_i and V_i satisfy:

$$k\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{H}\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & m_9 \end{bmatrix}. \tag{1}$$

Eliminating k from equation (1), we can get two equations. Combining these two equations, we get:

$$\mathbf{KM} = m_9\mathbf{U}, \tag{2}$$

where $\mathbf{K}$, $\mathbf{M}$ and $\mathbf{U}$ are defined as follows $(i = 1,...,5)$:

$$\mathbf{K} = \begin{bmatrix} x_i & y_i & 1 & 0 & 0 & 0 & -x_iu_i & -y_iu_i \\ 0 & 0 & 0 & x_i & y_i & 1 & -x_iv_i & -y_iv_i \\ & & & & \cdots & & & \\ & & & & \cdots & & & \end{bmatrix}, \tag{3}$$

$$\mathbf{M} = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{U}. \tag{4}$$

The process can be concluded as follows:

1) Capturing the current frame and rectifying it to remove the distortion.

2) Finding out the correspondent features between the current frame and the last frame, then computing the homography matrix H to reject outliers.

3) Capturing the next frame and repeat this process.

3 Pose estimation

3.1 POSE ESTIMATION

Assuming that the coordinate of point P is $[x_w, y_w, z_w, 1]^T$ in the world coordinate system, then the Equation (5) is satisfied, the Equation (5) is called external parameter model of camera.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} {}^cn_{wx} & {}^co_{wx} & {}^ca_{wx} & {}^cp_{wx} \\ {}^cn_{wy} & {}^co_{wy} & {}^ca_{wy} & {}^cp_{wy} \\ {}^cn_{wz} & {}^co_{wz} & {}^ca_{wz} & {}^cp_{wz} \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} R & p \\ 0^T & 1 \end{bmatrix}\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = {}^cT_w\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}. \tag{5}$$

As the Figure 3 shows, on the time i when the robot is moving, the reference coordinate system Or_i is set up in the cross point between the ceiling and the optic axes, And the axis direction of the reference coordinate system is the same as the world coordinate system, Similarly, the reference coordinate system Or_{i+1} is set up too.

When the camera moves from point Oc_i to Oc_{i+1}, the process can be decomposed into two parts: the one is the rotation of the robot around axis Zc_i and the other one is the translation from point Oc_i to point Oc_{i+1}. Then the relative position will be calculated if the position of the feature point P_j in the reference coordinate system is obtained between time i and time i+1.

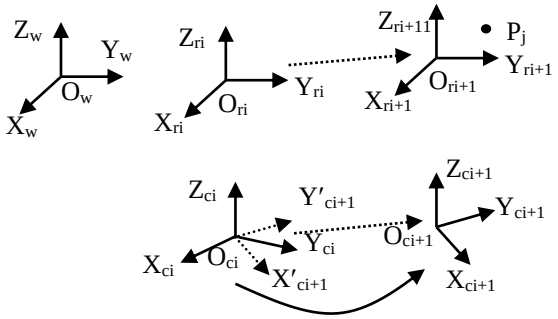


FIGURE 3 The coordinate system in different time

The transform of the reference coordinate can be described as equation 6 in the time *i*.

$$ {}^{i}T_{ri} = \begin{bmatrix} {}^{i}n_{wx} & {}^{i}o_{wx} & {}^{i}a_{wx} & 0 \\ {}^{i}n_{wy} & {}^{i}o_{wy} & {}^{i}a_{wy} & 0 \\ {}^{i}n_{wz} & {}^{i}o_{wz} & {}^{i}a_{wz} & {}^{c}p_{rz} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6} $$

The distance between the camera axis and the feature points is required in the localization. If the camera plane is parallel to the ceiling, it would be a constant. Considering the camera axis O_c is the same both in the reference coordinate system and in the world coordinate system, this can be calculated based on the Equations (5) and (6).

$$ \begin{cases} ({}^{i}n_{wx} - {}^{i}x_{1cj}\,{}^{i}n_{wz})\,{}^{i}x_{rj} + ({}^{i}o_{wx} - {}^{i}x_{1cj}\,{}^{i}o_{wz})\,{}^{i}y_{rj} = {}^{i}x_{1cj}\,{}^{c}p_{rz} \\ ({}^{i}n_{wy} - {}^{i}y_{1cj}\,{}^{i}n_{wz})\,{}^{i}x_{rj} + ({}^{i}o_{wy} - {}^{i}y_{1cj}\,{}^{i}o_{wz})\,{}^{i}y_{rj} = {}^{i}y_{1cj}\,{}^{c}p_{rz} \end{cases}. \tag{7} $$

In the above equations, ($^{i}x_{rj}$, $^{i}y_{rj}$) represents the P_j coordinate and ($^{i}x_{1cj}$, $^{i}y_{1cj}$) represents the normalized coordinate of the image focus.

Then, the coordinate of feature P_j can be calculated in the reference coordinate system of different time.

Because in time *i* and time *i*+1, the transform matrix between the references coordinate system and the camera ordinate system can be represented as Equation (8):

$$ {}^{ri}T_{i} = \begin{bmatrix} {}^{i}n_{wx} & {}^{i}n_{wy} & {}^{i}n_{wz} & -{}^{i}n_{wz}\,{}^{c}p_{rz} \\ {}^{i}o_{wx} & {}^{i}o_{wy} & {}^{i}o_{wz} & -{}^{i}o_{wz}\,{}^{c}p_{rz} \\ {}^{i}a_{wx} & {}^{i}a_{wy} & {}^{i}a_{wz} & -{}^{i}a_{wz}\,{}^{c}p_{rz} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{8} $$

The localization of the robots can be calculated by iteration based on the equation (9):

$$ \begin{cases} p_{xi+1} = p_{xi} + {}^{i}x_{rj} - {}^{i+1}x_{rj} - {}^{i+1}n_{wz}\,{}^{c}p_{rz} + {}^{i}n_{wz}\,{}^{c}p_{rz} \\ p_{yi+1} = p_{yi} + {}^{i}y_{rj} - {}^{i+1}y_{rj} - {}^{i+1}o_{wz}\,{}^{c}p_{rz} + {}^{i}o_{wz}\,{}^{c}p_{rz} \end{cases}. \tag{9} $$

3.2 ERROR ANALYSIS AND FILTERING

To analyze the error, we set up the moving model of the mobile robot at first, and then the Kalman filter is designed to decrease the error. The robot used in this paper is a wheel robot, its model is shown in Figure 4, assuming that SL represents the distance of the left wheel moves and SR is that of right wheel. The rotation angle of the wheel axis is $\Delta\theta$, θ is the intersection angle, then the relation between the current pose and the last pose can be referred, assuming that $X_i(x_i, y_i, \theta_i)$ is the current pose and $X_{i-1}(x_{i-1}, y_{i-1}, \theta_{i-1})$ is the last pose.

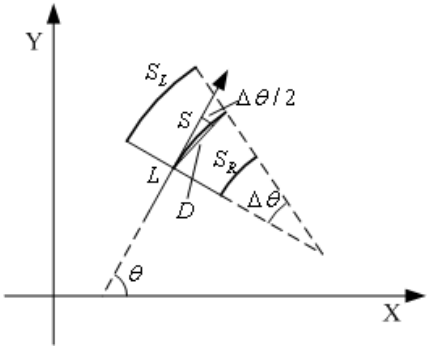


FIGURE 4 The moving model of robot wheels

Then the relation between the rotation angle and the distance the robot moves is:

$$ \Delta\theta = (S_L - S_R)/L\,. \tag{10} $$

The arc length, which is the distance the robot centre has moved, can be represented as follows:

$$ S = (S_L + S_R)/2\,. \tag{11} $$

And *D* can be referred as:

$$ D/2 = S/\Delta\theta \times \sin(\Delta\theta/2)\,. \tag{12} $$

Considering that S_L and S_R are very close, it can be assumed that:

$$ \Delta\theta = 0\,, \tag{13} $$

$$ \begin{aligned} \Delta x &= S \times \cos\theta, \\ \Delta y &= S \times \sin\theta. \end{aligned} \tag{14} $$

Updating the current pose:

$$ X_i = X_{i-1} + (\Delta x, \Delta y, \Delta\theta)\,. \tag{15} $$

Considering the noise, the formula of the pose can be changed as:

$$ X_i = F(X_{i-1},\, S_{Li-1},\, S_{Ri-1}) + W_i\,, \tag{16} $$

where W_{i-1} is assumed as gauss noise and its average value and the variance are:

$$ \mathrm{E}[W_i] = 0,\ \mathrm{E}[W_i, W_i^{\mathrm{T}}] = Q_i\,. \tag{17} $$

Covariance can be modeled as diagonal matrix, the diagonal entries are:

$$Q_{11} = K_x |S\cos\theta|,$$
$$Q_{22} = K_y |S\sin\theta|, \quad (18)$$
$$Q_{33} = K_{S\theta}|S| + K_{\theta\theta}|\Delta\theta|,$$

where K_x and K_y are drift coefficients the robot moves along the axis X and axis Y, similarly, $K_{S\theta}$ and $K_{\theta\theta}$ is the drift coefficients of angles. The values of the coefficients can represent the error. The more the coefficients, the more the errors.

For the more accurate localization, we choose the extended Kalman filter, in EKF framework; the moving model can be represented as:

$$x_t = f(x_{t-1}, u_{t-1}, w_{t-1}). \quad (19)$$

And the observation model is:

$$z_t = h(x_t, v_t). \quad (20)$$

Then, state prediction and observation prediction can be determined as:

$$\hat{x}_t^- = f(\hat{x}_{t-1}, u_{t-1}, 0), \quad (21)$$

$$\hat{z}_t^- = h(\hat{x}_t^-, 0). \quad (22)$$

Using tailor Equation to linear the models, the updating equation in EKF is:

$$\hat{x}_t^- = f(\hat{x}_{t-1}, u_{t-1}, 0),$$
$$P_t^- = A_t P_{t-1} A_t^T + W_t Q_{t-1} W_t^T. \quad (23)$$

And the observation model updating equation is:

$$K_t = P_t^- H_t^T (H_t P_t^- H_t^T + V_t R_t V_t^T)^{-1},$$
$$\hat{x}_t = \hat{x}_t^- + K_t(z_t - h(\hat{x}_t^-, 0)), \quad . \quad (24)$$
$$P_t = (I - K_t H_t) P_t^-.$$

In the design, we set up the observation model according to the character of the vision sensors and decrease the error further.

4 Experiments and analysis

Various experiments have been carried out using the mobile robot. The camera is precisely calibrated beforehand and is mounted on the center of the robot fixedly.

The inner calibration parameter is shown in Table 1 and Table 2. They represent the focus and the center respectively of left camera and right camera.

TABLE 1 The inner parameter of calibration (mm)

Inner parameter	**kx**		**ky**	
	result	*variants*	*result*	*variants*
Left camera	568.27	2.44	566.21	2.76
Right camera	569.35	3.18	569.58	2.98

TABLE 2 The center of the camera (mm)

u0		**v0**	
result	variants	result	variants
228.77	2.19	183.25	2.35
232.12	2.91	177.98	1.98

The external parameter is shown in Tables 3 and 4, which are the parameter of the rotation and the translation respectively. And the unit is mm. the values represent the relative angle and translation between the two cameras.

TABLE 3 Translation vector of the external parameter (relative to the left camera)

Translation vector P(mm)	result	445.83	6.54	8.25
	variants	4.23	3.15	2.86

TABLE 4 Rotation vector of the external parameter (relative to the left camera)

Rotation matrix R	**results**	0.0127; 0.9875; -0.0643 0.9876; -0.0354; -0.1768 -0.1920; -0.0675; -0.9326
	variance	0.05342 0.03425 0.00978

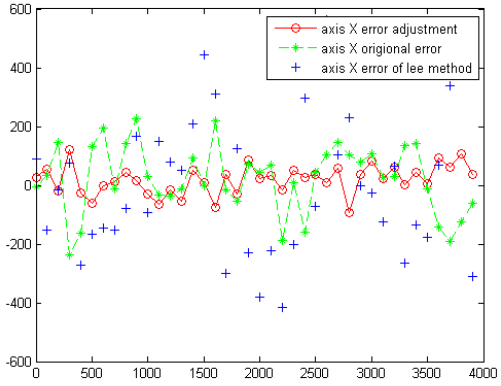


FIGURE 5 The error comparison of axis X

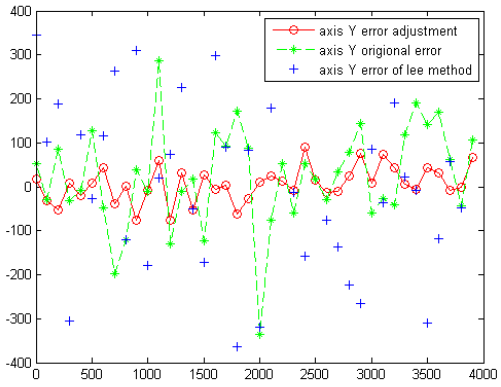


FIGURE 6 The error comparison of axis Y

The comparison is implemented between the method we proposed and the famous Lee method [13], because the accurate feature extraction, the error of our method is about 100mm or so, while the Lee algorithm is about 200mm, the error is decreased further through the Kalman

filter, which is about 50mm after filter. Figure 5 gives the error of axis *X* and Figure 6 gives that of axis Y. The figure shows that the effect of robot localization has been greatly improved.

5 Conclusions and future work

In the various applications, Vision sensor is one of the most important sensors because it can provide larger amount of information comparing with other traditional sensors. In this paper, an effective method is proposed for accurate robot localization. The experiments show its accurateness. Vision sensor has its shortcoming that the features in image would be affected when the environment has great changes. Future work will consider more about the robustness of the Vision sensors, the combination of the traditional sensor and the Vision sensors maybe a possible way to resolve this problem.

References

[1] Kiryati N, Kälviäinen H, Alaoutinen S 2000 Randomized or probabilistic Hough transform: unified performance evaluation *Pattern Recognition Letters* **21**(13-14) 1157-64.
[2] Nistér D, Naroditsky O, Bergen J 2006 Visual odometry for ground vehicle applications *Journal of Field Robotics* **23**(1) 3-20
[3] Yang Z-F, Tsai W-H 1999 *IEEE Transactions on Industrial Electronics* **46**(3) 653-61
[4] Olson C G, Matthies L H Schoppers M, Maimone M W 2003 Rover navigation using stereo ego-motion *Robotics and Autonomous Systems* **43**(4) 215-29
[5] Cumani A, Guiducci A 2010 Comparison of feature detectors for rover navigation *Mathematical Modelling and Applied Computing* **1**(1) 245-50
[6] Schweighofer G, Pinz A 2006 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(12): 2024-30.
[7] Benhimane S, Malis E 2004 Real-time image-based tracking of planes using efficient second-order minimization *Intelligent Robots and Systems* **41**(4) 1943-8
[8] Jepson A D, Fleet D J, El-Maraghi T F 2003 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(10) 1296-311
[9] Kalal Z, Mikolajczyk K, Matas J 2012 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(7) 1409-22
[10] Mei X, Ling H 2011 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(11) 2259-72
[11] Avidan S 2007 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(2) 261-71
[12] Ross D A, Lim J, Lin R-S, Yang M-H 2008 Incremental learning for robust visual tracking *International Journal in Computer Vision* **21**(3) 125-41
[13] Jang G, Kim S, Lee W, Kweon I 2003 Robust self-localization of mobile robots using artificial and natural landmarks *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation 2003* **1** 412-7

Author

Xiaoming Dong, born in November, 1977, Anqing County, Anhui Province, P.R. China

Current position, grades: Associate Professor at the School of Anqing teachers college, China.
University studies: MSc at Beijing science and technology University in China.
Scientific interests: machine vision, pattern recognition.
Publications: more than 10 papers.
Experience: teaching experience of 9 years, 2 scientific research projects.

Identification of the hatching egg before the incubation based on hyperspectral imaging and GA-BP network

Zhihui Zhu[1, 2], Ting Liu[1, 2], Lirong Xiong[1, 2], Meihu Ma[1, 3*]

[1]*National R&D Center for Egg Processing, Huazhong Agricultural University, Wuhan, Hubei, China, 430070*

[2]*College of Engineering, Huazhong Agricultural University, Wuhan, Hubei, China, 430070*

[3]*College of Food Science and Technology, Huazhong Agricultural University, Wuhan, Hubei, China*, 430070

Received 26 August 2014, www.cmnt.lv

Abstract

The removal of the unfertilized egg from the hatching egg before the incubation could improve the efficiency of incubation. The identification of the unfertilized and fertilized eggs by hyperspectral imaging technology combined with GA-BP algorithm was proposed. The comparative analysis for the unfertilized and fertilized eggs was implemented by different pretreatment and principal component. In order to improve the performance of BP neural network, GA algorithm was used to optimize the network parameters. The application of GA-BP network was established the qualitative detection model. The results of the study showed that the MSC + SD pretreatment method was the most suitable for the model. The determination coefficient was 0.95, which indicated the optimized network model had a good generalization ability and high prediction precision with unfertilized eggs accuracy being 93%, fertilized eggs accuracy being 94%, the overall accuracy being 93.5%. The results indicated that the method of non-destructive identification for fertilized and unfertilized eggs based on hyperspectral imaging technology combined with GA-BP algorithm was feasible.

Keywords: hyperspectral imaging technology, before incubation, hatching eggs, identification, GA-BP

1 Introduction

Identification of unfertilized eggs before the incubation is one of the difficult problems in the hatchery industry, which has no reasonable solution yet. Hatchery statistics show that about 8%-9% of all incubated eggs do not hatch due to unfertilized eggs [1], which lead to huge waste of time, space, labour force and energy. In practical applications, candling eggs after 7 to 12 days of incubation are always used, but the pick-out unfertilized eggs have lost edible value, which lead to economic loss. The non-destructive identification of unfertilized eggs before incubation as early as possible can optimize space, save labour and energy, avoid contaminating of other eggs and bring better profits to hatcheries.

In recent years, many scholars have worked with the hatching eggs non-destructive detection using machine vision [2], acoustic pulse vibration [3], optical method [4], nuclear magnetic resonance imaging [5], hyperspectral and so on [6, 7], who have also made certain research progress, but mainly concentrated in the middle of hatching egg quality detection. Our group have used near infrared diffuse reflection for hatching egg quality detection [8], which made some achievements, but need to improve precision. High spectral technology as a new detection technology has a super multichannel, high spectral resolution, narrow band, wide spectral range. It is the perfect combination of image and spectrum technology, which can obtain both image information and spectral information. So more and more people get the attention of hyperspectral image technology in non-destructive testing of the quality of agricultural products, especially widely used in detecting the quality and safety of livestock products, fruits and vegetables [9, 10]. In recent years, some scholars used hyperspectral image technology for detection of unfertilized and fertilized eggs, but they are mainly research on the detection during the early period. The main purpose of this study was to explore the feasibility of hyperspectral image technology on the unfertilized and fertilized eggs detection before the incubation.

2 Materials and method

2.1 EGG SAMPLES

Two hundred eggs including 100 fertilized eggs and 100 unfertilized eggs were collected from Huazhong Agricultural University Hatchery within one week, which were obtained from 50-55-wk-old Single Comb White Leghorn chickens. We randomly chose, for each class, 60% of the subset to build the training set (60 fertile eggs and 60 infertile eggs) and the remaining 40% was put aside for testing set (40 fertile eggs and 40infertile eggs). Corresponding training set and testing set were 120 and 80 samples respectively. After cleaned, all the samples were numbered.

[*]***Corresponding author*** e-mail: zzh@mail.hzau.edu.cn

2.2 TESTING SYSTEM

The Headwall Hyper spectral imaging spectrometer (HS-VNIR N-series) was used, which had 400 nm to 1000 nm spectral range. The hyperspectral image acquisition system was shown in Figure 1. In order to ensure that the sample uniform illumination, six halogen tungsten lamp (Philips, 35W) were displaying asymmetric distribution to form DC circular lighting system. The main working parameters were set as following: the exposure time was 5 ms; the interval of the conveyer was 120 μm; spectrum sampling interval was 0.8 nm; scanning length was 70 mm.

ENVI V. 4.3 (Research System, Inc., USA) and MATLAB V. 7.0 (MathworksCo., USA) software were used to process the data of hyperspectral image. Hyperspectral images were 3D data blocks, each pixel in the image corresponded to one spectrum. 10556 pixels of the elliptical area of egg image were selected to be region of interest (ROI) (the green part in the Figure 2a). The average of the spectrum, which was obtained by the calculation of all pixels in the ROI, was regards as the reflection spectrum of eggs (the white part in the Figure 2b).

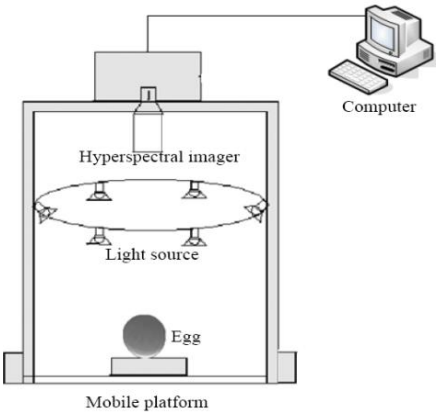


FIGURE 1 Hyperspectral imaging system

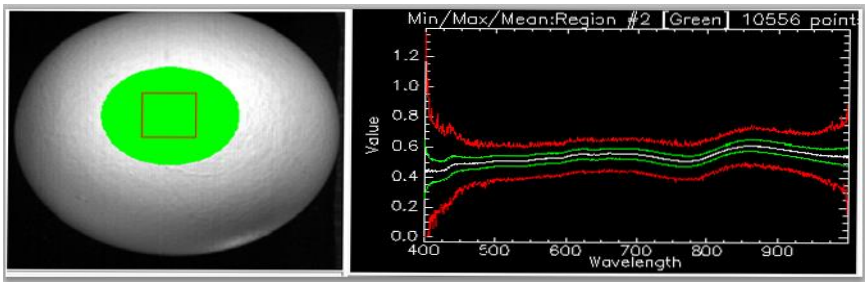


FIGURE 2a Marking of ROI

FIGURE 2b Spectral information of ROI

2.3 THE CALIBRATION OF IMAGES

The uneven light intensity distribution and dark current noise, which existed in the working process of hyperspectral image acquisition system, would cause that the image with the weak light intensity distribution of band contained larger noise. Therefore black and white calibration should be implemented for the hyperspectral image [11]. Before collecting all sample images, the black calibration image *B* was collected with covering lens, and white calibration image *W* was collected by scanning calibration standard white plate, then according to the formula (1), absolute image *I* was converted relative image *R*.

$$R = \frac{I - B}{W - B}. \quad (1)$$

2.4 EGGS ASSESSMENT

The eggs were sent to incubator for hatching after scanned by hyperspectral image acquisition system. The incubator was sterilized by formaldehyde solution fumigation and worked for 24 hours in predefined temperature and humidity conditions before the eggs were put in. The incubator temperature was 38.5 ℃ and the relative humidity was 65%. It was difficult to distinguish unfertilized eggs from fresh hatching eggs by their morphology. Only after hatched for four days, they could be distinguished by candling. The fertilized eggs would appear blood spots through candling in the fourth day, at the same time, unfertilized eggs had no embryonic development phenomenon. According to candling at the fourth day, the actual type of unfertilized eggs and fertilized eggs were determined. If it could not be distinguished clearly, the egg would be broken and observed whether the embryo has developed or not.

2.5 GA-BP NETWORK ALGORITHM

Back-Propagation (BP) neural network has strong nonlinear mapping ability and flexible network structure. But it is easy to fall into local minimum, slow convergence speed and so on. Genetic algorithm (GA) is a kind of adaptive global optimization probability search algorithm based on natural selection in the formation of biological and genetic processes. GA was used to optimize the initial weights and threshold of BP neural network. The optimal initial weights and thresholds were obtained through selection, crossover and mutation operation of GA, which could overcome the random defects of connection weights and thresholds of BP neural network [12]. GA-BP neural network not only played the mapping ability of BP neural network generalization, but also made BP neural network has fast convergence and strong learning ability. The specific algorithm flowchart was shown in Figure 3.

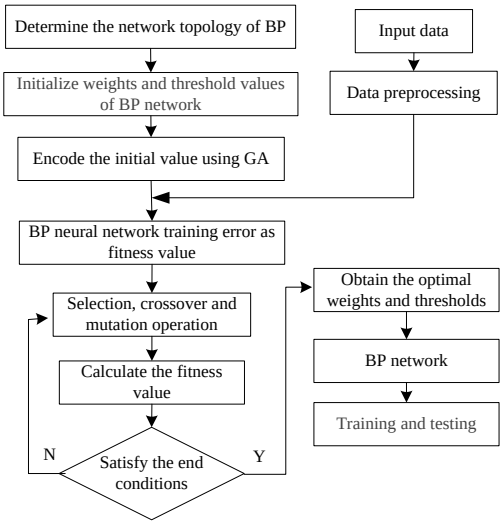


FIGURE 3 The flow chart of GA-BP neural networks

3 Results and discussion

3.1 PREPROCESSING AND DIMENSIONALITY REDUCTION

The spectrum extracted from eggs contained not only the information of the eggs themselves, but also information and noises from eggshell, background and instruments. Therefore, some pretreatments must be taken for the egg spectra to remove the useless information. In this paper, 10 kinds of pretreatment methods were used for the spectra data, which included the first order derivative(FD), the second order derivative (SD), multiplicate scatter correction (MSC), the standard normalization(SNV) and the combination of MSC+FD, MSC+SD, SNV+FD, SNV+SD, FD+SM, SD+SM. The following only gave the original spectra (Figure 4) and pre-treated spectra with MSC+SD (Figure 5). From Figure 4, two kinds of egg original spectra had the same general trend and small differences, while after pretreatment (Figure 5), the features of image peaks and valleys, lifting trend were more obvious. At the same time, the differences of fertilized and unfertilized eggs in some range of peaks and valleys were widened so that it was easier to distinguish them. Further it could be seen that the original spectral values range in each band was larger, but the spectral values of the samples after the pre-processing were concentrated in a small interval, which made the feature more obvious and was beneficial to classification. After analysis and comparison, it could be seen that the differences of spectra between the fertilized and unfertilized eggs were mainly concentrated within the wavelength range of 400 nm~500 nm, 700 nm~900 nm.

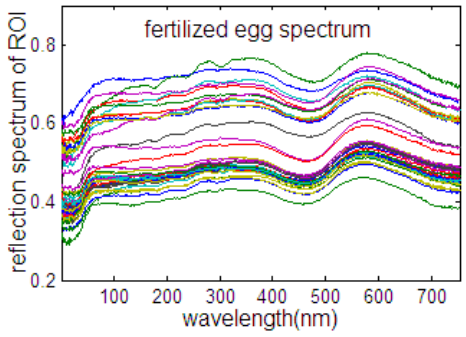


FIGURE 4a Spectrum of Fer.

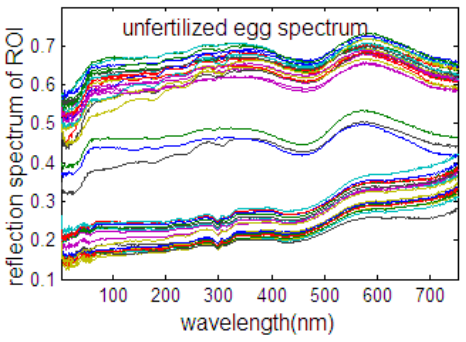


FIGURE 4b Spectrum of Unfer.

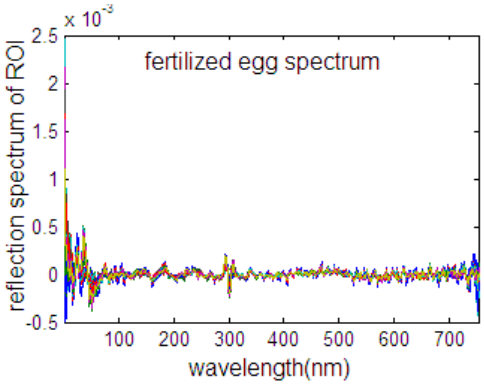


FIGURE 5a Fer. spectrum after MSC+SD

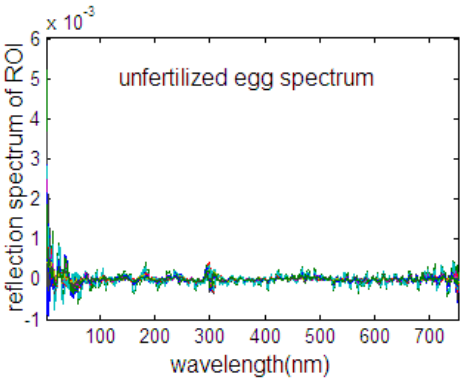


FIGURE 5b Unfer. spectrum after MSC+SD

Note: Fer. means fertilized egg; Unfer. means unfertilized egg

TABLE 1 Number of principal component by various spectroscopic data pre-processing

PM	FD	SD	SNV	MSC	MSC+FD	MSC+SD	SNV+FD	SNV+SD	SM+FD	SM+SD
NPC	4	7	3	2	4	7	4	7	4	6

In order to reduce the dimension of hyperspectral data, principal component analysis (PCA) was used to analysis above ten kinds of pretreatment of spectrum respectively. When the cumulative contribution rate of the eigenvalues of the covariance matrix exceeded 90%, the corresponding principal components were selected. The statistical results were shown in Table 1. From Table 1, it could be seen that the number of the variables were greatly reduced by using PCA so that it could improve of training speed and precision of the neural network model.

3.2 CLASSIFICATION OF BP NETWORK

BP neural network was worked as a classifier in this experiment. The 200 samples were randomly divided into a training set and a test set. First, the data that was pretreated in different ways should do principal component analysis. The principal component variables instead of the original variables were used as the input of the neural network, and then built the BP neural network for training. The training times was 2000; error performance objectives was 10^{-4}, and the learning rate was 0.01. The average of accuracy and coefficient of determination was used to evaluate the neural network in the number of this hidden layer. The programs had run 10 times and the results were shown in Table 2. As could be seen from Table 2, the model worked best after MSC + SD data preprocessing. So MSC + SD could be determined as the best pretreatment way. Corresponding network structure were one hidden layer and three hidden layer nodes. But average accuracy rate of the network prediction was only 83% and decision coefficient is only 0.57 after the optimal pretreatment. Also, classification accuracy of the network was not satisfactory. In the absence of initial weights threshold value, BP network were randomly assigned to each training weights threshold, causing instability in the models and turbulence in the training results. Therefore, it was necessary to further optimize the parameters of the network in order to improve the network performance.

TABLE 2 Model performance of various pre-processing pattern

PM	FD	SNV	MSC	SD	MSC+FD	MSC+SD	SNV+FD	SNV+SD	SM+FD	SM+SD
NPC	4	3	2	7	4	7	4	7	4	6
DC	0.53	0.53	0.43	0.41	0.44	0.57	0.51	0.56	0.56	0.44
PA	82%	86%	66%	74%	76%	83%	82%	83%	84%	74%

Note: PM means Pretreatment Methods; NPC means Number of Principal Component; DC means Determination Coefficient; PA means Prediction Accuracy

3.3 GA-BP NETWORK OPTIMIZATION

The neural network structure that previously constructed was established by the 7-3-1. The total number of weights was 24(7 * 3 + 3 * 1 = 24) and the total number of threshold was 4(3 + 1 = 4), so the encoding length of individual of genetic algorithm was 28 (24 + 4 = 28). We randomly chose from original training set of 120 samples, for each class, 80 eggs of original training set to build the training subset (40 fertile eggs and 40 infertile eggs) and the remaining all was put aside for testing subset (20 fertile eggs and 20 infertile eggs). The absolute value of the training data prediction error was regard as the individual fitness value, the smaller the fitness value, the more preferably individuals.

The genetic algorithm parameters were set as follows: population size was 10, evolution was 100 times; crossover probability was 0.4; the variation coefficient was 0.2. In the process of the genetic algorithm to

optimize, the best individual fitness value changed as shown in Figure 6.

The optimal weight and threshold value of BP neural network were obtained by the genetic algorithm, which were shown in Table 3.

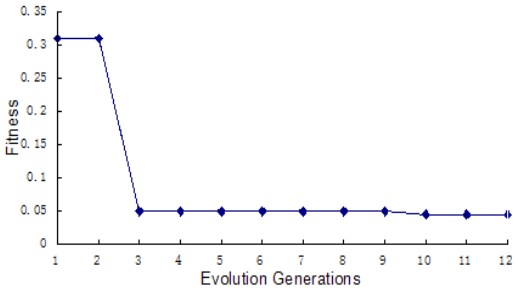


FIGURE 6 The best single adaptive by GA

TABLE 3 The best weights and threshold of BP net optimized by GA

Weights between input layer and hidden layer			Hidden layer node threshold	Weights between hidden layer and output layer	Output layer node threshold
-2.2145	-0.9389	0.9784	2.7922	-2.0859	-2.8262
0.0972	1.78	0.4596	-1.0909	-0.6995	
-2.5988	-1.5093	2.4521	2.7336	-2.9215	
2.3371	2.9388	-1.0581			
2.9245	0.6832	0.9125			
0.6081	1.7868	0.6627			
-0.7366	-1.9521	-2.6405			

3.4 TESTING OF OPTIMIZED BP NETWORK

The optimized BP network was used to test the 80 samples of original test set. Ten replicates were performed. Table 4 showed that using the BP network model predicted 10 times of classification accuracy and determination coefficient. It could be seen that unfertilized egg accuracy was 93%; fertilized eggs accuracy was 94%; among the 10 replicates testing, the extreme maximum accuracy of fertilized and unfertilized egg was 100% and 97.5% respectively; the extreme minimum accuracy of fertilized and unfertilized egg was 90% and 87.5% respectively; the overall accuracy of 93.5%, and the determination coefficient was 0.95. The result showed the network had high prediction precision and good generalization ability.

TABLE 4 The prediction results by 10 times tests

Egg type	Extreme maximum accuracy	Extreme minimum accuracy	Average accuracy	Total accuracy	Determination Coefficient
Fertilized egg	100%	90%	94%	93.5%	0.95
Unfertilized egg	97.5%	87.5%	93%		

4 Conclusions

In this research, an attempt has been made to preprocess spectral data, reduce data dimension and classify the unfertilized eggs using GA-BP network. Classification accuracy was 93.5%, which was encouraging. The results showed that MSC+SD pretreatment method was the most suitable for the model, PCA reducing data dimension was effective and hyperspectral technology combined with GA-BP network to identify unfertilized eggs and fertilized eggs was feasible.

Acknowledgments

This work was supported by Special Fund for Agro-scientific Research in the Public Interest(Project Code No. 201303084), the Fundamental Research Funds for the Central Universities of Huazhong Agriculture University (Grant No. 2662013JC005), Natural Science Foundation of Hubei Province of China (Grant No.2011CDB137) and the National Natural Science Foundation of China (Grant No. 51105160). The authors would like to thank the reviewers of this paper for their constructive comments and advice, which greatly helped us to improve the quality of the paper.

References

[1] Das K, Evans MD 1992 Detecting fertility of hatching eggs using machine vision I: Histogram characterization method *Trans. ASABE* **35**(4) 1335–41

[2] Wang Q H, Ma M H, Zhu Z H, Zhu T, Li M 2012 Non-destructive Detection of Hatching Egg's Survival Based on Machine Vision *Journal of Food, Agriculture & Environment* **10**(1) 578-81

[3] Coucke P M, Room G M, Decuypere E M, de Baerdemaeker JG. 1997 Monitoring embryo development in chicken eggs using acoustic resonance analysis *Biotechnology Progress* **13**(4) 474-8

[4] Bamelis F R, Tona K, De Baerdemaeker J G, Decuypere E M 2002 Detection of early embryonic developmentin chicken eggs using visible light transmission *British Poultry Science* **43**(2) 204-12

[5] Klein S, Rokitta M, Baulain U, Thielebein J, Haase A, Ellendorff F 2002 Localization of the fertilized germinal disc in the chicken egg before incubation *Poultry Science* **81**(4) 529-36

[6] Smith D P, Lawrence K C, Heitschmidt G W 2008 Fertility and embryo development of broiler hatching eggs evaluated with a hyperspectral imaging and predictive modeling system *International Journal of Poultry Science* **7**(10) 1001-4

[7] Zhang W, Pan L Q, Tu K, Zhang Q, Liu M 2014 Comparison of spectral and image morphological analysis for egg early hatching property detection based on hyperspectral imaging *PLOS ONE* **9**(2) 1-10
[8] Zhu Z H, Wang Q H, Wang S C, Dai M Y, Ma M H 2012 The detection of hatching eggs prior to incubation by the near infrared spectrum *Spectroscopy and Spectral Analysis* **32**(4) 962-5
[9] Lu R F 2007 Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images *Sensing and Instrumentation for Food Quality and Safety* **1**(1) 19-27
[10] Naganathan G K, Grimes L M, Subbiah J, Calkins C R, Samal A, Meyer G E 2008 Partial least squares analysis of near-infrared hyperspectral images for beef tenderness prediction *Sensing and Instrumentation for Food Quality and Safety* **2**(3) 178-88
[11] Zhao J W, Wang K L, OuY Q, Chen Q S 2011 Measurement of chlorophyll conent and distribution in tea plant's leaf using hyperspectral imaging technique *Spectroscopy and Spectral Analysis* **31**(2) 512-5
[12] Pan H 2005 Application of BP neural network based on genetic algorithm *Computer application* **25**(12) 33-45

Authors

Zhihui Zhu, born in April, 1975, Hubei, P.R. China

Current position, grades: associate professor.
University studies: PhD degree in electrical engineering and automation, Wuhan University in 2008.
Scientific interest: nondestructive testing of agricultural products.
Publications: 20 papers.
Experience: teaching experience of 6 years, 3 scientific research projects.

Ting Liu, born in August, 1987, Hubei, P.R. China

Current position, grades: postgraduate student.
University studies: Huazhong Agricultural University.
Scientific interest: nondestructive testing of agricultural products.
Publications: 2 papers.
Experience: teaching experience of 1 year, 3 scientific research projects.

Lirong Xiong, born in January, 1977, Hubei, P.R. China

Current position, grades: associate professor.
University studies: PhD degree in agricultural engineering, Huazhong Agricultural University in 2010.
Scientific interest: nondestructive testing of agricultural products.
Publications: 20 papers.
Experience: teaching experience of 4 years, 3 scientific research projects.

Meihu Ma, born in June, 1957, Hubei, P.R. China

Current position, grades: professor, vice president of College of Food Science and Technology, Huazhong Agricultural University.
University studies: PhD degree in food engineering, Hunan agricultural University in 2005.
Scientific interest: food science.
Publications: 100 papers.
Experience: teaching experience of 10 years, 3 scientific research projects.

Mining multiple level association rules under weighted concise support framework

Haiyan Zhuang*, Gang Wang

Department of public security technology, Railway Police College, Zhengzhou 450053, Henan Province, China

Received 30 July 2014, www.cmnt.lv

Abstract

Association rules tell us interesting relationships between different items in transaction database. Traditional association rule has two disadvantages. Firstly, it assumes every two items have same significance in database, which is unreasonable in many real applications and usually leads to incorrect results. Secondly, traditional association rule representation contains too much redundancy which makes it difficult to be mined and used. This paper addresses the problem of mining weighted concise association rules based on closed itemsets under weighted support-significant framework, in which each item with different significance is assigned different weight. Through exploiting specific technique, the proposed algorithm can mine all weighted concise association rules while duplicate weighted itemset search space is pruned. As illustrated in experiments, the proposed method leads to better results and achieves better performance.

Keywords: weighted concise association rule, transaction database, closed itemset, support-significance

1 Introduction

Extensive studies have been devoted into association rules mining in data mining area [1-4]. Association rule is an important knowledge representation and it tells us significant relations among itemsets present in large number of transactions [5-8]. Association rules generation has two steps [9-12]: one is to mine all the itemsets, and the other is to enumerate association rules. Reference [13] focused on the need for generating hierarchical minimal rules that provide maximal information and an algorithm has been proposed to derive minimal multilevel association rules and cross-level association rules that has made significant contributions in mining the minimal cross-level association rules, which express the mixed relationship between the generalized and specialized view of the transaction itemsets. In Reference [14], in order to solve the difficult concept association mining of query keywords in the semantic search, it raises a method of extracting concept association based on hierarchical clustering and association rules. Firstly, quick updating algorithm of association rules is adopted to extract non-categorization relationships of key words in the paper. Meanwhile, improved method of hierarchical clustering is used for the extraction of categorization relation while how to find father node for leaf nodes of the hierarchical clustering tree is provided as well. Reference [19] proposes a new algorithm for fast frequent itemset mining, which scan the transaction database only once which all the frequent itemsets can be efficiently extracted in a single database pass. To attempt this objective, it define a new compact data structure, called ST-Tree (Signature Transaction Tree), and a new mining algorithm ST-Mine to extract frequent itemsets.

But above association rule mining methods has two disadvantages as follow [13-16]. First, it assumes every two items have same significance in transaction database. Obviously it is unreasonable in real applications, which usually leads to incorrect results biased with users' expectation in real applications. Second, traditional representation contains too much redundancy which often makes the mining process and results are flooded in the combinatorial explosion of insignificant relationships.

In fact, one item may be different greatly from the other for the semantics in different real context. It is impossible to express all the differences absolutely between them. Attaching different weights to items with different significances is a good alternative in many circumstances. In real circumstance, general association rule must be re-defined to derive the result selectively referring to the weights on items and make more reasonable results [17-20].

In this paper, association rule under weighted framework and its mining methods are respected. Weighted concise association rule is re-defined through exploiting weighted support-significant framework. We find that it is possible to define weighted concise association rule based on closed itemset when the "downward closure property" holds. We can mine the whole set of weighted closed itemsets. The first step of association rule generation can be completed. Through exploiting the depth-first mining strategy and exactly itemset checking approach, duplicate closed itemsets can be identified early. In the second step, weighted concise association rules are enumerated. Weighted concise

* ***Corresponding author*** e-mail: zhuanghiyan@rpc.edu.cn

itemsets are arranged according to the item sequences based on tree structure. We gear the rule generating sequence to concise itemsets generating sequence, so that they can be processed concurrently. Additionally much space can be saved. Experimental studies are conducted to show that the method in this paper is reasonable and achieves good results.

2 Basic theory

Let $I=\{i_1, i_2, \ldots, i_m\}$ be a set of items and $TDB=\{T_1, T_2, \ldots, T_n\}$ be a transaction database, where $T_i(1\leq i\leq n)$ is a transaction which contains a set of items in I. A set of items is called an itemset or a pattern. A transaction T is said to contain itemset X if and only if $X\subseteq T$. The number of transactions in TDB that contain X is called the support count of X, denoted as count(X), and the support of X is denoted as support(X), which equals to count(X)/|TDB|, where |TDB| is the total number of transactions in TDB [21-24].

Given the form $X\Rightarrow Y$, the confidence of X is denoted as confidence($X\Rightarrow Y$), which equals to support ($X\cup Y$)/support(X) [21].

Definition 1. Weighted item. Given a transaction database TDB, and the item sets $I=\{i_1, i_2, \ldots, i_n\}$ which appear in TDB. We attach a value w_m to each i_m representing its significance. Such an item is called **weighted item**. And its weight is denoted weight(i_m) or weight($\{i_m\}$), which equals to w_m.

The items' weights may be defined respectively in different areas to balance the significance between items in transaction database [22].

Definition 2. Weighted itemset. Given a itemset X, for each item in it has a weight, itemset X may take different significance beyond the other. We denote the significance of X weight(X), which varies according to the itemset X.

Given an itemset X, from above definition, we define the weighted support of X as ws(X) [23].

Definition 3. Weighted closed itemset. Given a transaction database TDB and a minimum support threshold ξ_1, an itemset X is a frequent weighted closed itemset if both of the following conditions are true [24]:

1) $ws(X)\geq\xi_1$;

2) $\forall Y\supset X$ such that $ws(Y)<ws(X)$.

The problem of weighted closed itemsets mining is to find the complete set of frequent weighted closed itemsets in a given transaction database with respect to the given support threshold ξ.

Definition 4. Weighted concise associateion rule based on closed itemset. Given a transaction database TDB and a minimum support threshold ξ_1 and a minimum confidence threshold ξ_2, and rule $X\Rightarrow Y$ is a weighted concise association rule based on closed itemset if both of the following conditions are true [20]:

1) X, Y($Y\supset X$) are frequent weighted closed itemsets;

2) $ws(Y)/ws(X)\geq\xi_2$.

The problem of **weighted concise association rule mining** is to find the complete set of weighted concise association rule based on closed itemsets in a given transaction database with respect to the given support threshold ξ_1 and confidence threshold ξ_2.

In order to improve the efficiency of rule mining algorithm, we often employ the following two basic properties [9]:

1). Any loophole of frequent itemsets collection is also frequent itemsets.

2). Any superset of the frequent itemsets is also frequent itemsets.

The above two properties are basic pruning strategy of Apriori algorithm, called them Apriori properties.

From the above two properties of the Apriori known, if an itemset contains the infrequent k - itemsets, so the itemset must be infrequent itemsets. The process of mining in the subsequent can delete them. From this perspective, by means of connecting all frequent k-itemsets in the process of mining, we can obtain all candidate (k + 1) - frequent itemsets. This can reduce the size of the candidate itemsets. Apriori algorithm scans affairs library for the first time, calculates the support degree of each item to obtain the set of frequent 1 - itemsets which compliance with minimum support threshold. In after the first k times scanning library, it firstly obtained a new potential a set of candidate frequent k – itemsets based on the collection of frequent k-1-itemsets. Then calculates the support degree of each item, finally selected from a candidate set of frequent k – itemsets, which is in line with the minimum support threshold, and will serve as a collection of transactions on the basis of the next scan, and cycle to repeat the process until the termination conditions to achieve algorithm.

In above process, the procedure delete (L_{k-1}), the procedure candidate_gen (L'_{k-1}), the procedure has_infrequent_subset (c, L'_{k-1}) and the procedure join (l_1,l_2) were adopted [11, 12]:

Procedure Delete (L_{k-1})

//delete the item-sets whose item number is less than k-1 from L_{k-1} ;
Begin
Given $\forall$item Initial value i.count=0
{
For all l in L_{k-1} item-set part
If $i\in l$
Then i.count++// calculates the emerging times of item i in l;
}
If (i.count<k-1) then
{
For all l in L_{k-1} item-set part
If $i\in l$
Then delete the frequent item-sets of l from L_{k-1} ;
}
Return L_{k-1} .
End

Procedure Candidate_gen(L'_{k-1})

//generate k-item candidate set
Begin
For ∀item-set $l_1 \in L'_{k-1}$
For ∀item-set $l_2 \in L'_{k-1}$
If the first k-2 items of l_1 *and* l_2*are the same while those of the first k-1 item-sets are different*
Then c=jion(l_1*,*l_2*)// connect them to generate candidate item-set set;*
If has_infrequent_subset(c, L_{k-1} *)*
Then delete c;// prune to delete the non-frequent candidate item-sets
Else put c into C_K *;*
return C_K *;*
End

Procedure jion(l_1,l_2)

Begin
Command that:
The item-set part of l_1 *and* l_2 *are respectively are* l'_1、l'_2 *;*
The indicating transition sets are l_1^1、l_2^2 *;*
$l_c = l_1' \infty l_2'$ *is the item-set part after connecting c,* $l_c^c = l_1^1 \cap l_2^2$ *is the transition set part of c;*
Return c;
End

Procedure has_infrequent_subset(c, L'_{k-1} **)**

//judge whether the subsets of L'_{k-1} *are non-frequent sets;*
Begin
for each (k-1)-subset s of c item-set part
if s $\notin L'_{k-1}$ *item-set part*
Then return true;
Return false;
End.

3 Weighted concise association rules based on closed itemsets

In this section, we explore properties of weighted concise association rules based on closed itemsets. First, weighted closed itemsets should be enumerated. In definition 3, frequent weighted closed itemset is defined. But it is difficult to decide how to calculate the weights of itemsets reasonably to mine the complete set of weighted closed itemset and rules properly. Different definitions of itemset's weight may lead to different results, but properties of traditional closed itemset may not hold. The total itemset search space will be viewed, which is impossible. And we must ensure that all weighted association rules generated can be derived from concise ones, and duplicate itemsets and search space can be pruned efficiently while mining weighted concise association rules [21].

Traditional itemsets will come less frequently when being added to items while being enumerated. Then the enumerating process can be stopped when infrequent itemsets are found. If the "Anti-monotone" property of itemset does not hold on weighted itemset, we may not know when the enumerating process will be stopped. Otherwise, weighted concise association rules can be generated in depth–first like manner properly. A weighted support-significant framework is adopted to keep "downward closure property". In the mining process, the method of calculating weighted support of an itemset is adopted. We prove that weighted closed itemset keeps the anti-monotone properties. All weighted concise association rules can be mined and all weighted association rules can be derived from concise ones.

Definition 5. Weighted support of itemset. Given a transaction database TDB and each transaction T_i in it is attached a weight tw_i. Then the weighted support of itemset X is defined as follow:

$$ws(X) = \frac{\sum_{T_i \in TDB \cap X \subseteq T_i} tw_i}{\sum_{T_i \in TDB_i} tw_i}. \quad (1)$$

For any $T_i \in$TDB, we can define ws(T_i) reasonably. Generally weighted support of itemset X is defined as follow:

$$ws(T_i) = \frac{\sum_{I_j \in T_i} weight(I_j)}{|T_i|}. \quad (2)$$

Lemma 1. (downward closure property). For an itemset X, $\exists Y \supset X$, ws(Y)≤ws(X).

Lemma 2. For an itemset X, $\exists Y \supset X$, ws(Y)=ws(X), then X can't be a weighted closed itemset.

Proof: From definition 4, we can have the lemma 2.

Lemma 3. For two itemsets X, Y($Y \supset X$), ws(Y)=ws(X), then ∀ Z ($Y \supset Z \supset X$), Z isn't a weighted closed itemset.

Proof: Let TS(X)={T_i|$T_i \in$TDB, $T_i \supset X$} be the set of transactions containing itemset X, then we can have that:

$$ws(X) \sum_{T_i \in TDB_i} tw_i = \sum_{T_i \in TS(X)} tw_i. \quad (3)$$

For $Y \supset Z \supset X$, we have that $TS(X) \supseteq TS(Z) \supseteq TS(Y)$, then

$$\sum_{T_i \in TS(X)} tw_i \geq \sum_{T_i \in TS(Z)} tw_i \geq \sum_{T_i \in TS(Y)} tw_i. \quad (4)$$

From Equation (3), we can have:

$$ws(X) \sum_{T_i \in TDB_i} tw_i \geq ws(Z) \sum_{T_i \in TDB_i} tw_i \geq ws(Y) \sum_{T_i \in TDB_i} tw_i. \quad (5)$$

Then ws(X)≥ws(Z)≥ws(Y), and ws(Y)=ws(X), so ws(Y)=ws(Z) =ws(X), from definition 4, we have the lemma 3.

Lemma 4. For itemsets X, Y($Y \supset X$), ws(Y)=ws(X), then ∀ Z ($Z \supset X$, $Y \not\subset Z$), Z can't be a closed itemset.

Proof: For $Y \supset X$, $TS(X) \supseteq TS(Y)$, and ws(Y)=ws(X) in lemma 4, then TS(X)=TS(Y).From the property of set, we

can have TS(X∪Z)=TS(Y∪Z).Additionally, Z⊃X, Y⊄Z, we have

TS(Z)=TS(Y∪Z) and Y∪Z⊃Z,From the inference in lemma 3, we have that ws(Y∪Z)= ws(Z),Then we have the lemma 4.

From above lemmas, we can explore the weighted closed itemsets search space in specific manner, and can prune much search space early.

Theorem 1. All weighted association rules can be derived from weighted concise association rule based on closed itemset.

Rationale: Weighted support of any frequent itemset can be calculated by weighted closed itemset.

Proof: For any itemset X, there must be a weighted closed itemset Y⊃X, and ws(Y)=ws(X), otherwise from lemma 1, itemset X must be a weighted closed itemset. Thus we have weighted support of itemset X from itemset Y or X itself.

Then for any weighted association rule X⇒Y, there must be two itemsets V(V⊃X) and W(W⊃Y), that ws(V)= ws(X) and ws(Y)= ws(W). We can have support and confidence rule X⇒Y from rule V⇒W. Then we have theorem 1.

4 Enumerate weighted concise association rules

In this section, we try to enumerate weighted concise association rules while weighted closed itemsets are being mined. Several techniques are adopted to realize the process.

4.1 WEIGHTED CLOSED ITEMSET TREE

In this section, we arrange weighted closed itemsets in tree to decrease space occupation in weighted concise association rules generation. The tree is called weighted closed itemset tree (WCIT), which is like FP-tree. But items in the link from one node with weighted support to root form a weighted closed set while node without weighted support doesn't (Figure 1).

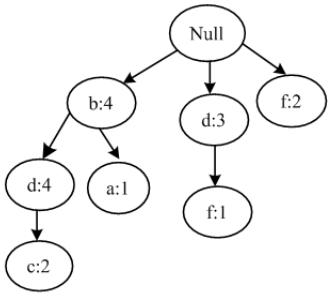


FIGURE 1 Weighted closed itemset tree

Then we have following property:

Property 1. One node without weighted support must have descendant nodes.

In the weighted concise association rules generation, we can index weighted closed itemsets or just candidate ones. If the tree is too large to be stored in memory, we can serialize it in external storage partially by branches.

4.2 MINING WEIGHTED CLOSED ITEMSET

In order to gear to mine the weighted concise association rules concurrently, we enumerate weighted closed itemsets in specific order.

We assume that there is a partial order on itemset I, denoted as $\prec$. Without loss of generality, we assume in the remainder of the paper $\prec$ is the ascending supports order of items in I. During the mining process, weighted closed itemsets will be enumerated according to partial order on itemset I. And duplicate search space will be pruned at the same time.

In this section, we exploit the conditional weighted closed set [21] to facilitate the mining process, which is denoted as CWCS($i_1,i_2,\ldots,i_k$), while following conditions are satisfied:

1) $i_k \prec \ldots \prec i_2 \prec i_1$;

2) $\forall X \in CWCS(i_1,i_2,\ldots,i_k)$ such that $\exists T \subset TDB$, $X \cup \{i_1,i_2,\ldots,i_k\} \subset T$;

3) Let i_m be the largest item in X, then $i_m \prec i_k$;

4) CWCS() is the conditional weighted closed set of ∅.

First, transactions' weights in database are calculated. Then they are projected into the conditional weighted closed set as the pattern grows.

Second, we initialize CWCS(). We eliminate all the items whose weighted support is less than ξ in all transactions and insert itemsets left into CWCS(). If there is a same itemset, then eliminate it and update weight of the other, otherwise initialize its weight. Then for every frequent item i_m in a CWCS($i_1,i_2,\ldots,i_k$), we do as follows: For every itemset t containing i_m in CWCS($i_1,i_2,\ldots,i_k$), we collect all the frequent item i_n in t where $i_n \prec i_m$, as a new itemset to be inserted into CWCS($i_1,i_2,\ldots,i_k,i_m$). Then we set CWCS ($i_1,i_2,\ldots,i_k$) to ∅. With the mining process recursively going on, all the conditional weighted closed sets are initialized.

But the weighted closed sets can be generated in different orders. We must choose one to facilitate weighted concise association rules generation. Generally the itemsets can be enumerated in depth-first order as following (Figure 2).

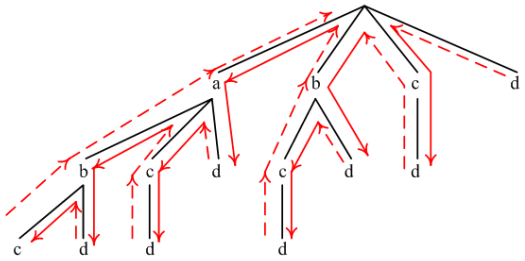


FIGURE 2 Itemset search space with four items

Itemset search space is traversed as following:

-{a}

-{b},{b,a}

-{c},{c,a},{c,b},{c,b,a}

We call it in right-depth-first order, in which latter items are always respected first, just as following:

-{d}

-{c},{c,d}

-{b},{b,d},{b,d},{b,c,d}

In right-depth-first mining process, the items nearer to root are always dealt with first. For an item i_n in CWCS $(i_1,i_2,\ldots,i_k)$, we assume all CWCS$(i_1,i_2,\ldots,i_k,i_m)$ $(i_n \prec i_m \prec i_k)$ have been dealt with, then for every i_m $(i_n \prec i_m \prec i_k)$, we do Weighted Checking (CWCS$(i_1,i_2,\ldots,i_k,i_n)$, CWCS$(i_1,i_2,\ldots,i_k,i_m)$).

After all items in CWCS$(i_1,i_2,\ldots,i_k)$ have been dealt with, we do as follows: for every item i_e in CWCS$(i_1,i_2,\ldots,i_k,i_n)$, i_e is inserted into every element of CWCS$(i_1,i_2,\ldots,i_k,i_n)$, then elements of CWCS$(i_1,i_2,\ldots,i_k,i_n)$ are removed to CWCS$(i_1,i_2,\ldots,i_k)$.

4.3 ENUMERATING WEIGHTED ASSOCIATION

In this section, we present the general mining method, which can generate weighted concise association rules while weighted closed itemsets are being enumerated as in Figure 3.

Through exploiting Weighted Checking technique and right-depth-first mining order, the proposed algorithm prunes itemset search space efficiently and integrates weighted closed itemsets and rules enumerating process properly [22,23].

For the given database and minimum support threshold ξ, all the weighted concise association rules based on closed itemsets can be get by calling procedure WeightedConciseRulesEnumerate () as following:

Procedure *WeightedConciseRulesEnumerate*(α)

(Here α represents $i_1,i_2,\ldots,i_k$ $(i_k \prec \ldots \prec i_2 \prec i_1)$)

Begin
Recalculate all weighted support(i) with the new occurrence frequencies in CWCS(α);
If (all support(i)< ξ)
{
Set CWCS(α) ={∅};
Return;
}
for item i=max to min contained in CWCS(α) do
{
if(support(i) ≥ξ)
{
project itemset in CWCS(α) to CWCS(α, i);
WeightedConciseRulesEnumerate (α, i);
for j=i+1 to max WeightedChecking(CWCS(α,i), CWCS(α, j));
}
}
if(there is an item i in CWCS(α) and support(α)= support(i))
Set CWCS(α)=RS(α)= ∅;
Else set CWCS(α) =∅;
RS(α)= {∅⇒{i}};
for(item i=max to min contained in CWCS(α))
{
for(every itemset is in CWCS(α,i))
Set CWCS(α)=CWCS(α)∪{is∪{i}} ;
Serialize Y={α,}∪is∪{i} into WCIT ;
For (every subset s⊂sY)
Generate rule s⇒sY with confidence ws(Y)/ws(X);
}
End.

4.4 WEIGHTED-SUFFIX-PROJECTION: MINING WEIGHTED CLOSED ITEMSETS DIRECTLY

When an item i_n is concerned, we can tell whether current itemset is closed by projecting larger items than i_n in its suffix link through adjusting its descendant linked branches. After all the descendant linked branches are combined, only the items with the same weight of item i_n are linked to its suffix link.

In the right-depth-first mining process, the larger items are always considered first. For items $i_3 \prec i_2 \prec i_1$, if ws(i_3) in CWCS(i_1) is equal to that in CWCS(i_1, i_2), item i_3 need not to be considered in CWCS(i_1). And item i_3 needn't to be projected on larger items in suffix links.

From above two main optimizations, we can just tell whether a weighted itemset I is closed through respecting the suffix of the minimum item in I. Additional calculation and space are not needed.

TID	items	*ws*
100	c,e	1.1
200	a,b,c,e	0.8
300	b,c,d	0.9
400	b,c	0.8
500	a,b	0.5
600	b,f	0.95
700	b,e	0.9
800	a,c,e	0.8667

item	*weight*	*support*	*ws*
a	0.4	0.375	0.3178
b	0.6	0.75	0.7115
c	1	0.625	0.6553
d	1.1	0.125	0.1320
e	1.2	0.5	0.5379
f	1.3	0.125	0.1395

FIGURE 3 Transaction database *TDB* and weight settings

Example 1. Given the transaction database TDB with the item weight settings, the minimum weighted support threshold ξ_1 =0.2 and confidence threshold ξ_2 =0.3, the created global FP-tree like structure is shown in Figure 4. Item e is removed for its support is below ξ. The ascending support order on I is $b \prec c \prec e \prec a \prec d \prec f$.

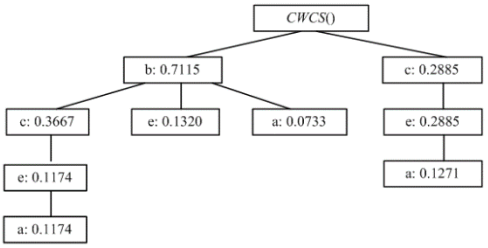


FIGURE 4 Data structure created for *CWCS*()

We can get following weighted closed itemsets as Figure 5:

-{a}:0.3178
-{b}:0.7115
-{c}:0.6533
-{e}:0.5379
-{b,c}:0.3667
-{b,e}:0.2494
-{c,e}:0.4059
-{c,e,a}:0.2445

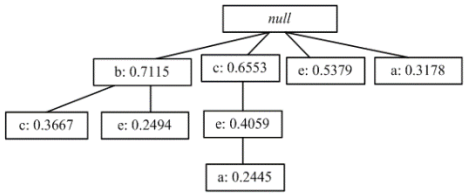


FIGURE 5 WCIT for generated concise itemsets

And we get weighted concise association rules based on weighted closed itemsets as following:

{b}⇒{b,c}:0.5154,
{c}⇒{b,c}:0.5154
{b}⇒{b,e}:0.3505
{e}⇒{b,e}:0.4637
{c}⇒{c,e}:0.6194
{e}⇒{c,e}:0.4637
{c}⇒{c,e,a}:0.3731
{e}⇒{c,e,a}:0.4545
{a}⇒{c,e,a}:0.7694
{c,e}⇒{c,e,a}:0.6024
{c,a}⇒{c,e,a}:1
{e,a}⇒{c,e,a}:1

We can see that the count of weighted concise rules is larger than that of weighted closed itemsets. After attaching weights to itemsets, the mining result is more reasonable in real applications than traditional ones. Additionally the number of useless concise itemsets and rules generated decreases greatly. And all general association rules can be derived from weighted concise ones.

5 Experimental results

In this section, we conduct experiments on public simulated data sets. The target platform is a Lenovo PC equipped with 2.6G clock rate CPU and 1024M main memory. The operation system is Windows XP Professional. The proposed algorithm is implemented in C++. The experimental data subsets include mushroom which contains all kinds of mushroom information with different properties which decide whether this kind of mushroom is poisonous, chess, pumsb, kosarak, T10I4D100K, Segmentation, etc. which can be downloaded from UCI Repository or Frequent Itemset Mining Implementation Repository [22].

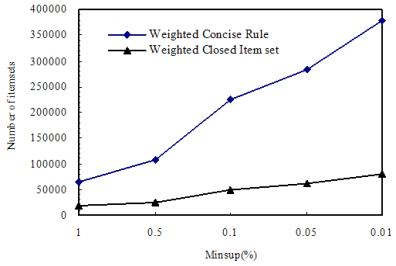


FIGURE 6 Size of rules and closed itemsets generated comparison

Here we carried out general association rule mining method and the proposed weighted concise association rule mining method in this paper under specific weights setting, selected size of rules and closed itemsets and runtime as assessment criteria. When attaching high weight to specific properties, the corresponding properties are more likely to be mined. Following Figure 6 and 7 show the experiment results.

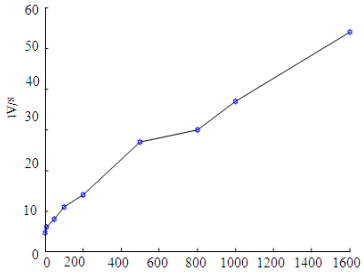


FIGURE 7 The saving time trend of the proposed weighted concise association rule mining

From above results, we can see that weighted concise association rules generation consumes no more time than just weighted closed itemsets generation only. But it can derive more reasonable result than traditional ones.

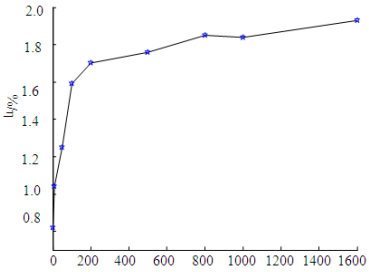


FIGURE 8 The saving time rate trend of the proposed weighted concise association rule mining method

From Figure 7 and 8,we can see that: when the data set size is bigger, the proposed weighted concise association rule mining method will save more time in constructing decision tree by contrast with the traditional one, thus the more efficient it is. Therefore, we can conclude that: in mining weighted concise association rules, the bigger the data set are, the more advantageous the proposed weighted concise association rule mining method is.

6 Conclusions

In this paper, we proposed a method for mining weighted concise association rules effectively. Based on specific search space exploring order, we gear rules generating process to itemsets exploring process properly, and derive all rules while enumerating closed itemsets concurrently. It can derive more reasonable results than general ones in real circumstance, but with less time consuming. The test results show that the algorithm has good time scalability.

Acknowledgments

This work was supported in part by the Foundation Projects of Analysis system of railway public security information based on cloud computing (No. 201202ZDYJ017).

References

[1] Aoudjit R, Belkadi M, Daoui M 2013 Mobility Prediction Based on Data mining *International Journal of Database Theory and Application* **6**(2) 71-8
[2] Sheikhan M, Rad M S 2013 Gravitational search algorithm-optimized neural misuse detector with selected features by fuzzy grids-based association rules mining *Neural Computing and Applications* **123**(8) 2451-63
[3] Tahrima H, Chowdhury Farhan A C, Samiullah Md, Akther S, Jeong B S, Jeon S 2012 An efficient approach for mining cross-level closed itemsets and minimal association rules using closed itemset lattices *Expert Systems with Applications* **40**(22) 7334-42
[4] Wei C 2013 Concept association mining based on clustering and association rules *International Journal of Applied Mathematics and Statistics* **47**(17) 302-10
[5] Noman I, Zubair S, Rehman A, Siddiqui M 2013 HANDY: A hybrid association rules mining approach for network layer discovery of services for mobile ad hoc network *Wireless Networks* **19**(8) 1961-77
[6] Alpar P, Winkelsträter S 2013 Assessment of data quality in accounting data with association rules *Expert Systems with Applications* **40**(22) 7343-50
[7] Wang G, Song Q 2013 A novel feature subset selection algorithm based on association rule mining *Intelligent Data Analysis* **17**(5) 803-5
[8] Agrawal S, Pandey N K 2011 A Near Optimal Approach for Top-K Frequent Itemset Mining *International Journal of Database Theory and Application* **4**(4) 43-56
[9] El Hadi Benelhadj M, Arour K, Boufaïda M, Slimani Y 2011 A New Compact Structure to Extract Frequent Itemsets *International Journal of Database Theory and Application* **4(**4) 25-42
[10] Salleb-Aouissi A, Vrain C, Nortet C, Kong X, Rathod V, Cassard D 2013 Quantminer for mining quantitative association rules *Journal of Machine Learning Research* **4**(7) 3153-57
[11] Hsiao H-W, Sun H-M, Fan W-C 2013 Detecting stepping-stone intrusion using association rule mining *Security and Communication Networks* **6**(10) 1225-35
[12] Li M, Shen C, Zhu X, Chen X 2011 Urban and rural integrated cadastral association alternation rule and model *Wuhan Daxue Xuebao (Xinxi Kexue Ban)* **38**(10) 1253-56 *(in Chinese)*
[13] Zian B, Benmlouka A, Ouinte Y 2013 Improving index selection accuracy for star join queries processing *Advances in Intelligent Systems and Computing* **220** 67-74
[14] Sarath K N V D, Ravi V 2013 Association rule mining using binary particle swarm optimization *Engineering Applications of Artificial Intelligence* **26**(8) 1832-40
[15] Wang M, Wu S, Cai R 2013 Two novel interestingness measures for gene association rule mining *Neural Computing and Application* **23**(3) 835-41
[16] Paranjape-Voditel P, Deshpande U 2013 A stock market portfolio recommender system based on association rule mining *Applied Soft Computing Journal* **13**(2) 1055-63
[17] Cagliero L, Fiori A, 2012 Discovering generalized association rules from Twitter *Intelligent Data Analysis* **17**(4) 627-48
[18] Sowan B, Dahal K, Hossain M A, Zhang L, Spencer L 2013 Fuzzy association rule mining approaches for enhancing prediction performance *Expert Systems with Applications* **40(**17) 6928-37
[19] Rodríguez-González A Y, Martínez-Trinidad J F, Carrasco-Ochoa J A, Ruiz-Shulcloper J 2013 Mining frequent patterns and association rules using similarities *Expert Systems with Applications* **40**(17) 6823-36
[20] Nabwey H A, El-Paoumy M S 2013 An Integrated Methodology of Rough Set Theory and Grey System for Extracting Decision Rules *International Journal of Hybrid Information Technology* **6**(1) 57-66
[21] Huang Y-F, Wu C-M 2011 Preknowledge-based generalized association rules mining *Journal of Intelligent and Fuzzy Systems* **22**(1) 1-13
[22] Luna J M, Romero J R, Ventura S 2013 Grammar-based multi-objective algorithms for mining association rules *Data and Knowledge Engineering* **86**(7) 19-37

Authors

Haiyan Zhuang, born in May, 1976, Zhengzhou County, Henan Province, P.R. China

Current position, grades: associate professor at the department of public security technology, Railway Police College, China.
University studies: BSc in Application of Computer Technology at Zhengzhou Industrial University in China. MSc at Zhengzhou University in China.
Scientific interests: information retrieval, data mining.
Publications: more than 10 papers.
Experience: teaching experience of 8 years, 3 scientific projects.

Gang Wang, born in December, 1958, Zhengzhou County, Henan Province, P.R. China

Current position, grades: professor at the department of public security technology, Railway Police College, China.
University studies: BSc in Optical Physics at Changchun University of Science and Technology in China.
Scientific interests: information security.
Publications: more than 20 papers.
Experience: teaching experience of 20 years, 7 scientific projects.

Data oriented workflow using semantic technologies

Hui Bu*, Ran Liu

School of Software, North China University of Water Resources and Electric Power, China

Received 1 March 2014, www.cmnt.lv

Abstract

Scientific workflows are a topic of great interest in the Grid community that sees in the workflow model an attractive paradigm for programming distributed wide area Grid infrastructures. Scientific workflows have recently emerged as a new paradigm for scientists to formalize and structure complex and distributed scientific processes to enable and accelerate many scientific discoveries. In contrast to business workflows, which are typically control flow oriented, scientific workflows tend to be dataflow oriented, introducing a new set of requirements for system. In this paper, we consider a general workflow setting in which input data sets are processed by a graph of transformations to produce output results. Our goal is to perform efficient selective refresh of elements in the output data, i.e., compute the latest values of specific out-put elements when the input data may have changed. The data provenance is investigated to be used to enable efficient refresh. The proposed approach is based on capturing one level data provenance at each transformation when the workflow is run initially. Then at refresh time provenance is used to determine (transitively) which input elements are responsible for given output elements, and the workflow is rerun only on that portion of the data needed for refresh. The reported preliminary experimental results are developed on the overhead of provenance capture, and on the crossover point be-tween selective refresh and full workflow computation development.

Keywords: scientific workflows, scientific workflow management system, semantic technologies

1 Introduction

Scientific workflows have recently emerged as a new paradigm for scientists to integrate, structure, and orchestrate a wide range of local and remote heterogeneous services and software tools into complex scientific processes to enable and accelerate many scientific discoveries. Significant scientific advances are increasingly achieved through complex sets of computations and data analyses. These computations may com-prise thousands of steps, where each step might integrate diverse models and data sources that different groups develop. The applications and data might be also distributed in the execution environment. The assembly and management of such complex distributed computations present many challenges, and increasingly ambitious scientific inquiry is continuously pushing the limits of current technology.

Workflows have recently emerged as a paradigm for representing and managing complex distributed scientific computations, accelerating the pace of scientific progress. Scientific workflows orchestrate the dataflow across the individual data transformations and analysis steps, as well as the mechanisms to execute them in a distributed environment. Each step in a workflow specifies a process or computation to be executed according to the data flow and dependencies among them. The representation of these computational workflows contains many details required to carry out each analysis step developed within the context of an earthquake science application.

Workflow systems exploit these explicit representations of complex computational processes at various levels of abstraction to manage their life cycle and automate their execution. In addition to automation, workflows can provide the information necessary for scientific reproducibility, result derivation, and result sharing among collaborators. By providing automation and enabling reproducibility, they can accelerate and transform the scientific analysis process.

This traditional best effort execution model commonly does not fully consider the dynamic and course-grain nature of Grid environments dominated by a broad set of performance overheads such as large latencies (several seconds), unpredictable queuing times, sudden unavailability of existing resources, external load on shared-memory computers, unexpected jobs on shared local queues, and inaccurate predictions for new processor architectures, which can severely deteriorate the workflow execution that is likely to deviate from the expected schedules. In addition, although the Grid provides in theory an unlimited amount of compute power, currently existing Grid test beds offer a limited amount of high-performance resources that are important to be used efficiently.

Over the years, performance analysis tools have emerged as an important means for detecting bottlenecks in high performance computing applications for which optimizing compilers rarely deliver satisfactory results. Grid computing is sensitive to similar problems. Running workflows based on pure scheduling techniques without understanding what really happens during the execution

***Corresponding author** e-mail: buhuiemail@126.com

may easily lead to poor execution times and inefficient usage of computing resources. Performance analysis tools are there-fore an important asset for understanding the behaviour and reasons for performance losses required to improve runtime middleware environments and, ultimately, the workflow executions in dynamic Grid infrastructures. Such a tool for high-level performance analysis of scientific Grid work-flows is currently missing.

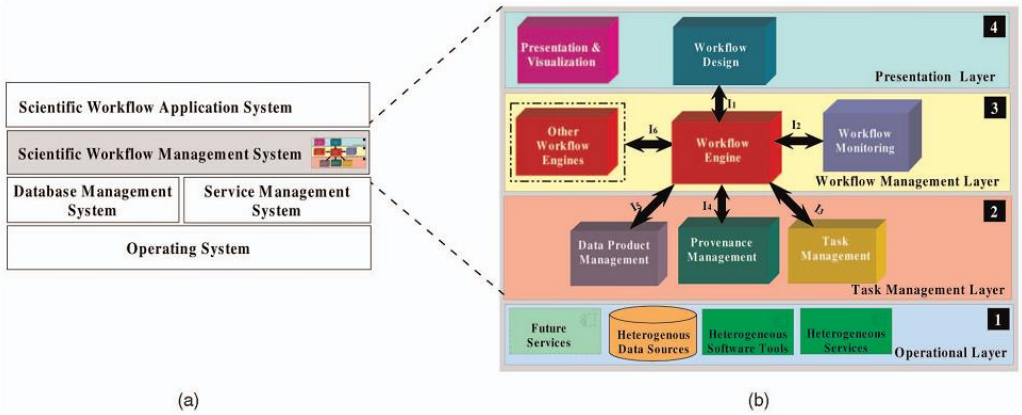


FIGURE 1 a) The position of an SWFMS within a software stack; b) zoom-in view of the reference architecture for SWFMSs

In this paper, we propose a systematic approach to building a tool for performance analysis and steering of scientific workflow applications in heterogeneous and distributed Grid environments. We introduce a theoretical reference parameter, called ideal execution time, that provides a realistic bound for the lowest (that is, "fastest") execution time achievable by a workflow in a certain Grid infrastructure. To address this issue, we propose the first reference architecture for SWFMSs based on a comprehensive survey of the literature and identification of key requirements. According to the proposed reference architecture, we further propose a service-oriented architecture for the VIEW system. Leveraging SOA, VIEW consists of six loosely coupled service components, each of which corresponds to a functional component that is identified in the reference architecture, whose functionality is exposed as a Web service. We implemented the VIEW system to validate the feasibility of the proposed architectures. We present a VIEW based scientific workflow application system (SWFAS), to demonstrate the capabilities of VIEW in support of user interaction intensive, visualization intensive, and compute intensive scientific workflows in a heterogeneous and distributed computing environment.

The rest of the paper is organized as follows. Section 2 identifies the related work of the proposed method. Section 3 proposes our reference architecture for SWFMSs. Section 4 we evaluate five representative SWFMSs using the proposed reference architecture. Finally, Section 5 concludes the paper and comments on future work.

2 Related works

Although the term "scientific workflows" were first coined by Vouk and Singh in 1996 [1] for workflow applications in scientific PSEs, only recently, there is an increasing momentum for the research and development of SWFMSs and their applications, due to the increasingly demanding requirements of many compute-intensive and data intensive scientific applications, enabled by the underlying advances of computing technologies, notably Services computing [2], Grid computing [3], and Cloud computing [4]. Scientific workflows leverage existing techniques developed for business workflows but deviate from them as a result of a different set of requirements raised from a wide range of science and engineering problems [5]. While business workflows are control flow oriented with the mission of carrying out business logic to achieve a business goal, scientific workflows tend to be dataflow oriented and aimed at enabling, facilitating, and speeding up the derivation of scientific results from raw data sets.

There has been a large body of work in lineage and provenance over the past two decades. Surveys are presented in, e.g., [6-8], and formal models for provenance are presented in, e.g., [9-12]. Provenance in the context of schema mappings is studied in [13-15]. None of these papers exploits provenance for selective refresh in a general workflow environment. There also has been a large body of work in incremental view maintenance: the efficient propagation of base data modifications, usually in a relational setting [16, 17]. Our work considers general workflows, rather than relational views. Also, in contrast to the view maintenance problem, selective refresh considers efficiently computing the up-to-date value of individual output elements, rather than keeping the entire view up to date by propagating changes made to the base data. Reference [18] provides a framework to explain "missing" answers in queries. There is some high-level

similarity between how explanations provided by their framework are created and how we support efficient refresh using data provenance. However, the details are quite different, and their framework supports SQL queries rather than general workflows. Reference [19] considers the problem of "update exchange" be-tween data peers linked by mappings. A problem they address is determining when a derived data element is no longer valid, but they do not provide a means to selectively refresh out-of-date values. Also, transformations in [19] are restricted to those that can be expressed in data log.

3 Incorporating semantic to data oriented workflow

Although the reference architecture proposed by WfMC, has been widely used for BWFMSs, this reference architecture does not satisfy key requirements R1-R5 for SWFMSs identified in the previous section. In this section we propose a reference architecture for SWFMSs. As shown in Figure 1b, the reference architecture consists of four logical layers, seven major functional subsystems, and six inter-faces. Figure 1a shows a typical software stack of a scientific workflow application: on top of an operating system, a data management system and a service management are used by an SWFMS for data management and service management, respectively. An SWFAS is developed over an SWFMS by the introduction of additional domain specific application data and functionalities.

3.1 DIFFERENT LAYERS OF THE PROPOSED METHOD

The first layer is the Operational Layer, which consists of a wide range of heterogeneous and distributed data sources, software tools, services, and their operational environments, including high-end computing environments. The separation of the Operational Layer from other layers isolates data sources, software tools, services, and their associated high-end computing environments from the scope of an SWFMS, thus satisfying requirement R5.

The second layer is called the Task Management Layer. Tasks are the building blocks of scientific workflows. Tasks consume input data products and produce output data products. At the same time, provenance is captured automatically to record the derivation history of a data product, including original data sources, intermediate data products, and the steps that are applied to produce the data product. This layer abstracts underlying heterogeneous data into data products, services, and software tools into tasks, and provides efficient management for data products, tasks, and provenance metadata. Therefore, the Task Management Layer satisfies requirements R2, R3, and R4. Moreover, the separation of the Task Management Layer from the Operational Layer promotes the extensibility of the Operational Layer with new services and new high-end computing facilities, and localizes system evolution due to hardware or software advances to the interface between the Operational Layer and the Task Management Layer. The task-level interoperability requirement (R7: level 1) should be addressed in this layer.

The third layer is the Workflow Management Layer, which is responsible for the execution and monitoring of scientific workflows. At this layer, the building blocks of a scientific workflow are the tasks provided by the underlying Task Management Layer. In this layer, an execution of a scientific workflow is called a workflow run, which consists of a coordinated execution of tasks, each of which is called a task run. Therefore, the Workflow Management Layer addresses requirements R6 and R7. The separation of the Workflow Management Layer from the Task Management Layer concerns two aspects as follows:

1) it isolates the choice of a workflow model from the choice of a task model, so changes to the workflow structure do not need to affect the structures of tasks and

2) it separates workflow scheduling from task execution, thus improves the performance and scalability of the whole system. The interoperability of workflows (requirement R7: level 2) has to be addressed by standardizing workflow models, workflow run models, and workflow languages.

The fourth layer is the Presentation Layer, which provides the functionality of workflow design and various user interfaces and visualizations for all assets of the whole system. The Presentation Layer has interfaces to each lower layer (not shown in the figure for simplicity). The separation of the Presentation Layer from other layers provides the flexibility of customizing the user interfaces of the system and promotes the reusability of the rest of system components for different scientific domains. Thus, this separation supports requirement R1. The interoperability of workflows (requirement R7: level 2) should be addressed by standardizing the workflow layout (e.g., look-and-feel) at this layer.

3.2 MIDDLEWARE OVERHEAD

The middleware overhead is due to the work performed by the middleware services to support the proper execution and completion of the workflow, which we further divide into several components based on the service functionality.

Resource brokerage. This represents the time required by the Resource Broker to query the information service and provide to the Scheduler the processors and activity deployments needed to execute the application. Addition-ally, this overhead has an important latency component (few seconds), mostly due to the mutual host authentication. This service latency is a common overhead component present in all our middleware services.

Performance prediction. This represents the time to provide forecast information about the execution time of individual activities on the Grid sites indicated by the Scheduler, for example, using a polynomial fitting heuristic based on historical or training data.

Scheduling. This represents the time to appropriately map the workflow activities onto the Grid resources, which includes the following two sub overheads:

1) a scheduling algorithm, which represents the time required to compute a schedule (often using time-consuming optimization heuristics if the scheduling problem is NP-complete), and

2) rescheduling, which represents the time needed to make a new scheduling decision, for example, because of a performance contract violation or if the workflow changes its runtime structure.

3.3 SUBSYSTEMS

The seven major functional subsystems correspond to the key functionalities required for an SWFMS. Although the reference architecture allows the introduction of additional subsystems and their features in each layer, this paper only focuses on the major subsystems and their essential functionalities.

The Workflow Design subsystem is responsible for the design and modification of scientific workflows. Workflow Design produces workflow specifications represented in a workflow specification language that supports a particular workflow model. One can design and modify a scientific workflow using a standalone or Web-based workflow designer, which supports both graphical- and scripting-based design interfaces. The interoperability of workflows (requirement R7: level 2) should be addressed in this subsystem by the standardization of workflow languages.

The Presentation and Visualization subsystem is very important especially for data-intensive and visualization-intensive scientific workflows, in which the presentation of workflows and visualization of various data products and provenance metadata in multi-dimension is the key to gain insights and knowledge from large amount of data and metadata. These two subsystems are located at the Presentation Layer to meet requirement R1. In this subsystem, the interoperability of workflows (requirement R7: level 2) should be addressed by the standardization of scientific workflow layout.

The Workflow Engine subsystem is at the heart of the whole system and is the subsystem that provides management and execution environments for workflow runs. The Workflow Engine creates and executes workflow runs according to a workflow run model, which defines the state transitions of each scientific workflow and its constituent task runs. The interoperability of workflows (requirement R7: level 2) should be addressed by the standardization of interfaces, workflow models, and workflow run models, so that a scientific workflow or its constituent subworkflows can be scheduled and executed in multiple Workflow.

Engines that are provided by various vendors. In SWFMSs, multiple Workflow Engine subsystems can be distributed, and each Workflow Engine can execute several workflows in parallel.

The Workflow Monitoring subsystem meets requirement R6 and is in charge of monitoring the status of workflow execution during workflow runtime and if failures occur, provides tools for failure handling [18].

The Task Management subsystem addresses heterogeneity and distribution issues (requirement R3) and provides management and execution environment for tasks, according to a task model and task run model, respectively. The interoperability of tasks between various workflow environments (requirement R7: level 1) can be addressed in this subsystem.

The Provenance Management subsystem meets requirement R2 and is mainly responsible for the management of scientific workflow provenance metadata, including their representation, storage, archival, searching, and visualization. The Data Product Management subsystem meets requirement R4 and is mainly responsible for the management of heterogeneous data products. One key challenge for data product management is the heterogeneous and potentially distributed nature of data products, making efficient access and movement of data products an important research problem. The interoperability of data products between various workflow environments (requirement R7: level 1) can be addressed in this subsystem.

3.4 SYSTEM PROTOTYPE

We have built a prototype system that implements all features and algorithms presented in this paper. This system, built initially to support refresh, has been evolving into a more ambitious sys-tem we call Panda (for Provenance and Data), supporting several other aspects of provenance and data in addition to refresh [17].

In this paper we focus our system description on features relevant to refresh. For the time being, all data sets handled by Panda are encoded in relational tables, but as we have seen, our formal under-pinnings and algorithms do not rely on the relational model. The high-level architecture of the Panda system. The main backend is a SQLite server, storing all data sets, relational (SQL) transformations, provenance, and workflow information. The Panda system supports “opaque” transformations programmed in Python; they are stored separately in files.

Users interact with Panda through a simple command-line inter-face; we intend to build a GUI in the near future. There are three types of user commands:

1) Creating or modifying input data sets;

2) Creating transformations that generate newly-defined data sets from existing ones, to build up workflows;

3) Refresh commands.

The Panda Layer processes all user commands: it stores workflow graphs and their transformations, creates and maintains auxiliary provenance tables, generates provenance predicates and forward filters for output elements, and runs the refresh algorithms.

The Panda system also supports transformations specified as SQL queries, including

queries/transformations involving multiple input tables. Provenance predicates are created automatically for SQL transformations, following known definitions and techniques [5, 7, 10]: Single table Select statements are one-one, so their output provenance predicates can select on declared keys from the input data set. Multi-table Select statements generate provenance predicates for each input table separately, again relying on declared keys. Finally, Group-by queries generate provenance predicates based on the grouping attribute(s). The command to create a new SQL transformation is similar to the command shown in Section 8.1, except followed by a SQL query, whose from clause must refer to already-defined tables. The steps performed by the Panda Layer are also similar to those outlined in Section 8.1; forward filters are never needed since SQL queries cannot produce many-many transformations.

When workflows are created and run, Panda stores everything needed to support selective refresh: provenance predicates and intermediate data sets for backward tracing; transformations and forward filters for forward propagation. The Panda system assumes that all transformations, provenance, and workflows satisfy the requirements specified in this paper. Automatically detecting when the requirements are satisfied—particularly the most interesting requirement of workflow safety is an important area of future work. Under the assumption of all requirements being satisfied, Panda supports selective refresh using the exact algorithms given in this paper.

4 Experimental results

We implemented our approach as a distributed online performance analysis tool within the ASKALON programming and computing environment for the Grid. We implemented P-C and P-SC correlators in Python as WSRF-compliant Web services exposing all unprocessed events as reference properties. We translate our formal rule correlation algebra to a lower level open source correlation engine called the Simple Event Correlator.

We employ the PyGnuplot module that is a Python wrapper to the Gnuplot program to display two different kinds of execution graphs in real time, as presented in this section. We describe experiments of applying our tool for online analysis of the main overheads, using the WIEN2k workflow. We chose a problem case that we solved using 193 parallel k-points and a problem size of 8.5, which represents the number of plane waves that is equal to the size of the eigenvalue problem (that is, the size of the matrix to be diagonalized).

We created three groups of peers, one for each city location. The group in Innsbruck consists of four P-Cs and one P-SC, whereas in Linz and Salzburg, we only started one P-SC as we only had one Grid site available. We elected the Innsbruck P-SC as the coordinator.

The first histogram in Figure 2 illustrates a generic technique that we use to represent the major online phases that occur during the workflow execution. We defined for all overheads a performance contract membership function with a low step threshold that generates a contract variable with a critical value at every polling instance (we disabled the rescheduling action). The horizontal line indicates the overhead that holds at any execution instance, whereas the vertical lines are drawn for readability purposes only. The histogram shows that at the beginning of the execution, we experienced some delay due to the operations performed by the middleware services, more precisely the Resource Broker to retrieve the list of available Grid sites and the Scheduler to compute the mapping of the workflow onto the Grid sites. One characteristic of this workflow is that the number of activities in the parallel regions is unknown until the first activity completes its execution. Since this cardinality port is statically unknown, the Scheduler assumed one activity in each case and serialized all workflow activities onto one Grid site. After another preparatory step to create remote directory structures required to run the legacy applications that implement the workflow activities, the Enactment Engine submitted the first activity and therefore added some queuing overhead. After this first activity completed, the output files were broadcasted to all sites, which added a small data transfer overhead.

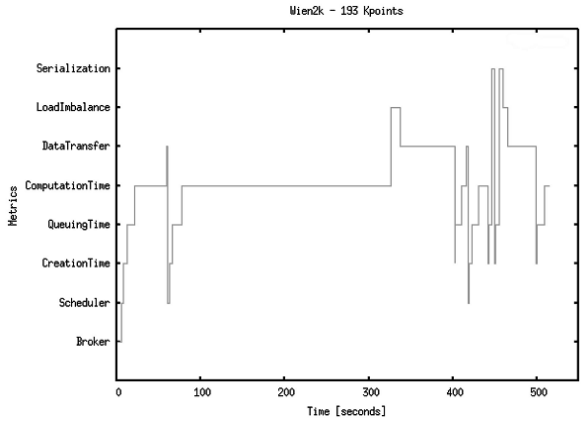


FIGURE 2 Online overhead histogram

After the cardinality port was instantiated, a rescheduling event was triggered upon contract violation, and the Scheduler mapped the new workflow onto the Grid using the Heterogeneous Earliest Finish Time heuristic which added a corresponding overhead. Thereafter, all 193activities were submitted in parallel, since the number of processors available on the Grid test bed was large enough to accommodate them. Since our Grid contains heterogeneous processors, some of the activities completed before the others, which produced a slight load imbalance. As lapw1_k is the most time-consuming activity type of the workflow, we did not experience any significant overhead during this phase, which means that the workflow was performing a useful computation. After this positive execution phase, a large number of files had to be gathered onto one single site where a small activity

called lapp1_fermi processed them, which caused a rapid increase in the severity of the data transfer.

Figure 3 depicts another online-generated histogram that represents the aggregated severity of each analysed over-head at every time instance during execution. The last four sites in our Grid test bed are workstation networks that are automatically restarted in Linux Grid mode during night, weekend, and public holidays and are manually rebooted by students and lecturers during the weekdays in Windows mode for their laboratory classes when they are no longer available for our Grid experiments. After this first positive phase in the workflow execution, we decided to exploit the dynamic characteristic of our Grid test bed and removed the availability of the last four Grid sites by manually turning off the workstations that run the GRAM gatekeepers. This eliminated 51 per cent from the total number of available processors, and as a consequence, the Resource Broker created an appropriate event signalling that at time instance, an important number of computational resources had been lost. The arrival of the (simplified) event and, therefore, the serialization overhead, which indicated that some of the lapw2_k parallel activities had to wait in the queue and were executed sequentially due to the lack of available Grid sites.

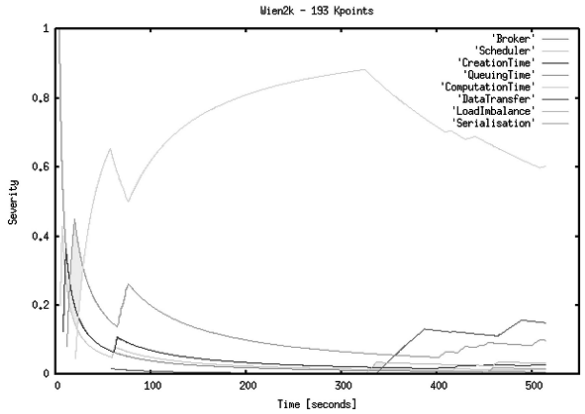


FIGURE 3 Online overhead severity histogram

Our tool is able to automatically generate real-time graphs that represent the chain of events that led to this serialization overhead using the Pydot wrapper to the GraphViz program, as illustrated in Figure 4. At the same time, our tool can automatically produce an online over-head correlation tree for any workflow region at any execution instance. Figure 5 shows this tree for the workflow region indicating that 76.6 per cent of its time is due to the total identified overhead, which is further split into five overheads: queuing, load imbalance, job preparation, data transfer, and serialization, each of them with its own severity value.

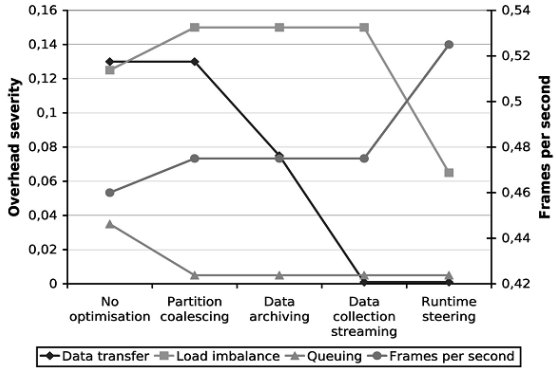


FIGURE 4 Overhead severities for various optimizations.

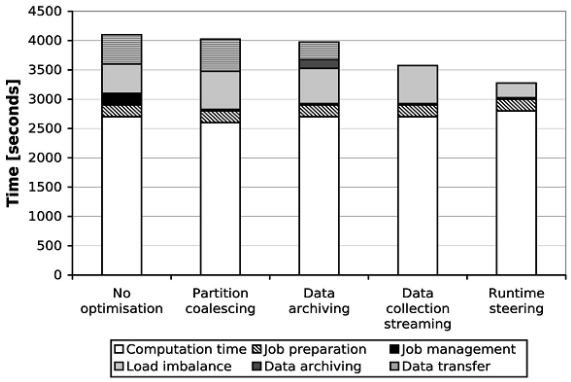


FIGURE 5 Runtime comparison for various optimizations

6 Conclusions

In this paper, we presented a systematic analysis model consisting of a theoretical ideal execution time and a detailed hierarchy of overheads that help the application developers understand the sources of bottlenecks that affect the distributed execution of scientific workflows in heterogeneous Grid infrastructures. We carefully defined the overheads to be as little overlapping as possible, which gives us an important indication of whether any performance loss remained unidentified. We adjusted well-known normalized metrics from parallel processing to the Grid computing scope, including overhead severity, speed-up, and efficiency, which are invaluable parameters to be considered before scheduling when the efficient use of resources is an important issue.

We implemented our approach as a distributed online performance analysis tool within the Grid programming and computing environment. We proposed a distributed super architecture for performance analysis, in which individual peers local to Grid sites correlate large numbers of events to find small sets of meaningful overheads at a higher level of abstraction. To the best of our knowledge, this is the first attempt of applying the event correlation technology, highly successful in the networking field, to the performance analysis of workflow applications in Grid environments. Additionally, we extended the current best effort practice in executing scientific Grid workflows by

defining performance contracts as QoS parameters to be enforced during execution through event-correlation-based fuzzy logic rules. We illustrated experiments for the postmortem and online analysis of two real-world workflow applications with a dynamic structure (that is, statically unknown before the execution) in a real and dynamic Grid environment. The postmortem analysis presents the advantage of detail and rigor through repetitive executions and measurements, whereas the online analysis is invaluable for runtime adaptation, QoS enforcement, and steering. We learned that the serialization of independent activities, load imbalance, job preparation and management, and transfer of large numbers of data dependencies are the most severe overheads for our case-study applications that have to be carefully tuned for achieving good speedup and the efficient use of Grid resources. In this context, we introduced several generic optimization and tuning techniques including workflow partitioning, data archiving, data collection streaming, and runtime steering

References

[1] Berman et al. 2005 New Grid Scheduling and Rescheduling Methods in the GrADS Project *Parallel Programming* **33**(2-3) 209-29
[2] Deelman E et al 2003 Mapping Abstract Complex Workflows onto Grid Environments *Grid Computing* **1**(1) 25-39
[3] Mayer A, McGough S, Furmento N, Lee W, Newhouse S, Darlington J 2003 ICENI Dataflow and Workflow: Composition and Scheduling in Space and Time *Proc. UK e-Science All Hands Meeting* 627-34
[4] Taylor I, Shields M, Wang I, Rana R 2003 Triana Applications within Grid Computing and Peer to Peer Environments *Grid Computing* **1**(2) 199-217
[5] Oinn T, Addis M, Ferris J, Marvin D, Senger M, Greenwood M, Carver T, Glover K, Pocock M, Wipat A, Li P 2004 Taverna: A Tool for the Composition and Enactment of Bioinformatics Workflows *Bioinformatics* **20**(17) 3045-54
[6] Erwin D W 2002 UNICORE - a Grid Computing Environment *Concurrency and Computation: Practice and Experience* **14**(13-15) 1395-410
[7] Fahringer T, et al 2007 Askalon: A Development and Grid Computing Environment for Scientific Workflows *Workflows for e-Science: Scientific Workflows for Grids, I J Taylor, E Deelman, D B Gannon, M Shields eds Springer* http://www.askalon.org
[8] Alves A, et al. 2006 Web Services Business Process Execution Language, Specification 2, Organization for the Advancement of Structured Information Standards ftp://www6.software.ibm.com/software/developer/library/ws-bpel11.pdf
[9] Wolski R, Spring N T, Hayes J 1999 The Network Weather Service: A Distributed Resource Performance Forecasting Service for Metacomputing *Future Generation Computer Systems* **15**(5-6) 757-68
[10] Czajkowski K, et al 2001 Grid Information Services for Distributed Resource Sharing *Proc 10th IEEE Int'l Symp High Performance Distributed Computing (HPDC)*
[11] D Nurmi, Mandal A, Brevik J, Koelbel C, Wolski R, Kennedy K 2006 Evaluation of a Workflow Scheduler Using Integrated Performance Modelling and Batch Queue Wait Time Prediction *Proc ACM/IEEE Supercomputing Conf (SC)*
[12] DAGMan: Directed Acyclic Graph Manager 2007 *Univ ofWisconsin,Madison* http://www.cs.wisc.edu/condor/dagman/ *Condor Project*
[13] Vaarandi R 2002 SEC—A Lightweight Event Correlation Tool *Proc. Workshop IP Operations and Management (IPOM)*
[14] Liu G, Mok A K, Yang E J 1999 Composite Events for Network Event Correlation *Proc Sixth IFIP/IEEE Int'l Symp Integrated Network Management (IM)*
[15] The Austrian Grid Consortium 2007 http://www.austriangrid.at
[16] PBS: The Portable Batch System 2007 http://www.openpbs.org
[17] Sun Microsystems Sun Grid Engine 2007 http://gridengine.sunsource.net/
[18] Czajkowski K 1998 A Resource Management Architecture for Metacomputing Systems *Proc Fourth Workshop Job Scheduling Strategies for Parallel Processing* 62-82
[19] Foster I, Kesselman C 1997 Globus: A Metacomputing Infrastructure Toolkit *Supercomputer Applications and High Performance Computing* **11**(2) 115-28

Authors

Hui Bu, born in August, 1979, Zhengzhou, Henan Province, China

Current position, grades: master, a lecturer in School of Software, North China University of Water Resources and Electric Power, China.
University studies: computer science and technology.
Scientific interest: computer software, database and software test automation.
Publications: 12 papers.

Ran Liu, born in December, 1979, Zhengzhou, Henan Province, China

Current position, grades: master, a lecturer in School of Software, North China University of Water Resources and Electric Power, China.
University studies: Computer science and technology.
Scientific interest: computer software, database and computer network.
Publications: 12 papers.

A boundary knowledge field based data mining method

Shan Feng*

School of Computer Science and Technology, Hubei Polytechnic University, China

Received 1 March 2014, www.cmnt.lv

Abstract

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. Knowledge is the source of getting and keeping the strength of competence, knowledge becomes an important strategic resource at the age of knowledge economy. The way to facilitate knowledge transfer smoothly plays a critical role in determining the competence of an organization or organizational system. To explore the mechanism of knowledge transfer, the characteristics of sticky transfer (flow) is compared with the flow characteristics of viscous fluid, and the knowledge on the theories of field in physics, viscous fluid mechanics and boundary layer is employed and analysed in this paper. Firstly, the concept of boundary layer in the knowledge field is proposed to analyse the difference of knowledge stickiness in the knowledge field and describe the contradictory relationship between knowledge stickiness and liquidity. Secondly, based on the analysis of the knowledge transfer in the boundary layer, the dynamic mechanisms of sticky knowledge transfer are analysed from the three aspects of the knowledge potential difference force, the viscous force and the extern driving force. The rotation mechanism of knowledge in the knowledge field is discussed, and the dynamic model of the boundary layer of knowledge field is built. Finally, the phenomenon of knowledge flowing into and flowing out of boundary layer is discussed, and the knowledge transfer conservation equation in the boundary layer is constructed to describe the updating and appreciation of knowledge in the boundary layer of knowledge field.

Keywords: data mining, knowledge transfer, knowledge field, boundary layer

1 Introduction

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Knowledge is distinguished from data and information. The data is organized to produce general information [1], and information has little value in itself until it becomes knowledge as a result of processing by the human mind [2]. For knowledge, it involves a person using his or her perception, skills and experience to process information - thus converting it into the knowledge in the mind of the individual [3]. Knowledge is a vital source for the organizations to gain the advantage in the competition [4]. Ever since the conception of knowledge economy is brought forward by Drucker [5]. The position of knowledge in the development of economy is increasingly gaining more and more attention, and researches have shown that the world economy is driven by the production, transfer, and use of knowledge [6, 7]. In the time of knowledge economy, knowledge transfer and creation are growing more and more important, no matter for individuals, organizations, or even countries [8, 9]. Taking its fluidity into consideration, knowledge is considered to be one fluid and it can transfer among organizations. Therefore, the concept of knowledge transfer is propounded by Teece [10]. He pointed out that it is a process in which knowledge is sent and received back and forth between knowledge senders and receivers [10]. Knowledge transfer is a communication process in which knowledge flows between demanders and producers [11]. It is an important factor for an organization's success [12]. Accordingly, many studies analyzed the factors that affect knowledge transfer, and showed that the knowledge characteristics, such as tacitness, expressiveness and complexity, directly change the process of knowledge transfer (Szulanski, 1996; Zander, 1991; Simonin, 1999) [13-15]. Afterwards Cummings and Teng [16] studied the influences of the expressiveness and embedded-ness of knowledge on knowledge transfer. Then Ajmal & Koskinen (2008) [1] elaborated the knowledge transfer from three perspectives: first, as a solution, knowledge emphasizes the real-time transfer of knowledge among practitioners who are seeking to solve problems or enhance operations; second, as an experience, knowledge describes the progressive process of acquisition and accumulation for the possible use in the future; third, as a social product, knowledge is created and shared through interpersonal relationship and social network.

Due to the fact that knowledge transfer is an essential role for the organizations to gain the advantage in the

***Corresponding author** e-mail: shanfenghbpu@126.com

competition, it is very important to accurately describe the transfer mechanism of knowledge in an organizational system [17, 18]. Since the concept of knowledge field (Ba) was proposed by Nonaka and Takeuchi in 1995 [19], the concept gave researchers lots of inspiration on knowledge transfer. Nonaka and Takeuchi [19], additionally, structured the SECI model of knowledge transfer and creation, meanwhile they pointed that knowledge field could provide energy for knowledge transfer and creation and affect the quality of knowledge production. Based on the SECI model of knowledge creation, Ba, which is conceptualized as a shared context in motion, can transcend time, space, and organization boundaries to create knowledge [20]. The theory of knowledge field is widely recognized [21, 22]. Against this background, many researchers have conducted deep research on the transfer mechanism of knowledge based on the concept of "knowledge field". The process and the influence factors of knowledge transfer (flow) are analysed [23, 24]. The view of Nonaka is accepted by Eastern researchers, but the Western pay attention to research the process of knowledge share and based on the organizational theory and the social capital theory analyse the construction of communities of practice and the impact of the knowledge-sharing field (communities of practice) on the organizational performance [25-27].

Throughout the above researches, the knowledge transfer (knowledge flow) and knowledge stickiness have been analysed, which attempt to explain the mechanism of knowledge transfer and the influence mechanism of the stickiness on knowledge transfer. However, most of these investigations mainly focus on qualitative discussion and empirical analysis based on questionnaires. Since the mechanism of knowledge transfer is still not fully understood, there remains a lack of systematically theory and mathematical model to elaborate the process and mechanism of knowledge transfer. As discussed earlier, it is known that knowledge has the properties of stickiness and field. In fluid dynamics, when the fluid flows in the flow field, it has the properties of liquidity and stickiness which make the phenomenon of the boundary layer generated [45, 46]. In 1904 it is the great achievement of Ludwig Prandtl who set forth the concept of boundary layer at the Heidelberg mathematical congress. Prandtl showed that the flow past a body can be divided into two regions: a very thin layer close to the body (boundary layer) where the viscosity is important and the remaining region outside this layer where the viscosity can be neglected [47]. It is confirmed that the appearance of the boundary layer phenomenon is caused by the liquidity, the viscosity and the properties of field. It should be mentioned that knowledge also has these properties in the process of knowledge transfer (flow). It is believed that the mechanism of knowledge flow (knowledge transfer) is similar to that of viscous fluid flow.

Therefore, the study is inspired by viscous fluid mechanics, and the viscous fluid-flow model and the boundary layer theory is employed to describe the process of knowledge transfer in this article. The flow of the sticky knowledge in the knowledge field is explained by the phenomenon of the boundary layer. On this basis, the rotation of knowledge field through the boundary layer phenomenon of knowledge stickiness is analysed. Meanwhile, based on the partial differential theory and the analogy method of viscous fluid dynamics, the dynamic equation and conservation equation of knowledge field are derived in order to analyse the dynamic mechanism of knowledge transfer under the action of different effects and the updating and appreciation of knowledge in the boundary layer of the knowledge field.

2 Boundary layer model of knowledge transfer

In a great deal of previous researches, knowledge transfer is widely recognized as an essential process in which an organization can acquire valuable knowledge to produce knowledge and boost economic performance [28-31]. However, organizational members in their own interest are unwilling to share core information they possess with other members [32]. Thus, it is difficult to effectively achieve the transfer of knowledge among different organizations [14, 33, 34], even among the different departments in an organization. This is, to a large extent, because knowledge has stickiness which severely hinders the transfer of knowledge between knowledge providers and knowledge users [34-36]. The knowledge stickiness can be defined as the extent to which knowledge is unremovable from its originating source [37]. A series of investigations on knowledge stickiness have been conducted by many researches [38-40]. Knowledge stickiness comes from the concept of "sticky information", which was introduced by Von Hippel [38] to describe information that is difficult to transfer. On this basis, Szulanski [14] extend the concept of information stickiness and defined knowledge stickiness as difficulties encountered in the process of knowledge transfer. The above knowledge which is hard to transfer is called sticky knowledge, and it is hard to transfer smoothly than general knowledge (namely, its stickiness is negligible). And then researches have further explored how properties of knowledge influence the transfer of sticky knowledge. Simonin [15] analysed the phenomenon of stickiness in the process of knowledge transfer from the aspects of tacitness, assets specificity, complexity, protection of its owners, cultural and organization distance etc. The "causal ambiguity" of the knowledge or the extent to which it was not well understood predicted the difficulty of transfer throughout all phases of the transfer process [4]. According to the process of knowledge transfer (initiation, implementation, ramp-up and integration), Szulanski [41] also discussed the nature of difficulty caused by stickiness at each stage.

When acknowledged, knowledge can be categorized into tacit knowledge and explicit knowledge [19], the stickiness of tacit knowledge is greater than that of explicit knowledge, and the transfer of tacit knowledge is more difficult. Thus, simple, codified knowledge is easier and

more efficient to transfer than complex, tacit knowledge among members in organizations, and meanwhile to strengthen members' tie can increase the ease and efficacy of transferring complex, tacit, and private knowledge [42, 43]. According to the research of Szulanski and Cappetta et al [44], knowledge stickiness means the difficulty of intra-organization knowledge transfer, and it can weaken the organizational flexibility and reduce the capacity of making full utilization of the knowledge which transfers between the enterprise and its partners. Thus, for an organizational system, the way to transfer the sticky knowledge smoothly plays an important role in improving the efficiency of innovation and keeping strength of competence.

Knowledge has the properties of field, and its distribution is uneven in time and space. In the knowledge field, the potentials of knowledge owners are related to the knowledge stocks that they own, that is, the more knowledge stocks they have, the higher knowledge potential they locate (Wang, Xing & Tang, 2010) [22]. Driven by the fluidity and inhomogeneous distribution of knowledge in the knowledge field, knowledge flows from the area with higher knowledge potential to the area with lower knowledge potential [24], which makes the knowledge providers interacting with knowledge users. In the interacting process, the share, transfer, and production of knowledge are conducted. However, because of its stickiness, knowledge fails to transfer smoothly, namely there are obstacles during the process of sticky knowledge transfer [36, 48, 49].

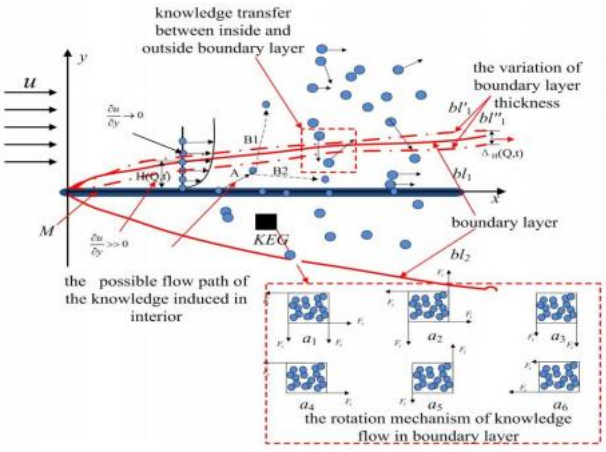


FIGURE 1 Boundary layer model of knowledge transfer

To describe the process of sticky knowledge transfer and to characterize the mechanism of knowledge transfer, the boundary layer model of knowledge field is established when sticky knowledge flows between knowledge providers and knowledge users in the knowledge field. To simplify this model, a two-dimensional knowledge field is employed to describe the boundary layer phenomenon of knowledge field, as the Cartesian coordinate of X-O-Y shown in Figure 1.

In Figure 1, we suppose that the blunt body M represents an organization in the knowledge field, and there is a variety of knowledge in the two-dimensional knowledge field. The circular marks indicate a variety of knowledge that the organization has in the knowledge field. The size of circles mark indicates the stock of one type of knowledge, and the arrows for the circle marks represent the direction of knowledge motion in the knowledge field. The circle marks is also called one knowledge unit in this paper, which is so small that it cannot be divided. Different knowledge units have different speeds. Some knowledge units can almost transfer freely in the knowledge because of the tiny sticky, some may not transfer smoothly and even the speed of it is near zero due to the strong viscous resistance. The speed of knowledge transfer in the knowledge field is expressed in u. According to the viscous fluid dynamics theory, the phenomenon of boundary layer will be produced near the surface of an object, when viscous fluid passes the surface of an object at a certain speed [46, 50]. Similarly, due to the stickiness of knowledge, it is believed that, when knowledge flow around an organization in the knowledge field, the phenomenon of the boundary layer will appear near the organizational boundary. The phenomenon of the boundary layer in the knowledge field is referred to as the boundary layer of knowledge field in this paper. As shown in Figure 1, the knowledge field is divided into two regions: the knowledge field inside the boundary layer and outside the boundary layer. The knowledge field inside the boundary layer is the area between the red solid line bl1 and bl2, and the other area is the knowledge field outside the boundary layer. Similar to the theory of fluid mechanics [51], the effect of stickiness on the knowledge field inside the boundary layer is so great that the velocity gradient of knowledge transfer is very large, and the velocity near the areas of the surface of an object (the object is M in Figure 1) tends to zero. That is, the effect of the stickiness of knowledge on knowledge transfer in boundary layer is significant, and the stickiness of knowledge inside boundary is not ignorable. However, the stickiness of knowledge outside the boundary layer is small, and even it can be ignored, it is approximatively considered that knowledge in the area is in the free-flow state without viscous force. It is known that with the value of knowledge increasing, the will of share becomes weak and the difficulty of knowledge transfer increases. Therefore, it is believed that the magnitude of the stickiness is directly related to the value of knowledge. In the current discussion, the knowledge in the knowledge field is divided into two categories according to its value or importance: sticky knowledge and general knowledge. Sticky knowledge is usually core and important knowledge for an organization, and general knowledge is not core knowledge and can be obtained freely, and its stickiness is small.

In the previous studies, there are arguments that knowledge field is rotational or not. Some researchers suggested knowledge field is rotational [23, 24, 52], and

others believed that knowledge field is irrotational or did not considered the rotation of knowledge field [22, 53]. According to viscous fluid mechanics, the area inside the boundary layer is the rotational area due to the stickiness of fluid in the flow process [54]. However, owing to the tiny viscous shear force, the rotation of flow field outside the boundary layer is negligible; namely, the area outside the boundary layer can be considered to be irrotational area [54]. Similarly, it is believed that knowledge field is rotatable owing to the knowledge stickiness. Due to the fact that the stickiness of knowledge outside boundary layer is small, the rotation of knowledge field outside boundary layer is small and even negligible. However, inside the boundary layer, the rotation of knowledge field is large because of the large stickiness of knowledge. That is, the knowledge field inside boundary layer is rotational, and the knowledge field outside boundary layer is irrotational. To clearly show this point, the microscopic subunit inside the boundary layer of knowledge is called knowledge element group (KEG) in this paper, as shown in Figure 1. In the microscopic subunit, there are interactions between different kinds of knowledge under the influence of strong viscous force. As the stickiness of each kind of knowledge in the knowledge element group is different, and the level and the direction of the viscous shear force on the different boundary of knowledge element group are also different, so the phenomenon of the rotation occurs under the action of viscous shear force. The six rotation situations caused by the viscous shear force with different levels and directions on the boundary of KEG are illustrated, as the marks a shown in Figure 1. Whereas, when the KEG is located outside boundary layer, the interactions between the different kinds of knowledge in KEG under the influence of viscous force is tiny, and the situation which is similar to the interaction between ideal inviscid gas molecules exists only in KEG. Meanwhile, the viscous shear force on the boundary of KEG is also tiny. Therefore, the rotation of knowledge field outside boundary layer is tiny and even negligible.

3 The dynamic model of knowledge transfer

The blunt body M on the x axis represents an organization in the knowledge field, which is the same as the Figure 1. In Figure 2, BC shows the boundary line of the boundary layer; represents the velocity of knowledge transfer in the boundary layer; H expresses the thickness of boundary layer at x=x0, and H1shows the thickness of boundary layer at x=x1. The variable dH shows that the variation of thickness when x changes the distance of dx.

The variables Fp, Fτ and Fe are the powers in the direction of x which represent the knowledge potential difference force, the viscous force and the extern driving force, respectively. The knowledge potential difference force, the viscous force and the extern driving force are caused by the knowledge potential difference, knowledge stickiness and the extent of the external driving, respectively. To analyse how sticky knowledge transfers under the interaction of internal and external force in the boundary layer, the dynamic model of sticky knowledge transfer is built, as shown Figure 2.

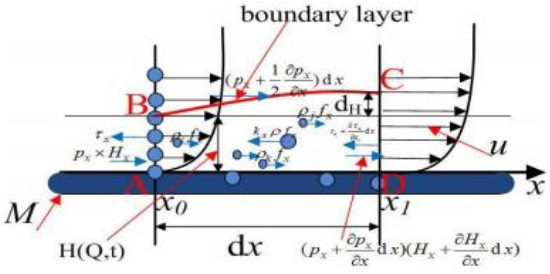


FIGURE 2 The dynamic model of knowledge transfer

Finally, to clearly describe the dynamic model of the boundary layer of knowledge field, the example of the supplier management of Japan's Toyota Motor Corporation is employed.

In the knowledge field, it is believed that the knowledge transfer in the boundary layer is mainly influenced by the three forces:

(i) Due to the difference of knowledge stocks among organizations, knowledge pressure (similar to water pressure and the voltage) is caused, and then the first field force, that is, the potential difference of knowledge force that lead to knowledge flow is generated [22, 56].

(ii) The influence of knowledge stickiness on knowledge transfer and sharing causes the second field force, namely the viscous force [36, 37].

(iii) The organizations within and outside the knowledge field, such as government and core enterprise, can make scientific incentive mechanisms to promote knowledge sharing and flowing.

Thus, the third field force is caused, which is referred as the external force [57, 58]. Under the joint action of the above three field forces, all kinds of knowledge unit in the knowledge field will move directionally.

The case of Japan's Toyota Motor Corporation is also introduced to discuss the sticky knowledge transfer in knowledge field. Japan's Toyota Motor Corporation (hereafter, referred to as simply Toyota) was founded in 1933, which is the biggest car company in Japan and world's top ten car companies, and it plays an important role in the world's auto industry. The success of Toyota to some extent comes from effective supplier management way. Toyota pioneered the management style of knowledge sharing among suppliers, which has had some great effects on supplier management, and has been adopted in many other companies. Toyota has always attached importance to its cooperation with suppliers, and it and 270 suppliers form a knowledge field by cooperation, in which the knowledge stocks of Toyota and its suppliers is significantly different, so there are the potential difference forces (Fp) prompting sticky knowledge transfer among Toyota and its suppliers. With the above force, Toyota would open and share sticky knowledge that can help suppliers improve productivity, and the suppliers also share their sticky knowledge with Toyota and others

so that suppliers and Toyota grow together through their mutual cooperation. Nevertheless, given their own interest consideration, suppliers are reluctant to share their core knowledge inside the boundary layer to others, so the transfer of the core knowledge in the knowledge field is hindered. The above core knowledge is called sticky knowledge, such as patents and core technologies. Generally speaking, the transfer of patents and core technologies is purchased at the expense of equal knowledge or cost of the users. The value of the equal knowledge or cost can be considered as the viscosity of the transferred knowledge. The force generated by sticky knowledge (Fτ) has the reverse force on the knowledge transfer.

Due to the viscous force, it is difficult to achieve sticky knowledge transfer smoothly in the knowledge field. Therefore, to be successful in implementing sticky knowledge transfer according to the wishes of the core organization (Toyota), some policies (Fe) should be introduced by Toyota to drive high efficiency knowledge transfer. For this, Toyota formulated two efficient governance mechanisms which is respectively reputation and commitment mechanism to make suppliers share their own core knowledge to others, as follows:

(i) During the process of cooperation, if the supplier positively shares the core or valuable knowledge (sticky knowledge) with others in the knowledge field, his reputation will be raised, and Toyota will offer more valuable technology or knowledge to him. Otherwise, if the supplier only obtains valuable knowledge from others in the knowledge field but is not reluctant to shares his own core knowledge or valuable knowledge (sticky knowledge), his reputation will be reduced, and Toyota will not offer key technology or knowledge to him and even break up the partnership.

(ii) Since the suppliers fear that their interests will be hurt when they share their core knowledge with others, Toyota establishes commitment mechanism to protect their interests.

This is, Toyota makes the commitment that the orders of the above suppliers will be increased and the supply prices will not be also raised due to their higher productivity within 2 to 3 years. Therefore, under the combined action of Fp, Fτ and Fe in the knowledge field, Toyota and its suppliers successfully achieve the goals of mutual cooperation, promotion and common development.

4 The conservation model of knowledge transfer

To describe the variation of knowledge stocks in the process of knowledge transfer in the boundary layer, the research object M is also employed, as shown in Figure 3.

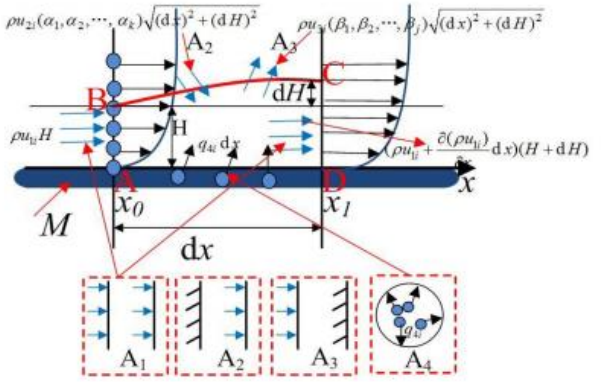


FIGURE 3 The conservation model of knowledge transfer

The blunt body M and BC represents an organization in the knowledge field and the outer horizon of the boundary layer, respectively; ui, H, H1 and dH represents the speed of knowledge transfer in the boundary layer, the thickness of boundary layer on the x0 points, the thickness of boundary layer on the x1point and the variation of thickness is caused as the point in the direction of x changes the distance of dx, respectively. The above is the same as the Figure 2.

In the boundary layer of knowledge field, the sticky knowledge of organizations do have is not to be constant, which flows constantly among organizations in the knowledge field. According to fluid dynamic theory, for a research object, generally speaking, if it has inflow variables, it also has outflow variables. Analogously, for the research object M in the boundary layer of knowledge field, the inflow variable on the boundary AB can be written as shown in Figure 3.

Unlike the flow dynamics, there are the other three ways due to knowledge specific attributes in the knowledge field, as the marks A2, A3 and A4 shown in Figure 3. From the mark A2, it has outflow variables but no inflow variables in the boundary layer. For example, when participating in the league and cooperation, the organization, in the knowledge field, shares its core knowledge (namely, the sticky knowledge) with others, but the spill over effect of the knowledge which is absorbed from others in the organization is not obvious because of the organization's low learning, low absorbing knowledge ability and low efficiency of utilizing knowledge gained from others. Thus, there might be a situation where the organization only shares core knowledge but not obtain valuable knowledge from the knowledge field. The process of knowledge transfer is achieved by overflowing the boundary of the boundary layer in the knowledge field. It should be mentioned that new knowledge will be created due to the novel combination of existing knowledge [61]. The extent to which actors can efficiently search for, access, transfer, absorb, and apply knowledge influences their ability to create knowledge [62, 63]. Another case is that new knowledge created by an organization itself is be stored in

the boundary layer and does not outflow the boundary layer, as the mark a4 shown in Figure 3.

As analysed above, we know that the variation of the thickness of boundary layer (the knowledge stocks of boundary layer) is characterized by the knowledge flow conservation equation in the boundary layer. The thickness of boundary layer of the knowledge field for each organization is dynamic, which changes with time and location. The factors, such as knowledge transfer between inside and outside boundary layer, knowledge creation, and usefulness of knowledge declining, can cause the variation of boundary layer thickness. Meanwhile, knowledge transfer should be in relative balance situation under the action of the function and mechanism in the boundary layer, and the mutual transformation process of "balance unbalance balance" promotes the updates and appreciation of knowledge.

5 Conclusions

The flow characteristics of viscous fluid are extended to the analogy analysis of the knowledge management in this paper. Some conclusions can be summarized as follows:

1) The contradictory relationship between knowledge stickiness and knowledge transfer in the knowledge management is illustrated based on the boundary layer theory in fluid mechanics. The boundary layer theory of knowledge field is put forward to analyse the stickiness and fluidity of knowledge, which extends the conception of knowledge field, and enriches the theory of knowledge management.

2) The boundary layer dynamic model of knowledge field based on the features of sticky knowledge transfer is constructed to discuss the dynamic mechanism of sticky knowledge transfer. The knowledge transfer conservation equation according to the law of conservation of knowledge stocks is built to further explore the updates and appreciation of knowledge in the transfer process of knowledge.

3) The thickness of boundary layer for each organization is dynamic. On the one hand, the features of knowledge, such as value, importance, specificity, and complexity, are constantly changing with the variety of time and the development the society, which changes the viscosity of knowledge and the thickness of boundary layer; on the other hand, there are knowledge transfers between inside the boundary layer and outside the boundary layer, knowledge creation and usefulness of knowledge declining in knowledge field, which can also make the thickness of boundary layer transformed.

4) The arguments about the rotation or not of knowledge field are explained through the boundary layer phenomenon of knowledge stickiness. It is believed that knowledge field is rotatable owing to the knowledge stickiness. Due to the fact that the stickiness of knowledge outside boundary layer is small, the rotation of knowledge field outside boundary layer is small and even negligible. However, inside the boundary layer, the rotation of knowledge field is large because of the large stickiness of knowledge. We can also maintain that, the knowledge field inside boundary layer is rotational, and the knowledge field outside boundary layer is irrotational.

References

[1] Ajmal M M, Koskinen K U 2008 Knowledge Transfer in Project-Based Organizations: An Organizational Culture Perspective *Project Management Journal* **39**(1) 7-15

[2] Ash J 1998 Managing knowledge gives power *Communication World* **15**(3) 23-6

[3] Kirchner S R 1997 Focus on: Database integration and management for call centers *Telemarketing* **16**(2) 22-4

[4] Argote L, Ingram P 2000 Knowledge Transfer A Basis for Competitive Advantage in Firms **82**(1)150-69

[5] Drucker P F 1969 The age of discontinuity Guidelines to our changing economy *New York: Harper & Row*

[6] Matusik S F, Hill C W L 1998 The Utilization of Contingent Work, Knowledge Creation, and Competitive Advantage *Academy of management review* **23**(4) 680-97

[7] Powell W, Snellman K 2004 The knowledge economy *Annual Review of Sociology* (30) 199-220

[8] Blundell R, Dearden L, Meghir C, Sianesi B 1999 Human capital investment The returns from education and training to the individual, the firm and the economy *Fiscal Studies* (20) 1-23

[9] Furman J L, Porter M E, Stern S 2002 The determinants of national innovative capacity *Research Policy* (31) 899-933

[10] Teece D J 1977 Technology transfer by multinational firms: the resource cost of transferring technological know-how *The Economic Journal* (87) 242-61

[11] Kiefer L, Frank J, Ruggerio E D, Dobbins M, Manuel D, Gully P, Mowat D 2005 Fostering Evidence-Based Decision-Making in Canada: Examining the Need for a Canadian Population and Public Health Evidence Centre and Research Network *Canadian Journal of Public Health* (96) I-1–I-19

[12] Snider K F, Nissen E 2003 Beyond the body of knowledge: A knowledge flow approach to project management theory and practice *Project Management Journal* **34**(2) 4-12

[13] Zander U 1991 Exploiting a technological edge: voluntary and involuntary dissemination of technology *International Business* **37**(2) 233-46

[14] Szulanski G 1996 Exploring internal stickiness: Impediments to the transfer of best practice within the firm *Strategic Management Journal* (17) 27–43

[15] Simonin B L 1999 Ambiguity and the process of knowledge transfer in strategic alliances *Strategic Management Journal* **20**(7) 595-623

[16] Cummings J L, Teng, Bing-Sheng 2003 Transferring R&D knowledge: the key factors affecting knowledge transfer success *Journal of Engineering and Technology Management* (20) 39-68

[17] Chua A L, Pan S L 2008 Knowledge transfer and organizational learning in IS off shore sourcing *Omega* **36**(2) 267-81

[18] Nguyen T,Burgess SA 2014 Case analysis of ICT for knowledge transfer in small businesses in Vietnam *International Journal of Information Management* **34**(3) 416-21

[19] Nonaka I, Takeuchi H 1995 The knowledge-creating company *London: Oxford University Press*

[20] Nonaka I, Toyama R 2003 The knowledge-creating theory revisited: knowledge creation as a synthesizing process *Knowledge Management Research & Practice* **1**(1) 2-10

[21] Nonaka I, Konno N 1998 The concept of 'ba': building a foundation for knowledge creation *California Management Review* **40**(3) 1-15

[22] Wang G H, Xing R, Tang L Y 2010 Research on the Integrative Industrial Innovation of Based on Knowledge Field *China Soft Science* (9) 96-107

[23] Nonaka I, Toyama R, Konno N 2000 SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation *Long Range Planning* **33**(1) 5-34
[24] Dang X H, Li L 2005 Research on Knowledge Creation Model in the Technology Innovation Cooperation Based on the Perspective of Knowledge Potential *China Soft Science* (11) 143-8
[25] Brown J S, Duguid P 1991 Organizational Learning and Communities-of-Practice Toward a Unified View of Working, Learning, and Innovation *Organization Science* **2**(1) 40-57
[26] Wenger E 2000 Communities of Practice and Social Learning Systems *Organization* b(2) 225-46
[27] Lesser E L, Storck J 2001 Communities of practice and organizational performance *IBM Systems Journal* **40**(4) 831-41
[28] Rynes S L, Bartunek J M, Daft R L 2001 Across the Great Divide: Knowledge Creation and Transfer Between Practitioners and Academics *Academy of management Journal* **44**(2) 340-55
[29] Liao S H, Hu T C 2007 Knowledge transfer and competitive advantage on environmental uncertainty: An empirical study of the Taiwan semiconductor industry *Technovation* **27**(6-7) 402-11
[30] Keil T, Maula M V J, Schildt H A, Zahra S A 2008 The effect of governance modes and relatedness of external business development activities on innovative performance *Strategic Management Journal* (29) 895-907
[31] Nonaka I 2009 Perspective - Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory *Organization science* **20**(3) 635-52
[32] Stasser G, Titus W 1987 Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion *Journal of Personality and Social Psychology* (53) 81-93
[33] Riusala K, Suutari V 2004 International knowledge transfer through expatriates: A qualitative analysis of international stickiness factors *Thunderbird International Business Review* **46**(6) 743-70
[34] Kang J, Rhee M, Kang K H 2010 Revisiting knowledge transfer: Effects of knowledge characteristics on organizational effort for knowledge transfer *Expert Systems with Applications* (37) 8155-60
[35] Szulanski G 1994 Intra-firm transfer of best practice project: Executive summary of the findings (Report) APQC
[36] Li C Y, Hsieh C T 2009 The impact of knowledge stickiness on knowledge transfer implementation, internalization, and satisfaction for multinational corporations *International Journal of Information Management* (29) 425-35
[37] Andersen P H 1999 Organizing international technological collaboration in subcontractor relationships: an investigation of the knowledge-stickiness problem *Research Policy* **28**(6) 625-42
[38] Von Hippel E 1994 Sticky information' and the locus of problem solving: Implications for innovative *Management Science* **40**(4) 429–39
[39] Ford D N, Sterman J D 1998 Expert knowledge elicitation to improve formal and mental modes *Systems Dynamics Review* (14) 309-40
[40] Chermack T J 2004 Improving decision-making with scenario planning *Futures* (36) 295-309
[41] Szulanski G 2000 The Process of Knowledge Transfer: A Diachronic Analysis of Stickiness *Organizational Behavior and Human Decision Processes* **82**(1) 9-27
[42] Centola D, Macy M 2007 Complex contagions and the weakness of long ties *American Journal of Sociology* (113) 702-34
[43] Reagans R, McEvily B 2003 Network structure and knowledge transfer: The effects of cohesion and range *Administrative Science Quarterly* (48) 240-67
[44] Szulanski G, Cappetta R, Jensen R J 2004 hen and How Trustworthiness Matters: Knowledge Transfer and the Moderating Effect of Causal Ambiguity *Organization Science* **15**(5) 600-13
[45] Tang S 1975 A Boundary Layer Theory-Part I: Laminated Composites in Plane Stress *Journal of Composite Materials* **9**(1) 33-41
[46] Young, Alec D 1989 Boundary layer *Washington, American Institute of Aeronautics and Astronautics*
[47] Schlichting H, Gersten K 2000 Boundary Layer Theory *Berlin: Heidelberg*
[48] Montazemi A, Pittaway J J, R, Saremi H Q, Wei Y 2012 Factors of stickiness in transfer to know-how between MNC units *The Journal of Strategic Information Systems* **21**(1) 31-57
[49] Li C Y 2012 Knowledge stickiness in the buyer-supplier knowledge transfer process: The moderating effects of learning capability and social embeddedness *Expert System with Applications* **39**(5) 5396-408
[50] Tang S, Levy A 1975 A Boundary Layer Theory-Part II: Extension of Laminated Finite Strip *Journal of Composite Materials* **9**(1) 42-52
[51] Chorin A J, Marsden J E 1979 A Mathematical Introduction to Fluid *New York, Springer US*
[52] Nonaka I, Toyama R, Konno N 2000 SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation *Long range planning* **33**(1) 5-34
[53] Li T T, Liu Z Z 2010 The Internal Dynamic Mechanism of Disciplines Integration of under the Perspective of "knowledge market *Meitan Higher Education* **28**(5) 32-5
[54] Anderson J D 1991 Fundamentals of aerodynamics *New York, MacGraw-Hill*
[55] Martin X, Salomon R 2003 Knowledge transfer capacity and its implications for the theory of the multinational corporation *Journal of International Business Studies* (34) 356-73
[56] Willcocks L, Hindle J, Feeny D, Lacity M 2004 It and Business Process Outsourcing: The Knowledge Potential *Information Systems Management* **21**(3)7-15
[57] Ernst D 2000 Inter-organizational knowledge outsourcing: what permits small taiwanese firms to compete in the computer industry *Asia Pacific Journal of Management* (17) 223-55
[58] Ernst D, Kim L 2002 Global production networks, knowledge diffusion, and local capability formation *Research Policy* **31**(8-9) 1417-29
[59] Marjolein C J, Verspagen Caniels B 2001 Barriers to Knowledge and Regional Convergence in all Evolutionary Model *Journal of Evolutionary Economies* (11) 307-29
[60] Attewell P 1992 Technology diffusion and organizational learning: the case of business computing *Organization Science* **3**(1) 1-19
[61] Fleming L 2001 Recombinant uncertainty in technological search *Management Science*(47) 117-32
[62] Nahapiet J, Ghoshal S 1998 Social capital, intellectual capital and the organizational advantage *Academy of Management Review* (23) 242-66
[63] Phelps C, Heidl R, Wadhwa A 2012 Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda *Journal of Management* **38**(4) 1115-66

Author

Shan Feng, born in January, 1980, Hubei, China

Current position, grades: a lecturer in School of Computer Science and Technology, Hubei Polytechnic University.
University studies: computer Science and Technology.
Scientific interest: computer science and technology, electronic communications
Publications: 10 papers.

Design of real-time clock based on ARM embedded system

Lijuan Liao*

College of Computer Science and Technology, Taiyuan University of Technology, 030024, China

Received 1 March 2014, www.cmnt.lv

Abstract

Real-time clock of ARM processor has disadvantages such as dependence on the processor, low interrupt level and unadjustable accuracy. Linux embedded system based on ARM processor is provided with independent real-time clock using X1227 real-time clock chip. In addition, the work presents transplant of uCLinux system in S3C2440A, procedure of system boot loader and frame of Linux character device driver, thus achieving design of RTC driver based on I2C protocol of uCLinux system.

Keywords: S3C2440A, RTC, X1227

1 Introduction

Real-time system (RTS) is that correct calculation depends on logical validity of procedure and resulting time. If time constraint of system is not satisfied, the system will go wrong. Then real-time clock is applied to RTS. Software counting can be used to realize function of real-time clock by the timer in ARM processor. However, it is not universal because of the disadvantages including data loss after power down, low interrupt level, unadjustable accuracy, etc. Real-time clock, with independent crystal oscillator and power supply system, can constantly run, thus providing reliable time for RTS and avoiding trouble. By using I2C bus, X1227 real-time clock chip has wide supply voltage range, high interrupt level and adjustable accuracy, thus promoting achievement and correct operation of RTS.

2 S3C2440A microprocessor

S3C2440A microprocessor of Samsung Corporation is applied to embedded system using a new bus structure—Advanced Micro controller Bus Architecture (AMBA). The CPU is a 16/32-bit ARM920T RISC processor designed by Advanced RISC Machines (ARM). ARM920T has a cache architecture consisting of MMU, AMBA BUS and Harvard. The architecture has independent 16KB instruction and 16KB data caches comprising eight-byte lines. S3C2440A is provided with a complete set of common system peripherals to reduce cost of the whole system and additional components [1].

The real-time clock integrated in S3C2440A is controlled by INT_RTC and INT_ADC, which have the lowest priority in all the 26 interrupt sources. Besides, this real-time clock is not universal, and cannot run with support of 3.3V voltage and microprocessor. Therefore, external RTC chip X1227 is applied.

3 RTC chip X1227

3.1 INTERNAL STRUCTURE OF X1227

X1227 is a real-time clock with functions of clock/calendar, CPU monitoring circuit and two-way query and alarm. Dual port clock and alarm register can ensure the accuracy of clock work, even during read and write operation. Clock/calendar can be provided with functions of controlling and reading by registers. Clock can precisely display the time by the unit of second, minute, hour, day, week, month and year using a 32.768 kHz crystal with low cost. There is a watchdog timer in X1227. If the watchdog timer overtimes, the reset pin $\overline{RESET}$ will be activated. X1227 has an input pin V_{BACK}, using a non-rechargeable battery as the back-up power. A 4K-bit EEPROM array in X1227, used as configuration data memory, will be safe and not affected when the main and standby power supplies lose efficiency. Figure 1 shows block diagram of X1227 consisting of control registers, SRAM, EEPROM, I2C serial interface decoder, status registers, standard crystal, watchdog timer, etc [2].

* ***Corresponding author*** e-mail: lijuantaiyuan@126.com

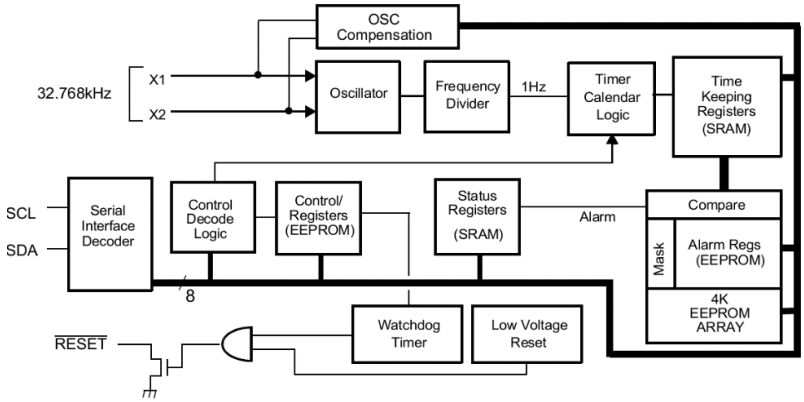


FIGURE 1 Block diagram of X1227

3.2 MAIN FUNCTIONS OF X1227

RTC of X1227 uses a external quartz crystal with a frequency of 32.768 kHz to keep accuracy of year, month, day, week, hour, minute and second. RTC, with century byte, can adjust leap years, the months less than 31 days, and the form of time by one bit. After format conversion, data of CCR in form of BCD code should be read and written by I2C interface in X1227. There are two alarm registers in X1227. If the time set by alarm register is the same with RTC, the corresponding position of SR register in CCR will be set to 1, thus realizing timing alarm. There are four KEEPROMs in X1227. They will be used to store key data of self-check program and system when system is power off. WatchDog can be set to control the shortest time of feeding the dog by writing WD1, WD0 of BL register in CCR. When the procedure comes into endless loop because of external disturbance, and timing register reaches maximum, WatchDog will emit a signal to reset SCM, thus the procedure is in control again.

X1227 is connected with S3C2440A using I2C interface (See Figure 2). Serial data pin (SDA), as a bi-pin connected with external 330Ω pull-up resister, is used to input and output data to the device. Input of serial clock pin (SCL) is used to time the whole data of input and output devices, thus providing serial clock signals of data transmission. VCC and VBACK inputs are received by the power control circuit of X1227. If VCC< VBACK-0.2V, then the power will be switched to VBACK by the power control circuit; if VCC＞VBACK, then it will return to VCC.

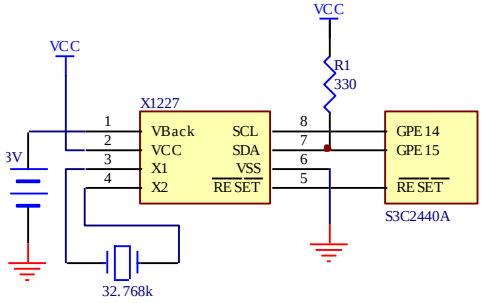


FIGURE 2 X1227 connected with S3C2440A

4 Transplant of uclinux system to S3C2440A

4.1 UCLINUX SYSTEM

Uclinux (micro-control linux), as a Linux system in micro-control field, is a major product of Lineo Corporation and a model for embedded Linux of open source. Aiming at the target processor of embedded system without MMU (Memory Management Unit), uclinux is designed and successfully transplanted to platforms. The users of uclinux operation system can apply the whole Linux API functions based on GNU general license. After clipping and optimization, uclinux forms an embedded Linux with high optimization and compact codes. Uclinux has advantages such as small size, stability, good transplant, excellent network function, perfect support to all the file systems and abundant API functions. Therefore, uclinux has good compatibility with Linux. API functions of uclinux, except fork(), are the same as those of standard Linux [3].

4.2 ESTABLISHMENT OF CROSS-COMPILING ENVIRONMENT

Cross-compiling environment used for target machine is established in PC because of limited storage of common embedded system. The executable files, got from compilation, connection and location of procedure in PC, are loaded to the target machine through serial port. Cross-compiling environment is required to establish kernel header files, binutils, bootstrap gcc, glibc, etc. The work applies the method as follows. Firstly, GCC is installed in PC aiming at compiler of ARM (arm-elf-gcc). Secondly, the configured cross tool chain "arm-elf-tools-20040427.sh" provided in "www.uclinux.org" is convenient to use. In the root directory, after adding the executed authority by running command "#chmod 755 arm-elf-tool-20040427.sh", the script can be operated to install cross tool chain in correct position. At last, the address "/user/local/bin" is checked to identify whether the compiling environment file started with "arm-elf-" is existed or not. If so, then cross-compiling environment

will be successfully installed. The program above is actually procedure of establishing uclinux library file.

5 Driving X1227 in unlinux system

5.1 SYSTEM BOOTSTRAP ROUTINE

BOOT, as the first running line of code in the chip after power on, is applied to initialize running environment of hardware and software in system for application program operation. System bootstrap routine boots up operation system and hands the control power over to operation system core. This operation depends on the type of CPU core and the resource applied to develop embedded system software in CPU chip. Figure 3 shows the flow of system boot loader based on this chip and application program.

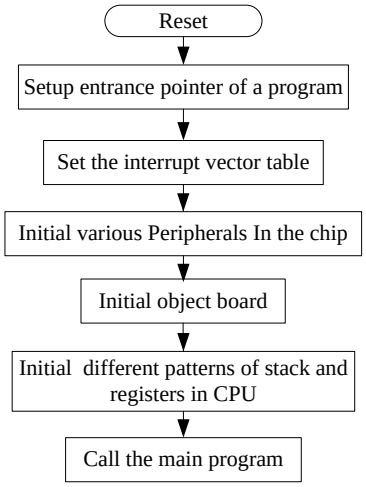


FIGURE 3 Flow of the system boot loader

5.2 DRIVER FRAMEWORK OF LINUX CHARACTER DEVICE

X1227 as well as alarm clock is set and read by compiling software. As a hardware device of system running in Linux, X1227 is operated by drivers in Linux operation system. The devices in Linux include character and block. X1227 is a character device. Programs of user mode can run character devices like common files. Therefore, the driver of character device type should at least realize system call functions including open(), release(), read() and write(). In Linux kernel, X1227 character driver is designed based on the framework as follows [4].

```
struct file_operations {
ssize_t(*read)(struct file *, char *, size_t, loff_t *);
sszie_t(* write)(struct file *, const char *,size_t, loff_t *);
int(* ioct1)(struct inode *, struct file *, unsigned int,Unsigned long);
int(* open)(struct inode *, struct file *);
int(* release)(struct inode *, struct file*);
...};
```

The driver should realize the following functions:

1) open( )

Initialize the devices supported by the driver.

2) release( )

Close the devices supported by the driver after use.

3) read( )

Read data from character devices by application program or Linux.

4) write( )

Write multiple byte data to character devices.

5) ioctl( )

Provide application program with some special operation which cannot be easily realized by read() and write() methods.

The above methods derive the functions as follows: initialization, load and release of hardware devices; management of equipment including setting real-time parameter and providing unified operation interface for devices; reading data of device files obtained from (or responding to) application program; detecting or processing device errors.

5.3 DESIGN OF X1227 DRIVER

X1227 can communicate with ARM by I2C bus. Using signal wires SCL and SDA, I2C bus can achieve data interaction between devices, thus simplifying occupation of hardware resources and PCB wiring space. I2C bus realizes data transmission and command control by timing signals including start, stop and ACK. If SCL is at the high level and SDA switches from high to low level, I2C bus will start sending signals; if SCL is at the high level and SDA switches from low to high level, it will stop sending signals; if the device receives 8-bit data, it will send ACK to the sender. Then the sender will set SDA at the high level, and the receiver will set SDA at the low level during responsive clock pulse, thus achieving signal response. Figure 4 shows start condition, stop condition and ACK of I2C bus [5].

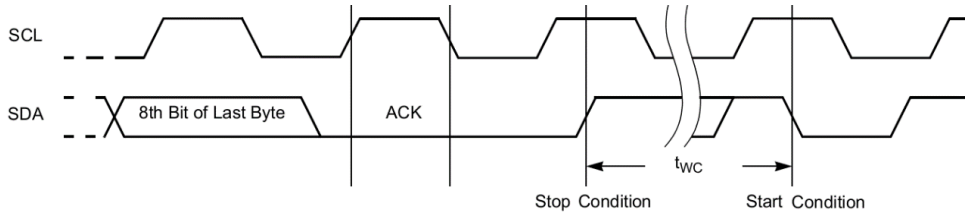


FIGURE 4 I2C bus

X1227 has slave device addresses including 1010 (access to 4KB EEPROM) and 1101 (access to CCR) used to control RTC and WatchDog. Figure 5 shows operation of a byte access including a slave address byte, an address

word (16 bits) to be accessed and an 8-bit operand. The last bit of the slave address byte determines this operation. If the last bit is 1, the reading will be conducted; if 0, writing will be operated. The communication process of I2C bus is simulated using I/O interface. Firstly, slave address of I2C bus is defined to read and write devices by program.

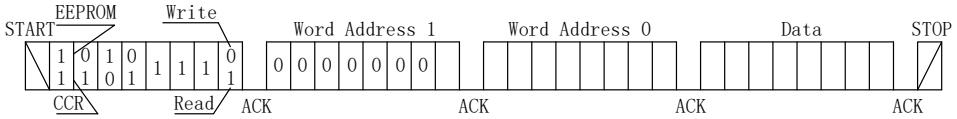


FIGURE5 Access format of X1227

 # define RTC_I2C_READ 0xaf //Read the address by I2C bus
 # define RTC_I2C_WRITE 0xae // Write the address by I2C bus

RTC driver module applies character devices achieved by "struct file_operation" framework structure:

```
static struct file_operations x1227_fops = {
.owner = THIS_MODULE,
.ioctl = x1227_ioctl,
.open = x1227_open,
.release = x1227_release,
};
```

In this structure, the whole functions are accomplished by upper call function ioct1 [6]:

int x1227_ioctl(struct inode *inode, struct file *filp, unsigned int cmd, unsigned long arg)

In the function, reading and writing transmission commands of cmd are "RTC_RD_TIME" and "RTC_SET_TIME"; arg is the structure pointer directing to "struct rtc_time" for time storage. At first, the function reads transmission command cmd to determine whether the operation is writing or reading. If the result is reading, then Function rtc_read() will be called to read data from the register and keep them in arg; else if writing, then Function rtc_write() will write data to the register.

Function rtc_read() can be used to read the time and date in RTC register, achieved by the following function:

```
rtc_r ead (struct file * filep, cha r * buffer, size_t length)
{ ...
Init_Iic( ) ; / / Initialize I2C bus
RdBy _Iic( 0xde, 0x30, & (rdata) ) ; / / Read second register
dbu f[0] = rdata ;
RdBy_Iic(0xd e,0x31 , &(rdata)) ; / / Read minute register
dbuf[1] = rdata;
...
RdBy_Iic(0xd e,0x35,&(rdata)) ; / / Read year register
dbuf[5] = rdata;
copy_to_user (buffer,dbu, f6) ; / / Change data from kernel to user mode
...}
```

Function rtc_write() is mainly used to set time and date.

The compiled program can be re-compiled using cross-compiling tool of uclinux for operation in S3C2440A. In Windows, uclinux kernel, root file system and executed file are read and written through serial port of hyperterminal. Basic settings are as follows: baud rate is 115200; data bit number 8; no parity checking; stop bit 1; no data flow control.

6 Conclusions

RTC is one of the typical applications of I2C in embeded products. Even if there is no special I2C interface to control processor chip, I2C bus time sequence will be simulated by software to achieve communication between chips with two I/O interface pins. Therefore, RTC is suitable for embedded devices. The work aims at designing driver of X1227 clock chip to achieve real-time clock combined with RTC device driver model in uclinux. The design realizes real time of system by replacing the integrated clock in CPU. The designed clock has a good performance in running and function, thus proving practicability and stability of this driver design.

References

[1] Samsung Electronics Co Ltd 2004 S3C2440A 32-BIT RISC Microprocessor User's Manual *Korea Revision 0.12* Samsung Electronics Co Ltd
[2] Intersil Corporation 2005 2-Wire RTC Real Time Clock/Calendar/CPU Supervisor with EEPROM X1227 Datasheet [EB/OL] http://www.intersil.com/
[3] Corbet J, Kroah-Hartman G, Rubini A 2005 Linux Device Drivers 3rd Edition *O'Reilly* 49-53
[4] Bovet D P 2002 Understanding the Linux Kernel 2nd Edition *O'Reilly*
[5] Philips Semiconductors 2003 AN10216-01 I2C Manual 13-7
[6] Nilsson J, Rytterlund D 2000 "Modular Scheduling in Real-Time Linux" *MSc Thesis Department of Computer Engineering* Idt Malardalen University

Author

Lijuan Liao, born in September, 1963, Hengyang, Hunan Province, China

Current position, grades: associate professor in College of Computer Science and Technology, Taiyuan University of Technology, China.
Scientific interests: computer control and embedded technology.
Publications: 20 papers.

Spatial and temporal mining method using GPS data

Xiaolei Li*

School of Electronic and Information, Ningbo Dahongying University, Ningbo, Zhejiang, 315175, China

Received 1 March 2014, www.cmnt.lv

Abstract

Geographic information has spawned many novel Web applications where global positioning system (GPS) plays important roles in bridging the applications and end users. Learning knowledge from users' raw GPS data can provide rich context information for both geographic and mobile applications. However, so far, raw GPS data are still used directly without much understanding. Spatial-temporal data analysis plays an important role in many applications, including transportation infrastructure, border security and inland security. To analyse the moving patterns of vehicles on a road network, a measure for determining the similarity of vehicle trajectories with respect to space and time has to be defined. Although previous research has addressed the trajectory similarity problem, most of the studies focus on Euclidian distance instead of network distance. This paper deals with the variations in applying a spatial-temporal similarity measure with given Points of Interest (POI) and Time of Interest (TOI), treating spatial similarity as a combination of structural and sequence similarities that is evaluated using the techniques of dynamic programming. The similarity set thus formed will be used by the remote database to broadcast trigger-based messages to participating vehicles in a neighbourhood for future route- and information-sharing activities. The performance of the scheme is evaluated using experiments on standard real-life data.

Keywords: K-means, clustering algorithm, error rate, iteration, reduction, stable

1 Introduction

The increasing pervasiveness of location-acquisition technologies, such as GPS and GSM network, is leading to the large collection of spatial-temporal datasets. Such datasets have supported a variety of novel Web applications, in which locality and mobility usually connect to one another closely. For instance, people can tag user-generated contents like photos with locations; trace their outdoor mobility; and use location-based services. Recently, a branch of GPS-track-sharing applications using Web maps appeared on the Internet. In this category of Web applications, people can record their travel routes using a GPS-equipped device and then share travel experiences among each other by publishing these GPS tracks in a Web community. GPS-track-sharing offer a more fancy and interactive approach than text-based articles to better express people's travel experiences, which provide users with valuable references when planning a travel itinerary.

However, so far, these applications require people either to manually label their own trajectories or to use raw GPS data such as GPS coordinates and timestamps without much understanding. Neither of these methods is optimal to the development of such applications. Actually, users become easily frustrated by the additional data labelling effort, and then give up uploading their data. Moreover, people intend to understand an individual's mobility, and learn information about user behaviours as well as user intentions behind the raw data. Being an important kind of human behaviour, transportation modes, such as walking, driving, and taking a bus, can enrich their mobility with knowledge and provide pervasive computing systems with more contexts.

In recent years, information technology has significantly penetrated surface transportation. The transportation environment is embedded with various mobile sensors, including on-board GPS receivers, sensors mounted on public transportation vehicles and pedestrian cell phones. These sensors continuously generate spatial-temporal data and enable applications such as vehicle tracking and environmental monitoring. Studying people and vehicle movements within a certain road network is both interesting and useful, especially if it can be used to understand, manage and predict the traffic flows. By studying the massive flow of traffic data as a trajectory, the traffic flow can be monitored and traffic-related patterns can be discovered. The development of Intelligent Transportation Systems (ITS) allows better monitoring and traffic control to optimize traffic flow.

Advances in social networking and location-based services are increasingly creating new, sophisticated mechanisms that can foster a seamless integration of information among travellers to provide alternatives and support sustainable economic and social policies. All of these issues raise challenges to the dissemination of information, such as safety, traffic, entertainment, service and content, in a vehicular network. In all of the above cases, a major challenge arises from the fact that the relevance of information changes with time and with the location of the vehicle. The quality of data is measured by their spatial-temporal relevance, which indicates the probability that the resource will still be available when the vehicle reaches it. For example, in the case of the

***Corresponding author** e-mail: xiaoleilizj@yeah.net

dissemination of service information (e.g., the availability of parking spaces), the temporal and spatial delivery of information must be carefully considered. In the moving object database, which will have the complete history of the vehicle's trajectory, a spatial-temporal trajectory similarity search process will identify vehicles with very similar movement patterns with respect to space and time. Based on such a similarity set, useful information can be disseminated among neighbouring vehicles in the same similarity set containing extensive applications (e.g., knowing customer preferences, re-routing and scheduling of further travel, and so on). One advantage of this system is that because the broadcast message came from an official central database, it will be more authentic, and thus the nodes can be confident in the messages they have received. However, when internal nodes share information based on the central broad-cast, the system has to go through a confident management process to ensure the accurate dissemination of information. This paper relates how the spatial-temporal similarity measure of vehicle trajectory will help to better disseminate information in vehicular networks for data mining applications.

In this article, we aim to automatically infer transportation modes, including driving, walking, taking a bus, and riding a bicycle, from raw GPS logs based on supervised learning. It is a step toward recognizing human behaviour and understanding user mobility for pervasive computing systems. Also, it is a step toward improving local/mobile applications on the Web and enhancing the connection between mobility and locality by mining knowledge from raw GPS data with minimal user efforts. The contributions of this work lie in the following three areas.

First, we propose a change point-based segmentation method. This method aims to partition each GPS trajectory into separate segments of different transportation modes, while maintaining a segment of one mode as long as possible. In addition, this segmentation method is capable of enhancing the reliability of our methodology facing the variable traffic conditions.

Second, from each segment, we identify a set of sophisticated features, such as direction change rate, velocity change rate, and stop rate. These features have few correlations with the velocity, hence are not affected by differing traffic conditions. These set of features can also be extended to other pervasive computing systems aiming to recognize human behaviour and understand user mobility.

Third, we conduct a graph-based post processing algorithm to further improve the inference performance. In this algorithm, we mine the common sense constraints of the real world and typical user behaviours on a location from user-generated GPS logs. Therefore, we are able to leverage this location-constrained knowledge as probabilistic cues, while maintaining our methodology being in-dependent of an additional database of road networks or points of interests.

Overall, the advantages of our method over the related works include two parts.

1) Our method is independent of other sensor data like GSM signal and heart rate, and map information, for example, road networks and bus stops, etc. Thus, it is generic to be deployed in a broad range of Web applications.

2) The model learned from the dataset of some users can be applied to infer GPS data from others; that is, it is not a user-specific model.

2 Related Works

In conventional approaches, the location information of a moving vehicle is expressed as a geometric coordinate (x, y) in two-dimensional space. However, Lee et al. (2007) express location information using both the hierarchical administrative district and road network in one-dimensional space, more accurately fitting the real world. For instance, if a moving object is in a building at latitude of 125.58 and longitude of 37.34, then it can be expressed as a set of fields according to an administrative district, such as the city, road name, or block (e.g., Seoul, Main Road, 165th block). The dimension reduction of spatial-temporal data management (Abraham and SojanLal, 2008) discusses two algorithms for binary encoding pro-cess: one for location encoding and one for converting a position represented as geometric coordinate into an equivalent binary string. Because the proposed similarity scheme is based on this encoding method, these basic algorithms are briefly discussed below.

Many recent studies regarding data management in vehicular networks and related topics discuss the VANET, while very few focus on the processing of spatial-temporal information. In the case of traffic and service information, Xu et al. (2004) suggest aging information with time and distance. Information is opportunistically pulled from neighbouring vehicles as a target vehicle moves in a given area. Such information needs to have time and location stamps. Because either the vehicle moves farther away from where the information is relevant or time elapses, the information is aged and eventually purged. This enables vehicles to maintain up-to-date information without taxing memory and other resources. Xu et al. and Wu (2004) discuss the propagation and survival of information in time and space. Because of the spatial-temporal relevance of information, a piece of information tended not to propagate beyond a specific boundary. In time, a given piece of information would propagate very quickly until it reached a maximum number of copies, at which point it would also rapidly decline. Wolfson et al. (2005) discuss the dissemination of resource discovery (e.g., parking spaces) using spatial-temporal information. Each disseminated report represents information about a spatial-temporal event, such as the availability of a parking slot or a cab request. Reports are disseminated by a peer-to-peer broadcast paradigm in which an object periodically broadcasts the reports it carries to encountered objects. The

authors evaluated the value of resource information in terms of how much time is saved when using the information to discover a resource.

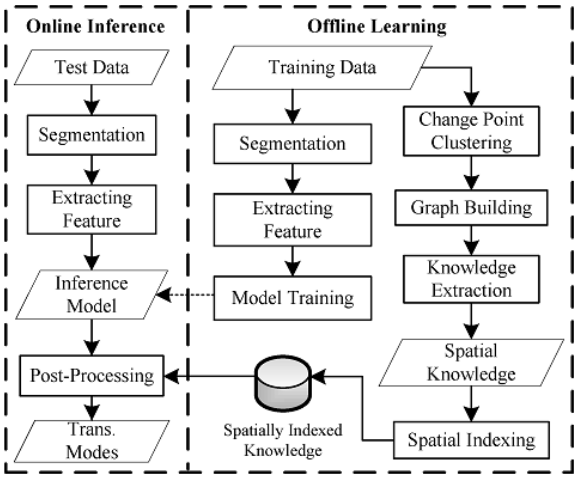


FIGURE 1 Architecture of our approach

The efficiency and reliability of information dissemination will be impacted by how vehicles are clustered when they form and leave groups, how apart they are under different traffic conditions, the density of vehicles and so on. Naumov et al. (2006) simulated realistic vehicular traces, which form trajectories. The authors demonstrated that the realistic traces are noticeably different than those in widely used mobility models. Szczurek et al. (2010) provide another promising approach to the dissemination of spatial-temporal information, such as the current traffic condition of a road segment or the availability of a parking space. Ranking becomes critical in this situation by enabling the most important information to be transmitted under the bandwidth constraint. Delot et al. (2010) proposed a vehicle-to-vehicle communication system for data sharing in vehicular networks using the concept of Encounter Probability to share information. The objective of this paper is to facilitate the dissemination of information between vehicles when they meet each other, taking into account the relevance of the data to the drivers and the type of event (e.g., available parking spaces, obstacles in the road and information relative to the coordination of vehicles in emergency situations) in the network.

Recently, some works (e.g., Delot et al., 2011) have addressed the problem of processing queries in a highly dynamic vehicular network in order to share information between vehicles. Queries interesting in pervasive and mobile computing environments (Vargas-Solar et al., 2010) are usually location-dependent (Ilarri et al., 2010). In vehicular networks, vehicles usually receive data from its neighbours and decides whether they are relevant enough to be stored in a local data cache. Then, a query processor can use the data to retrieve relevant data for the driver. Delot et al. (2011)discuss the challenges and possible solutions of multi-scale query processing techniques to exploit, at the mobile device's level, different access modes (e.g., wireless networks, push, pull) and various data sources (e.g., locally cached data, data stored by vehicles nearby and remote Web services) to provide users with results for their queries. Wolfson and Xu (2010) discuss various research issues related to the methods of routing, navigation, and tracking in transportation networks (the spatial-temporal database) that may involve multiple modes (e.g., train, bus, private car and bicycle).

3 The proposed spatial-temporal mining method

The algorithm for mapping locations in binary code is a recursive procedure, which will successively divide the entire region into sub-regions, map each district into a two-dimensional space and assign a binary string to each district. The total conversion time increases almost linearly with the number of two-dimensional locations, and more than seven million locations can be pro-cessed per second. The characteristics of the binary encoding scheme discussed above, which is used as a baseline in the proposed similarity scheme, are the following: It will be easy to determine the lowest common administrative district by extracting the longest common prefix of a given set of binary strings, and a district containing a set of lower districts can be represented by the range of binary string (i.e., county ''A'' will be represented by the range [00000, 00111]). These advantages make it easy to address the whole country, whole district or a single road by identifying the common prefix in the binary string representing the location of the object. For example, an object's location on the road may be encoded as 001010000110010, where first two bits denote the country, the next four bits denote the district, the following five bits denote the road and the final four bits denote the relative location on the road. Thus, if the first two bits in all of the locations are 00, they all belong to the same country.

3.1 ARCHITECTURE OF OUR APPROACH

As shown in Figure 1, the architecture of our approach includes two parts, offline learning and online inference. In the offline learning section, on one hand, we first partition GPS trajectories into segments based on change points and extract features from each segment. Then, the features and corresponding ground truths are employed to train a classification model for online inference. On the other hand, using a density-based clustering algorithm, we group the change points detected from all users' GPS logs into clusters. Subsequently, a graph based on these clusters and user-generated GPS trajectories is built.

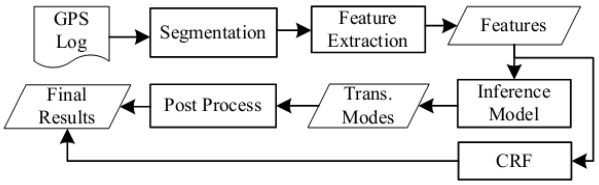


FIGURE 2 Procedure of inferring transportation mode

From this graph we can mine some location-constrained knowledge, such as the probability distribution of different transportation modes on each edge. The knowledge can be employed as probabilistic cues to improve the inference performance in the post processing. In addition, a spatial index is built over the detected spatial knowledge to enhance the processing efficiency. In the online operation, when a GPS trajectory comes, like the offline training process, we first partition it into segments and extract the same features from each segment. Second, given the features, the generative inference model will predict the transportation mode of each segment in a probabilistic manner. Third, given the probabilities of a segment being different transportation modes, a post-processing algorithm is used to improve the inference accuracy by leveraging the spatial knowledge mined from the training data. Finally, the transportation mode with maximum posterior probability will be selected as the ultimate result.

3.2 INFERENCE STRATEGY

As shown in Figure 2, when a GPS Log file comes, first, we divide the GPS track into trips and then partition each trip into segments by change points. Then, we extract the features from each segment and send these features to the inference model. Two alternative ways are considered when we attempt to learn a user's transportation mode. In one way, we regard the segments of GPS tracks as independent instances. General classifiers like Decision Tree are employed to perform inference. After the inference, a post-processing, which takes the transition probability between different transportation modes into account, is implemented to improve the prediction accuracy. In the other way, GPS data are deemed as a kind of sequential data. Conditional random field (CRF) [13], a framework for building probabilistic models to segment and label sequence data, is leveraged to perform the inference. Since the conditional probabilities between different transportation modes have been considered in the CRF graphical model, in this way, we do not take the post-processing. In the inference, the mode of transportation can take four different values including Bike, Bus, Car and Walk. At the same time, we do not discriminate driving private car from taking taxi. Both of them are deemed as Car.

3.3 SPATIAL KNOWLEDGE EXTRACTION

Figure 3 illustrates the four steps toward mining spatial knowledge from users' GPS logs. The knowledge includes a change point-based graph and the probability distribution on each edge of the graph. First, given GPS logs with labeled ground truths, we can get the special points consisting of change points and the start/end points of each GPS trajectory. These special points were subsequently grouped into several nodes (clusters) using a density-based clustering algorithm. The reasons we prefer to use density-based clustering instead of agglomerative methods, such as K-Means, lie in two aspects.

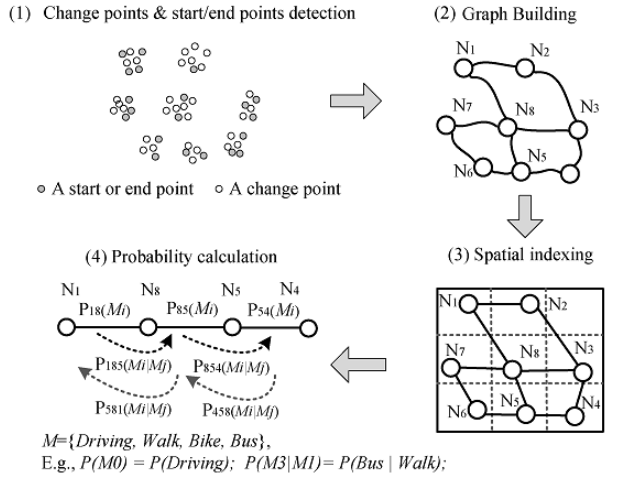


FIGURE 3 Mining spatial knowledge from GPS logs

First, a density-based approach is capable of detecting clusters with irregular structures, which may stand for bus stops or parking places. Second, it can discover popular places where most people change their transportation modes while removing sparse change points representing places with low access frequency.

Second, with the GPS trajectories from multiple users' GPS logs, we can construct an undirected graph. In such a graph, a node represents a cluster of the special points mentioned above, and an edge denotes users' transitions between two nodes. Here, we do not differentiate various trajectories with similar start/end points; that is, all the trajectories passing two graph nodes are regarded as similar trajectories.

Third, we build a grid-based spatial index over the graph to improve the efficiency of accessing the information of each node and each edge. The space covered by the graph is partitioned into many disjoint grids. Then, the graph nodes falling in different grids are associated with the grids. Therefore, when a new GPS trajectory comes to be inferred, we only need to match the special points detected from the trajectory against the graph nodes pertaining to the grids where these special points falling in. Of course, this step is optional unless the scale of the GPS dataset is quite large.

Fourth, we are able to calculate the probability distribution of different transportation modes on each edge. For instance, as depicted in the fourth step of stands for the likelihood of the event that people take buses on the edge between node 1 and node 8. Further, the conditional probability be-tween different transportation modes can also be computed based on the graph, for example, (Bus |Walk) represents the transition probability from Walk to Bus between edge 18 and edge 85. In other words, it denotes the likelihood of the event that a user walks from node 8 to node 5 based on the observed occurrence that the user takes a bus from node 1 to node 8.

Such previously mentioned knowledge is promising in improving the inference accuracy due to the following reasons:

1) This implies people's typical behaviors among different places. The clusters of change points represent the places many people change their transportation modes. Usually, these places could be bus stops, parking lots, and railway stations. We can take into account user behavior among these nodes as probabilistic cues when we infer other trajectories passing these two nodes.

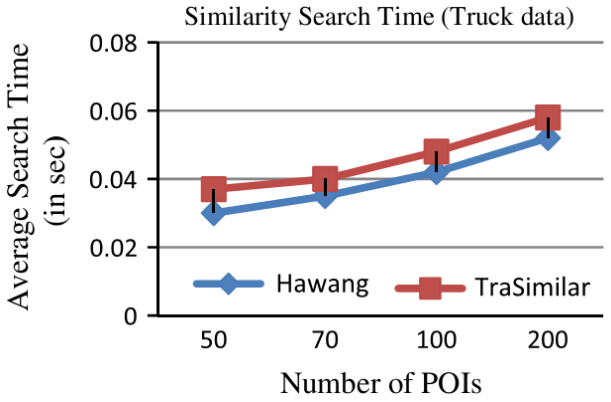


FIGURE 4 Comparison of the average search times for the truck dataset

Naturally, for example, if most users take a bus between two nodes, we can suggest that the two nodes could be bus stops, and the edge between them could have a very high probability of being a bus line.

2) The probability on each edge implies constraints of the real world. For instance, buses only take passengers on at bus stops, cars are left in parking lots, and cars and buses only travel on streets, etc. This knowledge mined from multiple users' GPS logs take advantages of the location constraints while keeping our method independent of an additional map database. In such a way, it is not necessary to match each GPS trajectory against the road network. Mean-while, we do not need to maintain a database of bus stops, railway stations and parking lots.

4 Experimental Results

This experimental setup focuses on vehicle trajectories in which two common characteristics are assumed. First, vehicle trajectories typically follow road networks (i.e., they are not free movements in two-dimensional space). Second, vehicle positions are measured at a reasonably good temporal resolution (e.g., one GPS measurement per minute). Many existing vehicle trajectory datasets satisfy the above resolution requirement, such as the truck data set and the INFATI data derived from the INFATI Project (Jensen http://www.infati.dk/uk). Both of these datasets were used in this research, resulting in one GPS measurement for every 30 s.

In this study's experiments, the performance and scalability of similarity search techniques have been evaluated. All experiments have been conducted on a Intel Core 2 Duo machine running Windows XP with 2 GB of RAM and a 320 GB.

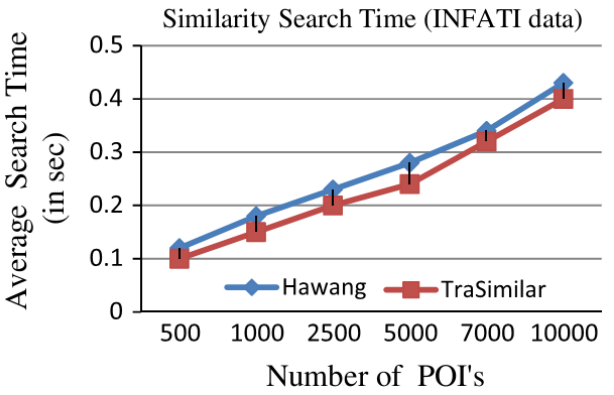


FIGURE 5 Comparison of the average search times for the INFATI dataset

SATA2 16-MB hard drive. The performance of the similarity search technique of the proposed algorithm is measured by comparing the average search time with that of the Hawang algorithm for different numbers of POIs, ranging from 50 to 300 for small datasets (truck data) and from 500 to 3000 for large datasets (INFATI data). Scalability is measured in terms of how the search time grows with respect to an increase in the number of POIs. The truck dataset consists of 50 trucks delivering concrete to several construction places around the Athens metropolitan area in Greece for 33 different days. The data set has 276 trajectories and 112,203 GPS points (approximately one GPS measurement for every 30 s in most trajectories). This study's approach can be used to analyse other vehicle trajectory datasets with similar temporal resolutions, such as the Milan data set. The structure of each record in truck data set is as follows: {obj-id, traj-id, date (dd/mm/yyyy), time (hh:mm:ss), lat, lon, x, y}, where (lat, lon) is in the WGS84 reference system and (x, y) is in the GGRS87 reference system.

As shown in Figure 4, the experimental results confirm that the average search time of the proposed spatial-temporal similarity algorithm increases linearly with an increase in the number of POIs, supporting the fact that the algorithm is scalable. As both of the graphs are similar in terms of average search time, the search performance of TraSimilar is con-firmed to be similar to that of the earlier algorithm even when the proposed spatial similarity extends the structural and sequence similarity measures.

The similarity search performance was tested with a large GPS log dataset (INFATI) of car movements to consider a large number of input trajectories and a POI number greater than 300. The INFATI data are derived from the INFATI Project, which is an intelligent speed-adaptation project carried out by a team of researchers at the Department of Development and Planning, Aalborg University. For each car that delivered data, the INFATI data contains one file with GPS log data.. As shown in Figure 5, the experimental results for a large number of POIs confirm the scalability and search performance of the

algorithm in agreement with the results for a smaller number of POIs. The results of the experiment also shows that the TraSimilar search performance is better than that of the Hawang method when there is a larger number of POIs. The authors claim that this is due to the proposed method's advantage of reduced dimensions when representing location and trajectory data in the binary-encoded scheme.

5 Conclusion

Our approach consists of three parts: a change point based segmentation method, an inference model and a post-processing algorithm based on conditional probability. The similarity set thus formed will be used by the remote database to broadcast trigger-based messages to participating vehicles in a neighbourhood for future route and information-sharing activities. The performance of the scheme is evaluated using experiments on standard real-life data.

In this paper, by using knowledge mined from raw GPS data, we aim to improve geographic applications on the Web and build closer connections between locality and mobility. The knowledge we gained as well as the connections enable more novel applications and improve user experience in a variety of tasks. An approach has been proposed to automatically learn transportation mode from raw GPS data. The inferred transportation mode can help Web users more deeply understand their own experience while better sharing other users' knowledge. It also enables context-aware computing based on a user's present transportation mode and creation of innovative user interface for Web users. The proposed approach is independent of other information and devices. Therefore, it is universal to be performed in both mobile devices and servers.

References

[1] Tietbohl Palma A, Bogorny V, Kuijpers B, Otavio Alavares L 2008 A clustering based approach for discovering interesting places in trajectories *Proceedings of SAC'08* 429-32

[2] Andrienko G, Andrienko N 2008 Spatio-temporal aggregation for visual analysis of movements *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST '08)* 51-8

[3] Chang J W, Bista R, Kim Y C, Kim Y K 2007 Spatio-temporal Similarity measure algorithm for moving objects on spatial networks *Proceedings of ICCSA* 1165-78.

[4] Chen, L, Özsu M T, Oria V 2005 Robust and fast similarity search for moving object trajectories *Proceedings of ACM SIGMOD* 491-502

[5] Cuevas-Vicenttin V 2008 Towards multi-scale query processing *Proceedings of IEEE 24th International Conference on Data Engineering (ICDE'08) Workshops* 137-44

[6] Delot T, Cenerario N, Ilarris S 2010 Vehicular event sharing with a mobile peer-to-peer architecture *Transportation Research Part C* **18**(4) 584-98

[7] Delot T, Ilarri S, Thilliez M, Vargas-Solar G, Lecomte S 2011 Multi-scale query processing in vehicular networks *Journal of Ambient Intelligence and Humanized Computing* 2(3) 213-26

[8] Gunduz S, Özsu M T 2003 A web page prediction model based on click-stream tree representation of user behavior *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 535-40

[9] Hwang J R, Kang H Y, Li K J 2005 Spatio-temporal Similarity Analysis between Trajectories on Road Networks *Perspectives in Conceptual Modeling Lecture Notes in Computer Science* **3770** 280-9

[10] Hwang J R, Kang H Y, Li K J 2006 Searching for Similar Trajectories on Road Networks using Spatio-Temporal Similarity *Advances in Databases and Information Systems Lecture Notes in Computer Science* **4152** Springer-Verlag LNCS 282-95

[11] Ilarri S, Mena E, Illarramendi A 2010 Location-dependent query processing: where we are and where we are heading *ACM Computing Survey* **42**(3) 1-73

[12] Leontiadis I, Mascolo C 2007 Opportunistic spatio-temporal dissemination system for vehicular networks *Proceedings of the 1st International Mobisys Workshop on Mobile Opportunistic Networking (MobiOpp'07)* 39-46

[13] Salhi I, Cherif M O, Senouci S M 2009 A New Architecture for Data Collection in Vehicular Networks *Proceedings of IEEE/ICC'09* Dresden DE 1-6

[14] Li Q, Zheng Y, Xie X, Chen Y, Liu W, Ma W Y 2008 Mining user similarity based on location history *Proceedings of GIS 08* 1-10

[15] Naumov V, Baumann R, Gross T 2006 An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces *Proceedings of ACM Mobihoc06* 108-19

[16] Orenstein J A, Merrett T H 1984 A class of data structures for associative searching *Proceedings of 3rd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems* Waterloo Ontario 181-90

[17] Krishnamurthy P 2008 Information dissemination and information assurance in vehicular networks: a survey *Proceedings of ACM Schools Conference 08*

[18] Abraham S, Sojan-Lal P 2008 Trigger based security alarming scheme for moving objects on road networks *Proceedings of IEEE International Conference on Intelligence and Security Informatics (ISI 08) LNCS* Springer-Verlag Berlin 92-101

[19] Abraham S, Sojan-Lal P 2010 Trajectory similarity of network constrained moving objects and applications to traffic security *Proceedings of Pacific Asia International Workshop on Security Informatics (PAISI 10)* Springer-Verlag Berlin LNCS 31-43

[20] Sakurai Y, Yoshikawa M, Faloutsos C 2005 FTW: Fast Similarity Search under the Time Warping Distance *PODS 05* 326-37

[21] Lee S, Park S, Kim W 2007 An efficient location encoding method for moving objects using hierarchical administrative district and road network *Information Sciences* **177** 832-43

[22] Shim C B, Chang J W 2003 Similar sub-trajectory retrieval for moving objects in spatio-temporal databases *Proceedings of the 7th EECADIS* 308-22

[23] Szczurek P, Xu B, Lin J, Wolfson O 2010 Spatio-temporal information ranking in VANET applications *Inaugural Issue of the International Journal of Next-Generation Computing (IJNGC)* **1**(1) 52

[24] Theodoridis, 2007. R-Tree Portal. <http://www.rtreeportal.org> (viewed 28.02.07)

[25] Tiakas E, Papadopoulos A N, Nanopoulos A, Manolopoulos Y 2006 Trajectory similarity search in spatial networks *Proceedings of the 10th IDEAS* 185-92

[26] kas E, Papadopoulos A N, Nanopoulos A, Manolopoulos Y, Stojanovic D, Djordjevic-Kajan S 2008 Searching for similar trajectories in spatial networks *Journal of Systems and Software* **82**(5) 772-88

[27] Vargas-Solar G, Ibrahim N, Collet C, Adiba M, Petit J-M, Delot T 2010 Pervasive Computing and Communications Design and Deployment: Technologies, Trends and Applications *IGI Global*

[28] Abraham S, Sojan-Lal P 2012 *Spatio-temporal similarity of network-constrained moving object trajectories using sequence alignment of travel locations* Transportation Research Part C **23** 109-23

[29] Cherfaoui V, Denouex T, Cherfi Z L 2008 Distributed data fusion: application to confidence management in vehicular networks *Proceedings 11th International Conference of Information Fusion 08* 1-8
[30] Vlachos M, Gunopulos D, Kollios D 2002 Robust similarity measures of mobile object trajectories *Proceedings of the 13th International Workshop on DEXA* 721-8
[31] Weinan W, Osmar R Z 2002 Clustering web sessions by sequence alignment *Proceedings of the 13th International Workshop on Database and Expert Systems Applications* 394-8
[32] Wolfson O, Xu B 2010 Spatio-temporal databases in urban transportation *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 18-25
[33] Wolfson O, Xu B, Yin H, Rishe N 2005 Resource discovery using spatio-temporal information in mobile ad-hoc networks *Proceedings of the 5th International Workshop on Web and Wireless Geographical Information Systems W2GIS'05* 129-35

Authors

Xiaolei Li, born in December, 1981, Ningbo, Zhejiang Province, China

Current position, grades: lecturer at the School of Electronic and Information, Ningbo Dahongying University, China.
University studies: Master's degree in Information Engineering.
Scientific interest: information engineering and data mining-algorithms.
Publications: 12 papers.

An improved watershed algorithm for image segmentation

Wenhong Wu[1*], Hengmao Niu[1, 2]

[1]*Inner Mongolia University of Technology, Inner Mongolia, 010051, China*

[2]*Inner Mongolia Technical College of Construction, Inner Mongolia, 010070, China*

Received 1 March 2014, www.cmnt.lv

Abstract

Watershed transform is a key operator in image segmentation algorithms. However, the computation load of watershed transform is too large for real-time applications. Previously published watershed segmentation algorithms required at least three global synchronization points: minima detection, labelling and flooding. This paper presented an algorithm of watershed transformation based on opening-closing operation and distance transform. It improved the classical watershed segmentation algorithm based on distance transform, overcoming over segmentation. The experiment result demonstrated that this method for segmentation inherits the advantage of watershed algorithm based on distance transform that it successfully segment out each dowel in the image bringing convenience to computer vision and auto-counting of dowels. It also overcame over-segmentation existed in traditional watershed segmentation preserving the original edges of each dowel in the image completely. This algorithm can be combined with any image segmentation algorithm to give more precise segmentation results. An example is also shown by combining a background registration and change-detection-based segmentation algorithm with Watershed. This new video segmentation algorithm can give accurate object masks with acceptable computation complexity.

Keywords: watershed transform, image segmentation, automatic segmentation method

1 Introduction

Image segmentation is the process of building a partition of the image into connected regions, such that picture elements (pixels) of the region are homogenous according to some criterion (gray value, motion, etc). In this paper, we will focus on the watershed transform which is a mathematical morphology method for segmenting images. Most of the time, the watershed transform is applied to a gradient image for extracting homogeneous regions with respect to luminance. This gradient image is considered as a topographic relief, and watershed segmentation amounts to extracting significant basins in this relief. Watershed is a fundamental approach in a variety of fields, like image compression, coding, and analysis. It is used in telecommunication, biomedical, physics, satellite picture analysis, fast imaging system, etc. However, watershed segmentation is a computationally intensive task.

Watershed transform, which can separate an image into many homogeneous non overlapped closed regions, has been widely applied in image segmentation algorithms. It is also applied to image sequences as a core operator of video segmentation, which is a key technique in MPEG-4 content-based encoding systems. Video segmentation algorithms with watershed transform are taken as mainstream since they can generate object masks with accurate boundaries.

Many watershed algorithms have been proposed. Vincent and Soille proposed a watershed algorithm using immersion simulations. With sorting before the flooding process and with priority queue, this algorithm is dramatically faster than any former ones. Beucher and Meyer's algorithm also uses immersion simulations. Two types of algorithms are included: one creates watershed pixels and the other produces a complete tessellation of an image. An ordered queue is used in this algorithm, whose concept is similar to that of Vincent and Soille's algorithm; however, the minima of the input image need to be detected and labeled first, thus increases the complexity of this algorithm. Dobrin et al. proposed a fast watershed algorithm named split-and-merge algorithm. It can solve the isolated area problems of the former two algorithms when they are employed to create watershed pixels. Although the results obtained are more correct, it is more complex than the other two algorithms. Moreover, watershed transform for video segmentation is often required to produce a tessellation of an image, where the isolated area problem would not occur. Moga et al. proposed a watershed algorithm suitable for parallel implementation. With parallel computation, the watershed algorithm can be further accelerated. However, it is also complex and requires a powerful platform, which is impractical for general cases.

Several implementations of the watershed algorithm can be found in the literature: Sequential algorithms by immersion and other algorithms based on topographical distance, like the Hill-Climbing algorithm. Efficient implementations and parallelization of these algorithms have to cope with the problem of non-minimum plateau, which can be solved by using hierarchical FIFO (First-In-First-Out) queues, or by a lower completion of the input image. A recent implementation of a Hill-Climbing

***Corresponding author** e-mail: wwh801225@163.com

algorithm using FIFO queues is presented. The paper surveys approaches for a parallel implementation of the watershed algorithm: most of them are based on the Hill-Climbing algorithm, distributing the image in blocks over the set of processors implemented on a Parsystec supercluster 128 or a Cray3D. The lack of scalability, of these implementations is mainly due to the need for a global synchronization between processors for local minima detection, non-minimum plateau flooding, or lower completion. A fine grain parallelization was also proposed, using an associative net model, although the minimum detection is a prerequisite.

This paper proposes an improved watershed algorithm for image segmentation on the basis that watersheds are highly related to original frame data. It updates watersheds instead of recalculating watershed transform frame by frame. The watershed process can be accelerated, and the results are almost the same as those of conventional watershed algorithms. Furthermore, the segmentation results can be further improved with the intra-inter watershed scheme. In this paper, the watershed algorithm is programmed as a set of concurrent communicating iterative programs that are efficiently mapped onto an asynchronous parallel architecture. This idea was previously suggested and has also more recently been studied with the paradigm of Asynchronous Cellular Automata (ACA), i.e., a lattice of finite state machine interconnected to its nearest neighbours. The general specification of the watershed algorithm we propose is mapped onto an asynchronous processors array which enables us to derive a set of architectures, ranging from a fine grain processing (one finite state machine per pixel), to coarser implementations (a complex finite state machine or program which is able to process several pixels). In the case of fine granularity, this ACA leads to an implementation similar to the associative mesh.

2 Related works

The watershed segmentation technique has been widely used in medical image segmentation. Examples include the work presented in [1, 2], which make use of the watershed transform to segment grey and white matter from magnetic resonance (MR) images. The algorithm originated from mathematical morphology that deals with the topographic representation of an image [3, 4]. The set of pixels with the lowest regional elevation corresponds to the regional minimum. The minima of an image are the groups of connected pixels with their grey level strictly lower than their local neighbouring pixels. The rainfall simulation [3] describes that when rain falls onto the surface, any rain drop reaching a point in the surface will flow along its steepest descent until it reaches a minimum. The paths of pixels, which converge towards a common minimum, constitute a catchment basin. Watersheds are the elevated areas that divide the different catchment basins. The partitions, which we aim to obtain, are the catchment basins, and the boundaries between the partitions are the watersheds. Advantages of the watershed transform include the fact that it is a fast, simple and intuitive method. More importantly, it is able to produce a complete division of the image in separated regions even if the contrast is poor, thus there is no need to carry out any post-processing work, such as contour joining. Its drawbacks will include over-segmentation and sensitivity to noise [2]. There has also been an increasing interest in applying soft segmentation algorithms, where a pixel may be classified partially into multiple classes, for MR images segmentation [5-7]. The fuzzy C-means clustering algorithm (FCM) is a soft-segmentation method that has been used extensively for segmentation of MR images [8]. However, its main disadvantages include its computational complexity and the fact that the performance degrades significantly with increased noise. K-means clustering algorithm [9, 10], on the other hand, is a simple clustering method with low computational complexity as compared to FCM. The clusters produced by K-means clustering do not overlap.

In past few decades, various segmentation techniques have been pro-posed. Generally, they are based upon two basic properties: similarity and discontinuity. Pixel similarity gives rise to region-based segmentation [7-9], whereas pixel discontinuity gives rise to edge-based segmentation [10-12]. These two traditional techniques may be able to get good result for some simple images. However, it is always difficult to achieve desired result for HSRI segmentation due to the complexities of the landscape structure on the image. While textures in objects, such as forest, are always detected as spurious object boundaries, which cause over-segmentation; the edges between the farm fields may appear to be obscure that may induce under-segmentation. With regard to these problems, many works have been done. Some researchers integrated texture in segmentation to avoid the over segmentation of spectral heterogeneous objects [13]. In this strategy, the texture calculation before segmentation is time-consuming. Moreover, the texture boundary effect [14] may induce the location uncertainty of the extracted object boundaries. Some researchers proposed to segment image at multistage, then select and merge the optimal scale of segmentations to achieve good result [15]. However, how to select the optimal scale for different type of ground objects and merge them is still a difficult task that needs further research. Some researchers tried to integrate edge information into segmentation to get more accurate object boundaries [16]. This strategy takes the advantage of both region-based and edge-based segmentation techniques and is more practical for the quick and effective processing of large size data.

3 The proposed method

3.1 BASIC IDEA OF THE PROPOSED METHOD

Watershed algorithm is usually applied to the gradient image. Imagine the gradient image is a topographic surface,

a hole is drilled in each minimum of the surface, and water is flooded into different catchment basins from the holes. As a result, the water starts filling all catchment basins, which have minima under the water level. If two catchment basins would merge as a result of further immersion, a dam is built all the way to the highest surface altitude and the dam represents the watershed lines. This flooding process will eventually reach a stage when only the top of the dam is visible above the water line. The result is a tessellation of the input image into its different catchment basins, each one characterized by a unique label.

The method is a two-stage process, including the extraction of marker image and the labelling of pixels (flooding). Markers are a set of components marking flat regions of an image, i.e., each marker indicates the presence of an object. If the object interiors (markers) are set to 1, and the uncertainty areas are set to 0, we get a binary marker image. It contains a set of components (markers) marking the core regions and a large number of pixels may remain unassigned. The next step is then to label the unassigned pixels by the extended watershed algorithm dealing with markers to get the final partition. It has the advantage that segmented results can have coherent regions, link edges, no gaps due to missing edge pixels. However, applying this method to HSRI segmentation, the noises or textures on the image are usually labelled as the pseudo-local minimum regions and result in over-segmentation. To reduce over-segmentation, we make some improvements to the marked-based watershed segmentation algorithm. First, a regional adaptive marker extraction method is proposed. The marker image is created from gradient image by binary processing. Instead of using a fixed threshold, such as H-minima algorithm, a threshold image, on which the threshold of each pixel is statistically estimated and used for binarization. Because it considers the complex grey level distribution of HSRI, the extracted makers are more coincide with the real objects. Then, the image labelling scheme in the Meyer's algorithm is implemented by using one queue and one stack data structure. This scheme can largely save memory cost and make it applicable to large image. The overall framework is shown in Figure 1.

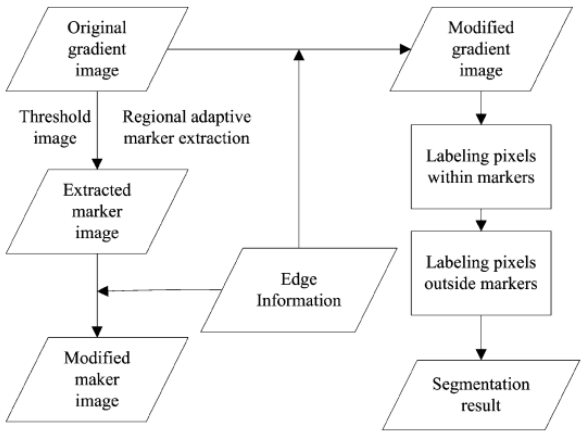


FIGURE 1 Framework of segmentation

3.2 HYBRID SCHEME

Although the segmentation results are almost the same as those of the conventional watershed algorithms, the proposed algorithm may introduce error in the following conditions. First, change detection is not sensitive enough. Second, the minimum of a catchment basin is included in UA, and a part the catchment basin is not included. Third, the scenes change a lot. The error will propagate in the proposed predictive watershed algorithm. The error propagation can be interrupted by inserting a frame where watersheds are generated with the conventional algorithm. These watersheds do not need information from the previous frame, so we have the term "intra watershed" (I-Water-shed). The I-Watershed and P-Watershed hybrid scheme, called IP-Watershed, is illustrated in Figure 2.

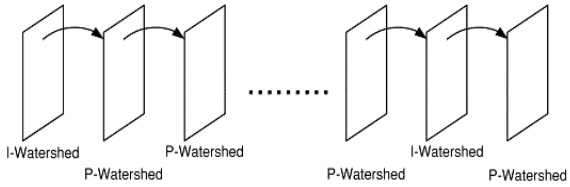


FIGURE 2 I-Watershed frames are inserted to avoid error propagation of Watershed

It can accelerate the watershed process and maintain accuracy at the same time. The time to insert an I-Watershed can be decided with the following two strategies to interrupt the error propagation. To deal with the first and the second error conditions described above, we can insert an I-Watershed frame after a fixed number of P-Watershed frames. The fixed interval to insert an I-Watershed frame depends on the threshold and the error rate required for the applications. After is decided, the error accumulation behaviour, which is the amount of increasing error for each frame, is also decided. An I-Watershed should be inserted when the accumulated error exceeds the required error rate. The less the error accumulated in each frame, the longer the interval to insert an I-Watershed frame; the tighter the error requirement, the shorter the interval. On the other hand, to deal with the third condition, scene change condition, the error rate is monitored in each frame. An I-Watershed frame is inserted dynamically when the monitored error rate exceeds a required value. These two strategies can be employed simultaneously to interrupt error propagation effectively.

3.3 LABELLING PIXELS

The labelling of pixels is the process of assigning each pixel a unique identity (ID). There are two often-used pixel labelling methods.

1) The first method uses the individual markers as the local minima in the gradient image. It filters out the undesired minima of the gradient image, and applies the traditional watershed segmentation on the revised gradient image.

2) The second method suppresses unwanted minima during labelling process. The marker based immersion type watershed algorithm proposed in literature, performs the flooding process directly on the original gradient image. It was realized by using the data structure of hierarchical circular queues. The hierarchical circular queues are a set of queues with different priorities, each queue is a first-in first-out data structure. The priority of each pixel in the gradient image is defined as the reciprocal of its gradient value. This implies that a high (low) priority is assigned to a pixel with low (high) grey-level value. To save memory cost for large image segmentation, we proposed to use a data structure of only one queue and one stack to store the temporary data in image labelling. It can largely reduce the memory cost. The pixels are first sorted in a descending order of priority. Then the possible neighbour labelled pixel of each unlabelled pixel is searched by seed tracing to identify the label ID for the current processing pixel. In the search process, one queue (QU) and one stack (ST) are used. In labelling pixels, the pixels within marker and the pixels outside makers are labeled successively.

During labelling the pixels outside the markers, the pixels labeled lastly are more possible to be on the watershed line. To make sure that the edges are labelled as the object boundaries, the gradient image is rectified by assigning the largest gradient magnitude value to the edge pixels first. The unlabelled pixels are processed in a priority value descending order. ST and QU are used to store the unlabelled pixels. For each priority level, there are two steps. In the first step, only pixels with priority higher than the current processing priority are pro-cessed. These pixels can be divided into two groups. The first group includes pixels with higher priority but not labelled in the previous labelling process, and they are stored in QU. The second includes all the unlabelled pixels with priority. The first group of pixels are processed in advance to the second one. After the first step, if the pixel cannot be labelled, it is stored in ST and processed in an reverse order in the second step. If after the second step, the pixel still cannot be labelled, it will be stored in QU and processed together with the pixels of which the priority is lower than marker-based watershed segmentation result. The labelling operation on pixels outside the markers is dependent upon the number of IDs labelled to the neighbouring pixels of the seed pixel. As same as the process of labelling the pixels within the markers, if two pixels are in the same region, their common neigh-boring pixels should not be edge pixels at the same time. This rule should be obeyed in the counting of the number of the ID; the same ID is counted only once, moreover the neighbouring pixel is not counted if it breaks the rule.

3.4 WATERSHED SEGMENTATION

In order to improve the computing speed of segmentation, in this paper a new algorithm is used. The new algorithm is called defined intuitive watershed algorithm that based on the "immersion" watershed algorithm proposed by Vincent and Soille.

The two algorithms have the same process, including sorting and flooding. To be specific, in the same gradient level, Vincent-Soille algorithm uses a FIFO queue to expand gradually from the inside out of the catchment basin. The new algorithm is to scan each pixel according to spatial relationship (from top left to bottom right), and to determine each pixel to belong to an existing catchment basin or a new one. The fundamental basis of the judgment is whether the pixel has marked adjacent pixels. If it has, it belongs to the catchment basin of its marked adjacent pixels. Otherwise it is to be a new minimum region and assigned a new regional marker. In order to avoid the generation of error minimum regions, two concepts of "error catchment basin" and "combined array" are introduced. Specific steps of segmentation are as follows:

1) Sorting: the new algorithm in the calculation of the sorting position takes the spatial location of pixels into account, so the elements of the sorted array meet a certain spatial relationship. The form of values of sorted array is: (X coordinates, Y coordinates, the gradient value).In this array, the smaller the gradient value is, the more forward the position will be. If the gradient value is the same, the pixel in the upper left (of spatial position) is in front.

2) Flooding: for each current scanned pixels, check its 4 adjacent pixels (or 8 adjacent pixels).if one of the 349 adjacent has not yet been marked, then turn to the next adjacent. If the entire adjacent are not marked, a new catchment basin is found and given a new marker. If one of the adjacent has been marked and the current pixel is not yet, then assign the marker of the adjacent to the current. If one of the adjacent and the current are marked, it may find an error minimum region or a watershed ridge lines.

3.5 WATERSHED CONTAINERS

According to implementation of the watershed segmentation algorithm, water is made from the process of spreading each storage pelvic gradually, namely the marked regional internal gradually to the edge of the area adhesions recursive. If the regional in image containers is too finely, which the whole image will be partitioned an infinite number of adhesion areas, if each containers correspond corn grain, it can guarantee the smooth implement impoundment process in grain internal. Therefore build excellent containers and water-shed, will directly determine the adhesion of particle swarm segmentation quality and efficiency. Through the edge detection has won the ideal watershed border, it should also ensure that the pelvic area between each grain to separate and divide boundary adhesion, and with low value area for the entire area.

1) Implementation taking binary processing after purification grayscale image in adaptive threshold;

2) Adapting radius of 10 pixels by type of structure elements, disk, operating binary image corrosion;

3) Taking negative after marking corrosion image, while assure background region pixels for "0" still, corrosion area of pixels is negative;

4) It will take the normalized image edge detection and buy negative corrosion images obtained by, and take "0" as a background, based on the negative marker value for storage, with "1" at the image matrix for watershed.

Through the above treatment, which ensure each corn grain image to have separate water, and relatively clear watershed, pelvic value is negative marker value, fantastic value close to "1", other areas are "0". The recursive operation process that from high value low value area regional was labelled by image of each grain, which like pixel gradient are uniformity, does not exist small containers and omit the impoundment of the interference of computation quantity and containers. It could improve the reliability and efficiency of operation, what's more, because each grain is a relatively independent starting point (saucers). Even though the public boundary (adjacent grain watershed) is discontinuous (existing gap), it also can ensure the effective segmentation adhesion areas. But if grain region and background area are linked, water that rising to a certain height from pelvic will inflow the background. Now water district have not reached threshold. It will mistake the background region as part of the grain area. The original image background region is "0" (black), via the single grain area, only after influenced by impoundment computing, in central small area isn't effected by leak influence, other areas are marked the not "0" value (white). According to the type of (1) method to implement the marker images, taking purification processing to leaked background, it can take "0" background region, the marker value of each grain regional remain unchanged.

4 Experimental results

Block size analysis for determining the optimal block size is shown in Figure 3.

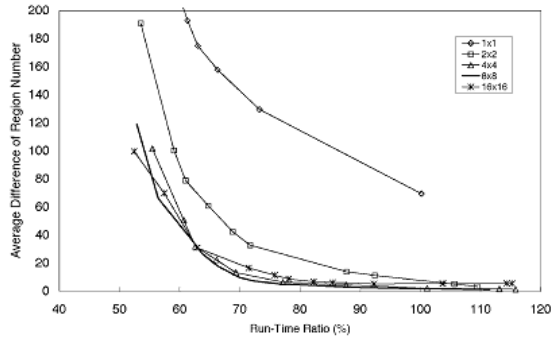


FIGURE 3 Block size analysis showing that 8X8 is the optimal block size.

The accuracy of the segmentation results is evaluated by the difference in region number, that is, the less average difference in region number between the results of the conventional algorithm and the proposed one, the higher the accuracy. Since regions in UAand outside UAare both processed with the conventional watershed algorithm, the positions of watersheds must be very close to those of the conventional algorithm, and only the over-segment problem may occur. The region number information is enough to evaluate the accuracy of watersheds, i.e., the closer the region number is between the result of the conventional watershed transform and that of the proposed one, the more accurate the proposed algorithm is. The ratio of the run time of the proposed algorithm to that of the conventional watershed algorithm is used as another criterion in this analysis. For each block size, a different threshold value is tested to decide the runtime-accuracy curve shown in Figure 3. The lower the threshold, the higher the runtime ratio, and the lower the average difference of region number. The test sequence is Children. Figure 3 shows that 8X8 is the optimal block size because it can provide higher accuracy with less run time than other choices. Several sequences were tested to give a similar conclusion.

The region number corresponding to a different run time is shown in Figure 4.

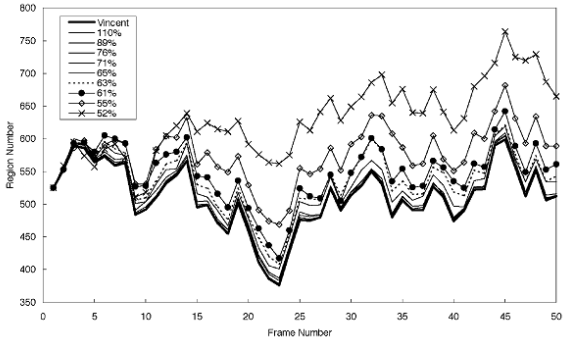


FIGURE 4 Region number of different run-time situations

As can be seen, when the run time increases, the similarity between the results of the conventional watershed and the proposed one also increases. When the run time is reduced to 61%, the region number is still very similar to the reference results. The execution time compared with Vincent and Soille's algorithm. Only core operations of watershed are recorded, namely, the gradient operation, sorting process, and flooding process. Note that the average difference in region number is fixed at 5% in these experiments. The simulation results show that the proposed algorithm can save 20%–50% of the computation. Moreover, the results of the conventional and proposed algorithms are similar. Consequently, the proposed algorithm can reduce computational intensity while maintaining segmentation quality.

5 Conclusions

Aiming at the limitation of watershed segmentation, this paper presents an algorithm of watershed transformation based on opening-closing operation and distance transform. It improves the classical watershed segmentation algorithm based on distance transform, overcoming over-segmentation. The experiment result demonstrates that this

method for segmentation inherits the advantage of watershed algorithm based on distance transform that it successfully segment out each dowel in the image bringing convenience to computer vision and auto-counting of dowels. It also overcome over-segmentation existed in traditional watershed segmentation preserving the original edges of each dowel in the image completely.

A new predictive watershed algorithm named Watershed for video segmentation is proposed in this paper. Taking into consideration the temporal coherence property of the video signal, the watershed algorithm can be accelerated. It updates watersheds in changing parts while keeping watersheds in other parts of a frame. The watershed process can be accelerated, and the results are almost the same as those of the conventional watershed algorithms. Moreover, the segmentation results can be further improved with the intra-inter watershed scheme. The proposed algorithm can be combined with any video segmentation algorithm to improve the segmentation results.

Acknowledgement

The work was funded by Research Project of Inner Mongolia University of Technology, "Method of PVA Fibre Identification Based on Image Processing", with Grant No.: X201423.

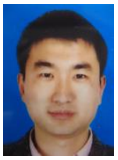
References

[1] Roerdink J B, Meijster A 2001 The Watershed Transform: Definitions, Algorithms and Parallelization Strategies *Fundamental Informaticae* 41 187-228
[2] Vincent L, Soille P 1991 *IEEE Trans Pattern Analysis and Machine Intelligence* **13**(6) 583-98
[3] Meyer F 1994 Topographic Distance and Watershed Lines *Signal Processing* 38 113-25
[4] Meyer F, Beucher S 1990 Morphological Segmentation *Visual Comm. and Image Representation* **1**(1) 21-45
[5] Moga A 1997 Parallel Watershed Algorithms for Image Segmentation *PhD dissertation, Tampere Univ of Technology Finland*
[6] Rambabu C, Rathore T, Chakrabarti I 2003 A New Watershed Algorithm Based on Hill-Climbing Technique for Image Segmenation *Proc. Conf. Convergent Technologies for Asia-Pacific Region, (TENCON '03)* 4 1404-8
[7] Ducourthial B, Merigot A 2002 Parallel Asynchronous Computations for Image Analysis *Proc. IEEE, special session on visual perception: technology and tools* **90**(7) 1218-29
[8] Bertsekas D, Tsitsiklis J N 1997 Parallel and Distributed Computation: Numerical Methods *Athena Scientific*
[9] Chandy K, Misra J 1988 Parallel Program Design: A Foundation *Addison-Wesley*
[10] Renaudin M 2000 Asynchronous Circuits and Systems: A Promising Design Alternative *Microelectronic Eng* **54**(1-2) 133-49
[11] Baudet G M 1978 Asynchronous Iterative Methods for Multiprocessors *ACM* **25**(2) 226-44
[12] Adachi S, Peper F, Lee J 2004 Computation by Asynchronously Updating Cellular Automata *Statistical Physics* **114**(1) 261-89
[13] Merigot A 1997 Associative Nets: A Graph-Based Parallel Computing Model *IEEE Trans Computers* **46**(5) 558-71
[14] Galile B, Renaudin M, Coulon P-Y, Mamalet F 2002 Watershed Parallel Algorithm for Asynchronous Processors Array *Proc ICME IEEE Int'l Conf Multimedia and Expo*
[15] Bieniek A, Moga A 2000 An Efficient Watershed Algorithm Based on Connected Components *Pattern Recognition* **6**(33) 907-16
[16] Bieniek A, Moga A 1998 A Connected Component Approach to the Watershed Segmentation *Math. Morphology and Its Applications to Image and Signal Processing* 12 215-22

Authors

Wenhong Wu, born in December, 1980, Inner Mongolia, China

Current position, grades: Master, a lecturer in Inner Mongolia University of Technology, China.
University studies: Computer science and technology.
Scientific interest: graphics and image processing and computer software.
Publications: 13 papers.



Hengmao Niu, born in April, 1980, Inner Mongolia, China

Current position, grades: PhD Candidate in Inner Mongolia University of Technology, lecturer in Inner Mongolia Technical College of Construction, China.
University studies: Fiber Reinforced Composites.
Scientific interest: fiber reinforced composites, computer science.
Publications: 15 papers.

A hybrid multi-class text categorization based on SVM-DT

Ying Fang[1, 2*], Heyan Huang[1]

[1]*College of Computer, Beijing Institute of Technology, Beijing, China, 100081*

[2]*School of computer & technology, ShangQiu Normal College, ShangQiu HeNan, China, 476000*

Received 6 October 2014, www.cmnt.lv

Abstract

How to improve the text categorization efficiency as well as keeping high speed is a research problem. Several factors are effected the processing of the decision tree construction, such as, the degree, the balancing degree, the constructing way, the group number and the division degree between groups etc. Considered the various roles between the above factors, a comprehensive algorithm to construct the SVM-DT (Support Vector Machine - Decision Tree) is proposed. In this method, three conditions are considered respectively. The text categorization experiments on massive corpus demonstrate that the algorithm can improve the efficiency in some degree and decrease the training and testing time largely at the same time. The algorithm to construct the SVM-DT is feasible and adaptable.

Keywords: Text Categorization; Support Vector Machine; Decision Tree; Multi-class category; Corpus; Positive Sequence Tree

1 Introduction

By constructing the best classified hyperplane, the Support Vector Machine (SVM, [1]) can rightly classify the samples, which is proved to be one of the most powerful text categorization methods. However, the SVM was originally developed for binary decision problems. If SVM is used to deal with multi-class classifier with the massive data, the calculating overhead is too huge. The popular methods for applying SVMs to multi-class classification problems usually decompose the multi-class problems into several two–class problems that can be addressed directly using several SVMs. The typical methods are the OAO (One-Against-One) method [2], OAA (One-Against-All) method [3], DAG-SVM (Directed Acyclic Graph-Support Vector Machine) method [4] and SVM-DT (Support Vector Machine-Decision Tree) method etc. Among them, DAG-SVM can solve the problem of "Rejecting Recognition" existing in OAO and OAA, but its generalization ability is limited. SVM-DT takes advantage of both the efficient computation of the decision tree architecture and the high classification accuracy of SVM, which is suitable to large categories [5].

There are many researches to constructing SVM-DT. Based on the information gain of the non-leaf node, Ramaswamy [6] proposed a method to construct a tree called Partial Sequence Tree; in this structure each SVM classifier separates one class from the remainders. The advantage of the method is that the classes at the top level of the tree had very high accuracy (more than 95% sometimes), but the ones at the lower levels (especially the last leaf node) had poor accuracy.

Madjarov etc. [7] applied ensemble learning techniques to constructed a tree structure called Positive Sequence Tree, in which each SVM classifier divides the classes into two groups. By contrast, the classifier with positive sequence tree was faster than the one with the partial sequence tree; and the accuracy between the classifiers on the different levels was relatively stable.

Takahashi and Abe [8] proposed four types of decision trees including the above two trees. But the reparability measure of the Euclidean distance or the Mahalanobis distance could not reveal the features of the texts because it only used the number of the words. And it did not explain which type was better under certain conditions. From these papers we can find that the type of the tree is an important effect to the category.

A balanced decision tree which included the tradeoffs between the sample number and the difficulties to divide was employed in [9]. Although it could decease training time in large degree while did not reduce the recognition rate, and it exaggerated the category effect of the "super class".

According to the massive text sets, in this paper we aim to find the effecting factors of SVM-DT construction and how to integrate many strategies while balancing the category precision and the working speed. The remaining parts of this paper are organized as follows. In Section 2, we introduce the strategies to construct the SVM-DT. Section 3 presents our algorithm and text categorization system based on the Section 2. Section 4 describes our experiments respectively done on English and Chinese corpus. Conclusions and further discussions are given in Section5.

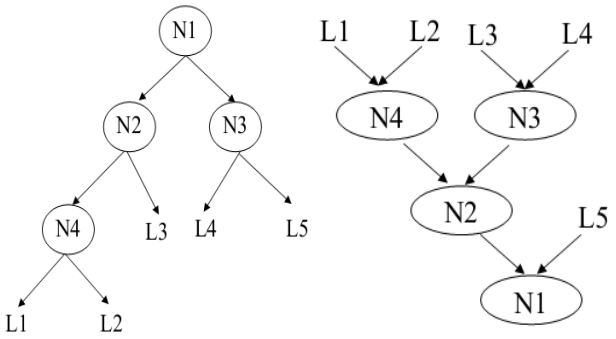
[*]***Corresponding author*** e-mail: yfang@bit.edu.cn

2 Constructing strategies of SVM-DT

In this section, we describe the constructing strategies of the SVM-DT, that is, the depth, the balance degree, the constructing way, the group number and the division degree between groups. These are the most important problems to consider when constructing the SVM-DT.

2.1 SVM-DT CONSTRUCTING MODE

According the constructing direction, the SVM-DT constructing way can be divided into: top-bottom division method and bottom-top accumulation method [10]. The former(Shown as FIGURE 1(a)) begins from the root and grows down to the leaves, which finds two classes which have the minimum distance and then divides the nodes. The similar dividing process will be done until each node points to one class. The latter(Shown as FIGURE 1(b)) first finds two nearest classes from all the samples and combines them into one new group, then recursively continues the finding and combining process until the last two groups are combined into one group.



(a) top-bottom mode (b) bottom-up mode

FIGURE 1 Constructing diagrams of SVM-DT

Both the division method and the accumulating method have $O(n^2)$ (n is the group number of the current notes) time complexity when finding two nearest groups. The calculating consuming is too large for multi-class category when the number of the texts is very big. So, one improving direction is how to reduce the unnecessary calculations.

2.2 EFFECT OF THE SAMPLES SCALE

The scale of the training and the testing sets tends to vary largely. To Reuters corpus, for example, the largest group is composed of 2,877 training documents, at the same time, there are 75 groups of the training documents are less than 10 articles. So when constructing a decision tree it is not appropriate to treat each group equally. Another example is on news sites, the category about the "business and economy" or "entertainment" will contain more pages than that about the "health and medicine" or "education" categories. Therefore, if the larger sets can be broken down earlier, the pressure to the latter work will be greatly reduced.

2.3 CONSTRUCTING STYLE OF DECISION TREE

We have known there are two kinds of constructing way for a decision tree [11]: (1) the Partial Sequence Tree, the two sub nodes of a non-leaf node describe the relation of one-vs.-many (1: n). (2) the Positive Sequence Tree, the relation of the child nodes of a parent node is many-vs.-many (m:n). It has been proved that during the categorization process the partial sequence tree having the bigger depth is slower than the positive sequence tree having smaller depth and higher paralleling degree. So when construct the tree we will try to build a positive sequence tree.

2.4 SIMILARITY BETWEEN THE CLASSES

If the classification performance is not good at the upper node of the decision tree, the overall classification performance becomes worse. So we should put the classes with bigger differences or higher classifying accuracy on the upper nodes of the decision tree, in order that the lower ones may be less affected. To realize that, we can calculate the similarity, then firstly place the groups with less similarity on the upper level.

Based on the above four problems, we present a hybrid strategies to construct SVM-DT: (1) the up-down dividing method is perfect; (2) large set in the samples should be divided out firstly to reduce the calculating overload; (3) the number of sets is tried to be balanced as far as possible to reduce the depth of the decision tree; (4) the dividing ability of the upper nodes should be made stronger as far as possible to reduce the accumulating errors.

3 A hybrid constructing algorithm of SVM-DT and text categorization system

In this section, we will introduce a kind of hybrid constructing algorithm of SVM-DT. The text categorization system based on the corresponding SVM-DT is to be provided then.

The symbols used in the following part are shown in Table 1.

TABLE 1 Notations used in the paper

Sym.	Description
S	the samples
K	the number of the set, $K \in [1,\infty)$
S_i	one sample set, $i \in [1,K]$
$Num(x)$	the number of the texts in set x
N_{dg}	the node on the dg ($dg \in [0,\infty)$) level of a decision tree
∂	the scale effecting factor, $\partial \in (0,1)$
β	the similarity effecting factor, $\beta \in (0,1)$s
SB	the buffer of the total processing set
S_A , S_B	the group centred as A or B
$Sim(A,B)$	the similarity of set A and set B
$IG(N)$	the Information Gain value of node N

3.1 THE HYBRID CONSTRUCTING ALGORITHM OF SVM-DT

In our algorithm, three conditions are considered when building the decision tree:

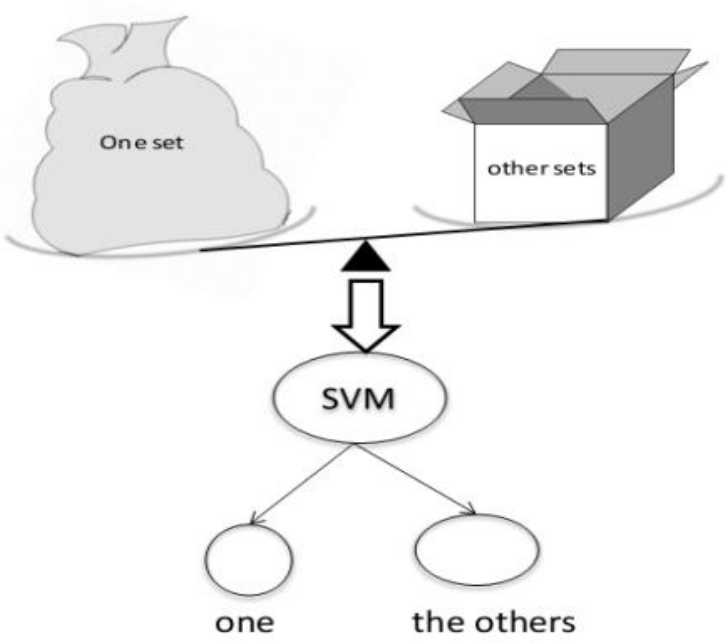


FIGURE 1 The description and the measure of condition (1)

If there is a very large class in the sample sets, we can recognize it only by the number of texts in the set. When the number of texts reaches a certain percentage (∂ here) of the total texts, the set will be put on the upper level first. The parameter ∂ is a value according to the sampling set, usually is set as 1/2 or 1/3 and so on. In our experiment we set it 1/3.

(2) The second condition is that one branch is not too far to divide from the other, which is described as Figure 3.

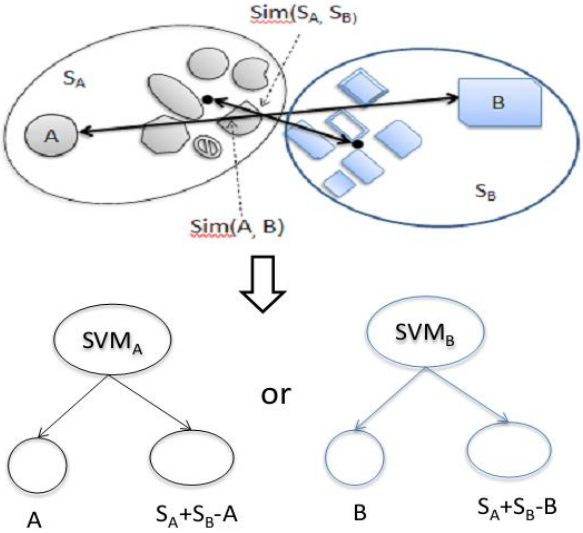


FIGURE 3 The description and the measure of condition (2)

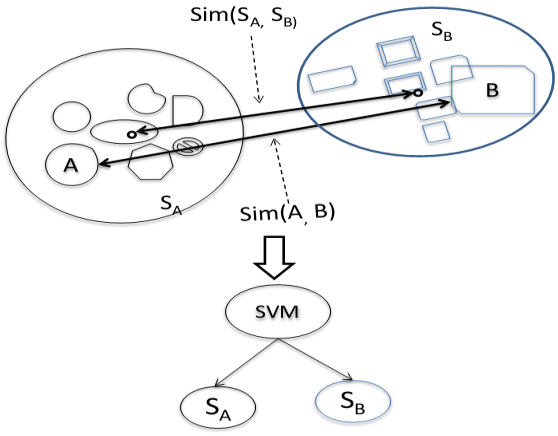


FIGURE 2 The description and the measure of condition (3)

If the gap between the similarity of two groups (S_A, S_B) and the similarity of the two farthest set (A, B) reaches a certain degree, we think A or B is not close enough to the centre of the group. The measure we take is to separate one of them out according to the Information Gain value of the parent node.

(3) The third condition is that two groups are far away enough, shown as FIGURE 4.

This condition should be the most common in the samples. The centre of the group (S_A , S_B) is close to the

dividing node (A , B). So we think two groups can be divided apart. Then a tree structure can be built on S_A and S_B.

The hybrid constructing algorithm for SVM-DT is shown on Algorithm 1.

Algorithm 1 The hybrid constructing algorithm for SVM-DT

Input: training set $S=\{S_1,S_2,...,S_K\}$

Output: a decision tree structure with the corresponding SVM classifiers

// initialization.

$SB=S$; $M=\sum_{i=1}^{K} Num(S_i)$; $dg=0$;

construct a root node N_{dg}.

Step1. let P is the biggest set of SB;

if $Num(P)/M>\partial$ then

{// process the super class P.

construct the sub tree of N_{dg} with the branches of P and $SB-P$;

build a SVM classifier between P and $SB-P$;

$dg++$; $M=M-Num(P)$; $SB=SB-P$;

turn to Step1; // make the sub tree of SB until there is only one set in SB;

}

else {turn to Step2;}.

Step2. //to divide the set into to groups

for each set pair C and D ($C,D\in SB, C\neq D$)

{calculate the similarity $Sim(C,D)$ using formula (3);}

let A and B are the sets which have the lest similarity among those pairs;

for each set P ($P\in SB-A-B$) {

if $Sim(P,A)>Sim(P,B)$

{put P into the set S_A;}

else {put P into the set S_B;} }

if $Sim(S_A,S_B)/Sim(A,B)\leq\beta$

{//to process the class with higher dividing degree

calculate $IG(N_{dg})_A$ with the branches A and $SA+SB-A$;

calculate $IG(N_{dg})_B$ with the branches B and $SA+SB-B$;

if $IG(N_{dg})_A<IG(N_{dg})_B$

{construct the sub tree of N_{dg} with the branches A and $SB-A$;

build a SVM classifier between A and $SB-A$;

$dg++$; $SB=SB-A$; $M=M-L_A$; $N_{dg}=N_{SB-A}$;}

else

{construct the sub tree of N_{dg} with the branches B and $SB-B$;

build a SVM classifier between B and $SB-B$;

$dg++$; $SB=SB-B$; $M=M-L_B$; $N_{dg}=N_{SB-B}$;}

}

else {turn to Step3;}

Step3. //build sub tree of two sets

let S_A and S_B as two branches of the node N_{dg}, construct the tree structure;

build a SVM classifier between S_A and S_B;

$SB=S_A$; turn to Step1;

//make the sub tree of SB until it can not be divided then.

$SB=S_B$; turn to Step1;

//make the sub tree of SB until it can not be divided then;

There are two main factors indicating the algorithm: the huge amount and the distributing character of the texts. So the hybrid algorithm gives an improvement on two sides: (1) the super class can be divided out at the early stage, which aims to speed the construction of the decision tree and reduce much similarity calculation; (2) the balance of two sub trees is taken into consideration, which aims to improve the degree of parallelism and reduce the attraction of the large groups in some degree.

3.2 TEXT CATEGORIZING SYSTEM BASED ON SVM-DT

The text categorizing system (FIGURE 5) is composed of two parts, training part and classifying part, which is connected with the core classifier.

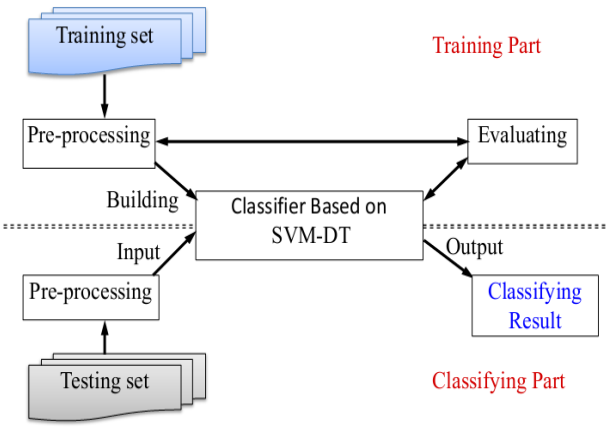


FIGURE 3 Function chart of the text categorizing system based on SVM-DT

Both the training texts and the testing texts should be pre-processed first, which includes the stemming for English texts, word segmentation and tagging for Chinese

texts, stop words removing, feature (words are the features in this paper) selection etc.

The selected feature is weighted by TF-IDF method. $W_{t,d}$ is the weight of word t in text d:

$$W_{t,d} = \frac{TF_{t,d} \times \ln(N / DF_t)}{\sqrt{\sum_{t=1}^{m} [TF_{t,d} \times \ln(N / DF_t)]^2}}. \quad (1)$$

With, $TF_{t,d}$ is the number of t appearing in d. DF_t is the number of the texts including t. N is the total number of the texts.

The text d can be described as a vector V composed of m feature and its weight:

$$V_d = \{(T_1, W_{1,d})(T_2, W_{2,d})...(T_m, W_{m,d})\}. \quad (2)$$

Then, the similarity between vector V_A and vector V_B is calculated as:

$$\mathrm{Sim}(V_A, V_B) = \frac{\sum_{t=1}^{m} (W_{t,A} \times W_{t,B})}{\sqrt{\sum W_{i,A}^2} \sqrt{\sum W_{j,B}^2}}. \quad (1)$$

with, m is the feature both in V_A and V_B.

The aim of the training part is building the classifier based on SVM-DT according to the algorithm 1. Because the classifying result is affected by the feature selection, so we should select the features repeatedly according to the evaluating result.

In training part, each text inputted into the SVM-DT classifier will be divided into one branch of the root node on the basis of the similarities. If the corresponding node is not the leaf node, then we will calculate the similarity and allocate the text to a node recursively until it point to a leaf node. Then the category of the last node is the one that the text to be classified into.

The similarity of two sets can be calculated between the category centres or according to the maximum /minimum distance, the reparability measure matrix. However, the complexity to calculate the reparability measure matrix of two sets is relatively high. Or, much valuable information will be cast away when calculating the similarity with the maximum /minimum distance. So we calculate the reparability between two sets according to the centre vectors similarity of two categories. Let the category L_i contains K texts, C_i is the centre vector: $C_i = \frac{1}{K}\sum_{k=1}^{K} V_k$. Sim_{ij} is the similarity between two categories L_i and L_j, $Sim_{ij} = \left\| C_i - C_j \right\|_2$.

4 Experiment and analysis

In this paper, our experiment system is built on the LIBSVM [12], which is an open source software. The kernel function is RBF. The running environment is 2.66GHz CPU clock speed, 2.0GB memory. We select two types of texts to do our experiments.

English data we used here is from the Reuters21578, which is an open corpus accepted in the field of text category. The total number of the classes is 80. To ensure there is at least one text both in training set and in the testing set, we select 8230 training test and 3200 testing texts. The words in the TITLE and BODY are composed of the texts. After the processes of stemming and stop word removal we get about 20,000 feature words.

To test the application ability of the text categorization system based on the SVM-DT, we collected about 10,000 texts from several famous Chinese news websites such as ifeng, netease, sohu, chnqiang et. al.. The time span of the dataset is a week. There are 36 topics classified by manual works.

4.1 METHOD COMPARISON

To compare with other method, we test three kinds of classifiers.

-- M1: method in Ref. [6] to construct a partial sequence tree.

-- M2: method of Ref. [7] to construct a positive sequence tree.

-- h-M: our method to construct a hybrid type of tree.

Ref.[13] used 8237 texts for testing purpose and 3186 texts to the SVM-DT, which was similar with our texts. The data shown in Ref [13] is our baseline. The precision is used to measure the classified result, calculated as formula (5).

The result of the experiment is shown in Table 2.

TABLE 2 Categorical result of the English texts

Methods	Training Time(s)	Testing Time(s)	Precision(%)
Ref[13]	67	27	94.3
M1	71	36	89.8
M2	62	19	93.7
h-M	56	23	95.9

Comparing with M1 and M2, we could infer that it is quicker to build a positive sequence tree (M2) than a partial one (M1), and the testing time of the former tree is less either. Our hybrid structure (h-M) is the fastest during training period mainly because of the early division of the super class, which avoid a series of the similarity calculation. What's more, testing based on M2 is faster than on h-M, which is reflected by the depth of the tree. Most important, the precision of our method is the best in all the methods. Comprehensively view, our implementation can perform well with relatively quick speed.

4.2 APPLICATIONS

To test the applying ability of the text categorization system based on SVM-DT, we make experiment on the self-built dataset then. The testing texts are randomly held out 1/3 of the total texts and the remaining texts are used to train the system. After 10-fold cross validation, we get the evaluating result. The evaluation index we used here is the standard Precision (P), Recall(R) and F1-measure (F1).

The result based on M1, M2, and h-M is shown on Figure 6.

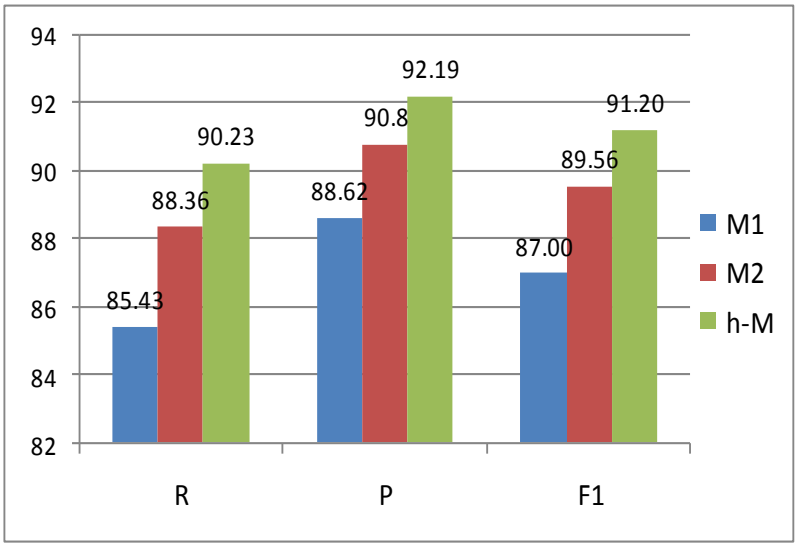


FIGURE 4 Text Categorization result of Chinese texts

On Figure 6, the highest columns are the best results. It is clear that h-M method have got the best values in all the evaluation data. So, we are sure that the text categorization system based on h-M method can distinguish the topics perfectly. Although the time consuming is not the least, the overall performance can meet the demands of the practical application.

5 Conclusion and future work

To SVM, there are two difficulties. One is the lower training speed; the other is the lower precision to deal with the big data. According to the problem, we do improvement on two sides. (1) We notice a hypothesis that the super sampling classes have big differentiation. So, when constructing the tree, we give priority to divide the super class, which can shorten the training time. (2) We make choice to build the branches depending on the disparity of the centre distance between two classes. Our experiment showed that the improvement is effective to reduce the training time, to compress the depth of the tree, to improve the precision comprehensively. Our attempt to balance the classification speed and the efficiency is efficient, which can achieve more feasible method.

Due to the lack of the compared corpus, we did not do the experiment on an open shared Chinese dataset. But our experiment has shown that the hybrid method is not confined to a kind of language. In the future, we can extend our method to the multilingual data. The ability to process the mass data needs to be verified too.

Acknowledgements

This work was supported by National Program on Key Basic Research Project (or 973 Program, No. 2013CB29606), Natural Science Foundation of China (No. 61202244), research funded project of Henan Province and ShangQiu Normal College (No. 2010A520031, No.2013GGJS013).

References

[1] Cortes C, Vapnik V 1995 Support-vector Network *Machine Learning* **20**(3) 273-97
[2] Weston J, Watkinns C 1998 *Multi-class support vector machines* Royal Holloway University of London
[3] KreBerl U PairWise 1999 Classification and support vector machines *Advances in Kernel Methods Support Vector Learning* 255-68
[4] Zhang C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34
[5] Xiong K, Zhang Y, Zhang Z, Wang, S, Zhong Z 2014. PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103
[6] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination. *Journal of Applied Research and Technology* **12**(2) 171-81
[7] Madjarov G, Gjorgjevikj D, Delev T. 2010 Ensembles of Binary SVM Decision Trees *ICT Innovations Web proceeding* 181-7

[8] Takahashi F, Abe S. 2002 Decision-tree-based multiclass support vector machines ICONIP'02 *Proceedings of the 9th International Conference on IEEE* **3** 1418-22
[9] Diao Z H, Zhao C J, Guo X Y, et al 2011 A new SVM multi-class classification algorithm based on balance decision tree *Control and Decision* **26**(1) 149-52, 156
[10] Qiao Z W, Sun W X. 2009 A multi-class classifier based on SVM decision tree *Computer applications and software* **26**(11) 227-30
[11] Zhu Y P, Dai R W 2005Text classifier based on SVM decision tree *Pattern recognition and artificial intelligence* **18**(4) 412-6
[12] Chang C C, Lin C J 2011 LIBSVB:A Library for Support Vector Machines http://www.csie.ntu.edu.tw/~cjlin/libsvm Mar.13
[13] Zhao T J 2010 Text classifier based on an improved SVM decision tree *Journal of intelligence* **29**(8) 141-3

Authors

Ying Fang, born in January, 1977, Haidian District, BeiJing, P.R. China

Current position, grades: Lecturer of ShangQiu Normal College and Ph.D student of Computer School, Beijing Institute of Technology, China.
University studies: She received her B.Sc in Computer Application from ZhengZhou technology College and M.Sc. in Computer Software and Theory from ShanXi University in China.
Scientific interest: Her research interest fields include Machine Learning, Natural Language Processing
Publications: more than 20 papers
Experience: She has teaching experience of 14 years, has completed three scientific research projects

Heyan Huang, born in October, 1963, Haidian District , BeiJing, P.R. China

Current position, grades: Professor and doctoral tutor of Computer School, Beijing Institute of Technology, China.
University studies: D.Sc from Chinese Academy of Sciences.
Scientific interest: Her research interest fields include Machine Translation, Natural Language Processing
Publications: more than 80 papers published in various journals.
Experience: She has teaching experience of 25 years, has completed more than 40 scientific research projects

Design of data acquisition node based on CAN bus

Xuewen Yu*

Shunde Polytenchnic, Shunde, Guangdong Province, 528300, China

Received 1 March 2014, www.cmnt.lv

Abstract

Taking single chip microcomputer (SCM) STC90C58AD as control core, a universal CAN bus data acquisition node was designed using CAN bus controller SJA1000 and CAN bus transceiver PAC82C251. Also, the hardware circuit and software design of CAN bus communication were given. Before data collection, current signal has been I-V transformed, and voltage signal has been filtering processed and amplified. Such program is suitable for multi-sensor field data collection in industry.

Keywords: CAN bus, SJA1000, data acquisition

1 Introduction

CAN (Controller Area Network) is the only field bus with international standard at present. Compared to normal communication bus, the CAN bus possesses excellent characters of reliability, real-time and flexibility. As CAN is a multi-master bus, the communication media can adopt twisted pair line, coaxial cable or optical fiber, whose communication rate can reach 1Mbps, and distance can be 10km. CAN protocol adopts encoding the communication data instead of coding the traditional address. Due to its powerful error-correcting capability and support for the transceiver of difference signal, CAN bus has farther transmission distance in highly disturbed environment [1]. This work presents a design of data acquisition node based on the CAN bus. Such node can output the data collected by sensor in the form of digital quantity, which can easily connect to the PC via CAN network, thus simplifying systemic structure and improving accuracy. Meanwhile, several multiple data acquisition nodes can be mounted on CAN bus, realizing the network of data acquisition node. This design is suitable for the data acquisition system whose sensors are dispersed.

2 Topology of CAN bus

Figure 1 shows the topology of CAN bus. System mainly consists of host computer (PC) and data collector. Bus transmission medium adopts the shielded twisted-pair. As the core of the whole system, PC's function is to monitor and manage the system. The function of CAN bus intelligent adapter is to collect the data from each node transferred to PC, and transmit PC's orders to each node. With the multiple host structure of CAN, each node can send data to the bus according to requirements. Because of this, while being monitored by PC, each node can also monitor the bus, thus receiving useful data. Data acquisition node is composed of microprocessors and programmable CAN controller chip, and its function is field data collection and transmission.

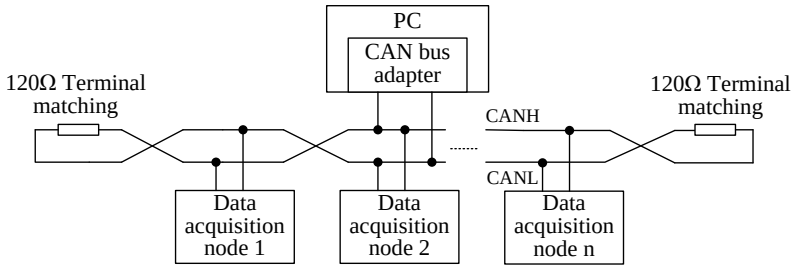


FIGURE 1 CAN bus topology

3 Hardware design of data acquisition node

The main function of data acquisition node includes: Changing the continuous analog signals into digital signals; converting the digital signals into information which is accord with CAN bus protocol; publishing the information on CAN bus. Figure 2 shows the structure diagram of data acquisition node. As shown, the information detected by sensor is sent to programmable amplifier consisted of instrument amplifier AD623 and digital potentiometer X9214. By changing the resistance of X9214, the amplification of signal can be regulated. After amplification, the information will be converted into digital signal by integrated 10-bit A/D converter inside

***Corresponding author** e-mail: xuewenyu_sd@163.com

SCM STC90C58AD. Then, the digital signal will be sent to the bus controller SJA1000. Finally, though photoelectric coupling based on CAN bus protocol, the information transferred from bus transceiver to CAN bus will be received by other nodes and PC.

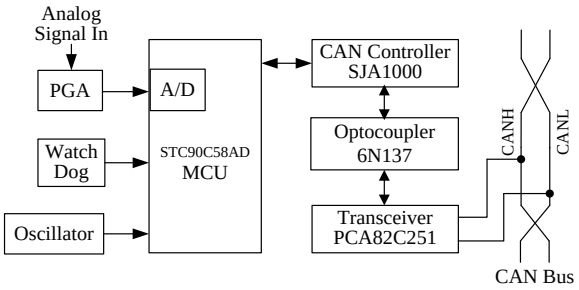


FIGURE 2 Overall block diagram of data acquisition node

3.1 SCM STC90C58AD

Figure 2 shows SCM STC90C58AD, the core processor of data acquisition node. As a new generation introduced by STC, SCM STC90C58ADSTC has the advantages of super anti-interference, high-speed and low-power. Such SCM can work under harsh industrial environments with the features of high speed data acquisition and long working hours. Besides, its instruction code is compatible with traditional 8051 SCM, with the frequency range of 0~40MHz. This SCM has the function of EEPROM for a 40K application space and integrated 256+4096 byte RAM. Moreover, this SCM also supports ISP and IAP for its inner couplers of Watchdog, eight-channel 10-bit A/D converter and three 16-bit timer/counters. In a word, such SCM includes all the unit modules required during data acquisition and controlling. So this SCM is a system with high reliability but low cost.

3.2 DESIGN OF I/V CONVERTING CIRCUIT

The standard current signal outputted by sensor is 4~20mA. To perform A/D conversion and achieve good conversion effects and accuracy, field current signal should firstly be converted into standard voltage 0~5V. It cannot eliminate common mode interference by using circuit network to directly change current signal into voltage signal, and the conversion accuracy cannot be guaranteed. Using the precision current-loop receiver chip RCV420 produced by U.S. company BURR-BROWN, this design can transfer the input signal 4~20mA into 0~5V output signal by current-loop technology. Figure 3 shows the converting circuit with second-order low-pass filter. Both the positive and negative electrodes of RCV420 should connect to a 1μF decoupling capacitor. Besides, two electrodes should be placed close to the amplifier. To avoid the gain and CMR errors introduced by external circuits, the electrodes should connect to ground as the method shown in Figure 3, thus ensuring the minimum grounding resistance. The unused Ref IN pins should be grounded to maintain a high CMRR.

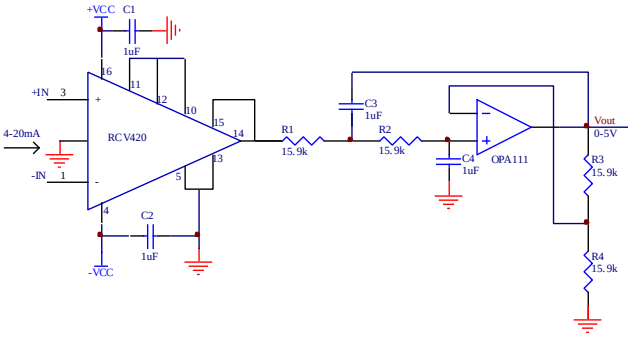


FIGURE 3 I/V converting circuit

3.3 PROGRAMMABLE GAIN AMPLIFIER CIRCUIT

Normal analog voltage signal transformed from industrial sensors includes: 0~5V, 0~10V, ±5V and ±10V. It is impossible for A/D transfer to match all the conditions. If single gain amplification is used, the range of A/D conversion chip cannot be maximized. Furthermore, the measured signal may be saturated, resulting in large measurement error or even damages to A/D chip. Therefore, the gain of preamplifier should be changed according to the output of sensor to obtain the maximum measuring accuracy. This design used the programmable gain amplifier consisted of instrument amplifier AD623 and digital potentiometer X9214. Figure 4 shows the schematic diagram of programmable control amplifier circuit.

AD623 has the advantages of high-accuracy, low-drift and low-noise. And this amplifier can change signal gain by only adjusting the resistance R_G between pin 1 and 8. Moreover, AD623 is easy to be used. The signal gain equation is as follows [4]:

$$G = \frac{49.4k\Omega}{R_G} + 1$$

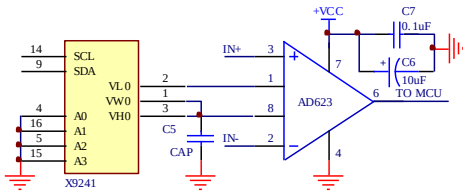


FIGURE 4 Programmable control amplifier circuit

Digital potentiometer X9214 has I2C interface. As shown in Figure 4, VL0 refers to the low end of potentiometer, VH0 the high end and VW0 the movable point. The resistance value of potentiometer between pin VL0 and VW0 can be changed through programming [5]. In specific use, SCM determines the signal range based on the signal value first collected. Then the highest magnification measurement accuracy will be obtained by certain algorithms. Finally, controlling the digital potentiometer X9241 by I2C bus, then the programmable amplification can be reached.

3.4 A/D CONVERTING CIRCUIT

Analog signal, after regulation, will be converted to digital signal by integrated 8-channel 10-bit A/D inside SCM STC90C58AD. The A/D converter interface of SCM STC90C58AD, whose speed can reach 250 KHz, lies in P1 interface. And the converter is a kind of successive approximations ADC. Users can set any one of the eight ways into A/D conversion through software, while the others continue acting as I/O interface. After conversion, the converter should be set at the A/D ending flag bit ADC_FLAG of control register ADC for program inquiring or interrupt application.

3.5 CAN CONTROLLING SJA1000

SJA1000 is an independent CAN controller, supporting the CAN2.0B protocol which has many new features. And the controller lies between the micro controller and transceiver. The CAN core module of SJA1000 is responsible for receiving and transmitting information frame of CAN, as well as realizing CAN protocol. Interface management logic manipulates the interface of the main controller outside. So the operation of STC90C58AD to SJA1000 is equal to that external RAM does [6]. After acceptance filter completes filtering the received information, the information frame that have passed acceptance filtering and without error will be sent to the receive FIFO buffers. Through read-write operation for built-in register of SJA1000, the master controller SCM can be set to CAN bus communication mode, thus achieving data transmitting and receiving. Then, the communication of SCM to CAN controller can be achieved by interrupt mode, which connects interrupt signal pin /INT of SJA1000 to the external pin /INT0 of STC90C58AD.

3.6 CAN TRANSCEIVER PCA82C251

The encoding and decoding of data transmission have been logically achieved by SJA1000. While to connect physical line, the driving capability of differential transmit and receive of CAN bus should be enhanced by means of bus transceiver PCA82C251. Such transceiver can transmit data between two differential voltage bus cables with a high speed of 1Mb/s. Both the CANH and CANL pins of transceiver can connect CAN bus by a thermistor, so as to protect itself from the impact of over-current when resistance heating value becomes larger. Between CANH and CANL, there parallel two small capacitors, which can filter the high frequency interference on bus, and shield electromagnetic radiation. As shown in Figure 5, TX0 and RX0 of SJA1000 are connected to PCA82C251 via high-speed optical coupler 6N137. It is to enhance the anti-jamming capability of CAN bus node and prevent crosstalk between lines. 6N137 is compatible with TTL and CMOS level, with a signal width of 10MHz. So it fully meets the requirement of a communication speed of 1Mb/s for CAN bus signals.

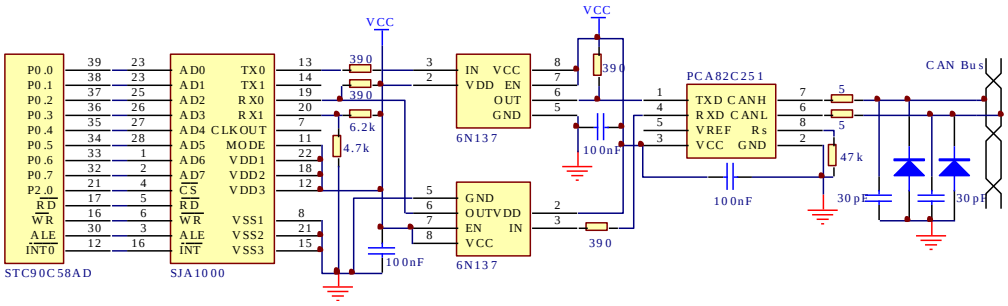


FIGURE 5 Schematic circuit of CAN communication node

4 Communication software design of CAN bus

Communication software design of CAN bus mainly includes SJA1000 initialization, CAN transmitting and receiving programs. Besides, there are some auxiliary functions, like interrupt processing, information processing, bus sleep and awaking treatment, bus error handling and overloading treatment. The modular and structured program design of software design enable the software has good portability.

4.1 FLOW CHART OF MAIN PROGRAM

Figure 6 shows the overall flow chart of software design for system node. When system powered up, part of the reset work would be first completed. Entering the initialization stage, the software would complete the initialization of the rest modules. Following that, system enters the wait state. When received the order of data acquisition from PC, note begins to collect the measured physical signal. Firstly, the analog signal received from the front end of sensor, after be amplified and filtered, should be sent to A/D converter inside SCM for analog/digital conversion. After completion of the signal conversion, results will be temporarily stored in RAM. Then, the data should be processed into CAN data frame until the SJA1000 CAN controller receives the command of data transmission. When PC requires data transmission, interrupt subroutine will start up to transmit data frame. Once transmission finished, system will jump out subroutine. If the data frames on bus need being received, system will receive it. And there is not confliction between

transmission and reception, namely the two processes can work simultaneously. While receiving the command of finishing data acquisition from PC, node will return to the wait state.

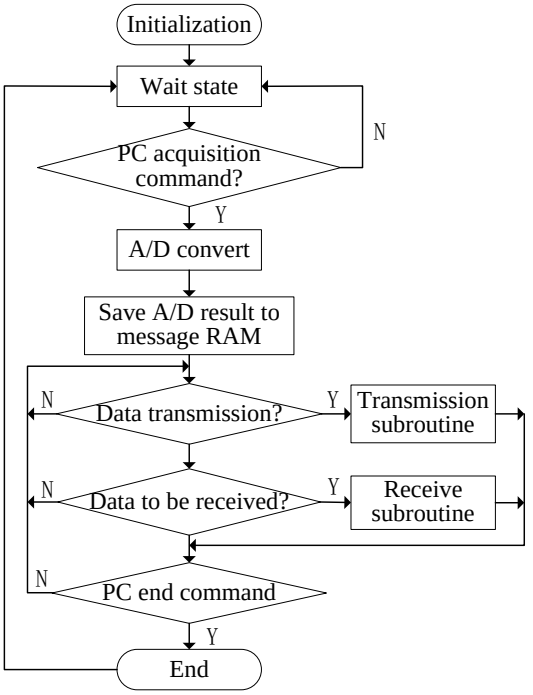


FIGURE 6 Flow chart of the main program

4.2 INITIALIZING PROGRAM OF SJA1000

When using the CAN controller, operator should allocate related registers according to the tasks, namely setting their initialization parameters. However, the initialization of SJA1000 performs only in reset mode. The initialization procedures of SJA1000 mainly includes: selecting work mode, setting interrupt system, setting the bite rate of bus, setting acceptance filter, and setting the output mode and configuration of CAN bus output pins. Actually, these settings refer to the write operations for some registers inside SJA1000. When initialization was completed, the reset request of SJA1000 should be removed. Then, SJA1000 can transmit and receive messages.

4.3 MESSAGE TRANSMISSION PROGRAM OF CAN

Message transmission, namely transforming the information into frame message with required formats, can be automatically accomplished by SJA1000 CAN based on bus protocol. While transmitting CAN message, SCM should first check the state of SJA1000 register. If the state register satisfies transmitting conditions, the data inside SCM will be sent to the transmission buffer of SJA1000. Then, the "transmission request bit" on register will be started by write command. SJA1000 can automatically set the "transmission completion bit" to 0, which means under transmitting. As long as there is another node on bus, such node will transmit acknowledgment (ACK) signal to the bus. If the ACK signal was received by bus, then the node will set to "status bit of transmission completed", which means transmitting successfully. If SJA1000 didn't receive the ACK signal, which means there is no other node on bus. Then SJA1000 will retransmit, and the error register will add one. Before each transmission, it is necessary to read the corresponding bit of state register. And the next frame data can only be loaded when transmission buffer is free. Figure 7 shows the transmission flow of CAN bus.

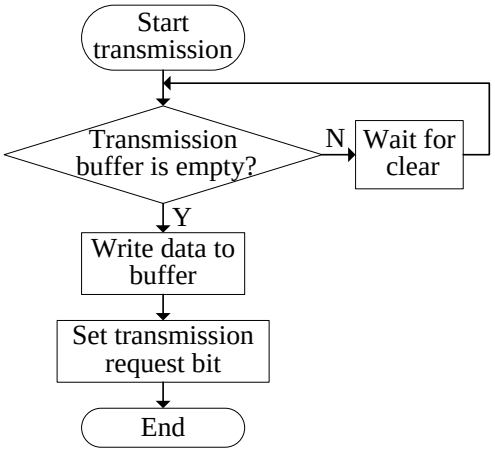


FIGURE 7 CAN bus transmission sequence

4.4 MESSAGE RECEIVE PROGRAM OF CAN

Receive subroutine is mainly responsible for reading the value of receive buffer and dealing with it. For message receive, both inquiry mode and interrupt mode can be used. This work adopts the interrupt mode, which is better than inquiry mode in terms of real-time. When the CAN bus controller SJA1000 receives a valid frame message (have passed filtering acceptance), the pin $\overline{INT}$ of SJA1000 will jump to low level, and the external interrupt 0 of microprocessor will be triggered. If the message was transmitted to microprocessor via interrupt subroutine, then the receive buffer would be released. Figure 8 shows the interrupt receive program sequence of SJA1000.

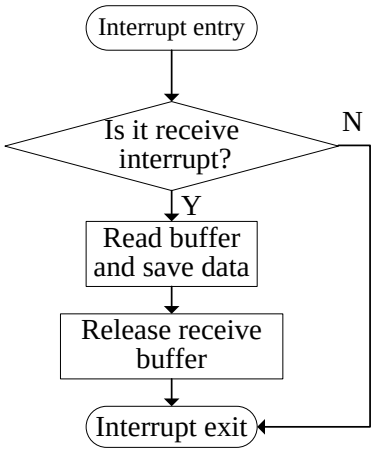


FIGURE 8 Interrupt receive program sequence of SJA1000

5 Conclusions

Due to the use of CAN bus, the system has the advantages of low-cost, small-capacity, multi-channel, high-precision, easy updating, high flexibility and strong extension ability. Analog signal acquisition front is compatible with the old industrial sensors. Therefore, it can well complete the acquisition of field information, as well as simple data processing and communication, with the features of noise immunity and low error rate. The system consists of such nodes can be used as the distributed control system, with complex working environment, various sensors and high real-time requirement.

References

[1] Cena G, Valenzanon A 1997 *IEEE Transactions on Industrial electronics* **44**(4) 553-64
[2] Hong J 2011 Technology Inc STC90C58AD series single chip microcomputer manual
[3] Burr-Brown 1997 Precision 4mA to 20mA Current Loop Receiver RCV420 data sheet
[4] Analog Devices Inc. 2004 Low Cost, Low Power Instrumentation Amplifier AD623 data sheet
[5] Xicor Inc 1996 Quad E2POT™ Nonvolatile Digital Potentiometer X9241 data sheet
[6] Philips Electronics 1997 SJA1000 Stand-alone CAN controller data sheet

Author

Xuewen Yu, born in December, 1978, Shunde, Guangdong Province, China

Current position, grades: experimentalist in Shunde Polytenchnic, China.
University studies: Master's degree in computer science and technology.
Scientific interests: computer science and technology, computer applications and software engineering.
Publications: 16 papers.

Reducer optimization design based on chaotic particle swarm optimization (CPSO)

Jiong Cai*

Panzhihua University, Panzhihua, Sichuan, 617000, China

Received 1 March 2014, www.cmnt.lv

Abstract

For the optimization design of two-stage gear reducer, an optimization mathematical model is built in this work to determine the objective functions and constraints. And chaotic particle swarm optimization (CPSO) is utilized to optimize these functions and constraints. Algorithm simulation is carried out based on CSPO algorithm steps, and the results are compared with particle swarm optimization (PSO). Simulation indicates that CSPO can optimize the results of PSO and achieve faster convergence rate.

Keywords: chaos particle swarm algorithm, optimization design, mathematical model, reducer

1 Introduction

As a conventional mechanical device, reducer is widely utilized in heavy machinery including mining machinery, construction machinery and transportation machinery. Its main role is to reduce motor speed or increase motor torque, so reducer optimization design has considerable theoretical and practical value.

2 Particle swarm algorithm

2.1 ALGORITHM PRINCIPLE

Particle swarm [4-7] optimization algorithm simulates predatory behavior of birds, make groups to achieve purpose through the collective cooperation among birds. In particle swarm optimization (PSO) algorithm, a bird is called "particle"; solving group is equivalently bird fauna; the migration from one location to another is equivalently evolution of the population; "good news" is equivalently the local optimization of population; food sources is equivalently the global optimal solution of population. In particle swarm model, the search space is D-dimension; and the total number of particles is n. Each optimization goal is the state of "particles" in the search space, including speed and position. Secondly every particle has a fitness value decided by the optimization function, and also a speed determines their flight direction and location. According to flying experience of oneself and the companions, the particles adjust dynamically the status, that is to say, update oneself through updating two positions. One is the individual best position id p found by particles themselves; another is the global best position found by entire population.

Particle swarm algorithm in operation process randomly generates an initial population and gives each particle a random speed, then update the particle speed and position according to

$$v_{id} = wv_{id} + c_1 r_1 \left(p_{id} - x_{id} \right) + c_2 r_2 \left(p_{gd} - x_{id} \right), \quad (1)$$

$$x_{id} = x_{id} + v_{id}, \quad (2)$$

$$v_{id}\left(t+1\right) = v_{id}\left(t\right) + c_1 rand_1 \left(p_{pid} - x_{id}\left(t\right)\right) + c_2 rand_2 \left(p_{gd} - x_{id}\left(t\right)\right). \quad (3)$$

Randomly generate the initial position and velocity of particle swarm, and then execute iteration according to Equations (1-3) until satisfactory solution was found. Convergence rate of particle swarm algorithm is fast; it is easy to implement and the number of the parameter needed to adjust is less. It has become a new hotspot of study in intelligent optimization evolutionary computation field. Its advantages aroused the attention of academic circles, such as simple implementation, high accuracy, fast convergence rate, strong approximation ability, and the algorithm shows its advantages in the solving actual problem.

3 Chaos particle swarm optimization

For the problem of local optimum in particle swarm algorithm, the work presents the chaos theory for improvement of PSO algorithm, and the process is as follows [8-10]:

1) In chaos initialization, supposing the variable to be optimized is D-dimensional, a D-dimensional vector $z_1 = \left[z_{11}, z_{12}, ..., z_{1D} \right]$ is randomly generated, and each component is within the range of $[0,1]$. Then *M* components are obtained according to the logistic equation [8], $z_1, z_2, ..., z_M$.

$$z_{n+1} = \mu z_n (1 - z_n), n = 0,1,2,...; 0 < z_n < 1; \mu \in [0,4]. \quad (4)$$

***Corresponding author** e-mail: jiongcaipzh@163.com

The chaotic interval will be mapped to the range of variable according to Equation (5).

$$x_{ij} = a_j + (b_j - a_j)z_{ij}, \tag{5}$$

where b_j, a_j are the upper and lower limits of optimized variable, respectively.

2) The fitness value of each particle is calculated using objective function. The *N* particle swarms with better performance are chosen as the initial solution from the *M* initial swarms, randomly generating particle velocity.

3) The initial individual and global extreme of particles are set: the current position of each particle is defined as individual extreme P_i, thereby calculating the corresponding fitness value of each individual extreme based on objective function; the position of particle with the optimal value is defined as global optimum P_g.

4) The flight speed and position of particles are updated according to velocity-position updating formula.

5) Chaos optimization is conducted on optimal position P_g: firstly, the optimal position is mapped to the defined domain of logistic equation [11] using Equation (6). Then, according to logistic equation, the iteration process generates *m* chaotic variable sequences. Finally, these sequences are mapped to the value interval of optimization variable, obtaining *m* particles. Fitness values of each particle are calculated for the optimal solution *p'*.

$$z_9 = \frac{P_g - a_i}{b_i - a_i}. \tag{6}$$

6) The current position of any particle in the swarm is substituted by *p'*.

7) The algorithm will return to Step 4 until the termination condition of particle swarm is fulfilled. Then it will stop calculating and output the results.

4 Reducer optimization design model

In this work, two-grade gear reducer is the object of design study, and the mechanism chart is shown in Figure 1 [12].

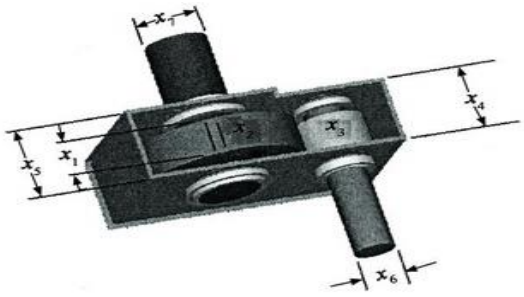


FIGURE 1 Reducer mechanism chart

This problem of design optimization has seven design variables, gear face width x_1, tooth mold x_2, tooth number of small gear x_3, bearing spacing of axis 1 x_4, bearing spacing of axis 2 x_5, the diameter of axis 1 x_6 and the diameter of axis 2 x_7. The ranges of these variables are as below:

$$\begin{aligned} &2.6 \le x_1 \le 3.6, 0.7 \le x_2 \le 0.8, 17 \le x_3 \le 28, \\ &7.3 \le x_4 \le 8.3, 7.3 \le x_5 \le 8.3, 2.9 \le x_6 \le 3.9, \\ &5.0 \le x_7 \le 5.5. \end{aligned} \tag{7}$$

The smallest volume of reducers can be found utilizing objective function.

$$\begin{aligned} \min f1(x) = {} & 0.7854x_1x_2^2(3.3333x_3^2 + 14.933x_3 - \\ & 43.0934) - 1.508x_1(x_6^2 + x_7^2) + \\ & 7.477(x_6^3 + x_7^3) + 0.7854(x_4x_6^2 + x_5x_7^2), \end{aligned} \tag{8}$$

$$A_1 = \left[\left(745x_2^{-1}x_3^{-1}x_4\right)^2 + 16.9\times10^6\right]^{0.5}, \tag{9}$$

$$B_1 = 0.1x_6^3, \tag{10}$$

$$A_2 = \left[\left(745x_2^{-1}x_3^{-1}x_5\right)^2 + 157.5\times10^6\right]^{0.5}, \tag{11}$$

$$B_2 = 0.1x_7^3. \tag{12}$$

There are 11 constraints including tooth bending stress, contact stress, axis transverse deviation, design size, etc. [13,14].

$$g_1(x) = 27x_1^{-1}x_2^{-2}x_3^{-1} - 1 \le 0,$$

$$g_2(x) = 397.5x_1^{-1}x_2^{-2}x_3^{-2} - 1 \le 0,$$

$$g_3(x) = 1.93x_2^{-1}x_3^{-1}x_4^3x_6^{-4} - 1 \le 0,$$

$$g_4(x) = 1.93x_2^{-1}x_3^{-1}x_5^3x_7^{-4} - 1 \le 0,$$

$$g_5(x) = x_2x_3 - 40 \le 0,$$

$$g_6(x) = 5 - x_1x_2^{-1} \le 0,$$

$$g_7(x) = x_1x_2^{-1} - 12 \le 0,$$

$$g_8(x) = 1.9 - x_4 + 1.5x_6 \le 0,$$

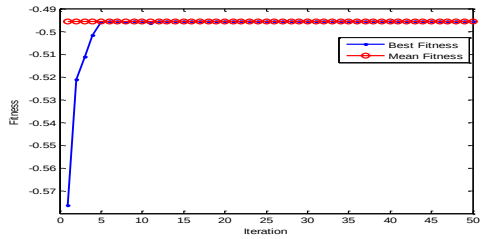
$$g_9(x) = 1.9 - x_5 + 1.5x_7 \le 0,$$

$$g_{10}(x) = A_1B_1^{-1} - 1800 \le 0,$$

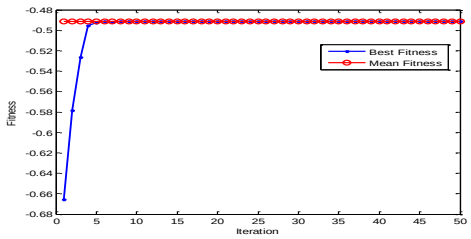
$$g_{11}(x) = A_2B_2^{-1} - 1800 \le 0.$$

5 Simulation

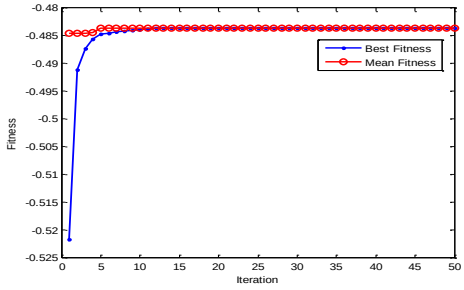
To verify the proposed algorithm, the above mathematical model is optimized and solved with the proposed method. The maximum of iteration is 50; the swarm size 20; *popmin* = –5.12; *popmax* = 5.12; *vmax* = 1; *vmin* = –1. Figure 2 shows the results of simulation.



a) Results of the first operation



b) Results of the second operation



c) Results of the third operation

FIGURE 2 CPSO simulation results

From the optimization results of CPSO algorithm, the design variables are 3.5, 0.7, 17, 7.30, 7.7153, 3.3502 and 5.2867. Figure 2 shows the convergence diagram of CPSO.

To highlight the difference between CPSO and PSO algorithms, parameters are set as follows: the maximum of iterations is 100; the swarm size 20; *popmin* = –5.12; *popmax* = 5.12; *vmax* = 1; *vmin* = –1. Figure 3 shows the convergence comparison before and after the improvement.

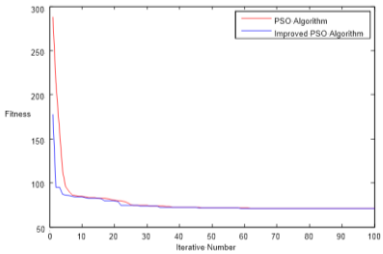


FIGURE 3 Convergence comparison of CPSO and PSO algorithm

In Figure 3, CPSO, with more stable performance, has a significantly faster convergence rate than PSO algorithm, which validates the stability and effectiveness of the proposed algorithm.

6 Conclusions

The work focuses on reducer design optimization. Chaos theory is introduced to particle swarm algorithm due to the problem of local optimum in PSO, thus proposing a CPSO algorithm for optimization design. The steps of CPSO algorithm are elaborated in detail, and the algorithm is combined with a specific case of reducer optimization design for simulation. The results indicate that the algorithm has faster convergence rate than PSO algorithm, so it is of great theoretical and practical value for engineering application.

References

[1] Donoho D L 2006 *IEEE Transactions on Information Theory* **52**(4) 1289-1306
[2] Candes E, Romberg J 2006 Quantitative robust uncertainty principles and optimally sparse decompositions *Foundations of Computational Mathematics* **6**(2) 227-54
[3] Laska J, Kirolos S, Duarte M, Ragheb T S, Baraniuk R G, Massoud Y 2007 Theory and Implementation of An Analog to Information Converter Using Random Demodulation *Proceedings of the IEEE International Symposium on Circuits and Systems 2007 (ISCAS 2007)* 1959-62
[4] Tropp J, Gilbert A 2007 *IEEE Transactions on Information Theory* **53**(12) 4655-66
[5] Donoho D, Tsaig Y 2006 Fast solution of ell-1-norm minimization problems when the solution may be sparse. Stanford University Department of Statistics Technical Report 2006-18
[6] Figueiredo M A T, Nowak R D, Wright S J 2007 Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing* **1**(4) 586-98
[7] Egiazarian K, Foi A, Katkovnik V 2007 *IEEE International Conference on Image Processing 2007* I-549-I-552
[8] Duarte M, Davenport M, Takhar D, Laska J, Sun T, Kelly K, Baraniuk R 2008 Single-pixel imaging via compressive sampling *IEEE Signal Processing Magazine* **25**(2) 83-91
[9] Wakin M, Laska J, Duarte M, Baron D, Sarvotham S, Takhar D, Kelly K, Baraniuk R 2006 *2006 IEEE International Conference on Image Processing* 1273-6
[10] Wakin Wakin M, Laska J, Duarte M, Baron D, Sarvotham S, Takhar D, Kelly K, Baraniuk R 2006 *Proceedings Picture Coding Symposium (PCS)* Beijing China
[11] Takhar D, Laska J, Wakin M, Duarte M, Baron D, Sarvotham S, Kelly K, Baraniuk R 2006 A new compressive imaging camera architecture using optical-domain compression *Computational Imaging IV at SPIE Electronic Imaging* San Jose California
[12] Haupt J, Nowak R 2006 Compressive sampling vs conventional imaging *2006 IEEE International Conference on Image Processing* 1269-72
[13] Maleh R, Gilbert A, Strauss M 2007 Sparse gradient image reconstruction done faster *2007 IEEE International Conference on Image Processing* **2** II-77-II-80

Author

Jiong Cai, born in October, 1970, Nanchong, Sichuan Province, China

Current position, grades: lecturer of Panzhihua University, China.
University studies: Master's degree in Computer Science and Technology.
Scientific interests: computer networks and communications.
Publications: 12 papers.

TV image enhancement technology based on particle swarm optimization

Lifang Yang*, Lin Liu

Chongqing College of Electronic Engineering, Chongqing, 401331, China

Received 1 May 2014, www.cmnt.lv

Abstract

Television images can be blurred and indistinct by noises in the acquisition and transmission process. Traditionally, control parameters of fuzzy enhancement algorithm are manually controlled, which leads to poor enhancement effect and efficiency. In this work, particle swarm optimization (PSO), due to its fewer parameters and global optimization capability, is combined with fuzzy enhancement algorithm for the optimization of fuzzy enhancement parameters. Simulation results show that PSO can make television images clearer and highlight in certain features, thus improving the visual effect of television images.

Keywords: particle swarm optimization, fuzzy enhancement, image processing, fitness function

1 Introduction

Television image can be blurred and indistinct by noises in the acquisition and transmission process. Television image enhancement can improve the quality of television image and highlight certain features, thus making images clearer. There are three categories of image enhancement methods, namely fuzzy processing, frequency domain method and spatial domain method [1-3]. For these methods, transit point and saturation point are manually determined, which limits their application.

2 Image fuzzy processing

In 1981, S.K. Pal, et al [4-6] proposed a new membership function and fuzzy enhancement operator to enhance image contrast, and the algorithm steps are as follows.

Step 1: Due to different purposes of image enhancement, parameters F_e, F_d, $g_{\max}$ in the membership function are adjusted according to Equation (2). The set of μ_{mn} is the fuzzy characteristic plane; g_{mn} the maximum pixel value; F_e exponential fuzzy factor; F_d reciprocal fuzzy factor. These parameters determine the size of fuzziness.

Thus, appropriate fuzzy parameters F_e and F_d can effectively enhance the quality of images. When $\mu_{mn} = G(g_c) = 0.5$, the point calls transit point. Selection of parameters has close relationship with transit point g_c as below [7].

$$G_{mn} = \begin{cases} <0.5 & g_{mn} < g_c \\ =0.5 & g_{mn} = g_c \\ >0.5 & g_{mn} > g_c \end{cases}, \tag{1}$$

F_d can be calculated through Equation (2) when transit point g_c and parameter F_e are determined.

Step 2: The image can be transformed from spatial domain to fuzzy domain by G-transformation.

$$\mu_{mn} = G(g_{mn}) = \left[1 + \frac{g_{\max} - g_{\min}}{F_d}\right]^{-Fe}, \tag{2}$$

$\mu_{mn} \to \mu'_{mn}$, the recursive correction membership of fuzzy enhancement operator, is calculated through Equation (3).

$$T(\mu_{mn}) = \begin{cases} 2\cdot[\mu_{mn}]^2 & 0 \le \mu_{mn} \le 0.5 \\ 1-2\cdot[1-\mu_{mn}]^2 & 0.5 \le \mu_{mn} \le 1 \end{cases}. \tag{3}$$

The key of fuzzy enhancement is that operators increase the membership values μ_{mn} greater than 0.5, and decrease those less than 0.5, thereby reducing the fuzziness. Then fuzzy enhancement operator will create another fuzzy set based on fuzzy set G.

Step 3: Inverse transform G^{-1} generates new gray level g'_{mn} and transforms the image from fuzzy domain to spatial domain [9-10].

$$g'_{mn} = G^{-1}(\mu'_{mn}) = g_{mn} - F_d\left[(\mu'_{mn})^{\frac{-1}{F_e}} - 1\right]. \tag{4}$$

***Corresponding author** e-mail: yanglifangcqedu@126.com

Original image

a) Original image

Enhance image

b) Fe=1,Fd=32

Enhance image

c) Fe=1,Fd=64

Enhance image

d) Fe=1,Fd=128

Enhance image

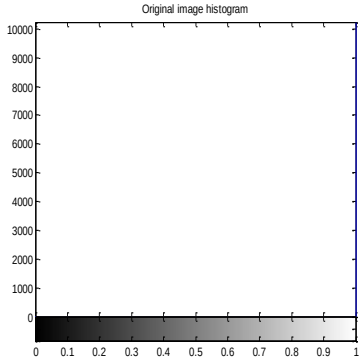
e) Fe=3,Fd=64

Enhance image

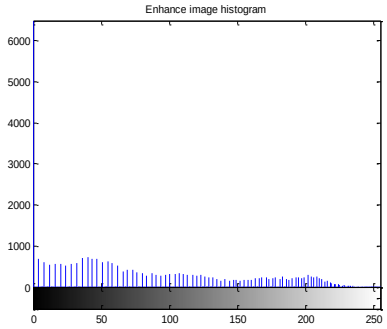
f) Fe=5,Fd=64

Enhance image

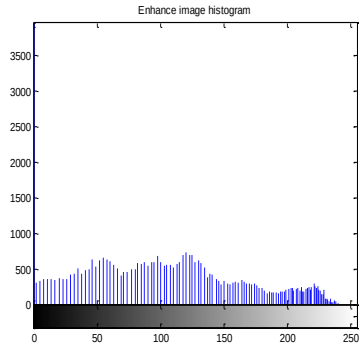
g) Fe=7,Fd=64



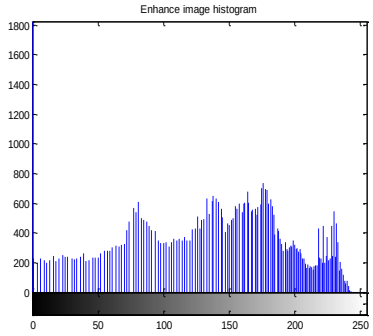
a') Original image



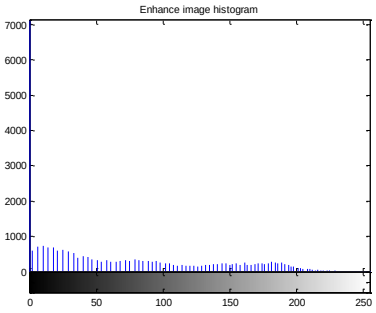
b') Fe=1,Fd=32



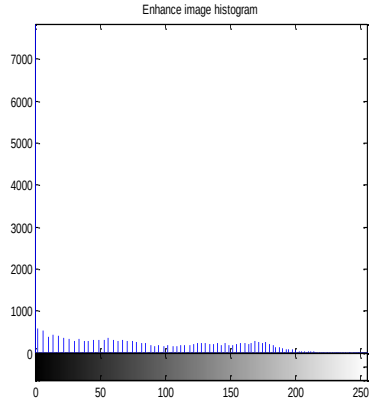
c') Fe=1,Fd=64



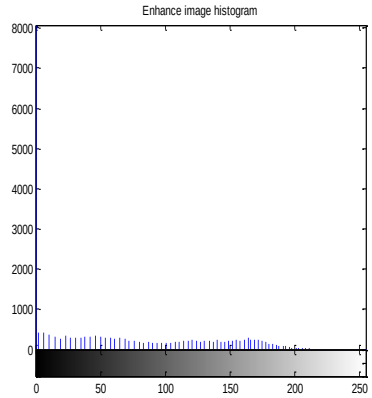
d') Fe=1,Fd=128



e') Fe=3,Fd=64



f')Fe=5, Fd=64



g') Fe=7, Fd=64

FIGURE 1 Different enhancement effect and histogram for different Fe and Fd

3 Particle swarm optimization (PSO)

3.1 ALGORITHM PRINCIPLE

PSO algorithm was firstly proposed by Kennedy, et al. for simulating the foraging behaviors of birds and achieving optimal results through collaboration and competition between groups. In PSO algorithm, each alternative solution calls a "particle", so after generating initial population, each particle is regarded as a feasible solution. Then fitness is determined by the objective function [11]. Each particle moves in the solution space, and the speed determines the direction and distance of movement.

Typically, the particle will follow the optimal particle and find optimal solution by iteration search. In each process of iteration, the particle follows two extremes to find the optimal solution for itself and the whole population.

3.2 MATHEMATICAL MODEL

For global optimization problem, a set of feasible solutions for the problem (p) is called as a population, where a

feasible solution as a particle, and population size as the number of particles.

$$(P)\min\{f(x): x\in\Omega\subseteq R^n\},\ f:\Omega\subseteq R^n\to R^l. \quad (5)$$

The *n*-dimensional vector $X_i=(x_{i1},x_{i2},...,x_{in})^T\in\Omega$ is the position of *i*-th particle, and vector $V_i=(v_{i1},v_{i2},...,v_{in})^T\in\Omega$ the speed of *i*-th particle. During particles' movement in the search space, its best position is $P_{pi}=(p_{pi1},p_{pi2,}....,p_{pin})^T$. The index number g indicates the best position of all the particles, namely P_g. Thus, the velocity of particles in each iteration and the position evaluation function can be transformed by Equations (6) and (7), respectively [12].

$$v_{id}(t+1)=v_{id}(t)+c_1 rand_1(p_{pid}-x_{id}(t))+ c_2 rand_2(p_{gd}-x_{id}(t)), \quad (6)$$

$$x_{id}(t+1)=x_{id}(t)+v_{id}(t+1), \quad (7)$$

where, $i=1,2,...,m$, $d=1,2,...,n$; $rand_1$ and $rand_2$ are subject to the distribution of $U(0,1)$; learning factors c_1 and c_2 are non-negative constants; for $v_{id}=[-v_{\max},v_{\max}]$, $v_{\max}$ is the speed limit set by users. Iterative algorithm is as follows:

Initialize population: random X_i
Repeat:
  For each particle $i\in[1,S]$
    If $f(X_i)<f(P_i)$
    $P_i=X_i$
  End
  If $f(P_i)<f(P_g)$
    $P_g=P_i$
  End
  Update the position and velocity of particle using Equations (6) and (7)
End
Until termination criterion is satisfied.

4 PSO fuzzy enhancement

4.1 Measurement of fuzzy enhancement

Main purpose of image fuzzy enhancement is to decrease information entropy, namely measuring the effect by comparing the information entropy before and after image enhancement. Entropy is defined in Equation (8).

$$H=-\sum_{i=1}^{256}p_i\log(p_i), \quad (8)$$

where p_i is the normalized histogram. Fuzzy entropy is defined in Equation (9).

$$1+\frac{1}{MN\ln 2}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[S_n(\mu_{ij})\right], \quad (9)$$

where MN is the image size; S_n is Shannon function:

$$S_n=-\mu_A(x_i)\ln(\mu_A(x_i))-(1-\mu_A(x_i))\ln(1-\mu_A(x_i)). \quad (10)$$

4.2 Fitness function

Fuzziness entropy is a parameter describing the effect of image enhancement. Only reflecting the brightness of images, it has low insensitivity to the contrast, so some improvements are needed. The fitness function of improvement is in Equation (11).

$$Fitness(\mu)=\frac{\max(\mu_{ij})-\min(\mu_{ij})}{1+\frac{1}{MN\ln 2}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[S_n(\mu_{ij})\right]}, \quad (11)$$

where, $\max(\mu_{ij})-\min(\mu_{ij})$ is the fuzzy contrast; $\max(\mu_{ij})-\min(\mu_{ij})$ are the maximum and minimum values of fuzzy feature plane, respectively. Larger $\max(\mu_{ij})-\min(\mu_{ij})$ makes the image clearer, while smaller fuzzy entropy Equation (9) makes the image clearer. Therefore, larger fitness function $Fitness(\mu)$ is good for the enhancement effect and image quality.

4.3 PSO FUZZY ENHANCEMENT ALGORITHM

A certain number of populations randomly generate during initialization, and corresponding $Fitness(\mu)$ of each population is calculated to find the maximum $Fitness(\mu)$ in the population. Then the velocity and position of particles can be updated according to PSO algorithm rules. After calculating for a given times of iteration, TV image fuzzy enhancement is conducted on parameters F_e, F_d corresponding to the obtained maximum fitness. The algorithm is as follows.

Step 1: Initialize the particle position and parameters;

Step 2: Calculate the corresponding $Fitness(\mu)$ of each population and compare the historical optimal value of individual particle with that of the population; if current value is superior to historical optimal value, then this value is retained, and the historical optimal value of individual

particle or the whole population will be updated; on the contrary, the historical optimal value is retained;

Step 3: Move particles to a new position according to the updation rules of particle position;

Step 4: If $Iteration < Maxgen$, then end the optimization; conversely, return to Step 2;

Step 5: TV image fuzzy enhancement is conducted with the fuzzy enhancement parameters F_e, F_d corresponding to the obtained maximum fitness $Fitness(\mu)$.

5 Simulation

In order to verify the proposed algorithm, the population size is set to 20, and $c_1 = c_2 = 2$, $v_{\max}$=5. Three pieces of standard test images Lena.jpg, Cameraman.jpg and Baboon.jpg are the test objects for verifying the effect of PSO image enhancement. Results are shown in Figures 2-4.

Original image

a) Lena.jpg

Enhanced image of PSO

a') Enhanced image

Original image

b) Cameraman.jpg

Enhanced image of PSO

b') Enhanced image

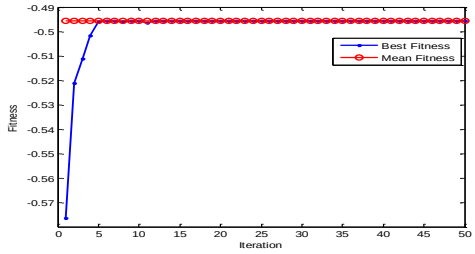
Original image

c) Baboon.jpg

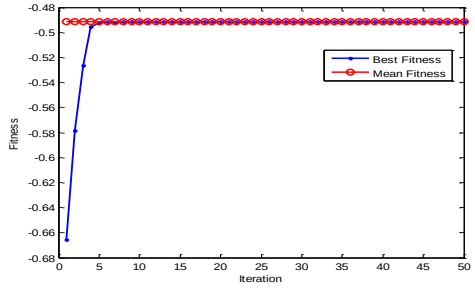
Enhance image of PSO

c') Enhanced image

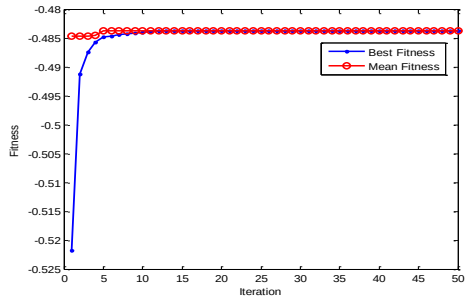
FIGURE 2 PSO fuzzy enhancement effect



a) Convergence curve of PSO fuzzy enhancement for Lena.jpg

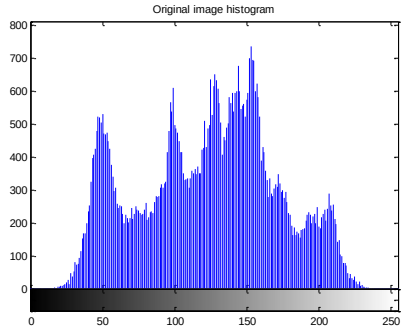


b) Convergence curve of PSO fuzzy enhancement for Cameraman.jpg

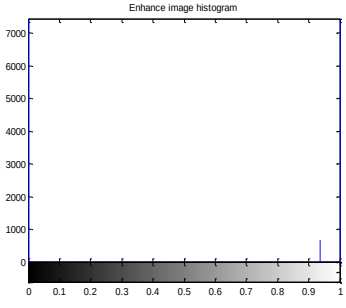


c) Convergence curve of PSO fuzzy enhancement for Baboon.jpg

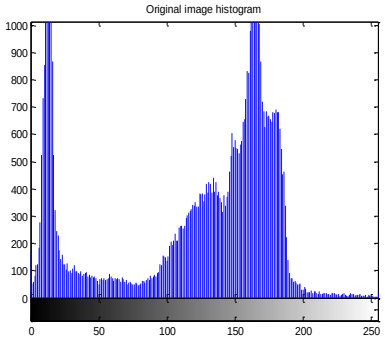
FIGURE 3 PSO Fuzzy Enhancement curve



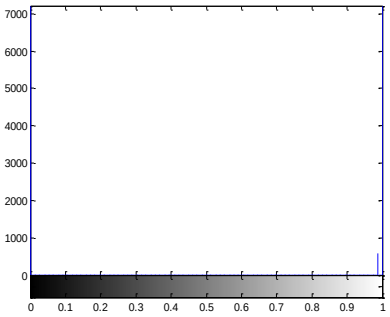
a) Initial histogram of Lena.jpg



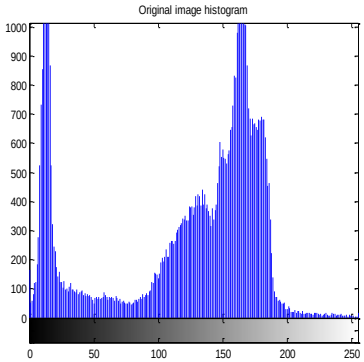
a') Histogram of enhanced Lena.jpg



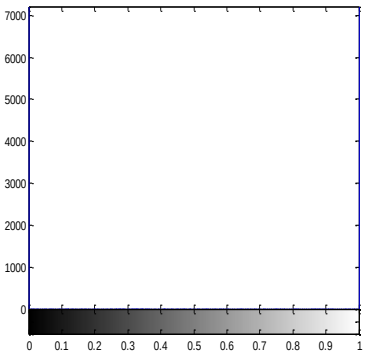
b) Initial histogram of Cameraman.jpg



b') Histogram of enhanced Cameraman.jpg



c) Initial histogram of Baboon.jpg



c') Histogram of enhanced Baboon.jpg

FIGURE 4 Comparison of the histogram before and after PSO fuzzy enhancement

Results of PSO fuzzy image enhancement prove that proposed algorithm can effectively highlight features of the image and improve visual effect. Besides, efficiency

can be improved by avoiding manual adjustment of parameters, and the optimal fuzzy enhancement parameters are set after ensuring the best quality of images.

6 Conclusions

PSO algorithm, due to its excellent search performance, is combined with image enhancement algorithm for the optimization of fuzzy enhancement parameters F_e, F_d. And appropriate fitness function is built to achieve fuzzy enhancement of television image. Simulation results show that PSO has better effect in image enhancement than traditional methods, so it has certain practical value. Meanwhile, PSO algorithm can achieve self-adaption adjustment of fuzzy enhancement parameters, thus greatly improving the efficiency.

References

[1] Li H, Yang H S 1989 *IEEE Transactions on Systems, Man and Cybernetics* **19**(5) 1276-81
[2] Peng D, WuT 2002 A generalized image enhancement algorithm using fuzzy sets and its application *Proceedings of the 1st International Conference on Machine Learning and Cybernetics* 820-82
[3] Tizhoosh H R, Krell G, Michaelis B 1997 On fuzzy enhancement of megavoltage images in radiation therapy *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems* **3** 1398-1404
[4] Pal S K, King R A *IEEE Transactions on Systems Man and Cybernetics* **11**(7) 494-501
[5] Pal S K, King R A 1983 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5**(1) 69-77
[6] Zheng C, Jiao L, Chen X, Yuan Z 2002 A new fast fuzzy processing method for B-scan image *Proceedings of the 4th World Congress on Intelligent Control and Automation* **1** 6-9
[7] Qiang P R, Meng X 2000 A method of local enhancement based on fuzzy set theory *Proceedings of the 3rd World Congress on Intelligent Control and Automation* 1751-3
[8] Tao W-B, Tian J-W, Liu J 2003 Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm *Pattern Recognition Letters* **24** 3069-78
[9] Li X, Ding R 1998 Fuzzy morphological operators to edge enhancing of images *Proceedings of International Conference on Signal Processing* **2** 1017-20
[10] Leou F-S, Wen K-A 1992 Image enhancement based on the visual model using the concept of fuzzy set *Proceedings of IEEE International Symposium on Circuits and Systems* **5** 2581-4
[11] Liu J, Tang J, Long T 2003 An Improved Fast Algorithm for Fuzzy Edge Detection *Journal of System Simulation* **15**(2) 273-4
[12] Sun W, Xia L, Pan H 2004 An Edge Detection Algorithm Based on Fuzzy Division *Journal of China Image and Graphics* **9**(l) 18-22 *(in Chinese)*

Authors

Lifang Yang, born in July, 1977, Wanyuan, Sichuan Province, China

Current position, grades: lecturer at the Chongqing College of Electronic Engineering, China.
University studies: Master's degree in computer science and technology.
Scientific interests: computer software, network multimedia and software test automation.
Publications: 15 papers.

Lin Liu, born in December, 1981, Luzhou, Sichuan Province, China

Current position, grades: lecturer at the Chongqing College of Electronic Engineering, China.
University studies: Master's degree in computer science and technology.
Scientific interests: computer software and multimedia technology.
Publications: 11 papers.

The automatic cutting algorithm of the plastic film based on image processing

Daode Zhang, Yanli Li*

School of Mechanical Engineering, Hubei University of Technology, Wuhan, China

Received 1 July 2014, www.cmnt.lv

Abstract

This paper proposes an algorithm based on image processing to solve the dislocation problem when cutting the edge, which is in the light of the uneven edge and the wrong side phenomenon. First of all, the images of the plastic bag edge captured by the camera are pre-treated; then the image is enhanced by fuzzy contrast enhancement algorithm, which is aim to extract the creases contour; finally the edge of the image is detected used by the Canny operator to get the crease on the edge of plastic bag, the cutter position can be obtained by offset some distance from the creases. The research results show that this algorithm is very effective to solve the dislocation problem when cutting the plastic bag edge automatically.

Keywords: fuzzy contrast enhancement, the Canny operator, edge detection, plastic film cutting

1 Introduction

In the industrial manufacturing field, a lot of sheet products are unable to guarantee the uniformity in the process of operation, the reasons are as follows:

1) The malalignment of the drive roller, roller, idler roller and guiding rollers all produce sheet deviation.

2) In the work process, if the sheet products are uneven and partial load, the sheet side uneven stress will be caused and the sheet offset will be resulted in.

3) The unevenness caused by the operator when they place the sheets and the unevenness of the back sheets' state can all make the sheet deviation.

4) When the sheets are coiled to the end, they have no much weight and roll both ways, which can also make the sheet deviation.

From the top, there many deviation factors in the process of coiling sheets, so adding automatic edge control system is necessary in the coiling process, which realizing the automatic alignment.

The edge mismatch and the wrong side are the problems of the plastic bag cutting in the production line, to address this phenomenon; this paper proposes a plastic film automatic cutting algorithm based on image processing to realize straight edge automatically when cutting the edge in the production line.

2 The preprocessing algorithm of plastic film automatic cutting

2.1 THE SUMMARY OF THE PREPROCESSING ALGORITHM

Image pre-processing is an important part of research based on image algorithm, effective treatment for the image can improve the resolution for the key information of the system,so then the stability and accuracy of the feature extracted will be improved, the pre-treatment technology mainly includes image filtering, image enhancement and object segmentation. The image filtering is to suppress the noise in the image, so as to reduce the influence of noise on the feature of the target and improve the stability of feature extracted; image enhancement is to make the target image more clear in the image, this is the focus algorithm in this paper, which is necessary to extract the feature points; object segmentation is to eliminate the impact of background, and improve the accuracy of feature extracted.

The images obtained from cameras are pre-processed used by the median filtering, fuzzy contrast enhancement and edge detection to get clear target contour [1].

2.2 THE REALIZATION OF THE PREPROCESSING ALGORITHM

1. Image filtering.

The objects processed in this paper is the edge image of plastic film in the production line, which have no much details, the purpose for preprocessing is that the complete edge can be extracted. The median filter is especially suitable for the situation in a strong pepper interference or

***Corresponding author** e-mail: author@domain.com

pulse type interference. Because of the interference values are very different from some of the adjacent gray-scale values of the pixels [2]. Therefore the median results after sorting is to make the interference values and some of the adjacent gray-scale values of the pixels same, which can remove the interference effect [3]. Therefore, the median filter will be taken to remove the original image in this paper.

2. Image enhancement

1) The definition of fuzzy contras t.

In recent years, many scholars devoted to introducing the contrast enhancement into the research of image processing, the concept of fuzzy contrast is introduced according to the fuzzy contrast enhancement algorithm proposed by Li Jiuxian et al.

The definition of fuzzy contrast: in the *M***N* image with *L* level grey-scales, $\mu_{ij} \in [0,1]$ ($i = 1,2,..., M$; $j = 1,2,\ldots, N$), μ_{ij} is the membership of the pixels' grey levels x_{ij} , $\bar{\mu}_{ij}$ is the membership of all the pixels gray mean values in the window, which regards the processed point as the centre, so the fuzzy contrast of the pixel x_{ij} is

$$F = \frac{|\mu_{ij} - \bar{\mu}_{ij}|}{|\mu_{ij} + \bar{\mu}_{ij}|}. \tag{1}$$

In the Equation, $|\mu_{ij} - \bar{\mu}_{ij}|$ is the absolute value of the difference between the membership of the pixels x_{ij} and the membership of its neighbour means, which can express the fuzzy contrast; F is the normalized relative fuzzy contrast, F not only considers the smoothing effects of the spatial neighbourhood averages, but also takes into account the tension of the fuzzy domain contrast [3].

2) The improved fuzzy contrast enhancement algorithm.

The improved algorithm of the image enhancement based on the fuzzy contrast is proposed on the basis of the fuzzy contrast enhancement algorithm in the paper, its realization process is as follows:

① The fuzzy enhancement algorithm based on grayscale use a fuzzy membership function that is a linear, that is:

$$\mu_{ij} = T(x_{ij}) = \frac{x_{ij}}{L-1}. \tag{2}$$

② In the Equation, *L* is the grayscale of the image pixel distribution;

③ Calculate the fuzzy image contrast with the 3*3 window;

④ Do a nonlinear transform for *F* to get

$$F' = \psi(F). \tag{3}$$

The selection of the convex function $\psi(g)$ directly affects the enhancement effect of the fuzzy contrast enhancement, if $|\psi(F) - F|$ is too small, the image details cannot be highlighted; if $|\psi(F) - F|$ is too large, the noise will be revealed.

After repeated experiments, this paper choose the polynomial function:

$$\psi(x) = 4x^3 - 12x^2 + 9x. \tag{4}$$

The following conclusion is obtained through the calculation and the verification: $\psi(x)$ is a convex function ($\psi''(x) = 24(x-1) \le 0, \quad \forall x \in [0,1]$, and $\psi(0) = 0$, $\psi(1) = 1$, $\psi(x) \ge x$.

⑤ Use F' to calculate the pixel gray membership degree and it's gray value after adjustment, the mathematical expressions is:

$$\mu'_{ij} = \begin{cases} \dfrac{\bar{\mu}_{ij}(1-F')}{1+F'} & \mu_{ij} \le \bar{\mu}_{ij} \\ 1 - \dfrac{(1-\bar{\mu}_{ij})(1-F')}{1+F'} & \mu_{ij} > \bar{\mu}_{ij} \end{cases}, \tag{5}$$

$$x'_{ij} = \mu'_{ij}(L-1). \tag{6}$$

⑥ The contrast enhancement method not only enhances the edge region but also sharpens the noise to a certain extent, therefore, the median filtering method is used to eliminate the isolated noise points [4].

In summary, the realization process is still that the image is mapped from the spatial domain to the fuzzy domain first， a local contrast operator is definited in fuzzy domain to represent the absolute value of the difference between the membership of the pixels and the membership of its neighbour means. Then, through the convex function to enhance and amplifier the difference among the neighborhood pixels; finally the image is append back to the spatial domain and is filtered, so as to complete the process of the image enhancement processing.

3) The experimental results and analysis.

Note: The red rectangular box marked in the following chart is the position of plastic film edge, the following enhanced image is enhanced used by the fuzzy enhancement two times in order to extract the feature of the edge folds solely.

FIGURE 1 Before enhancing (green)

FIGURE 2 After enhancing (green)

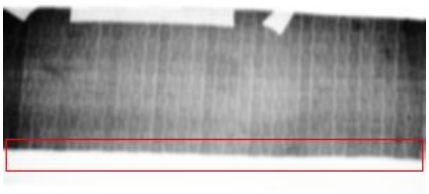


FIGURE 3 Before enhancing (yellow)

FIGURE 4 After enhancing (yellow)

FIGURE 5 Before enhancing (red)

FIGURE 6 After enhancing (red)

FIGURE 7 Before enhancing (white)

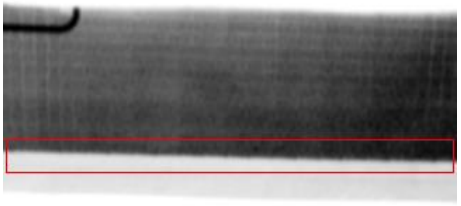


FIGURE 8 After enhancing (white)

By above knowable, after enhancing with fuzzy contrast**(**see Figure 2, Figure 4, Figure 6, Figure 8), the image contrast enhances significantly at the crease and decreases significantly at the edge and background, which lay the foundation for the crease extraction behind.

The image contrast enhancement method is that the spatial domain is enhanced everywhere by enhance the gray difference between neighborhood pixels by locally. But it not only enhances the contrast in the crease region but also sharpens the noise to a certain extent, so the image should be smoothed after enhancing, so as not to affect the quality of the feature extraction.

3. Image edge detection.

Edge detection is the fundamental topic of image processing and computer vision, the study of the human visual system show that the image edge is particularly important and an object can be identified only by a sketchy outline [7].

The essence of edge detection is to extract the boundary between a object and its background image using some algorithm. The process of general edge detection algorithm is that first the smoothing image is obtained by smoothing the original image; and then the image whose edge being enhanced using a variety of algorithms, whose gray levels is 256.

At this time, the flat grey regions disappears and the mutational grey regions remains alone in the image whose gray is enhanced, the 256 levels image is becomed to the binary image after the final threshold segmentation, the edge image is got if the mutant edges are displayed obviously [5].

According to the characteristics of the target image and the requirements of the edge feature, the edge is extracted by Canny operator in this paper, the selection of thresholds should be paid special attention to. After many experiments analysis, the ratio of the Canny operator's two parameter threshold is 2:3 in view of the identification target [6]. The experimental results are as follows:

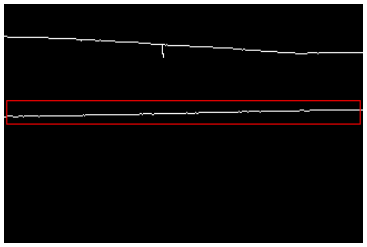


FIGURE 9 Edge detection (green)

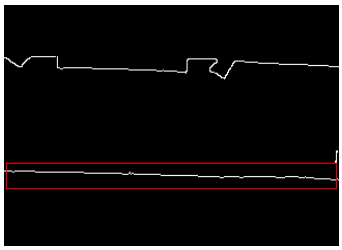


FIGURE 10 Edge detection (yellow)

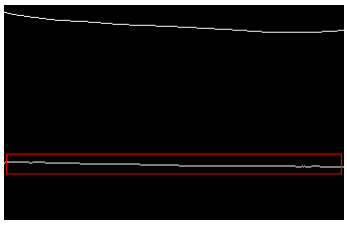


FIGURE 11 Edge detection (red)

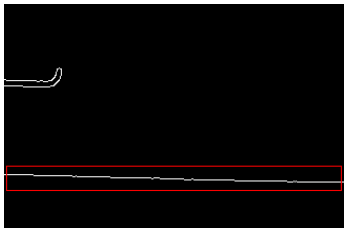


FIGURE 12 Edge detection (white)

3 The automatic cutting algorithm of the plastic film

3.1 THE SUMMARY OF AUTOMATIC CUTTING ALGORITHM

The accurate position of the cutter must be controlled timely in order to realize the automatic slitting of the plastic film. The crease contouring near the edge is obtained by the preprocessing of the plastic film image, including median filtering, image contrast enhancement, edge detection, then the cutter position will be obtained by offsetting some distance according to crease position.

3.2 THE RELIZATION OF THE AUTOMATIC CUTTING ALGORITHM

1. Select the reference position

A reference position needs to be selected as the standard in order to make the cutter regular while cutting the edge. Through the analysis of the plastic film image obtained by the backlight and the camera. The plastic film edge fold position is chose to be the standard position and then the position of the cutter will be obtained by offsetting some distance from the edge if the creases position is got in the image, so as to avoid the uneven edge [8].

2. Obtain the reference position.

A clear outline of the plastic film edge creases is obtained after pretreatment. The next is to determine the position of the crease. The average value of all feature points at the crease is selected as the crease's accurate position.

3. The steps of the algorithm:

1) A region of interest (ROI) is arranged at the crease in the image which includes all feature points in the region of interest.

2) Obtain the coordinate values of each feature point by traversing the ROI area and calculate their average value.

3) Traverse the image again and set a threshold according to the requirements. Then calculate the difference between the coordinate value of each feature point and the average value. If the difference is greater than the threshold, the feature point will be dropped; otherwise it will be remained.

4) Calculate the average coordinate value of each feature point that remained and regard it as the crease position.

5) The cutter position can be obtained by offsetting the distance after getting the standard position of the crease of the plastic film.

6) Select a scale by calibrating the camera and use the scale to transform the pixel position in the image into the actual distance of the controller.

3.3 THE RESEARCH RESULTS AND ANALYSIS

1. The experimental results:

Note: the red rectangular box is the ROI region which contains the crease, the green line is the reference position calculated by the upper algorithm.

FIGURE 13 Crease position (yellow)

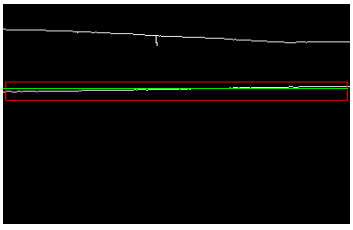


FIGURE 14 Crease position (green)

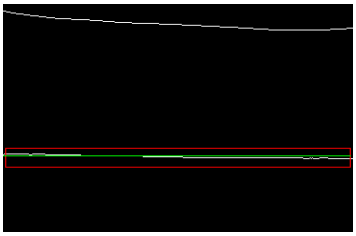


FIGURE 15 Crease position (red)

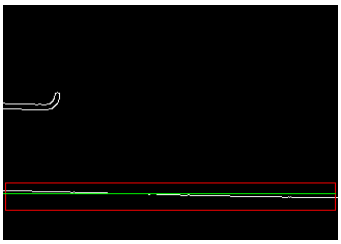


FIGURE 16 Crease position (white)

2. The analysis of the results

The crease position can accurately obtained by the algorithm proposed in this paper. It should be noted that the noise points that offset a long distance from the crease position will affect the accuracy of the algorithm, so it is necessary to eliminate the noise in the third step.

4 Conclusions

This paper aims at the deviation phenomenon of the plastic film automatic cutting in the production line, a lot of manpower and material resources are wasted because of artificial control deviations in the actual production, so the automatic cutting algorithm of the plastic film based on image processing are put forward which resolves the running deviation while cutting the edge and the wrong side phenomenon [9].It has the characteristics of high efficiency and high precision compared with manual visual method. The paper mainly completes the following work:

1. Preprocess the image of plastic film edge. Firstly, make the gray processing for the plastic film; then do smoothing for the processed image; finally the clear edge image can be obtained by the edge detection after getting relatively clear image.

2. Obtain the crease position of the plastic film edge. The clear outlines of the plastic film edge creases are got after pretreatment and the average value of all feature points at the crease is selected as the accurate position of the crease.

3. Get the actual position of the cutter by offsetting some distance from the position of the crease. The location of the cutter can be obtained by making the cutter offset the distance from the crease after getting the standard distance between the edge and the crease; the actual distance of the controller can be obtained by selecting a scale by calibrating the camera and use the scale to transform the pixel position in the image into [10].

The algorithm proposed in the paper also has shortcomings, when the noise points is too much nearby the crease of the plastic film edge image, which is easy to have a great influence on the reference position ,so the preprocessing work is very important.

Acknowledgments

This work is supported by Natural Science Foundation of Hubei Province (No. 2013CFB034)

References

[1] Zhou Z 2012 The algorithm and applied research of the image enhancement (D) *Beijing Beijing University of Technology* (*in Chinese*)
[2] Yang Shuyin 2003 Program Design on VC++ Image Processing(the Second Edition) *Tsinghua University Press* (*in Chinese*)
[3] Cao Yan 2013 Several contrast enhancement algorithms for low illumination images *Journal of Jingdezhen college* **28**(6) 11-2 (*in Chinese*)
[4] Zhao J, Zhao Y 2014 The image contrast enhancement algorithm based on histogram *Electronic technology* **27**(1) 49-51 (*in Chinese*)
[5] Li Xue 2007 Performance evaluation of edge detection algorithm of the gray image (D) *Shenyang: Shenyang University of Technology* (*in Chinese*)
[6] Zhang S 2013 The research and application of edge detection algorithm based on machine vision (D) *Shanghai: Shanghai Jiao Tong University* (*in Chinese*)
[7] Brink AD 1992 Threshold of Digital Image Using Two-dimensional Entropies *Pattern Recognition* **l8**(10)80-3
[8] Ye S, Zhu J 1999 Visual Inspection Technology and Application *China Engineering Science* 8 22-34 (*in Chinese*)
[9] http://blog.csdn.net/watkinsong/article/details/10212715
[10] Liu J 2007 Research Based on visual STM chip pin detection system *China Mechanical Engineering* 17(16)) 1908-12 (*in Chinese*)
[11] Ma Y, Yang Y 2001 Self-organizing RBF neural network to identify the driver's active safety factors *Control and Decision* 16(1)114-6 (*in Chinese*)

Authors

Daode Zhang, born on August 17, 1973, Wuhan Hubei

Current position, grades: professor.
University studies: doctor Huazhong University of Science and Technology
Scientific interest: mechanical and electrical integration.
Publications: 13 papers, 1 book.

Yanli Li, born on December 21, 1988, Xiangyang Hubei

Current position, grades: master student at Hubei University of Technology.
University studies: Hubei University of Technology.
Scientific interest: digital image processing

The method of measurement for buried pipeline centerline based on data fusion

Rui Li[1, 2*], Qingshan Feng[1], Maolin Cai[2], Haijun Li[3], Chenghai Liu[1], Xiaoming Zhao[1]

[1]*Petrochina Pipeline Company, Langfang 065000, China*

[2]*Automation Science and Electrical Engineering Collage of Beihang University, BeiJing, 100000, China*

[3]*China aviation thirty-three Department, Beijing, 100000, China*

Received 1 October 2014, www.cmnt.lv

Abstract

For the problem of long-distance oil and gas pipeline centerline measurement, a method of multi-sensor data fusion using the pipe centerline calculation is presented. The error model is set up by the system of navigation and nonlinear dynamic systems. Using the data of IMU and odometer to calculate the information of pig navigation. All of the errors were calculated by Kalman filter for estimation and compensation. It is concluded that the error for horizontal is 0.35m, the error for vertical is 0.74m for comparison of same pipeline centerline in different inspection time. One feature point is dug to verify the accuracy of the inspection which error is less than 1m. This method is effective for the buried pipeline to perform safely.

Keywords: in-line inspection of long distance pipeline, sensor fusion, inertia navigation, extended Kalman filter, error correction

1 Preface

With the continuous development of domestic oil and gas industry [1], more and more long-distance buried pipelines are used in the transportation of oil and gas products. Due to factors such as geological disasters that will cause pipe displacement or pipe deformation, thus great bending strain will be generated in local pipe body that which will result in additional stress leading to serious instability or material damage to the pipeline. To avoid pipeline failure accidents, Periodically and accurate pipeline centerline positioning measurements are needed in order to determine the specific location of the pipeline, and analysis situations such as pipe displacement, deformation and so on by calculating and comparing historical data, thereby making timely repair focusing on large deformation zone.

Pipeline inspection is the basic method to ensure pipeline safety and In-line Inspection is recognized as the most effective detect means of pipeline safety. Now, large foreign pipeline companies have been able to detect and analyze pipeline position using ILI tools mounted with IMU, due to the monopoly of foreign technology companies on this technology, the domestic research in this area started relatively late, and most research are still in the experimental stage.

In this paper, high-precision inertial measurement unit (IMU) oriented to long-distance gas pipeline centerline measurements was developed, and pipeline centerline location information was calculated through the use of multi-sensor data fusion technology, and a variety of errors generated were corrected by the sensors using the extended Kalman filter, measurement locus error correction was made by referencing Marker point information, and finally get a more accurate pipeline centerline.

2 Measurement systems in ILI

As shown in Figure 1, the Inertial Measurement Unit (IMU) was mounted within the sealed pipeline inspection gauge (PIG) [2-5], in order to the post processing of geographic information data, the master clock in IMU must sync to the GPS clock in above ground marker prior to the inspection. PIG move forward under the driven by the impetus of oil or gas in the pipeline. IMU system is composed of three-axis orthogonal high precision laser gyroscope and three-axis orthogonal accelerometers. Since the inertial navigation system error and odometer offset error accumulated over time, leading to the data migration, so transmitter with the ability to transmit low frequency signal to trigger the above ground GPS marker was also installed besides the odometer. By doing so, using these devices you can do inertial and non-inertial sensor information fusion to calibrate the pipe centerline position calculated from the PIG, Thereby obtaining high-precision pipe centerline data.

[*] ***Corresponding author*** e-mail: kjlimi@petrochina.com.ch

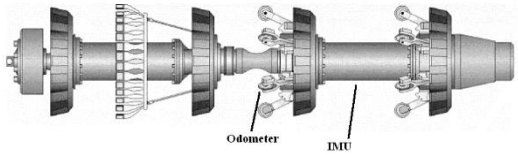


FIGURE 1 Measurement system for In-line inspection of Pipeline centerline

3 Pipeline centerline data fusion algorithm

When run in the pipeline, data collected by all sensors in PIG is stored in the built-in memory. Every movement was collected by the inertial navigation unit including the angular velocity and acceleration in three directions and so on. Odometer measure the mileage of the PIG, thus obtain speed of the PIG to compensate error of the inertial navigation unit [6-10]. The above ground GPS marker receive low frequency signals generated by the transmitter to obtain the exact current location information that error compensation can be done by post-processing. This paper adopt the extended Kalman filter to process the internal attitude, velocity and position information obtained by the PIG, Simultaneously merge the non-inertial odometer data and the ground coordinate data with inertial data [11-13]. The calculation process adopt both the state of forward and backward extended Kalman filter solver system and fix generated system errors. Where in the forward filter is used to calculate $\hat{x}_k$ the current state of the PIG, the backward filter is used to estimate $\delta\hat{x}_k$ the system state amount errors. Therefore the system current revision status is $\hat{x}_{C_k} = \hat{x}_k - \delta\hat{x}_k$. Pipeline centerline data fusion algorithm block diagram is shown in Figure 2.

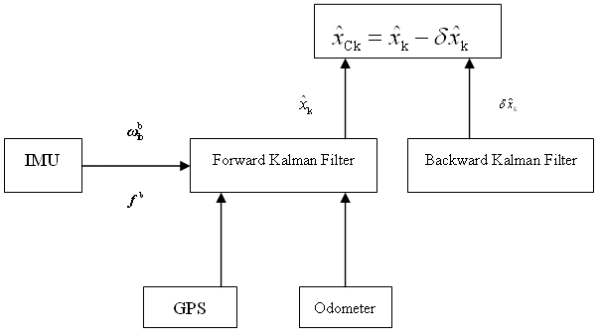


FIGURE 2 Data fusion algorithm diagram of pipeline centerline

3.1 INERTIAL NAVIGATION ALGORITHM

Strapdown inertial navigation system (SINS) and the PIG speed, position dynamic equations are as follows:

$$\dot{V}_e^n = C_b^n f^b - (2\omega_{ie}^n + \omega_{en}^n) \times V_e^n + g_l^n, \tag{1}$$

$$\dot{q} = \frac{1}{2} q \otimes \omega_{nb}^b, \tag{2}$$

$$\dot{L} = \frac{v_N}{(R_N + h)}, \tag{3}$$

$$\dot{l} = \frac{v_E}{(R_E + h)\cos L}, \tag{4}$$

$$\dot{h} = -v_D, \tag{5}$$

where $V_e^n = [v_N \ v_E \ v_D]^T$ is the velocity vector of the PIG, g_l^n is the gravitational acceleration, $q = [q_0 \ q_1 \ q_2 \ q_3]^T$ is the four elements, f^b is collected force ratio value, L and l are respectively latitude and longitude h is the elevation. R_E and R_N are the earth radius under the WGS1984 geography coordinator system.

C_b^n is the transmit matrix from the PIG carrier coordinate system into the local navigation coordinate system that can be obtained by calculating the four elements [14-17], that is:

$$C_b^n = \begin{bmatrix} q^2{}_0 + q^2{}_1 + q^2{}_2 + q^2{}_3 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q^2{}_0 - q^2{}_1 + q^2{}_2 - q^2{}_3 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q^2{}_0 - q^2{}_1 - q^2{}_2 + q^2{}_3 \end{bmatrix}, \tag{6}$$

where ω_{ie}^n is the Earth angular velocity in the navigation reference, ω_{en}^n is the angular rate relative to the Earth navigation reference, ω_{nb}^b is the platform's angular rate of the navigation system. These angular rate expressions are:

$$\omega_{ie}^n = \left[\Omega\cos L \quad 0 \quad -\Omega\sin L\right]^T, \tag{7}$$

$$\omega_{en}^n = \left[\frac{v_N}{R_E + h} \quad -\frac{v_N}{R_N + h} \quad -\frac{v_N}{R_E + h}\tan L\right]^T, \tag{8}$$

$$\omega_{nb}^b = \omega_{ib}^b - C_n^b\left(\omega_{ie}^n + \omega_{en}^n\right). \tag{9}$$

Constant Ω is the Earth's rotation angular rate. Vector $\omega_{nb}^b = [\omega_{nbx}^b \ \omega_{nby}^b \ \omega_{nbz}^b]^T$ is the angular rate of the carrier system to the Geography coordinate system using to update the attitude angle is system. The inertial navigation coordinates of the pipeline centerline can be calculated by integrating all the data in Equations (1)-(9). As the pure inertial navigation results has an error accumulation characteristic, and therefore require additional auxiliary sensor information to eliminate the accumulated error.

3.2 DEAD RECKONING ALGORITHM

Dead reckoning algorithm utilizes the carrier's heading and attitude angle provided by the IMU with odometer speed to calculate the carrier's relative position to the starting point. Assuming $v_m^b = [0 \ v_m \ 0]^T$ is the projection of the odometer output speed in the carrier

coordinate system, where the speed of the navigation odometer system can be expressed as: $v_m^n = C_b^n v_m^b = [v_{mN}^n \quad v_{mU}^n \quad v_{mE}^n]^T$, then the position information calculated using dead reckoning algorithm is:

$$\begin{cases} \dot{L} = v_{mN}^n / (R_M + h) v_m^n \\ \dot{\lambda} = v_{mE}^n \sec L / (R_M + h). \\ \dot{h} = v_{mU}^n \end{cases} \tag{10}$$

3.3 ERROR EQUATION OF NAVIGATION AND INERTIAL SENSORS

Run track of the PIG can be obtained by linearization the speed equation, and the speed error can be expressed as:

$$\delta \dot{v}^n = f^n \times \Psi + C_b^n \delta f^b - (2\omega_{ie}^n + \omega_{en}^n) \times \delta v^n - (2\delta\omega_{ie}^n + \delta\omega_{en}^n) \times v^n - \delta g_l, \tag{11}$$

$$\dot{\Psi} = -\omega_{in}^n \times \Psi + \delta\omega_{in}^n - C_b^n \delta\omega_{ib}^b, \tag{12}$$

where: $\delta v^n = [\delta v_N \quad \delta v_E \quad \delta v_D]^T$ is speed error vector of reference navigation $\Psi = [\delta\alpha \quad \delta\beta \quad \delta\gamma]^T$ describes the inclination and azimuth vectors, which can be approximately expressed as a smaller angle offset.

This gives a navigation error in the navigation coordinates:

$$\delta L = \frac{1}{R_N + h} \delta v_N - \frac{v_N}{(R_N + h)^2} \delta h, \tag{13}$$

$$\delta l = \frac{1}{(R_E + h)\cos L} \delta v_E + \frac{v_E \sin L}{(R_E + h)\cos^2 L} \delta L - \frac{v_E \sin L}{(R_E + h)^2 \cos L} \delta h, \tag{14}$$

$$\delta h = -\delta v_D, \tag{15}$$

where δL, δl and δh are the latitude, longitude, and elevation error.

3.4 ERROR EQUATION OF NON-INERTIAL SENSORS

Data fusion algorithm is used to process the inertial unit data and other non-inertial sensors data, in order to decrease inertial error brought by inertial navigation. The non-inertial sensors here are odometer and above ground GPS marking boxes. Continuous odometer data can be converted into an average speed used to represent v_m, the instantaneous speed of PIG. Assuming odometer ratio error factor is k_{fe}, and then the speed of the PIG can be expressed as:

$$v_m^b \approx v_x^b - k_{fe} v_x^b, \tag{16}$$

where v_x^b is the velocity component in the *x* direction of the PIG, v_v^n is the white Gaussian noise with zero mean generated in the measurement of the velocity. Therefore, error rate in the navigation frame can be expressed as:

$$\delta v_m^n = \hat{v}^n - C_b^n v_m^b = \delta\hat{v}^n + v_m^b k_{fe} + v_v^n. \tag{17}$$

The above ground marker (AGM) is used to receive the low frequency signal of transmitter, and can provide position information of the PIG. The position error model is:

$$\delta p_m = \hat{p}^n - p_m^n = \delta\hat{p}^n + v_p, \tag{18}$$

$\hat{p}^n$ is the estimated location coordinates (latitude and longitude, elevation), p_m^n is location coordinates vector, v_p is the zero mean Gaussian white noise generated by measuring the position.

3.5 EXTENDED KALMAN FILTER

The error of inertial navigation results is an accumulating amount over time, and the error will greatly affect the output accuracy of navigation and positioning system. Navigation and positioning algorithm is an iterative calculation, If the calculation errors cannot be corrected, Navigation and positioning system will not accurately reflect the operation situation of the PIG. So, after analysis all the errors generated by the system extended Kalman filter is used for the system error correction.

As the real-time state of the system is obtained by a series of non-linear formula, so the extended Kalman filter is used to calculate. The basic idea of Extended Kalman Filter is to distribute processing first, and then does the global integration; obtain global estimates based on all observations information data, it integrates speed, location measurement of INS and odometer, above ground GPS location marker boxes information. Extended Kalman filter is described by the following equation:

1) state and output equations

$$x_{k+1} = f(x_k, u_k, k) + w_k, \tag{19}$$

$$z_k = h(x_k, k) + v_k. \tag{20}$$

2) the state estimates partial derivative matrix after each calculation the time interval ($t = kT$) by the filter is:

$$A_k = \frac{\partial f(x_k, u_k, k)}{\partial x_k}\Big|_{x=\hat{x}_k}, \quad H_k = \frac{\partial h(x_k, k)}{\partial x_k}\Big|_{x=\hat{x}_k}. \tag{21}$$

3) filter update equation:

$$K_k = P_k H_k^T (H_k P_k H_k^T + R)^{-1}, \tag{22}$$

$$\hat{x}_{k+1} = f(\hat{x}_k, u_k, k) + K_k(z_k - h(\hat{x}_k, k)), \tag{23}$$

$$P_{k+1} = A_k (I - K_k H_k) P_k A_k^T + Q , \tag{24}$$

where v_k is the measurement error vector, z_k is the measurement vector.

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \end{bmatrix} = \begin{bmatrix} 0\,0\,0\,0\,1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,1 \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ v_N \\ v_E \\ v_D \\ L \\ l \\ h \end{bmatrix}_k + \begin{bmatrix} \delta_{v_N} & 0 & 0 & 0 & 1 & 0 \\ 0 & \delta_{v_E} & 0 & 0 & 0 & 1 \\ 0 & 0 & \delta_{v_D} & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_L & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_l & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta_h \end{bmatrix} \begin{bmatrix} v_{v_N} \\ v_{v_E} \\ v_{v_D} \\ v_L \\ v_l \\ v_h \end{bmatrix}_k . \tag{25}$$

By Equation (19)-(25), a system state equation was obtained from the system error model, which establishes the relationship between the speed, position error and the other error amount, between the previous error amounts. All the errors can be estimated through the forward Kalman filter, and then obtain the values of each state amount.

3.6 THE BACKWORD KALMAN FILTER

Filtering relevant information can be saved in the forward filtering and the error of each estimate can be got in the backward Kalman filtering, calibrated result $\hat{x}_{C_k}$ can be calculated finally. The backward Kalman filter using the following formula [18-20].

Assuming the estimated value and variance of each state are $\bar{x}_{k+1}$ and $\bar{p}_{k+1}$ at time $k + 1$, then: single stepping state predictive value:

$$\hat{x}_{k+1,k} = \phi_{k+1,k} \bar{x}_k . \tag{26}$$

Single stepping state prediction error covariance matrix

$$\hat{p}_{k+1,k} = \phi_{k+1,k} \bar{p}_k \phi_{k+1,k}^T + \Gamma_{k+1,k} Q_k \Gamma_{k+1,k}^T . \tag{27}$$

Time filter gain matrix at time *K*:

$$J_k = \bar{p}_k \phi_{k+1,k} (\phi_{k+1,k} \bar{p}_k \phi_{k+1,k}^T + \Gamma_{k+1,k} Q_k \Gamma_{k+1,k}^T)^{-1} = \hat{p}_k \phi_{k+1,k} \bar{p}_{k,k-1}^{-1} . \tag{28}$$

State estimation error at time *K*:

$$\delta \hat{x}_k = J_k (\hat{x}_{k+1} - \hat{x}_{k+1,k}) . \tag{29}$$

The estimated status value at time *K*:

$$\hat{x}_{C_k} = \hat{x}_k - \delta \hat{x}_k . \tag{30}$$

4 Field validation and data analysis

4.1 FIELD TEST EQUIPMENT AND PERFORMANCE INDICATORS

Most domestic oil pipeline are heating transportation, and pass through mountain, hills, rivers and other complex environmental areas, the technical and safety requirements of electrical equipment of which is extremely strict. To be able to safely detect the actual pipeline, IMU should not only meet the technical performance, but also consider the actual situation of the pipe and the external environment in order to ensure the detection successfully and safely conducted. Technical performances of inertial devices used in the field test are shown in Table 1; the environmental targets need to meet are shown in Table 2.

TABLE 1 Characteristic of Sensors

Sensor	Characteristics	Magnitude
Gyroscope	bias	<0.01°/h
	Random walk	0.002° / $\sqrt{h}$
Accelerometer	Bias Stability	<50ug
	Scaling factor	<50ppm
Odometer	Scaling factor	<0.3%
	White noise	<0.1m/s
Landmark	White noise	<±1m

TABLE 2 Characteristics of Circumstances

Characteristics	Magnitude
Temperature	0~60℃
Vibration	6g（RMS）
Impaction	30g
Pressure	3MPa

IMU inertial navigation unit is mounted on the PIG shown in Figure 1. The data collected by the system is stored in the built-in memory. The data in the built-in memory is downloaded at the end of the field test via the communication cable and then carry out post-processing.

4.2 DATA ANALYSIS

Two centerline measurement field tests were conducted on one crude oil pipeline of PetroChina Pipeline Company, and the two tests' relevant information are basically the same, here use the correlation results of the second trial as test instructions. Throughout the test, the PIG runs at an average speed of 1.16m/s in the pipe, the maximum speed is 1.27m/s, and the entire trial took about 54 hours and 15 minutes. An above ground GPS Marker is set just on the ground above buried pipeline centerline every 2Km to obtain the time that PIG pass through the Marker to perform error correction on the measurement results.

Integrate the IMU data, odometer data and the above ground GPS Marker data using the data fusion algorithm describes in this article can obtain a more accurate pipeline centerline. The following uses the first 30km trial data as an example of the data analysis, Figures 3 and 4 show the track and height of the pipeline centerline during inspection obtained by the PIG. As can be seen from the figure, the height of pipeline changes significantly and fluctuated considerably, while the PIG's speed traveling in the pipe is more stable and run in a relatively stable state.

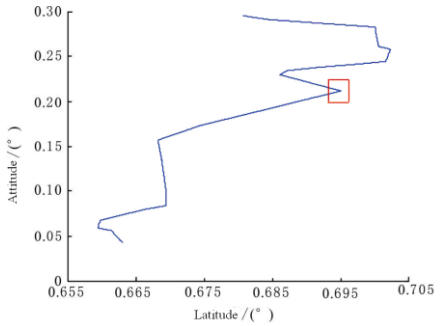


FIGURE 3 Curve of pipeline for measurement of In-line inspection

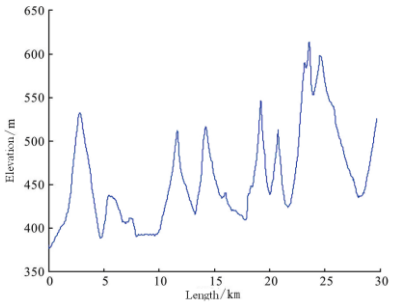


FIGURE 4 Height of the curve for measurement of In-line inspection

The twice measurement results of this pipe section were comparative analysed to calculate the repeatability between the two measurements. Since using 300m as an error analysis unit that there will be a total of 100 points in the first 30Km. As can be seen from Figures 5 and 6, the error curve and the error distribution of the two measurements as follows, the error between the two obtained pipeline centerline is small. The error statistics shows that the error in the horizontal direction is 0.35m (1σ), while in the height direction is 0.74m (1σ).

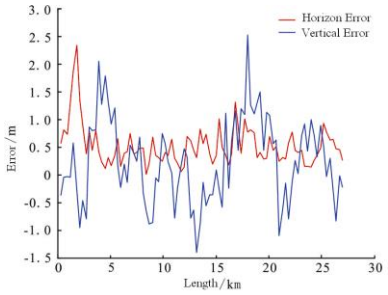


FIGURE 5 Horizon and height error for two measurements

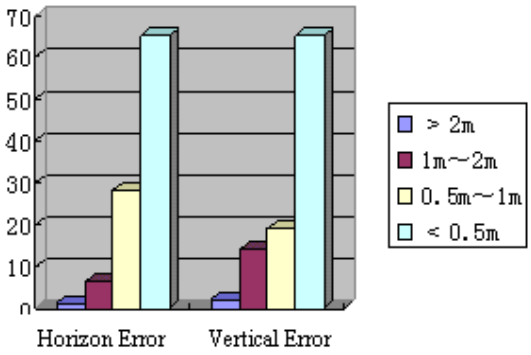


FIGURE6 Horizon and height error of the measurement distribution for two measurements

4.3 EXCAVATION VERIFICATION OF FEATURE POINTS

To further validate the accuracy of the measurement results, select the middle point of the right bend in Figure 3 as excavation verification point to verify the accuracy of IMU measurements. The location point calculated by IMU measurement result locates 0.4 meters downstream the target point while locates 0.7 meters right side in the vertical direction of the oil flow. This error is equal to the analysed repeatability error between two measurements in last section. From this we can see that the accuracy of IMU can achieve to less than 1m which meets the accuracy requirements in actual work.

5 Conclusion

1) This paper studied a buried pipeline centerline measurement method based on inertial navigation technology and multi-sensor data fusion, integrating inertial navigation systems, odometer data and ground GPS data to calculate the coordinates of buried pipeline centerline;
2) Through the use of high-precision, strong impact vibration resistance devices, improved system's security, reliability, and successfully applied to the site inspection of actual crude oil pipeline;
3) All the errors generated in the inspection process were effectively inhibited by this method and the accuracy of the measurement results was improved. Through repeatability error analysis between the two measurements and feature points excavations verify shows the actual positioning accuracy of PIG was within 1m.
4) The above ground GPS marker boxes trigger rate is not high due to factors such as the depth, conductivity of soil and so on. To improve GPS Marker boxes trigger rate then further improve positioning accuracy, a suitable ground tracking technology should be further researched.

References

[1] Qi A 2009 Chinese oil and gas pipeline transport development status and analysis of associated problems International Petroleum Economics **17**(12) 57-60 *(in Chinese)*

[2] Yang J, Li R, Gao S, et al. 2012 Navigating and positioning technique for inner detection of pipeline Journal of Shenyang University of Technology **34**(4) 427-32 *(in Chinese)*

[3] Zhang Y, Wang L, Zhan X, et al 2008 The New Advancement and Trend of Inertial Navigation Technology *Shipbuilding of China* 2008 **49**(S1) 134-44 *(in Chinese)*

[4] Liu B Qi Y, Yang Y 2010 Application of Kalman filtering in pipeline geographical coordinates positioning system *Journal of Shenyang University of Technology* **32**(5) 564-8 *(in Chinese)*

[5] Yue B, Tang Y, Zhang Y, et al 2008 Application of integrated navigation technology in inertial pipeline surveying system *Journal of Chinese Inertial Technology* **16**(6) 712-6 *(in Chinese)*

[6] Yu J, Lee J G, Park C G 2005 An off-line navigation of a geometry PIG using a modified nonlinear field-interval smoothing flitter *Control Engineering Practice* 13(11) 1403-11 *(in Chinese)*

[7] Hart J D, Powell G H, Hackney D, Zulfiqar N 2002 Geometry Monitoring of the Trans-Alaska Pipeline *Cold Region Engineering* 110-21

[8] Czyz J A, Falk J 2000 Use of Geopig for Prevention of Pipeline Failures in Environmentally Sensitive Areas *The Pipeline Pigging* Integrity Assessment and Repair Conference Houston February 2000

[9] Simmons G G, Alto J V 1988 Pipeline Curvature by Polynomial Approximation *Proceedings of the Conference on Pipeline Infrastructure* ASCE Pipeline Division Boston Massachusetts June 1988.

[10] Paeper S, Brown Bm Beuker T 2006 In-line Inspection of Dents and Corrosion Using "High Quality" Multi-Purpose Smart-Pig Inspection Data *2006 International Pipeline Conference* IPC2006-10157 **2** 243-8

[11] Wang Z 2008 Precision Location Technology of Pipeline Robot Based on Multi-Sensor Data Fusion *Robot* **30**(3) 238-41

[12] Reber K, Beller M, Willems H, Barbian O A 2002 A new generation of ultra-sonic in-line inspection tols for detecting, sizing and locating metal loss and cracks in transmission pipelines Proceedings of the 2002 IEEE Ultrasonics Symposium Piscataway NJ USA **1** 665-71

[13] Li M 2004 The study of accurate in-line inspection technology to offshore pipeline route *China Offshore Platform* **19**(6) 46-9

[14] Li J, Chen H, Zhang X 2006 Localization Technique of Pipeline Robot Based on Multi-sensor Data Fusion *Control and Decision* 2006 **21**(6) 661-5

[15] Hang L, Nassar S, El-Sheimy N 2010 Accurate Pipeline Surveying Using Two-Filter Optimal Smoothing of Inertial Navigation Data Augmented with Velocity and Coordinate Updates Proceedings of the 2010 International Technical Meeting of The Institute of Navigation 49-56

[16] Nassar S 2003 Improving the Inertial Navigation System (INS) Error Model for INS and INS/DGPS Applications *PhD thesis* UCGE Reports No. 20183 The University of Calgary Calgary Alberta Canada

[17] Liu H 2009 Optimal Smoothing Techniques in Aided Inertial Navigation and Surveying Systems *MSc thesis* UCGE Report No. 20298 University of Calgary Calgary Alberta Canada

[18] Wang F 2012 Inertial surveying internal detection method and its application in oil and gas pipeline *Oil & Gas Storage and Transportation* **31**(5) 372-5 *(in Chinese)*

[19] Sadovnychiy S, Ponomaryov V, Ramirez T, Herrera E 2001 Navigation system for automation of pipelines inspection missions *Instrumentation and Development* **5**(1) 3-15

[20] Zirnig W, Hausamann D, Schreier G 2001 A concept for natural gas transmission pipeline monitoring based on new high-resolution remote sensing technologies IGRC 2001 November Amsterdam

Authors

Rui Li, born on July 26, 1983, Urumqi China

Current position, grades: engineer, master.
University studies: automation science and electrical engineering collage of BeiHang University.
Scientific interest: automation science and electrical engineering

Qingshan Feng, born in October, 1974, JiangSu China

Current position, grades: PhD, senior engineer at BeiHang University.
University studies: BeiHang University.
Scientific interest: Material science.
Experience: management department of Petrochina Pipeline Company.

Maolin Cai, born in 1972, JiangXi China

Current position, grades: PhD.
University studies: PhD grade from Tokyo Institute of Technology.
Scientific interest: energy saving fluid power systems.
Experience: automation science and electrical engineering collage of BeiHang University.

Haijun Li, born on June 12, 1981, Huaian, China

Current position, grades: engineer.
Scientific interest: navigation, guidance and control.

ChengHai Liu, born on September 26, 1984, HuaiYang County, ZhouKou City, Henan Province

Current position, grades: section member.
University studies: Master of Geographic Information System Wuhan University.
Scientific interest: GIS, GPS, RS.
Publications: 1
Experience: integrity management department of Petrochina Pipeline Company.

A hand gesture interaction system based on Kinect

Weiqing Li*, Zehui Lu, Shihong Shen

School of Computer Science & Engineering, Nanjing University of Science & Technology, Nanjing, China, 210094

Received 1 October 2014, www.cmnt.lv

Abstract

We introduce a hand gesture interaction system using Kinect, which takes advantage of real-time dynamic motion capture, image recognition and so on, so that people can interact with computer by natural hand gestures. Five kinds of gesture are defined and can be recognized by the system. Kinect-based hand movements and gesture recognition algorithm is studied. A method for hand area image segmentation from the depth map is proposed, using the information of hand joints in skeleton map. We realized a hand gesture recognition algorithm with SVM and tested its stability and robustness. Finally, experimental results verified the feasibility of the algorithm, and a hand gesture interactive demonstration is implemented.

Keywords: hand gesture recognition, Kinect, SVM, interaction

1 Introduction

With the development of computer vision technology, interactive computer-based vision systems have gradually developed. The traditional interaction methods such as mouse and keyboard have been unable to play a facilitating role, in some specific areas. The interaction method based on computer vision will be a very good way. The hand gesture recognition based on computer vision is a hot spot of human-computer interaction research [1]. The hand gesture interaction is rich in expression and contains a lot of information, for example we can get lots of information through different gestures, different locations, and different directions. It is a much natural interaction method.

The Kinect sensor is a horizontal bar connected to a small base with a motorized pivot, and is designed to be positional lengthwise above or below the video display. The device features an RGB camera, depth sensor and multi-array microphone running proprietary software, which provide full-body 3D motion capture, image recognition and voice recognition capabilities [2]. It was developed at Microsoft Research Cambridge in collaboration with Xbox. Kinect gives completely hands-free control of electronic devices possible by using an infrared projector and camera and a special microchip to track the movement of objects and individuals in three dimensions.

In this paper, we propose a hand gesture interaction system based on Kinect. We define some hand gestures that can be used to interact with the system. Then we study the hand movement recognition algorithm, and set up a method of hand area image segmentation from the depth map background by using the information of hand joints in skeleton map. Also, we realize a hand gesture recognition algorithm with SVM. Finally some experimental results verify the validity of this algorithm.

2 Hand gesture definition

The basic idea of Kinect is to segment a single depth image into a dense probabilistic body part, labelling as a per-pixel classification task and then estimate this signal to give high-quality proposals for the 3D locations of body joints. And Kinect's depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures video data in 3D under any ambient light conditions. The sensing range of the depth sensor is adjustable, and the Kinect software is capable of automatically calibrating the sensor based on the player's physical environment, accommodating for the presence of furniture or other obstacles [3]. In order to recognize hand gestures and interact with the system, we define 5 gestures, show in Table 1.

TABLE 1 Hand gesture definition

Gesture	Meaning	Gesture	Meaning
Keeping single finger straight for 1 second	Left click	Spreading out the fingers and keeping moving at the same time	Mouse move
Keeping fist for 1 second	Right click	Spreading out the arms to the two sides	Zoom in
Holding the fingers straight and closing them	Click down	Making hands close to the chest	Zoom out
Spreading out the fingers on one hand for 1 second	Click release	Waving the right hand to right quickly	Forward
Putting hands over the head	Full screen	Waving the right hand to left quickly	Backspace

* ***Corresponding author*** e-mail: li_weiqing@139.com

3 Hand gesture recognition

In this section we will introduce two algorithms: hand movements tracking algorithm and hand gesture recognition algorithm with the depth map, let's discuss them in detail.

The hand movements refer to the gestures of waving our hands to right or left, up and down. We use these gestures to control the system such as zoom in, zoom out, full screen, forward, backspace and so on.

The hand gesture means the finger postures, such as making a fist, spreading out fingers and so on [4,8]. Kinect can't recognize the shape of fingers and palm at present, so we propose a hand gesture recognition method using SVM. The process is shown in Figure 1.

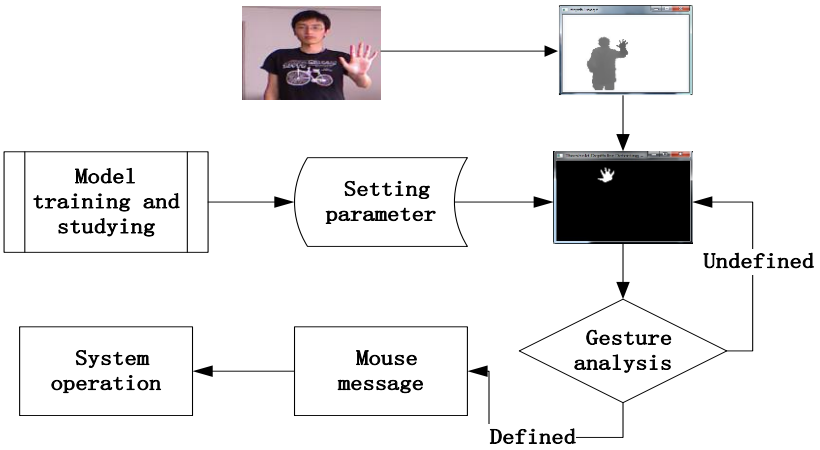


FIGURE 1 Hand gesture recognition process

3.1 HAND MOVEMENTS TRACKING

3.1.1 Joints data pre-processing based on the continuous space-time tracking

Kinect can get three kinds of images, they are color map, depth map and skeleton map. The skeleton map is consisted of twenty joints including our hand joints. The camera's coordinates are (0,0,0), the right direction is the positive direction of the x axis, the upward direction is the positive direction of the y axis and the positive direction of the z axis is facing to the operator. So every frame Kinect captures all the joints' three-dimensional coordinates.

Based on continuous space-time tracking of hand movements' characteristic [5], the movement tracking method can be separated into two parts. The first part is called preprocessing, as Kinect is capable of simultaneously tracking up to six people, but in fact the system can be operated just by one person. So the problem is whose hand gesture should be recognized when there are many people stand before Kinect. We use the bubble sorting algorithm for the head's coordinate on the z axis (if the people is not exist, we will set a maximum value for the coordinate). Kinect will consider the person who has the minimum coordinate as the operator and recognize his gesture.

The second part is called hand movement tracking. We set a sliding window and store 20 frames continuous skeleton image, then the three-dimensional coordinate data extraction and analysis. The sliding window slides backwards and inserts the next frame, the first frame will be deleted. Let's discuss the tracking method in next chapter.

3.1.2 Hand movements' characteristic definition

Hand gesture definition means the gesture's specific function, such as waving right hand to right means going to the next page, waving right hand to left means going back, waving left hand to left and right hand to right at the same time means zoom in, etc. If the left hand thrusts forward and then moving right hand, it means we have clicked the mouse and controlling the movement of the mouse now. These are defined as follows:

1. The definition of click. We get a series of values on the z axis, such as $z_1, z_2, ..z_n$, in a period of time (such as 3 seconds). If $z_1 \succ z_2 \succ ..z_n$ and $\delta_1 \succ z_1 - z_n \succ \delta_2$, then the system will consider the left hand has done the click action. The thresholds δ_1 and δ_2 can be set different values according to different conditions.

2. The definition of forward. We can get a series of right hand's two-dimensional coordinate values such as $x_1, x_2, ..x_n$ and $y_1, y_2 ..y_n$ in a period of time (such as 3 seconds). If $x_1 \prec x_2 \prec x_3 \prec ... \prec x_n$, $x_n - x_1 \succ \delta$ and $\max(y_i) - \min(y_j) < \varepsilon$, then the system will consider the operator has done the forward gesture. The thresholds δ and ε can be set different values according to different conditions. $\max(y_i)$ and $\max(y_j)$ represent the maximum and minimum value on the y axis when we wave our hands.

3. The definition of zoom in. We can get a series of values such as $R(x_1), R(x_2),...,R(x_n)$, $R(y_1), R(y_2),...,R(y_n)$, $L(x_1), L(x_2),..., L(x_n)$ and $L(y_1), L(y_2),..., L(y_n)$. If $L(y_1), L(y_2),...,L(y_n)$, $R(x_1) \prec R(x_2) \prec ... \prec R(x_n)$ and these values satisfy with the formulas as follows:

$$\left|R(x_n) - R(x_1)\right| \succ \delta, \tag{1}$$

$$\left|\max(R(y_i)) - \min(R(y_j))\right| \prec \varepsilon, \tag{2}$$

$$\left|L(x_n) - L(x_1)\right| \succ \delta, \tag{3}$$

$$\left|\max(L(y_i)) - \min(L(y_j))\right| \prec \varepsilon. \tag{4}$$

Then the gesture will be considered as zoom in. $R(x)$ and $R(y)$ are right hand's coordinates (x, y) . $L(x)$ and $L(y)$ are left hand's coordinates (x, y) . We can set different values for and according to different conditions. $\max(R(y_i))$ and $\min(R(y_j))$ represent the maximum and minimum value on the y axis when we wave right hand.

3.2 HAND GESTURE RECOGNITION BASED ON DEPTH MAP

3.2.1 Hand area image segmentation assisted by skeleton map

Except the skeleton map, Kinect can also get colour map and depth map. We can recognize the hand gestures with these two kinds of maps. The colour map has the advantage of clearness, but it is susceptible to interference and only contains the information in two dimensions. On the other hand, the depth map's resolution is lower than colour map, but contains information of 3D. We propose a method of hand seg-mentation from the depth map background using the information of hand joints in skeleton map.

As Kinect can track our hands' coordinates with the help of skeleton map, so we can easily get the precise location of our hands in the skeleton map. The hand's location can be mapped to the depth map with the help of the pixel ratio between the skeleton map and the depth map. As the depth map contains the three-dimensional information, so we can realize the segmentation from the depth map background according to the depth information [6].

For the image segmentation, threshold has a big influence. The definition of depth threshold is:

$$\psi = \min(z) + \delta, \tag{5}$$

$\min(z)$ represents a point's coordinate on the z-axis which is the nearest point to the camera of our body. δ is a experimental value, the range of the value is $\min(z)$ to ψ . This is the most precise value of the palm's location. In our experiment, we set 5cm for δ . In this case, we get the hand segmentation image as Figure 2.

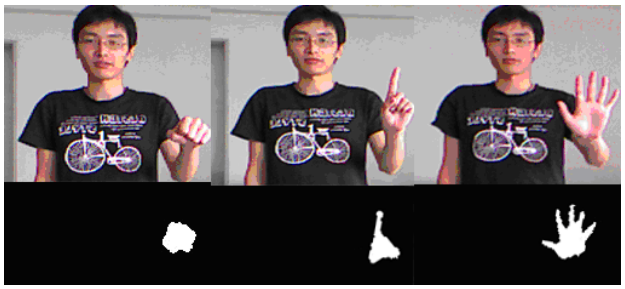


FIGURE 2 Hand area segmentation

3.2.2 Gesture recognition algorithm using SVM

SVM is a classifier, it is firstly proposed by Cortes and Vapnik in 1995. Compared with the traditional method, SVM doesn't depend on the selection of the model. It shows many special advantages in resolving the small sample, nonlinear pattern recognition [7]. The SVM algorithm references kernel function, and then the algorithm converts the practical problem to a high dimensional feature space through the nonlinear transformation. It creates the linear discriminate function in high dimensional space to realize the nonlinear discriminate function in the original space, its particular characteristics can certify the sys-tem has good generalization ability. In addition, SVM solved the dimension problem skillfully. Its algorithm complexity has nothing to do with the dimension of the sample, so the SVM classifier can adapt to the classification problem of different dimensionality. In this section, we use the SVM classifier to recognize the hand gestures according to the limitation of the sample and the accuracy of the recognition.

Firstly, in order to get the contour of the hand, we denoise and smooth the hand image which segmented from the depth map, and remove some isolated points. Then we will calculate some characters of the gesture, such as Circularity:

$$Circularity = \frac{L^2}{4\pi \times AL}. \tag{6}$$

This characteristic describes the closeness between the gesture and circle, the closer the value to one, the more like the circle. L represents the perimeter of the contour of the hand. A represents the area of the contour of the hand. The distribution graph is about the gestures' circularity. We can find that the circularity of the fist most close to one and the circularity of the palm is the biggest. Because the palm's opening in varying degrees so the value of L will be different and then lead to the fluctuant.

Filling Ratio:

$$FillingRatio = \frac{A}{A - R}. \tag{7}$$

This characteristic describes the area ratio of the hand occupies in the rectangle outside. The bigger of the value, the more gathered of the gesture. *A* - *R* represents the area of the smallest rectangle outside. The distribution graph is about the gestures' filling proportion. We can find from the graph that the fist's value is the biggest, and it means the fist has the highest polymerism.

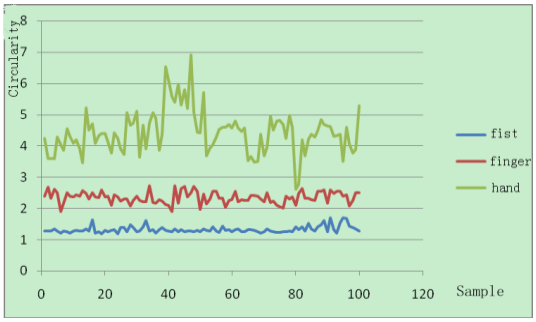


FIGURE 3 Circularity of gestures

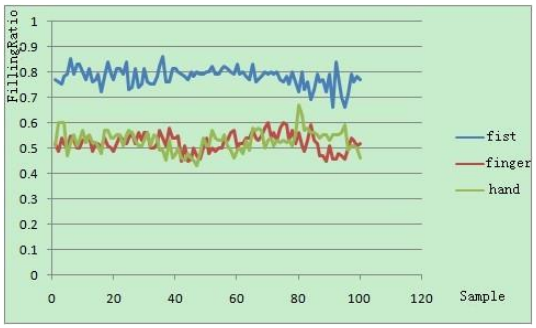


FIGURE 4 Filling ratio of gestures

Perimeter Ratio:

$$PerimeterRatio = \frac{L}{L-C}. \tag{8}$$

This characteristic describes the ratio between the perimeter of the contour of the hand and the mini-mum circumference of the circle outside, the bigger of the value, the more opening of the gesture. *L_C* represents the circumference of the smallest circle outside. The distribution graph shows three gesture's perimeter ratio. We can find that the palm's perimeter ratio is much bigger than the other two gestures.

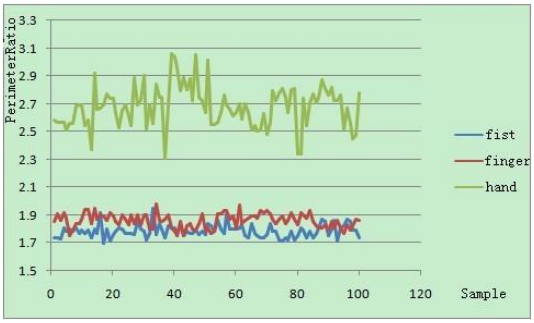


FIGURE 5 Perimeter ratio of gestures

4 Experiment

4.1 HAND MOVEMENTS RECOGNITION EXPERIMENT

Moving the right hand slowly means the movement of mouse, and moving the right hand to right or left quickly means going forward or going back. So it is necessary to set a threshold, if we want to distinguish the forward gesture from the back gesture. Because when we wave our arms, its orbit is an arc, so the co-ordinate on the z axis keeps changing all the time. In order to prevent from confusing with the click behaviour, we should set a threshold on the z axis. The specific value can be changed with different conditions, the method as follows:

As to the click gesture, δ_1 is set to 0.2 and δ_2 is set to 0.1, it means the left hand's range of movement should between 10cm and 20cm in a period of time, then the gesture will be recognized as click.

As to the forward gesture, δ is set to 0.3 and ε is set to 0.1, it means the right hand's range of movement should greater than 30cm and the amplitude should less than 10cm in a period of time, then the gesture will be recognized as going forward.

The gesture of zoom in has the same thresholds with the forward gesture, as the gesture of zoom in includes the forward gesture, so we introduce the priority in order to distinguish these two kinds of gestures. The gesture of zoom in has a higher priority than the forward gesture and the gesture of zoom out has a higher priority than the back gesture.

We perform 100 click operations and 100 forward operations with different thresholds, the recognition rates are showed in Table 2 and Table 3. When we perform the forward gesture, for the ε, a higher value is better, in order to have a better understanding of the gesture, ε should not get a too large value, 0.1 is a right value.

TABLE 2 Threshold of click operation

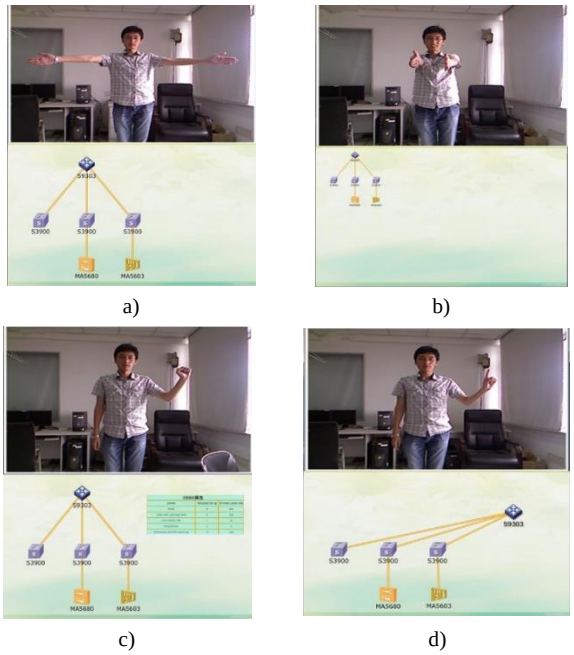
Threshold	$\delta_1 = 0, \delta_2 = 0.1$	$\delta_1 = 0.1, \delta_2 = 0.2$	$\delta_1 = 0.2, \delta_2 = 0.3$	$\delta_1 = 0.3, \delta_2 = 0.4$	$\delta_1 = 0.4, \delta_2 = +\infty$
Recognition rate	0.73	0.91	0.77	0.53	0.12

TABLE 3 Threshold of forward operation

Threshold	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$
Recognition rate	0.45	0.68	0.93	0.76

4.2 HUMAN-MACHINE INTERACTION DEMO

Figure 6a shows the gesture before zoom out, Figure 6b shows the gesture after zoom out and Figure 6c shows the gesture of catching and moving. Figure 6d shows the gesture of displaying some parameters.



a) b) c) d)

FIGURE 6 Human-machine interaction demo

5 Conclusions

In this paper, we introduce a hand gesture interaction system based on Kinect. We get the hand gesture image by Kinect, then the system will analyse and recognize both the movement and the gesture. In order to recognize the gesture precisely, we design an algorithm based on SVM. Also, we map gestures to mouse events to interaction with the computer. From the experimental results, we notice that the algorithm has well reliability and strong robustness.

On the other hand, this gesture interaction system has some work to go on, such as we need more gesture definitions and improve the system's sensitivity. These will be considered as future work.

References

[1] Chen F S, Fu C M, Huang C L 2003 Hand gesture recognition using a real-time tracking method and hidden Markov models *Image and Vision Computing* **21**(8) 745-58

[2] Lee S B, Ho Y S 2011 Real-time Stereo View Generation using Kinect Depth Camera *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* 2011 1153-6

[3] Laptev I 2005 On space-time interest points *International Journal of Computer Vision* **64**(2-3) 107-23

[4] Chen Y M, Zhang Y H 2009 Research on human-robot interaction technique based on hand gesture recognition *Robot* **31**(4) 351-6

[5] Lee M, Green R, Billinghurst M 2008 3D Natural Hand Interaction for AR Applications *23rd International Conference Image and Vision Computing* Christchurch New Zealand 6-12

[6] Fujimura K, Liu X 2004 Hand Gesture Recognition using Depth Data *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition* 529-34

[7] Agarwal A, Triggs B 2006 Recovering 3D human pose from monocular images *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(1) 44-58

[8] Zhu Y X, Ren H B, Xu G Y, Lin X Y 2000 Toward real-time human-computer interaction with continuous dynamic hand gestures *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition* 544-51

Authors

Weiqing Li, born in December, 1974, Luoyang, Henan, China

Current position, grades: associate professor at the School of Computer Science and Engineering, Nanjing University of Science and Technologies.
University studies: doctor's degree in Computer Application at Nanjing University of Science and Technologies in 2007.
Scientific interests: virtual reality, human machine interface, computer graphics.
Publications: 20 papers.

Zehui Lu, born in November, 1991, Xuzhou, Jiangsu, China

Current position, grades: academic master candidate in pattern recognition and intelligent system at Nanjing University of Science and Technologies.
University studies: bachelor's degree in Software Engineering at Changchun University of Science and Technologies in 2013.
Scientific interests: virtual reality and simulation systems.

Shihong Shen, born in June, 1987, Cixi, Zhejiang, China

Current position, grades: academic master candidate in computer simulations at Nanjing University of Science and Technologies.
University studies: Master's degree in Computer Science in Nanjing University of Science and Technologies in 2013.
Scientific interests: virtual reality and simulation system.

Robust volume data watermarking based on perceptual hashing

Yujia Li, Jingbing Li*

College of Information Science and Technology, Hainan University, Haikou, China, 570228

Received 1 March 2014, www.cmnt.lv

Abstract

This paper proposed a perceptual hashing algorithm of robust blind watermarking method for volume data. Which address the problems of authentication and protection of personal information. The scheme obtains the feature vectors of volume data and quantizes them to generate the hash value of the volume data. By combining the concept of zero-watermarking, the algorithm for watermarking of volume data that is robust to geometric attacks. The experimental results demonstrate that the proposed algorithm has good invisibility and robustness

Keywords: watermarking, perceptual hashing, volume data, DFT

1 Introduction

With the progress of computer technology, the application of multimedia has a qualitative leap. But security concerns over copyright violation of multimedia data have also increased at the same time. On the one hand multimedia data can faster and more efficient transmitted in public network [1], on the other hand intercept and manipulation of the multimedia information has also become very easy [2]. Thus, the information security of multimedia data has attracted a lot of attention during the last few years.

Digital watermarking is an efficient tool for multimedia information protection [3]. However, the directions of research have been mostly in image [4], audio [5] and video watermarking [6]. Currently, there are a lot of volume data in real life. Such as most of the medical image (CT, MRI, etc.) is volume data, so study how to embed the digital watermarking in the volume data is significant. The volume data is not allowed to modify the content in principle [7]. As a result, to embed watermarking in volume data is difficult. Discrete cosine transform (DCT) [8], Discrete Fourier transform (DFT) [9] and discrete wavelet transform [10,11] are used in embedding watermarking of volume data. But the robustness of the algorithms is not very ideal. Wu [12] develop an algorithm based on the spread-spectrum communication technique to watermarking a volume data which is invisible and robust. However, the original volume data is needed in watermarking detection. It's not blind watermarking.

In allusion to invisible and robust Image Watermarking, there are a series of requirements. First, we hope the embedding of watermarking could not change the original image [13]. What is more, the watermarking must have good robustness and invisible. At last, we wish the capacity of watermarking is bigger. Whereas to the traditional digital watermarking algorithm, difficult to achieve all requirements at a time.

In this paper, we propose a watermarking algorithm based on perceptual hashing which is a blind multi-watermarking algorithm and also can achieve a true embedded zero-watermarking [14]. The experiment results demonstrate that the algorithm has good robustness.

2 Theoretical background

2.1 THE 3D DISCRETE FOURIER TRANSFORM

The discrete Fourier transform is an important mathematical tool on the engineering application. Assuming the size of volume data is $X \times Y \times Z$. The corresponding 3D-DFT is done using:

$$F(u,v,w) = \sum_{x=0}^{X-1}\sum_{y=0}^{Y-1}\sum_{p=0}^{Z-1} f(x,y,z)\cdot e^{-j2\pi xu/X} e^{-j2\pi yv/Y} e^{-j2\pi zw/Z}, \quad (1)$$

$$u = 0,1,...,X-1;\ v = 0,1,...,Y-1;\ w = 0,1,...,Z-1.$$

The corresponding volume data's inverse discrete Fourier transform (IDFT) is computed using:

$$f(x,y,z) = \frac{1}{XYZ}\sum_{u=0}^{X-1}\sum_{v=0}^{Y-1}\sum_{w=0}^{Z-1} F(u,v,w)\cdot e^{j2\pi xu/X} e^{j2\pi yv/Y} e^{j2\pi zw/Z}, \quad (2)$$

$$x = 0,1,...,X-1;\ y = 0,1,...,Y-1;\ z = 0,1,...,Z-1.$$

Note that in our case $f(x,y,z)$ is the value at the point (x,y,z) of the volume data, $F(u,v,w)$ is the three-dimension DFT coefficient at the point (u,v,w) in frequency domain. Volume data is composed of many layers of slices, each slice is a two-dimensional image. The size of image is $M \times N$, the number of slice's layers is P.

* ***Corresponding author*** e-mail: Jingbingli2008@hotmail.com

2.2 PERCEPTUAL HASHING

In 2001, Kalker first proposed the concept of perceptual hashing. A majority of existing algorithms follow a three-step framework to generate a hash value [15], as illustrated in Figure 1.

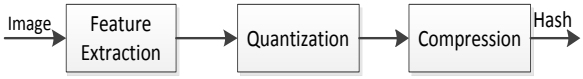


FIGURE 1 An example. Good quality with clear lettering

Perceptual hashing is a class of one way mappings from multimedia presentations to a perceptual hash value in terms of their perceptual content [16]. Perceptual hashing algorithm can transform volume data into binary sequences, and can be used to database search, content authentication, and watermarking [17-21]. The characteristics of the perceptual hashing function are robustness, collision resistance, compactness and one-wayness [16,22-25]. The specific method of perceptual hashing algorithm used in this paper was as follow:

Step 1: 3D-DFT to the volume data;

Step 2: 3D-IDFT to the precious 4×4×4 DFT coefficients;

Step 3: Calculate the average value of the real part of 4×4×4 IDFT coefficients.

Step 4 Compare the each IDFT real coefficient with the average. Greater than or equal to the average, recorded as 1. Less than the average, recorded as 0.

Step 5 Group the comparison result of the previous step together to constitute a 64 bits binary sequence, which is the hash value of the volume data. Group order is not important, as long as the order of all volume data using the same.

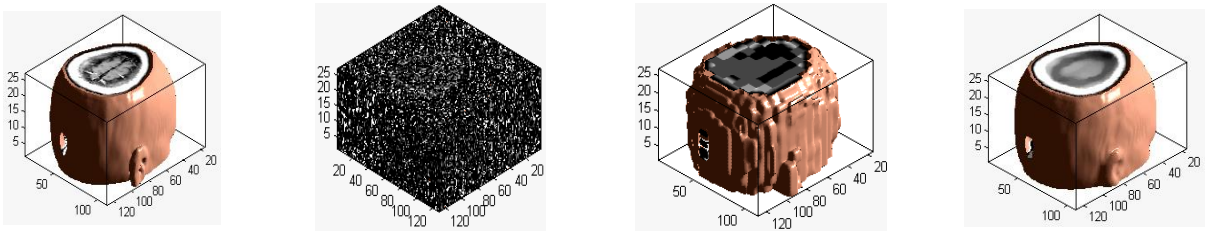
3 The watermarking algorithm

We choose a set of binary pseudo-random sequence Bg, Bg={bg(i)|bg(i)=0,1;1≤i≤L} represents information as multi-watermarking. Then, we select one MRI medical volume data as the original volume data which is describe as $F=\{f(i,j,k)|f(i,j,k)\in R;1\le i\le M,1\le j\le N,1\le k\le P\}$, where $w(j)$ and $f(i,j,k)$ represent the pixel gray-values of watermarking and the voxel values of the original volume data similar to the pixel gray-values of 2D image. To facilitate the operation, we assume $M1=M2=M$, $N1=N2=N$.

3.1 HASH VALUE EXTRACTION METHOD OF VOLUME DATA

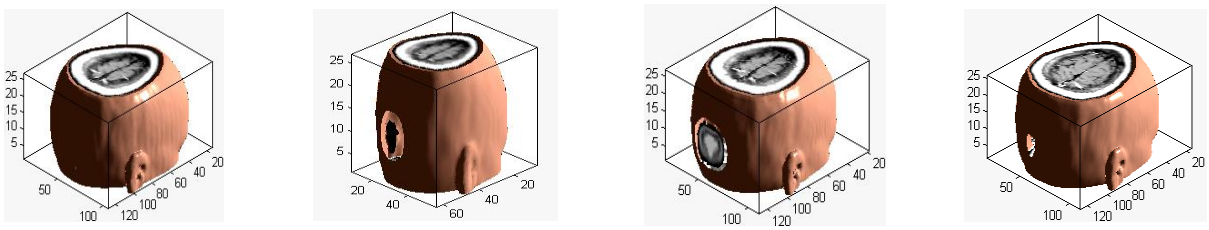
Though the above method of perceptual hashing, we can extract the hash value of the volume data. That is to say: first, we use the 3D-DFT to the original medical volume data by the geometric transformation on each slice. We select the previous 4×4×4 coefficients. Then, the selected coefficient is computed using 3D-IDFT. We recognize the real part of IDFT coefficients as the feature vector of the volume data. Then, through the comparison of feature vector and the real part coefficients average realize binary quantization. At last, we can generate the hash value of volume data.

In order to prove that the robustness of the hash value, which we select above. We randomly choose 8 coefficients of the feature vector ($F(1,3,4)$, $F(1,4,1)$,…,$F(4,3,4)$, $F(4,4,1)$), which show in Table 1 We can see that the value of the corresponding coefficients have changed after the volume data has undergone an attack (which is show in Figures 2 and 3) as illustrated in column "L2-L9". But the hashing value unchanged, as shown in Table 1. It can be seen that the normalized cross-correlation (NC) between the hashing values is equal to 1.0, as shown in column "L12". Therefore, the hash value can meet the perceptual robustness.



a) b) c) d)

FIGURE 2 Different common attacks: a) Original volume data; b) Gaussian noise (10%); c) JPEG compression 2%); d) Median filter ([5×5])



a) b) c) d)

FIGURE 3 Different geometric attacks: a) Rotation (10°); b) Scaling (0.5 times); c) Translation (3%, down); d) Cropping (3%, from *z* direction)

TABLE 1 The change that the real part of IDFT coefficients with different attacks

L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
Processing method	**F(1,3,4)**	**F(1,4,1)**	**F(2,3,4)**	**F(2,4,1)**	**F(3,3,4)**	**F(3,4,1)**	**F(4,4,4)**	**F(4,4,1)**	**average**	**Hash value**	**NC**
Original volume data	1.194	0.463	2.445	1.466	2.760	2.093	2.882	1.262	1.508	00101110	1.0
Gaussian noise (10%)	3.031	2.506	3.778	3.139	3.978	3.587	4.076	3.028	3.195	00101110	1.0
JPEG compression (2%)	1.239	0.525	2.520	1.548	2.784	2.132	3.008	1.314	1.575	00101110	1.0
Median filter [5×5]	1.177	0.463	2.478	1.474	2.756	2.139	2.911	1.212	1.508	00101110	1.0
Rotation (10°)	1.039	0.475	2.470	1.415	2.813	1.987	2.954	1.066	1.508	00101110	1.0
Scaling (×0.5)	0.290	0.113	0.613	0.368	0.695	0.514	0.726	0.304	0.378	00101110	1.0
Translation (down 3%)	1.400	0.469	2.356	1.340	2.687	2.100	2.828	1.373	1.507	00101110	1.0
Cropping (3% from x)	1.232	0.423	2.412	1.431	2.669	2.044	2.818	1.290	1.476	00101110	1.0
Cropping (3% from y)	1.462	0.481	2.393	1.399	2.648	2.053	2.758	1.348	1.504	00101110	1.0
Cropping (3% from z)	1.203	0.531	2.396	1.529	2.689	2.249	2.781	1.471	1.508	00101110	1.0

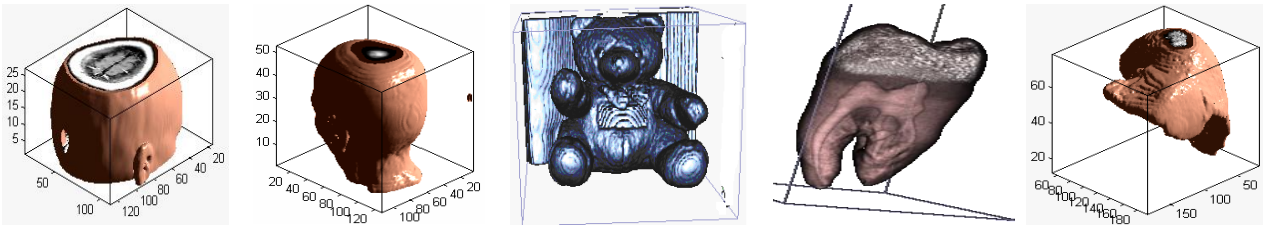
The unit of real part coefficients is 1.0e+005

Collision resistance is one of the important properties of perceptual hash value. Collision resistance means that there is large difference the hash value of two no similar images. In order to further illustrate the perception of the hash value, we select some different volume data as test objects which are shown in Figure 4. Calculate the NC between hash values of different test objects, as illustrated in Table 2.

In Table 2, we can see that the NC is largest between volume data itself, which is 1.00. The NC of a and b is larger, because a and b are brain volume data. The other NC values are small, which are all less than 0.5. That is observed with our human eyes. Therefore, collision resistance of the hash value in our algorithm is good.

TABLE 2 The NC between test volume data

NC	a	b	c	d	e
a	1.00	0.61	-0.04	-0.47	0.20
b	0.61	1.00	0.21	-0.39	0.21
c	-0.04	0.21	1.00	-0.05	0.12
d	-0.47	-0.39	-0.05	1.00	0.18
e	0.20	0.21	0.12	0.18	1.00



a) b) c) d) e)

FIGURE 4 Different volume data which are test objects

3.2 EMBEDDING WATERMARKING

Step 1: Acquire the hash value of the original volume data. In his section, we use the proposed perceptual hashing algorithm to the original volume data. Through volume data's feature extraction and binary quantization process, and then, we can get the hash value of the original volume data:

$$FF_4(i,j,k)=DFT3(F(i,j,k)), \tag{3}$$

$$FIF(i,j,k)=IDFT3(FF_4(i,j,k)), \tag{4}$$

$$RIF(i,j,k)=REAL(FIF(i,j,k)), \tag{5}$$

$$H(j)=BINARY(RIF(i,j,k)). \tag{6}$$

Step 2: Utilizing the HASH function of cryptograph, the logical sequence can got by us, $Key^g(j)$:

$$Key^g(j)=H(j)\oplus B^g(j), \tag{7}$$

where $H(j)$ is the hash value of volume data, B^g is the multi-watermark sequence. $Key^g(j)$ is a binary sequence. The $Key^g(j)$ is necessary to extract the watermarking, it would be saved. Moreover, we can take the $Key^g(j)$ for a secret key. In order to protect the copyright of the original volume data, we should registered the $Key^g(j)$ at the third part. In addition, during the watermarking embedding, the original volume data change a little. It's one of the zero-watermarking technologies.

3.3 EXTRACT WATERMARKING

Step 3: Acquire the hash value of the tested volume data. The perceptual hashing method is also performed on the test volume data $F'(x,y,z)$. As the same as Step 1, we can acquiring the hash value $H'(j)$.

$$FF'_4(i,j,k)=DFT3(F'(i,j,k)), \tag{8}$$

$$FIF'(i,j,k)=IDFT3(FF'_4(i,j,k)), \tag{9}$$

$$RIF'(i,j,k)=REAL(FIF'(i,j,k)), \tag{10}$$

$$H'(j) = BINARY\left(RIF'(i,j,k)\right), \quad (11)$$

Step 4: Utilizing the binary sequence $Key^g(j)$ and the hash value of tested volume data $H'(j)$, we can extract the watermarking $B^g'(j)$:

$$B^g'(j) = H'(j) \oplus Key^g(j), \quad (12)$$

$B^g'(j)$ describe the extracted multi-watermarking, can be also computed by the HASH function of cryptography. $Key^g(j)$ are obtained from Step 2.

3.4 WATERMARKING EVALUATING ALGORITHM

By calculating NC to determine whether it is embedded watermarking or not. We use NC to measuring the quantitative similarity between the extracted and embedded watermarking. The higher the value of NC is, the more approximation between the extracted watermarking $B^g(j)$ and the embedded watermarking $B^g'(j)$. Defined as:

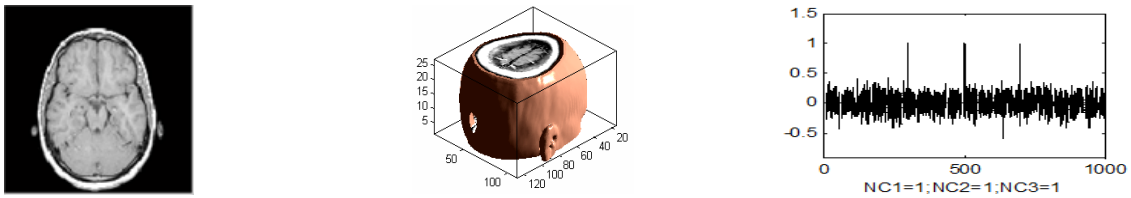
$$NC = \frac{\sum_j B^g(j) B^g'(j)}{\sum_j B^g(j) B^g(j)}. \quad (13)$$

4 Experiment results on robustness of the watermarking algorithm

We have implemented the volume data watermarking algorithm in Matlab2010a platform to verify the effectiveness. In these experiments, we should do these with the help of a thousand groups of independent binary pseudo-random sequences as used. Every sequence consists of 64 bits. In the experiment, the 300th, the 500th and the 700th group are selected at random from the thousand groups as the embedded watermarking. The size of the volume data (MRI.mat) is 128×128×27, which is offered in matlab.

It can be seen visually from Figure 5 that the quality of the volume data embedded has hardly any change. The quality of extracted watermarking is of high-quality with no difference with the original in normal case, all the NC are 1.0, which is shown in Figure 5a. (no attacking on watermarking)

The following are several types of common and geometric attacks to test the robustness of the algorithm. By attacks on each slice to achieve the purpose that an attack on the volume data in this experiment.



a) b) c)

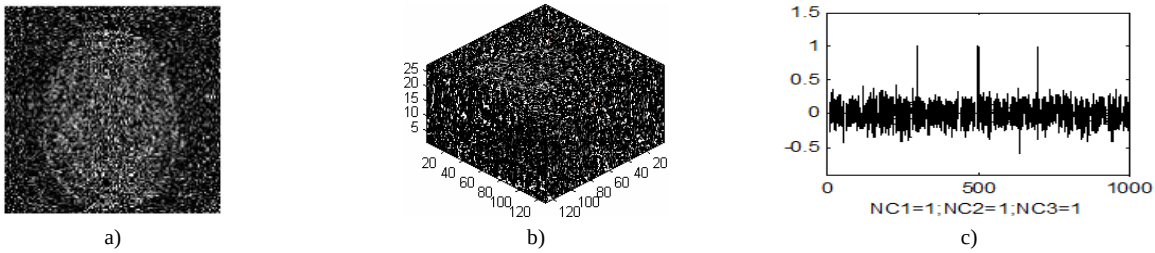
FIGURE 5 No attacking on watermarking: a) a slice of original volume data; b) the original volume data; c) watermarking detector

4.1 COMMON ATTACKS

In general, common attacks are common image processing. For example: Gaussian noise, digital to analogue and analogue to digital conversion, quantization and requantization, JPEG, median filtering and so on.

4.1.1 Gaussian noise

Figures 6a and 6b show the slice and volume data with the Gaussian noise (10%), respectively. As the Figure 6c shows, the PSNR is 3.30dB, and the multi-watermarking sequence can be detected, all of the NC are 1.00. Different parameters of noise impact on the volume data. The corresponding different PSNR and NC are given in Table 3. The result shows the value of NC is when the noise parameter top to 25%. Therefore, our algorithm has strong robustness against noise attacks.



a) b) c)

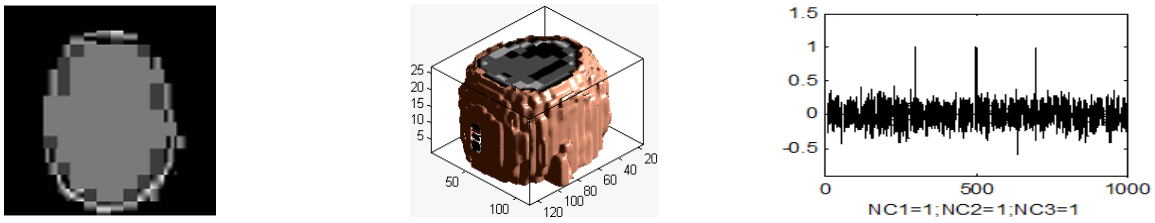
FIGURE 6 With noise attack: a) a slice with noise attack; b) the corresponding volume data; c) watermarking detector

TABLE 3 The PSNR and corresponding NC with Gaussian noise

Noise parameters (%)	1	3	5	10	15	20	25
PSNR	12.52	8.02	6.03	3.32	1.80	0.82	0.10
NC1	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NC2	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NC3	1.00	1.00	1.00	1.00	1.00	1.00	1.00

4.1.2 JPEG attacks

Figure 7a indicates the slice with JPEG attack (5%). Figure 7b indicates the corresponding volume data. As the Figure 7c displays the multi-watermarking sequence can be detected, NC1=1.00, NC2=1.00, NC3=1.00. Table 4 gives the PSNR and corresponding NC when different parameters of JPEG attacks are inflicted on the volume data. If the compression quality is down to 2%, the multi-watermarking still can be detected nevertheless. When the JPEG attacks, the results demonstrate that algorithm is robust.



a) b) c)

FIGURE 7 With JPEG attack (10%): a) a slice with JPEG attack; b) the corresponding volume data; c) watermarking detector

TABLE 4 The PSNR and corresponding NC with JPEG attacks

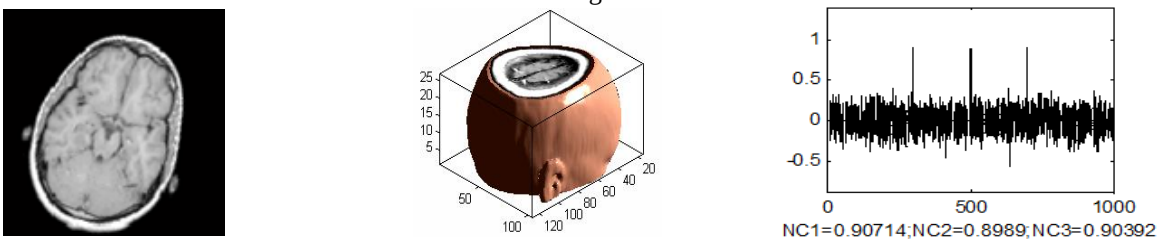
Compression quality (%)	2	4	8	10	20	40	60
PSNR	16.57	17.82	20.21	21.20	23.10	25.06	26.61
NC1	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NC2	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NC3	1.00	1.00	1.00	1.00	1.00	1.00	1.00

4.2 GEOMETRIC ATTACKS

Geometric attacks are refer to the watermarking image under rotation, scaling, cropping, translation and so on. It is almost the watermarking attack method which is hardest to solve. The implementation of the geometric attacks is very convenient. Just simple geometric attacks are often can cause the loss of watermarking. It affects the effectiveness of the watermarking algorithm greatly. Resistance to geometric attacks is a focus in the study of watermarking algorithm.

4.2.1 Rotation attacks

The slice with 20° rotated clockwise is indicated in Figure 8a. The corresponding volume data is indicated in Figure 8b. As the Figure 8c) displays the multi-watermarking sequence can be detected, NC1=0.91, NC2=0.90, NC3=0.90. Table 5 givers the PSNR and corresponding NC when different rotated angles are inflicted on the volume data. As the angle of rotation top to 35°clockwise, the multi-watermarking can be detected, too. When the rotation attacks, the results demonstrate that algorithm is robust.



a) b) c)

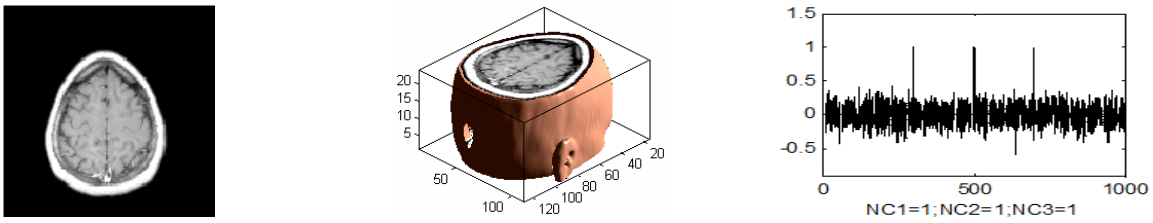
FIGURE 8 With rotation attack (20°): a) a slice with rotation attack; b) the corresponding volume data; c) watermarking detector

TABLE 5 The PSNR and corresponding NC with rotation attacks

Rotated (clockwise)	5	10	15	20	25	30	35
PSNR	16.54	13.97	12.98	12.44	12.04	11.68	11.33
NC1	0.97	0.94	0.91	0.91	0.88	0.84	0.84
NC2	0.96	0.93	0.90	0.90	0.87	0.83	0.83
NC3	0.97	0.94	0.90	0.90	0.87	0.85	0.85

4.2.2 Cropping attacks

The slice under cropping 10% from *Z* displays on Figure 9a. Then, the corresponding volume data is shown in Figure 9b. Figure 9c shows that the watermarking can be detected, NC1=1.00, NC2=1.00, NC3=1.00. Different cropping rates inflict on the volume data, the corresponding NC are given in Table 6. Hence, we can conclude the algorithm is robust to cropping attacks.



a) b) c)

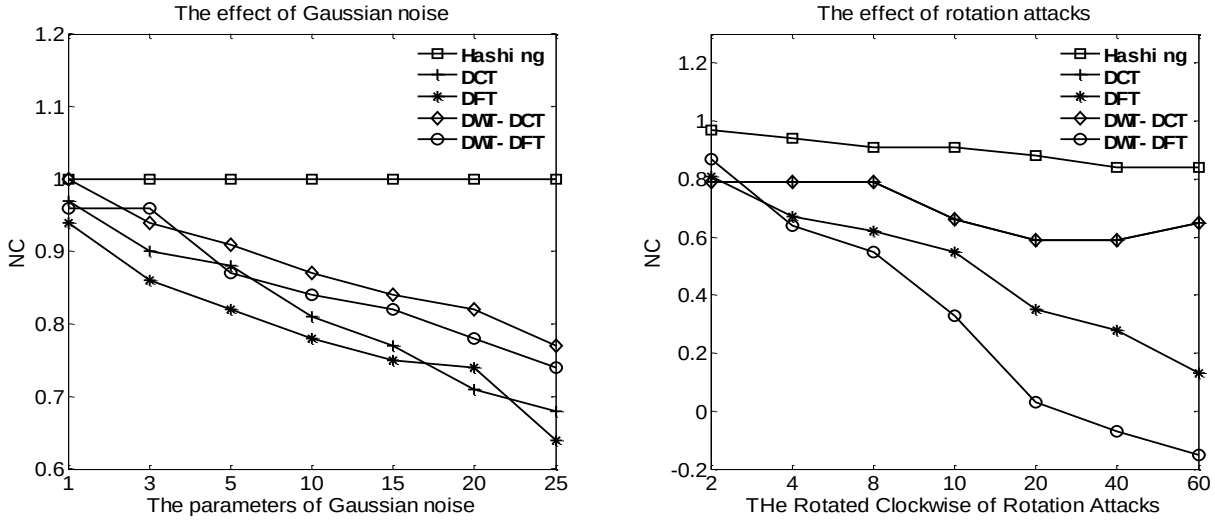
FIGURE 9 With noise attack: a) a slice with noise attack; b) the corresponding volume data; c) watermarking detector

TABLE 6 The PSNR and corresponding NC with cropping attacks

Cropping rate (from Z %)	**2**	**4**	**6**	**8**	**10**	**20**	**40**
NC1	1.00	1.00	1.00	1.00	1.00	0.97	0.88
NC2	1.00	1.00	1.00	1.00	1.00	0.97	0.87
NC3	1.00	1.00	1.00	1.00	1.00	0.97	0.88

4.3 ALGORITHM COMPARISON

In order to further demonstrate the robustness properties, we consider the performance of others volume data watermarking schemes, such as DCT, DFT, DWT-DCT, DWT-DFT [8], [10-11], [26]. We show the comparison results with different attacks in terms of NC in Figure 10. The results account for that the proposed volume data watermarking algorithm has good robustness. The performance for Gaussian noise and rotation attacks, are shown in Figure 10. In the case of Gaussian noise, the NC of the proposed perceptual hashing algorithm all are1.00. We can also observe that the NC of other volume data watermarking schemes are less than 1.00. We show the performance for rotation attacks is shown in Figure 10. In the case of rotation attacks, the NC of the perceptual hashing algorithms is very high. The NC of DCT scheme and DWT-DCT scheme are the same. In all, we observe that the proposed volume data watermarking scheme perform very well for both common attacks and geometric attacks. The algorithm has good robustness against common attacks and geometric attacks



a) b)

FIGURE 10 The performance of watermarking schemes under common attacks. a) Gaussian noise; b) JPEG attacks. To generate a point on the curve, X-axis is the volume data under different attacks; the corresponding watermarking of resulting volume data was detected. And the corresponding NC with the original watermarking is shown in the Y-axis.

5 Conclusion

A novel blind watermarking scheme appropriate for 3D volume data by using perceptual hashing was proposed in this paper. The method combines DFT transform, feature vector and database technology. The watermarking is a blind watermarking. In addition, the embedding of watermarking would not change the volume data. And the multi-watermarking could be embedded in volume data. It proved that the algorithm has a great capacity of watermarking embedding. The experiments results show that the algorithm has good robustness against common and geometric attacks, In a word, it is a efficiency algorithm.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No:61263033) and the NSF of Hainan Province of China(60894).

References

[1] Kaur S, Farooq O, Singhal R, Ahuja B S 2010 Digital watermarking of ECG data for secure wireless communication *Proceedings Information, Telecommunication and Computing (ITC) Conference* Kochi Kerala 140-4
[2] Hartung F, Kutter M 1999 *Proceedings of the IEEE* **87**(7) 1079-107
[3] Cox I, Miller M, Bloom J, Fridrich J, Kalker T 2007 Digital watermarking and steganography *Morgan Kaufmann* USA
[4] Lai C C, Tsai C C 2010 *IEEE Transactions on Instrumentation and Management* **59**(11) 3060-3
[5] Vivekanda B K, Sengupta I, Das A 2011 An audio watermarking scheme using singular value decomposition and dither-modulation quantization *Multimedia Tools and Applications* **52**(2) 369-83
[6] Preda R O, Vizireanu N D 2011 Quantization-based video watermarking in the wavelet domain with spatial and temporal redundancy *International Journal of Electronics* **98**(3) 393-405
[7] Rajendra A U, Niranjan U C, Iyengar S S, Kannathal N, Min K C 2004 Simultaneous storage of patient information with medical images in the frequency domain *Computer Methods and Programs in Biomedicine* **76**(1) 13-9
[8] Li J, Du W, Bai Y, Chen Y 2012 3D-DCT based zero-watermarking for medical volume data robust to geometrical attack *Wireless Communications and Applications* Springer Berlin Heidelberg 434-44
[9] Solachidis V, Pitas I 2007 *IEEE Transactions on Multimedia* **9**(7) 1373-83
[10] Li J, Du W, Bai Y, Chen Y 2011 Robust multiple watermarks for volume data based on 3D-DWT and 3D-DFT *2011 International Conference on Electronics, Communications and Control (ICECC)* Ningbo 446-50
[11] Li J, Du W, Bai Y, Chen Y 2011 3D DWT-DCT based multiple watermarks for medical volume data robust to geometrical attacks *2011 International Conference on Electronics, Communications and Control (ICECC)* 605-9
[12] Wu Y, Guan X, Kankanhalli M S, Huang Z 2001 Robust invisible watermarking of volume data using the 3D DCT *Proceedings of the Computer Graphics International Conf* Hong Kong 359-62
[13] Coatrieux G, Maitre H, Sankur B 2001 Strict integrity control of biomedical images *Proceedings of SPIE – The international Society for Optical Engineering* 229-40
[14] Li X, He G 2012 Efficient Audio Zero-Watermarking Algorithm for Copyright Protection Based on BIC and DWCM Matrix *International Journal of Advancements in Computing Technology* **4**(6) 109-17
[15] Swaminathan A, Mao Y, Wu M 2006 *IEEE Transactions on Information Forensics and Security* **1**(2) 215-30
[16] Niu X, Jiao Y 2008 An overview of perceptual hashing *Acta Electronica Sinica* **36**(7) 1405-11 *(in Chinese)*
[17] Lin S, Ozsu M T, Oria V, Ng R 2001 An extendible hash for multi-precision similarity querying of image databases *Proceedints of the 27th Very Large Data Bases (VLDB) Conference* Roma, Italy
[18] Wang L, Jiang X, Lian S, Hu D, Ye D 2011 Image authentication based on perceptual hash using Gabor filters *Soft Computing* **15**(3) 493-504
[19] Kailasanathan C, Safavi Naini R 2001 *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing* Baltimore
[20] Li C, Song H A 2009 geometrically robust watermarking scheme based on perceptual hashes and genetic Algorithm *Proceedings of the 4th International Conference on Computer Science & Education(ICCSE'09)* Nanning 673-8
[21] Holliman M, Memon N, Yeung M M 1999 On the need for image dependent keys for watermarking *Proceedings of the Content Security and Data Hiding in Digital Media* Newark
[22] Kozat S S, Venkatesan R, Mihcak M K 2001 *Proceedings of the IEEE Conference on Image Processing* **5** 3443-6
[23] Monga V, Banerjee A, Evans B L 2004 Clustering algorithms for perceptual image hashing *Proceedings of the 3rd IEEE Signal Processing Education Workshop* 283-7
[24] Monga V, Banerjee A, Evans B L 2006 A *IEEE Transactions on Information Forensics and Security* **1**(1) 68-79
[25] Monga V, Mıhçak M K 2007 *IEEE Transactions on Information Forensics and Security* **2**(3) 376-90 2007
[26] Li J, Du W, Bai Y, Chen Y 2011 3D-DFT Based Robust Multiple Watermarks of Medical Volume Data *Proceedings of the 3th Multimedia Information Networking and Security (MINES)* Shanghai 484-8

Authors

Yujia Li, born in December, 1990, Wannian, Jiangxi Province, China

Current position, grades: Master degree student in Information and Communication Engineering at University of Hainan, College of information science and technology.
University studies: BS degree in Information and Computing Science at Hainan University, Hainan, China, in 2012.
Scientific interests: multimedia information security, digital watermarking, image processing.
Publications: 4 papers.

Jingbing Li, born in June, 1966, Handan, Hebei Province, China

Current position, grades: doctor of Control Theory and control Engineering. Professor and doctoral supervisor in Hainan University
University studies: PhD degree at the Department of automation, Chongqing University, Chongqing, China in 2007.
Scientific interests: multimedia information security, digital watermarking, computer control.
Publications: 4 patents, 52 papers, 4 books.

The text image watermarking using arnold scrambling and DFT

Fan Wu, Jingbing Li*

College of Information Science and Technology Hainan University, Haikou, China

Received 1 October 2014, www.cmnt.lv

Abstract

With the popularization of Internet and the development at full speed of the multi-media technology, the copyright protection of digital works has already become the hot issue at present. Generally speaking, image, audio and video watermarking is comparatively similar in algorithm realization, with their redundancy, in which we can embed watermark. But other than the aforementioned, there is no redundancy to transfer secret info in text document. Nowadays, to embed watermark in text documents are limited to methods such as shifting the line and word, amending the characters' traits and disposing in the level of semantics. All these algorithms are not robust or lack of concealment, generally not serving the turn of Chinese text documents riffles. Based on the studies of the document digital watermarking methods and techniques, this dissertation presents that the problems of existed documents watermarking algorithms can be solved by Arnold Scrambling and DFT technique. The experimental results show that the scheme has strong robustness against common attacks and geometric attacks.

Keywords: Arnold scrambling, digital watermarking, DFT, zero-watermarking, text image

1 Introduction

With the rapid development of computer science and technology, and multimedia communication technology, digital media is becoming more and more universal. Digital watermarking is an important method for protecting digital media copyright. Most work focuses on audio, video, grayscale, and color images. However, binary images are very useful for security records, insurance information, financial document, fax images, case history, contract, e-business, e-Government, etc. Therefore, it may be very useful to embed and extract watermarking in binary images. For binary images, pixels take only two different values. Hence, embedding watermarking without causing visible perceivable changes in binary images is more difficult than in grayscale images. Brassial et al. proposed to change line spacing to embed the watermarking. Huang et al. presented a way to change word spacing to embed the watermarking. By changing a particular feature of an individual character can be embed with the watermarking. However there are some disadvantages to these watermarking approaches. The amount of data that can be hidden is few. The Process of embedding and extracting the watermarking is very complicated. Furthermore, these algorithms are vulnerable to many attacks, especially to geometric attacks.

This paper proposes an algorithm which combines DFT and Arnold scrambling. The algorithm combines both the first generation watermarking techniques and the second generation watermarking techniques. In the watermarking embedding process, firstly, the watermarking is encrypted by scrambling transformation technology, then the encrypted watermark information is embedded in the text of DFT transform image, to obtain a feature vector of the text image, then use the eigenvectors and watermarking information by HASH function to generate a binary logic series, as the sequence of keys. The result of experiment indicates that the algorithm can achieve a true embedded zero-watermarking. Meanwhile, it has a strong robustness against common attack and geometric attack.

2 The fundamental theory

2.1 ARNOLD SCRAMBLING TRANSFORM (AT)

Scrambling transformation as a means of encrypted technology is applied in the pretreatment stage of the watermarking, after scrambling transformation, one meaningful watermarking will become a meaningless, chaotic image. If you do not know the scrambling algorithm and keys, the attacker can not recover it even if he gets the watermarking from the embedded watermarking. And thus plays a role of secondary encryption. Additionally, after scrambling transformation, it will upset the relationship between the space locations of pixels and make it evenly distributed in all space of the carrier image. This will improve the robustness of the algorithm. Two-dimensional Arnold scrambling transformation is defined as follows:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \bmod N; x, y = \{0,1,2...,N-1\}, \quad (1)$$

where x, y is the pixel coordinates of the original space: x', y' is the pixel coordinates after iterative computation scrambling, N is the size of the rectangular image, also referred to as a step number.

By the above formula, the corresponding inverse

* ***Corresponding author*** e-mail: Jingbingli2008@hotmail.com

transform formula can be obtained:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \left(\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} N \\ N \end{bmatrix} \right) \bmod N, \quad (2)$$

$$x', y' = \{0,1,2...,N-1\}.$$

It is easy to restore the original initial state according to the corresponding iterations. Arnold transformation is cyclical, when iterate to a step, will regain original image. So if you do not know cycle and iterations, you will not be able to restore the image. Therefore, cycle and iterations can exist as a private key. Meanwhile, different image, because the desired effect is different, iterations should also be changed according to your need.

2.2 A METHOD TO OBTAIN THE FEATURE VECTOR OF TEXT IMAGE

First, the original image is computed using DFT. Then, we choose 5 low-frequency coefficients ($F(1,1)$, $F(1,2)$, …, $F(1,5)$) for formation of the feature vector, shown in Table 1. We find that the value of the low-frequent coefficients may change after the image has undergone an attack, particularly geometric attacks. However, the signs of the coefficients remain unchanged even with strong geometric attacks, as also shown in Table 1. Let "1" represents a positive or zero coefficient, and "0" represents a negative coefficient, and then we can obtain the sign sequence of low-frequency coefficients, as shown in the column "Sequence of coefficient signs" in Table 1. After attacks, the sign sequence is unchanged, and the Normalized Cross-correlation (NC) is equal to 1.0.

This means that the signs of the sequence can be regarded as the feature vector of the text image. Furthermore, it proves that the sequence of the DFT coefficient signs can reflect the main visual characteristics of text images.

TABLE 1 Change of DFT low-frequency coefficients with respect to different attacks

Image processing	PSNR	F(1,1)	F(1,2)	F(1,3)	F(1,4)	F(1,5)	Sequence of coefficient signs	NC
Original image		1.259	–0.006 + 0.005i	–0.022 + 0.017i	–0.034 + 0.010i	–0.045 + 0.009i	101010101	1.00
Gaussian interference (10%)	13.04	1.152	–0.005 + 0.002i	–0.017 + 0.010i	–0.026 + 0.006i	–0.035 + 0.007i	101010101	1.00
JPEG compression (10%)	17.71	1.219	0.002 + 0.006i	–0.016 + 0.019i	–0.030 + 0.012i	–0.044 + 0.012i	111010101	0.9
Median filter [3×3]	8.279	1.395	–0.035 – 0.001i	–0.046 + 0.006i	–0.047 – 0.001i	–0.048 + 0.001i	100010001	0.7
Rotation（clockwise, 5°）	6.319	1.250	–0.020 + 0.004i	–0.035 + 0.014i	–0.044 + 0.008i	–0.047 + 0.008i	101010101	1.00
Scaling(×0.5)		3.126	–0.011 + 0.014i	–0.054 + 0.042i	–0.085 + 0.023i	–0.113 + 0.020i	101010101	1.00
Translation (2% down)	7.389	1.259	–0.006 + 0.005i	–0.022 + 0.017i	–0.034 + 0.010i	–0.045 + 0.009i	101010101	1.00
Cropping (2%) (from Y direction)		1.259	–0.006 + 0.005i	–0.022 + 0.017i	–0.034 + 0.010i	–0.045 + 0.009i	101010101	1.00

DFT transform coefficient unit 1.0e+007

3 Robust watermarking algorithm based on digital text image

Use a meaningful binary image as the watermarking, Represented by W, F represents the original text image. $W=\{w(i, j)|w(i, j) = 0,1;\ 1\le i\le M1,\ 1\le j\le M2\}$ as digital watermarking, At the same time, we select a paragraph in an article as the original text image. It is describe as: $F = \{f(i, j)|f(i, j)\in R;\ 1\le i\le N1,\ 1\le j\le N2\}$, where $w(i, j)$ and $f(i, j)$ denote the pixel gray values of the watermarking and the original text image. Let $M1 = M2 = M$, $N1 = N2 = N$.

3.1 THE ALGORITHM OF THE EMBEDDED WATERMARKING

Step 1 Acquire the the encrypted watermarking image. The binary watermarking image is scrambled by Arnold scrambling transform, $BW(i, j)$.

$$BW(i, j) = AT(W(i, j)). \quad (3)$$

Step 2 Acquire the feature vector of the original text image. First, DFT of the whole $F(i, j)$ is computed as the DFT coefficient matrix, $FF(i, j)$. Then, after arranging the DFT coefficients from low to high frequency, the low-frequency sequence $Y(j)$ can be obtained. Finally, the feature vector $V=\{v(j)|v(j) = 0 \text{ or } 1;\ 1 \le j \le L\}$, can be achieved as a signs sequence of the top L values in the low-frequency $Y(j)$ by symbolic computation. Where the value of L can tune the robustness and capability of the embedded watermarking (in this paper we set $L = 32 = 4 \times 8$ bits).

$$FF(i, j) = DFT2(F(i, j)), \quad (4)$$

$$V(j) = -Sign(Y(j)). \quad (5)$$

Step 3 Generate the public key sequence. Utilizing the encrypted watermarking BW(i, j) and the feature vector V(j), we can generate the public key sequence, Key(i, j).

$$Key(i, j) = V(j) \oplus BW(i, j). \quad (6)$$

The public key sequence, $Key(i, j)$, can be computed by the HASH function of cryptography. The $Key(i, j)$ should be stored for extracting the embedded watermarking later. Furthermore, $Key(i, j)$ can also be regarded as a public key and registered to the third part to preserve the ownership of the original image registered to the third part to preserve the ownership of the original text image, so as to achieve the purpose of the protection of text images.

3.2 THE ALGORITHM OF THE EXTRACTED WATERMARKING

Step 1 Acquire the feature vector of the tested image. This process of acquiring the feature vector $T\_V'(j)$ is same tostep1 of the watermarking embedding process. The obtained feature vector, $T\_V'(j) = \{t\_v(j)|t\_v(j) = 0 \text{ or } 1; 1 \le j \le L\}$, also consists of the signs sequence of the DFT coefficients, where L has the same meaning as previously.

$$FF'(i,j) = DFT2(F'(i,j)), \tag{7}$$

$$T\_V'(j) = -Sign(Y'(j)). \tag{8}$$

Step 2 Extracting the watermarking BW'(i, j). According to the key, which generated in the embedded watermarking and the visual feature vector $T\_V'(j)$ of the being tested image, use HASH function properties to extract the watermarking $BW'(i, j)$. Extracting watermarking doesn't need original image, so it can protect the original image better.

$$BW'(i,j) = Key(i,j) \oplus T\_V'(j), \tag{9}$$

Step 3 Using the Arnold scrambling inverse transform to restore the extracted watermarking $BW'(i, j)$, get the watermarking of the being tested image, $W'(i, j)$.

$$W'(i,j) = IAT(BW'(i,j)). \tag{10}$$

3.3 DETECTION ALGORITHM OF THE WATERMARKING

Step 1 By calculating *NC* (Normalized Cross-Correlation) to determine whether there is the existence of the watermarking. The larger the value of *NC* is, the more approximation between $W'(i, j)$ and $W(i, j)$. Defined as:

$$NC = \frac{\sum_i \sum_j W(i,j) W'(i,j)}{\sum_i \sum_j W^2(i,j)}, \tag{11}$$

where $W(i, j)$ is the original watermarking, $W'(i, j)$ is the extracted watermarking.

Step 2 Evaluation of the quality of the text image after Embed watermarking by calculating the peak signal-to-noise ratio PSNR (dB), we often use peak value signal-to-noise ratio PSNR (dB) to reflect the quality of signal, defined as:

$$PSNR = 10 \lg \frac{MN \max_{i,j} (I(i,j))^2}{\sum_i \sum_j (I(i,j) - I'(i,j))^2}, \tag{12}$$

where $I(i, j)$, $I'(i, j)$ denote the pixel gray values of the coordinates (i, j) in the original image and the watermarking, respectively. M, N represent the image row and column numbers of pixels, respectively.

4 Experiments

To verify the effectiveness of our proposed algorithm, we carried out the simulation in Matlab2010a platform. We choose a significant binary image as the original watermarking and select a paragraph in an article as the original text image. The original watermarking $W = \{w(i, j)|w(i, j) = 0 \text{ or } 1; 1 \le i \le 32, 1 \le j \le 32\}$. The original text image $F = \{f(i, j), 1 \le i \le 128, 1 \le j \le 128\}$.

In the experiment, the parameter values: Arnold scrambling period is 24, and the number of transform times are 10, i.e. T=24, n=10.

Figure 1 represents the results when a binary watermarking is scrambled by Arnold scrambling transform. Figure 1a is the original watermarking. Figure 1b is the scrambled watermarking when n equaled to eight. Figure 1c shows the restored watermarking.

a) b) c)

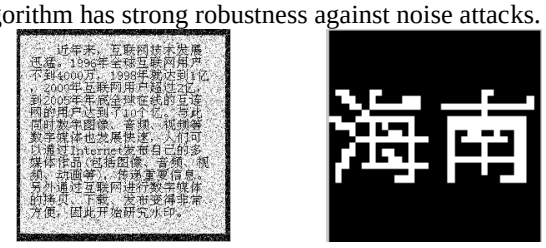
FIGURE 1 The watermarking is scrambled by Arnold scrambling transform: a) the original watermarking, b) the scrambled watermarking, c) the restored watermarking

In order to investigate this approach of embedding watermarking robust performance, the verification described below is chosen.

4.1 COMMON ATTACKS

4.1.1 Adding Gaussian noise.

In the watermarked text image, Gaussian noise is added by the noise function with different noise level. The text image under the attack of Gaussian noise (10%) with PSNR=12.96dB. At this time, the watermarked text image has been very vague, as shown in Figure 2a. The watermarking can obviously be extracted with NC=1.0. As shown in Figure 2b. The results prove that our proposed algorithm has strong robustness against noise attacks.



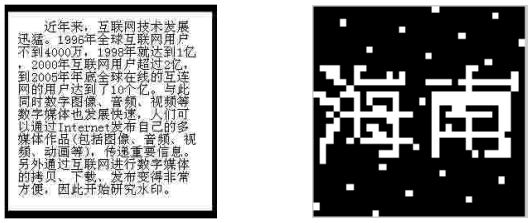
a) b)

FIGURE 2 The watermarked text image under Gaussian noise attacks (10%): a) the watermarked text image under noise attack, b) the extracted watermarking

4.1.2 JPEG attacks

JPEG compression process is done by using the percentage of image quality as a parameter to measure. The

watermarked text image with *PSNR* = 17.71dB under JPEG attacks (10%) is shown in Figure 3a. the watermarking can obviously be extracted with *NC* = 0.93 as shown in Figure 3b. The results show that the watermarking algorithm has strong robustness against JPEG attacks.



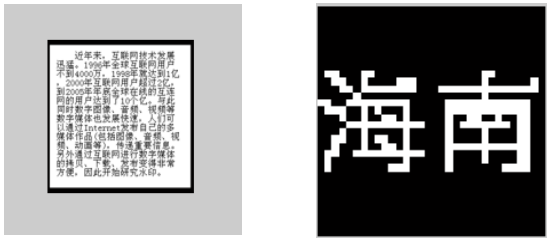
a) b)

FIGURE 3 The watermarked text image under JPEG attacks (10%): a) the watermarked text image under JPEG attacks, b) the extracted watermarking

4.2. GEOMETRICAL ATTACKS

4.2.1 Scaling attacks

We use the scaling factor as parameter to validate the effectiveness of our proposed algorithm on different scaling attacks. When the watermarked image is scaled 0.5 times, its pixel point has become the double of the original. Figure 4a shows that the watermarked image shrunk with a scale factor of 0.5. Moreover, Figure 4b shows that the watermarking can be extracted with *NC* = 1.00. It proves that our proposed algorithm has strong robustness against scaling attacks.

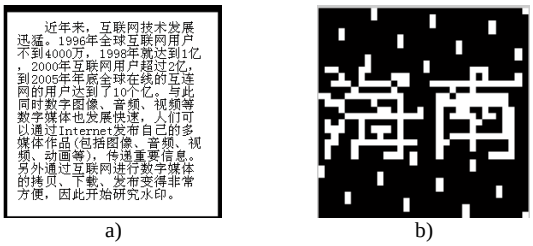


a) b)

FIGURE 4 The watermarked text image under scaling attacks: (0.5 times), a) the watermarked text image under scaling attacks, b) the extracted watermarking

4.2.2 Cropping attacks

The cropping attacks are added to the watermarked image for validating the effectiveness of our proposed algorithm. Figure 5a shows that the medical image cropping from *Y* axis with the ratio of 2%. Moreover, Figure 5b shows that the watermarking image can be extracted with *NC* = 0.87. The results show that the watermarking algorithm has strong robustness against cropping attacks.



a) b)

FIGURE 5 Under cropping attacks (from the *Y* axis, 2%): a) an image with cropping attack, b) the extracted watermarking image

5 Conclusion

This paper presents a watermarking encryption algorithm based on Arnold scrambling and DFT for text images, combing the visual feature vector of image, the encryption technology and the concept of third-party, and integrating Arnold scrambling. In watermarking embedding process, Arnold scrambling is employed to preprocess on the original watermarking. Without knowing the scrambling algorithm and key, the attackers cannot recover the images even after extracting the watermarking from the watermarked image. Such additional encryption provides double protection for text images. Our experiments show the proposed embedding watermarking scheme has robustness for common attacks and geometrical attacks while still keeping the quality of the original text image. Moreover, the proposed watermarking algorithm can be applied to protecting the other area's images.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (No: 61263033) and the NSF of Hainan Province of China (No: 60894).

References

[1] Pan J, Hiang H, Wang F 2002 A VQ-based multi-watermarking algorithm *Proceedings of 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering TENCON '02* **1** 117-20
[2] Solachidis V,Pitas I 2001 *IEEE Transactions on Image Processing* **10**(11) 1741-53
[3] Zhou Y, Jin W 2009 A novel image zero-watermarking scheme based on DWT-SVD *2011 International Conference on Multimedia Technology (IMCT)* 2873-6
[4] Niu X M, Lu Z M, Sun S H 200 Digital image watermarking based on multiresolution decomposition *Electronics Letters* **36** 1108-10
[5] Unoki M, Miyauchi R 2011 Reversible Watermarking for Digital Audio Based on Cochlear Delay Characteristics *Proceedings of the 2011 Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing* 314-7
[6] Kutter M, Bhattacharjee S K, Ebrahimi T 1999 Towards second generation watermarking schemes *ICIP'99* Kobe Japan **3** 320-3
[7] Kaur S, Farooq O, Singhal R, Ahuja B S 2010 Digital watermarking of ECG data for secure wireless communication *Proceedings of the 2010 IEEE International Conference on Recent Trends in Information, Telecommunication and Computing* 140-4
[8] Li X, He G 2012 Efficient Audio Zero-Watermarking Algorithm for Copyright Protection Based on BIC and DWCM Matrix *IJACT: International Journal of Advancements in Computing Technology* **4**(6) 109-17
[9] Zhang Y, Bi H, Zhu B, Dong B 2012 An Improved ICA-based Digital Watermarking *AISS: Advances in Information Sciences and Service Sciences* **4**(2) 167-74

Authors

Fan Wu, born in March, 1989, Haikou, China

University studies: Master's degree at Information Science & Technology College of Hainan University.
Scientific interests: digital watermarking technology.
Publications: 1 paper.

Jingbing Li, born in July, 1966, Haikou, China

Current position, grades: professor and doctoral supervisor at Hainan University.
University studies: PhD in control theory and control engineering at Chongqing University of China in 2007.
Scientific interests: medical image processing, artificial intelligence, digital watermarking, volume data.
Publications: 4 patents, 37 papers, 4 books.

Opposite degree algorithm and its application in engineering data processing

Xue-Chen Wang[1], Xiao-Guang Yue[1*], Mostafa Ranjbar[2], Sanjay Kumar Boddhu[3], Maia Viera Cañiv[4]

[1]*School Resources and Environmental Engineering, Wuhan University of Technology, 430070 Wuhan, China*

[2]*Department of Mechanical Engineering, Eastern Mediterranean University, Famagusta, TRNC via Mersin 10, Turkey*

[3]*Wright State University & Qbase Inc., Dayton, 45505 Ohio, USA*

[4]*Center for Environmental Studies of Cienfuegos, Cuba*

Received 1 March 2014, www.cmnt.lv

Abstract

In order to analysis and predict data, a new intelligent algorithm (opposite degree algorithm) is used for actual engineering example. The algorithm is based on concept of prior numerical, posterior values and opposite degree in the nature. The human's languages have positive words and negative words. The matrix method can calculate the opposite degree and predict data by considering data's relationship and opposite degree. At the same time, relevant results are obtained through the opposite degree calculation by using the data of coal and gas outburst. After the comparison of the actual results, the accuracy of prediction is 100%. The preliminary validation of opposite degree algorithm shows that the algorithm is basically feasible and effective. If this algorithm can be improved, it is expected to be applied in practical fields more widely.

Keywords: opposite degree, algorithm, engineering data processing, coal and gas outburst, prior value, posterior value, prediction method

1 Introduction

In China, coal and gas outburst accidents often happen in the disasters. Every year, a number of incidents caused casualties. Therefore, it has a very important realistic significance for the study of coal and gas prediction. There are lots of staff's deaths or injury event in China. So it has an important practical significance to research the forecasting methods in the field of mining engineering. In the study of engineering, there are many intelligent algorithms has been used. There are lots of new intelligent algorithms and improved algorithms for solving engineering problems, such as data driven nonlinear control strategy [1], improved GA-ANFIS [2], improved SVM [3] and CIP [4] and HCDCCMM [5] and so on. This research tries to put forward a new method for computing the opposite degree, and hope to be able to predict the classification. In order to verify the effectiveness of the algorithm, based on the coal and gas outburst instance data, we make a prediction results compared with other method. In the nature, the opposite degree is inseparable from the use of the synonyms, and antipodal opposites in people's daily life [6-8]. Judging from intuition, the opposite degree between "male" and "female" is relatively absolute, and the opposite degree should be very high. So, for intelligent information processing, especially in nature situation, it is very important to use a precise method to represent the relationship between the real relationships.

2 Opposite degree algorithm

Opposite degree is proposed by Xiao-Guang Yue [9], and it has many species can be divided into two or more than two kinds, called gender. Typically, a species can have two kinds of gender: male and female. Therefore, creatures with the opposition can be applied in the analysis of opposition a group of engineering data.

The definition of opposite degree are presented, mainly involves the following aspects:

1) A priori value refers to the value for the training.

2) A posteriori numerical refers to numerical prediction analysis.

3) Opposite degree is a priori and posteriori numerical value between degrees of difference, $(-\infty, +\infty)$ is the range of the values.

In general, the definition of opposite degree is C, a priori value is A, the posterior value is B:

$$C = \frac{B-A}{A} = \begin{cases} < 0, B < A \\ = 0, A = B \\ > 0, B > A \end{cases},$$

[*]***Corresponding author*** e-mail: xgyue@whut.edu.cn

C is close to +∞, it shows that *B* is more big, and indicates that the difference of *B* and *A* is greater. *C* is close to –∞, it shows that *B* is small, and indicates that the difference of *B* and *A* is greater. *C* is close to 0, indicating that *B* and *A* are closer. *C* is equal to 0, show that the *A* is equal to *B*.

There is a prior matrix $A_{m\times n}$:

$$A_{m\times n}=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

The classification column vector of $A_{m\times n}$ is *X*:

$$X=\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}.$$

There is a posterior matrix $B_{p\times n}$:

$$B_{p\times n}=\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{p1} & b_{p2} & \dots & b_{pn} \end{bmatrix}.$$

The opposite degree of $A_{m\times n}$ and $B_{p\times n}$ can be calculated by using the following equation:

$$C(a_{ij},b_{kj})=\frac{b_{kj}-a_{ij}}{a_{ij}}.$$

Based on the opposite degree calculation, the corresponding classification column vector *Y* of $B_{p\times n}$ can be calculated:

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_p \end{bmatrix}.$$

3 Calculation steps

1) According to the following formula, calculated separately for each line of the opposite, and returns the absolute value of the minimum value. $S(k)$ is the opposite degree calculation results of the *k* line data of $B_{p\times n}$ and each line of $A_{m\times n}$, the minimum absolute value can be got (if the results obtained with two or more than two of the same value, it returns the first value):

$$C(k)=\min\begin{cases} |\sum_{j=1}^{n}\frac{b_{kj}-a_{1j}}{a_{1j}}| \\ \dots \\ |\sum_{j=1}^{n}\frac{b_{kj}-a_{ij}}{a_{ij}}| \end{cases}.$$

2) Returns the corresponding x_i to the y_p, the *k* line classification of $B_{p\times n}$ can be obtained.

3) Repeat (1) and (2), until calculate all the results.

4) Output *Y*.

4 Analysis of calculation results

Firstly, the example of coal and gas outburst data comes from Yunnan Enhong coal mine [10]. This is a big coal mine in China. And 20 sets of data are selected as the training samples in Table 1. In order to do a comparison, 8 sets of data are selected as testing samples in Table 2. Initial speed, consistency coefficient, gas pressure, destruction of coal type and mining depth are selected as influencing factors in Table 2. We choose the factors. We choose the 5 factors based on effectiveness, independence and feasibility. The last column corresponds to the classification column vector.

Secondly, by using the OD algorithm, the opposite result is obtained (results accurate to four digits after the decimal point). Opposite degree calculation is based on each test data and 20 sets of training data; the results are shown in Table 3. In order to express the degree of deviation between the data, the data do not take absolute value. Where, the minimum data results indicate that is the minimum absolute opposite degree results, namely the forecast error data and the training data with minimum computational results.

Thirdly, the minimum opposite degree results and corresponding approximate data rows are shown in Table 4.

Finally, the experimental results show that the accuracy of OD algorithm is 100% (as shown in Table 5). The accuracy of OD and accuracy of Chen Zuyun et al's method [10] are the same (as shown in Table 6). This results show that the algorithm has a positive role in classification and prediction.

TABLE 1 Training samples

No.	Initial speed	Consistency coefficient	Gas pressure (mpa)	Destruction of coal type	Mining depth (m)	Coal and gas outburst
1	19.0	0.31	2.76	3	620	yes
2	6.00	0.24	0.95	5	445	yes
3	18	0.16	1.2	3	462	yes
4	5	0.61	1.17	1	395	no
5	8	0.36	1.25	3	745	yes
6	8	0.59	2.8	3	425	yes
7	7	0.48	2	1	460	no
8	14	0.22	3.95	3	543	yes
9	11	0.28	2.39	3	515	yes
10	4.8	0.6	1.05	2	477	no
11	6	0.24	0.95	3	455	yes
12	14	0.34	2.16	4	510	yes
13	4	0.58	1.4	3	428	no
14	6	0.42	1.4	3	426	yes
15	4	0.51	2.9	5	442	yes
16	14	0.24	3.95	3	552	yes
17	4	0.53	1.65	2	438	no
18	6	0.54	3.95	5	543	yes
19	7.4	0.37	0.75	4	740	yes
20	3	0.51	1.4	3	400	no

TABLE 2 Testing samples

No.	Initial speed	Consistency coefficient	Gas pressure (mpa)	Destruction of coal type	Mining depth (m)
21	11	0.37	2.1	3	412
22	12.1	0.49	2	3	412
23	11.5	0.28	1.9	3	407
24	11.8	0.36	2.3	3	403
25	10.8	0.3	2.2	3	396
26	12.4	0.38	1.8	3	410
27	11.8	0.57	1.6	3	408
28	10	0.55	1.5	3	405

TABLE 3 Opposite degree calculation results

No.	21st group OD	22nd group OD	23rd group OD	24th group OD	25th group OD	26th group OD	27th group OD	28th group OD
1	–0.8021	–0.3934	–1.1467	–0.7343	1.028	–0.8081	–0.3025	–0.5028
2	2.11137	2.68944	1.59794	2.39334	1.85568	2.06609	2.54273	2.04739
3	1.56539	2.29316	0.85317	1.69452	1.16548	1.45133	2.43451	2.11968
4	3.64447	3.97572	3.41333	3.93623	3.53468	3.67939	3.69486	3.20901
5	0.6358	1.02663	0.28159	0.85594	0.47488	0.59589	0.88598	0.5214
6	–0.2785	0.02671	–0.4517	–0.1452	–0.424	–0.1984	–0.0275	–0.3291
7	2.28791	2.64506	2.06097	2.4618	2.12873	2.3544	2.56017	2.20484
8	–0.2421	0.35664	–0.6753	–0.1963	–0.5787	–0.1763	0.59021	0.33989
9	0.00009	0.48682	–0.3693	0.10331	–0.2573	0.03367	0.57013	0.2874
10	2.27206	2.60599	2.02527	2.59367	2.17543	2.29049	2.28749	1.77763
11	2.49102	3.06909	1.97784	2.77343	2.23612	2.44584	2.92258	2.42739
12	–0.596	–0.2108	–0.9274	–0.4933	–0.8012	–0.6094	–0.1899	–0.4295
13	1.85055	2.26102	1.66584	2.15514	1.7139	1.99883	2.02889	1.46597
14	1.18142	1.57904	0.89588	1.41268	1.01529	1.21958	1.42441	0.99832
15	0.73175	1.20757	0.60001	0.96075	0.54278	0.99339	1.14245	0.61196
16	–0.3946	0.15866	–0.7936	–0.3448	–0.7042	–0.3325	0.36205	0.11939
17	2.16148	2.60229	1.98404	2.44328	2.00348	2.34396	2.42668	1.87148
18	–0.5911	–0.2108	–0.7343	–0.4422	–0.7582	–0.4189	–0.2213	–0.5892
19	1.59324	1.93288	1.14414	1.92883	1.48874	1.40676	1.56982	1.13514
20	2.92216	3.45269	2.757	3.28957	2.74966	3.18915	3.21384	2.49569

TABLE 4 The minimum opposite degree results and corresponding approximate data rows

Results	21st group	22nd group	23rd group	24th group	25th group	26th group	27th group	28th group
The minimum opposite degree results	0.00009	0.02671	0.28159	0.10331	–0.2573	–0.1763	–0.0275	0.2874
Corresponding approximate data rows' No.	9	6	5	9	9	8	6	9

TABLE 5 Comparison of the testing results and the actual results

Results	21st group	22nd group	23rd group	24th group	25th group	26th group	27th group	28th group
Testing Results	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Actual Results	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE 6 Comparison of different methods

No.	Method name	Accuracy
1	Chen Zuyun et al's method	100%
2	OD	100%

5 Conclusions

Human thinking with divergent characteristics, new things in the human brain will produce antagonism and association. We propose a new intelligent algorithm – opposite degree algorithm. The algorithm can calculate the opposite degree of values, so as to achieve the purpose of classification and prediction. Meanwhile, in order to verify the effectiveness of the algorithm, using coal and gas outburst instance data and OD algorithms to predict the classification, and do a comparison with other methods. Its prediction result is 100%. The study shows that the algorithm is feasible and effective. The next focus of the work is divided into two parts, on the one hand, to further verify the algorithm from the point of view of mathematics, and attempts to prove that the algorithm is applicable and feasible; on the other hand, to further improve the algorithm, especially in the numerical prediction to make supplement, expected application in actual field more.

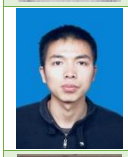
References

[1] Li Y G, Shen J, Lee K Y, et al. 2012 *Proc. Int. Conf. on the 18th IFAC World Congress*, Milano, Italy, 14778-83
[2] Xu Z, Mao Z Z 2012 *J Cent South Univ* **19**(9) 2520-7
[3] Yang S X, Cao Y, Liu D et al. 2011 *J Cent South Univ* **18**(6) 2074-9
[4] Yue X G, Zhang G, Ren Q G, et al. 2013 *Applied Mechanics and Materials* **340** 126-30
[5] Lu Y, Li X D 2011 *Safety Science* **49**(2) 279-85
[6] Mai F J, Wang T, Song R 2007 *Proc Int Conf on the 7th International Conference on Chinese Computing* Beijing China 204-9
[7] Mai F J, Wang T, Song R 2007 *Proc Int Conf on the 7th International Conference on Chinese Computing* Beijing China 90-4
[8] Mai F J, Wang T, Song R 2008 *Journal of Chinese Information Processing* **22**(4) 39-42 *(in Chinese)*
[9] Yue X G 2014 Sciencepaper Online http://www.paper.edu.cn/releasepaper/content/201405-173 *(in Chinese)*
[10] Chen Z Y, Zhang G Z, Wu C F, et al. 2010 *Industrial Safety and Environmental Protection* **36**(5) 33-6 *(in Chinese)*

Authors

Xue-Chen Wang, Guangdong, China

Current position: Doctor of Safety Engineering in Wuhan University of Technology, China.
University studies: BD at Sun Yat-sen University of China, MD at Sun Yat-sen University of China.
Scientific interests: safety engineering.
Publications: 3 papers, 3 books.



Xiao-Guang Yue, Henan, China

Current position: Doctor of Mining Engineering in Wuhan University of Technology, China
University studies: BD at PLA Information Engineering University of China. MD at Guilin University of Technology, China.
Scientific interests: safety engineering.
Publications: 30 papers

Mostafa Ranjbar, Turkey

Current position, grades: assistant professor at Eastern Mediterranean University, Turkey.
University studies: PhD in mechanical engineering.
Scientific interest: intelligent algorithms for engineering.
Publications: 10 papers.

Dr. Sanjay Boddhu, USA

Current position: Research Scientist at Qbase, Assistant Research Professor at Wright State University, Invited Visiting Researcher at Wright Brother's Institute's ICC Discovery.
Scientific interests: computational intelligence, natural language processing, deep learning, situational awareness and cyber physical system architectures, IOT.
Publications: 30 papers, 2 books, 1 book chapter.

Maia Viera Cañiv, Cuba

Current position, grades: Center for Environmental Studies of Cienfuegos, Cuba. Professor at the Cienfuegos University.
University studies: Santa Clara University, Cuba.
Scientific interests: environmental engineering.

A strategy of attribute reduction based on partition

Hui Wang[1*], Tao Zheng[1], Weiwei Zhang[2]

[1]*People's Public Security University of China, Beijing, 100038, China*

[2]*Tianjin Design & Reserarch Institute of Electric Drive Co.LTD. Tianjin, 300180, China*

Received 1 March 2014, www.cmnt.lv

Abstract

The attribute reduction is an important pre-processing step for data mining. In order to avoid striking equivalence classes repeatedly for positive region or information entropy reduction it is proposed to calculate attribute reduction by constructing partition directly. At the same time the judgments of the absolute reduction and the relative reduction based on the equivalent division are proved. And the data description quality for the relative reduction has been defined. It is shown that striking minimum relative reduction of decision table is in the cost of the relative decline of description quality for classification.

Keywords: attribute reduction, partition, rough sets, data mining

1 Introduction

The attribute reduction is an important pre-treating step in the process of data mining [1]. It can reduce the data scale for the following mining effectively [2, 3]. But finding a minimum attribute set is a NP-hard problem due to the combinatorial explosion of properties [4]. It has been the main problem of various reduction algorithms to get results quickly and accurately.

There are three classical methods for attribute reduction, such as distinguish matrix [5], positive region [6] and information entropy [7]. However, the reduction algorithm based on distinguish matrix is only suitable for processing small decision tables relatively because of the matrix calculation increasing the space overhead. So the reduction for large data tables is on the basis of the positive region or information entropy mainly [8]. But calculating equivalence classes is a basic and essential step whether it is positive attribute reduction or information entropy. To avoid duplication of obtaining equivalence classes, it is proposed to reduce directly in the process of equivalence classes obtained. This can decrease computational overhead for the positive region and information entropy reduction.

2 The concept of attribute reduction

The target of attribute reduction is to delete irrelevant or unimportant attributes. There is same basic category between the original set and the set having removed some attributes. Those attributes removed do not change the overall description of the domain. Any object of the information system is on behalf of a class with the same regularity properties after reduction. So the number and composition of the object have been simplified for the information system. This will reduce the time and space overhead for the subsequent data processing. The knowledge used in rough set and equivalence relation is shown as follows:

Definition 1 [9] The information system can be represented as a quadruplet $S = (\tilde{X}, R, V, f)$, where, $\tilde{X}$ represents a non-empty finite set of objects named domain. $\tilde{X} = \{X_1, X_2, \cdots, X_N\}$. X_i is an object. R is attribute set. V represents the set of the property value. V_a is range of the attribute $a \in R$. f is an information function $\tilde{X} \times R \to V$, $a \in R$, $X \in \tilde{X}$, $f_a(X) \in V_a$.

Definition 2 [9] Let A is an attribute set of the information table. a is on behalf of a particular attribute value. The objects X_i, X_j are named equivalence relation with respect to the property A. If X_i, X_j satisfies the relation as follows:

The equation $f_a(X_i) = f_a(X_j)$ founded for $\forall a \in A$, $A \subset R$, $X_i \in \tilde{X}$, $X_j \in \tilde{X}$. This relation is also known as indistinguishable relationship expressed as:

$$IND(A) = \{< X_i, X_j > \mid < X_i, X_j > \in \tilde{X} \times \tilde{X}, \forall a \in A, f_a(X_i) = f_a(X_j)\}. \quad (1)$$

Definition 3 [9] Let object $X \in \tilde{X}$. The set posed by the elements having the same values with X of $\tilde{X}$ on the attribute collection B will be named equivalence class of relation $IND(B)$ as $[X]_B$.

$$[X]_B = \{X_i \mid < X, X_i > \in IND(B)\}, \quad (2)$$

$[X]_B$ is also known as the basic concept or category of the attribute set B.

[*]***Corresponding author*** e-mail: wanghui_wwh@163.com

Definition 4 [9] The partition related to attribute set B on the basis of domain $\tilde{X}$ is expressed as:

$$\pi_B = \tilde{X} / B = \{E_i \,|\, E_i = [X_i]_B, i = 1,2,\cdots\}, \quad (3)$$

the partition has the properties as $E_i \neq \varnothing$. If $i \neq j$ then $E_i \cap E_j = \varnothing$ and $\tilde{X} = \bigcup E_i$.

Definition 5 [10] The entropy $H(P)$ of knowledge P defined as:

$$H(P) = -\sum_{i=1}^{n} p(X_i) \log_2(p(X_i)), \quad (4)$$

where, P is an attribute set of domain $\tilde{X}$.

Definition 6 [10] The partitions of the indistinguishable relationship about attribute set P and Q expressed as $\tilde{X} / IND(P) = \{X_1, X_2, \cdots, X_n\}$ and $\tilde{X} / IND(Q) = \{Y_1, Y_2, \cdots, Y_m\}$. The conditional entropy $H(Q|P)$ about Q related to P defined as:

$$H(Q|P) = -\sum_{i=1}^{n} p(X_i) \sum_{j=1}^{m} p(Y_j|X_i) \log_2(p(Y_j|X_i)), \quad (5)$$

where, $p(Y_j|X_i) = |Y_j \cap X_i| / |X_i|$. $i = 1,2,\cdots,n$. $j = 1,2,\cdots,m$.

Definition 7 [10] Let P and Q are equivalence relation clusters on the basis of domain $\tilde{X}$. The P positive domain of Q defined as $POS_P(Q)$.

$$POS_P(Q) = \bigcup_{X \in \tilde{X}/IND(Q)} P_-(X), \quad (6)$$

where, $P_-(X) = \bigcup \{Y \,|\, Y \in \tilde{X}/IND(P) \wedge Y \subseteq X\}$.

Definition 8 [10] $\tilde{X}$ is a domain. P and Q are attribute sets on the basis of $\tilde{X}$. $r \in P$. If $POS_P(Q) = POS_{(P-\{r\})}(Q)$ then relation r of P is absolutely unnecessary property related to Q. Or r of P is absolutely necessary property. If any $r \in P$ is necessary absolutely that the attribute set P is independent according to Q.

Lemma 1 [10] $\tilde{X}$ is a domain. P is conditional attribute set of $\tilde{X}$. The necessary and sufficient condition that r in P is necessary absolutely is $H(\{r\}|P - \{r\}) = 0$.

Lemma 2 [10] $\tilde{X}$ is a domain. P is conditional attribute set of $\tilde{X}$. d is decision attribute and domain $\tilde{X}$ is consistency with respect to $\{d\}$ on P. Then the necessary and sufficient condition that r with respect to $\{d\}$ is necessary absolutely is $H(\{d\}|P) = H(\{d\}|P - \{r\})$.

Lemma 3 [10] $\tilde{X}$ is a domain. P is conditional attribute set of $\tilde{X}$. d is decision attribute and domain $\tilde{X}$ is consistency with respect to $\{d\}$ on P. Then the necessary and sufficient condition that p with respect to $\{d\}$ is independent is $H(\{d\}|P) \neq H(\{d\}|P - \{r\})$ for any attribute $r \in P$.

Lemma 4 [10] $\tilde{X}$ is a domain. P is conditional attribute set of $\tilde{X}$. d is decision attribute and domain $\tilde{X}$ is consistency with respect to $\{d\}$ on P. Then the necessary and sufficient conditions that $Q \subseteq P$ with respect to $\{d\}$ is a reduction are two things as following:

1) $H(\{d\}|Q) = H(\{d\}|P)$.

2) Q with respect to $\{d\}$ is independent.

3 Reduction description based on partition

The relationship with equal value is an equivalence relation in an attribute information table. Any of equivalence relations will form a partition on the object belonged to the domain. This partition means a kind of classification model corresponding to the domain. The reduction result for attribute information table contains the minimum number of attributes. All of attributes in the reduction set are necessary absolutely. These attributes have same classification compared to all attributes on the domain objects. Generally there are two reduction forms as the absolute reduction and the relative reduction. The absolute reduction process applies to general information processing system. In this system all the properties aren't distinguish, and the reduction contains only part of properties. The relative reduction is suitable for handling special information table as the decision table. The properties are divided into condition attribute and decision attribute in this table. There are multiple properties but only one decision attribute in the table usually. For multi-attribute decision table, the process is needed to convert it into a single decision table at first. The condition attributes and the decision attributes formed a classification model for domain respectively. The reduction means to strike a condition attribute set containing the minimum attributes for the decision table. The result with respect to the decision classification is consistent with all the condition properties and expresses as a subset of condition attribute set. If the relativity of the decision attributes isn't considered the absolute reduction is the relative reduction also. Either the absolute reduction or the relative reduction has similarities to strike a minimum reduction. And the results of the reduction may be multiple normally.

Definition 9 $\tilde{X}$ is a domain. P and Q are attribute sets on the basis of $\tilde{X}$. And $Q \subseteq P$. If $\pi_{IND(Q)} = \pi_{IND(P)}$ and for either attribute $r \in Q$ the formula

$\pi_{IND(Q)} \neq \pi_{IND(Q-\{r\})}$ is founded. Then Q is an absolute reduction with respect to P.

Definition 10 $\tilde{X}$ is a domain. P and Q are attribute sets on the basis of $\tilde{X}$. And $Q \subseteq P$. d is decision attribute. If $POS_P(\{d\}) = POS_Q(\{d\})$ the Q is a relative reduction with respect to P.

In short, the main problem of attribute reduction is committed to strike less conditions meeting the consistent classification. The absolute reduction is suitable for general information table. The information table reduced can be used for finding association rules, clustering and other data mining process. The relative reduction is applicable to special information table like decision table. The classification rules can be generated with respect to the decision attributes through reduction. But the absolute reduction does not mean the relative reduction. The absolute reduction has strong constraint.

3.1 ABSOLUTE REDUCTION BASED ON PARTITION

The indistinguishable relation of the absolute reduction set will generate classification with objects on the domain. If a property a has r value that it will form a partition $\tilde{X}/\{a\}$ involving r equivalence classes for all objects of the domain. $\tilde{X}/\{a\} = \{X_1, X_2, ..., X_r\}$. When the rest of properties add to the equivalence classes as a part of the division the classes will be split likely. So the classification will be refinement further. When adding an attribute does not change the equivalence classes of division this property having no contribution to the classification is unnecessary. If the division generated by the distinguished relationship on the attribute set is smallest that the remaining properties can be reduced as unnecessary attributes.

Definition 11 The information system is a quadruplet $S = (\tilde{X}, R = C \cup D, V, f)$, where $IND(P)$ and $IND(Q)$ are indistinguishable relationships on C and $\tilde{X}/IND(P) = \{X_1, X_2, ..., X_m\}$, $\tilde{X}/IND(Q) = \{Y_1, Y_2, ..., Y_n\}$. For any $X_i \in \tilde{X}/IND(P)$ there is a $Y_j \in \tilde{X}/IND(Q)$ making $X_i \subseteq Y_j$ established. And there are $X_k \in \tilde{X}/IND(P)$ and $Y_r \in \tilde{X}/IND(Q)$ making $X_i \subseteq Y_j$ established at least. So the partition $\tilde{X}/IND(P)$ is more refined than $\tilde{X}/IND(Q)$. Or $\tilde{X}/IND(P)$ is a subdivision of $\tilde{X}/IND(Q)$.

Theorem 1 The information system $S = (\tilde{X}, C, V, f)$. $C' \subseteq C$ .The necessary and sufficient condition that the property $r \in C - C'$ is unnecessary absolutely is $\pi_{C'} = \pi_{C' \cup \{r\}}$.

Proof (Sufficiency) Let $\pi_{\{r\}} = \{B_1, B_2, \cdots B_k\}$ and $\pi_D = \{C_1, C_2, \cdots C_m\}$. According to the system S, if $\pi_{C'} = \pi_{C' \cup \{r\}}$ then $IND(C') = IND(C' \bigcup \{r\})$. Let $\pi_{C'} = \pi_{C' \cup \{r\}} = \{A_1, A_2, \cdots A_n\}$. And the partition $\pi_{C'}$ is more refined than $\pi_{\{r\}}$. From the definition 4 of equivalence division, either $A_i \subseteq B_j$ or $A_i \bigcap B_j = \varnothing$ has been established for any A_i or B_j. So $p(B_j|A_i) = 1$ or $p(B_j|A_i) = 0$. From the Equation (5) it is known that $H(\{r\}|C') = -\sum_{i=1}^{n} p(A_i) \sum_{j=1}^{m} p(B_j|A_i) \log(p(B_j|A_i)) = 0$.

According to Lemma 1 that attribute r is absolutely unnecessary relative to C'.

(Requirement) If the attribute r is absolutely unnecessary relative to C' then $IND(C') = IND(C' \bigcup \{r\})$. From one to one relationship between the equivalence relation and the division the equation $\pi_{C'} = \pi_{C' \cup \{r\}}$ is founded.

Theorem 1 illustrates that both equivalent division reduction and information entropy reduction are similar to for determining the absolute necessary properties. The process for seeking absolute reduction based on equivalent partition is shown according to the data in Table 1.

TABLE 1 Information table of sample (1)

$\tilde{X}$	a_1	a_2	a_3
x_1	1	1	1
x_2	0	0	0
x_3	2	1	3
x_4	1	0	3
x_5	1	1	1
x_6	3	2	2
x_7	3	0	0
x_8	0	0	0

From Table 1, $R = \{a_1, a_2, a_3\}$. Beginning from the attribute a_1 for division construction, the division formed by increasing properties a_2 and a_3 sequentially is shown as:

$$\tilde{X}/\{a_1\} = \{\{x_1, x_4, x_5\}, \{x_2, x_8\}, \{x_3\}, \{x_6, x_7\}\},$$

$$\tilde{X}/\{a_2\} = \{\{x_1, x_3, x_5\}, \{x_6\}, \{x_2, x_4, x_7, x_8\}\},$$

$$\tilde{X}/\{a_1, a_2\} = \{\{x_1, x_5\}, \{x_4\}, \{x_2, x_8\}, \{x_3\}, \{x_6\}, \{x_7\}\},$$

$$\tilde{X}/\{a_1, a_2, a_3\} = \{\{x_1, x_5\}, \{x_4\}, \{x_2, x_8\}, \{x_3\}, \{x_6\}, \{x_7\}\}.$$

From the result of division, it is seen that the distribution of equivalence classes division has not changed by adding a_3. So the attribute a_3 is unnecessary absolutely. $\{a_1, a_2\}$ is independent because of

$\tilde{X}/\{a_1,a_2\} \neq \tilde{X}/\{a_1\}$ by removing a_2. As same time, $\tilde{X}/\{a_1,a_2\} \neq \tilde{X}/\{a_2\}$ by removing a_1. An absolute reduction of R in Table 1 is $\{a_1,a_2\}$. By similar method $\{a_1,a_3\}$ is the other absolute reduction of R too. It can be seen that the absolute reduction of the information system is not the only.

From the above analysis, removing unnecessary property does not change the original classification ability of the information system because this attribute is not able to provide new information for further subdivision to the objects of the domain. And the results are the same on the contrary. So seeking absolute reduction process can be carried out by dividing the refinement. At first any of property is selected to make division according to their different values by scanning information system vertically. Then adding the remaining property individually will divide refinement in equivalence class set until the division does not change.

An absolute reduction set contains minimum attributes for all objects described on the field. But the relative reduction issues tend to be more concerned in the application process. This problem is what conditions would lead to the decision situation occurred.

3.2 RELATIVE REDUCTION BASED ON PARTITION

Both condition attribute set and decision attribute set will be formed division on the domain for special information system like decision table. At this time the relationship between one classification and another tends to be more concerned. At this point the attribute reduction goal is to focus on finding some necessary properties from condition set. The classification capability of these properties is same as all conditions relative to decision attribute. This leads to the concept of relative reduction.

In a decision table, P is condition attribute set and Q is decision attribute set. According to the definition 7 refining division will lead to changes in the positive region $POS_P(Q)$. The attributes that does not belong to the positive region originally are likely to fall in the region. When all equivalence classes refined fall within the equivalence classes corresponding decision this condition set is a reduction. The following theorem will be founded.

Theorem 2 The information system $S=(\tilde{X},R,V,f)$. $R=C\bigcup D$. The domain $\tilde{X}$ is consistent with respect to the decision attribute set D, $C^{'}\subseteq C$, π_D is a partition formed by decision attribute. $\pi_{C^{'}}$ and π_C are division formed by indistinguishable relationship on the condition attribute set $C^{'}$ and C. If both of conditions are true as following that the attribute set $C^{'}$ is a D reduction of C.

1) $\pi_{C^{'}}$ is a subdivision of π_D.

2) $\neg\exists r((r\in C^{'})\wedge\pi_{C^{'}-\{r\}}$ is a subdivision of π_D).

Proof:

Let $\pi_{C^{'}}=\{X_1^{'},X_2^{'},\cdots,X_m^{'}\}$, $\pi_C=\{X_1,X_2,\cdots,X_n\}$. $\pi_D=\{Y_1,Y_2,\cdots,Y_k\}$. The equation $POS_C(D)=\tilde{X}$ is founded because of the domain $\tilde{X}$ is consistent with respect to the decision attribute set D. So $\pi_{C^{'}}$ is a subdivision of π_D.

For any $X_i^{'}\in\pi_{C^{'}}$ there is $Y_j\in\pi_D$ making the equation $X_i^{'}\subseteq Y_j$ established because $\pi_{C^{'}}$ is a subdivision of π_D. At the same time knowing from Definition 4 that $X_i^{'}\bigcap Y_l=X_i^{'}$ or $X_i^{'}\bigcap Y_l=\varnothing$ is founded for any $Y_l\in\pi_D$. So $p(Y_l|X_i^{'})=1$ or $p(Y_l|X_i^{'})=0$. From the Equation (5) and $C^{'}\subseteq C$,

$H(D|C^{'})=-\sum_{i=1}^{n}p(X_i)\sum_{j=1}^{k}p(Y_j|X_i)\log(p(Y_j|X_i))=0$. In the same way $H(D|C)=0$ because $\pi_{C^{'}}$ is a subdivision of π_D.

Assuming there is a property $a\in C^{'}$ making the equation $H(D|C^{'})=H(D|C^{'}-\{a\})$ established. In other words the conditional entropy of attribute set $C^{'}$ is equal to the entropy of $C^{'}-\{a\}$ with respect to the decision attribute set D. And because the domain $\tilde{X}$ is consistent with respect to the decision attribute set D that the attribute a is unnecessary by Lemma 2. a in $C^{'}$ does not affect its classification with respect to D. Then $\pi_{C^{'}}=\pi_{C^{'}-\{a\}}$ or $\pi_{C^{'}-\{a\}}$ is a subdivision of π_D. This conflicts to the condition (2) that $\neg\exists r((r\in C^{'})\wedge\pi_{C^{'}-\{r\}}$ is a subdivision of π_D). So the assumption does not hold. And any $r\in C^{'}$ the formula $H(D|C^{'})\neq H(D|C^{'}-\{r\})$ is founded. By Lemma 3 it is shown that $C^{'}$ is independence. So $C^{'}$ is a reduction by Lemma 4.

Theorem 2 illustrates that the attribute reduction is a subset of the condition attribute set C with respect to the decision set D. And the result of the reduction is relatively consistent with the division to the domain forming by all attributes of C. But the reduction does not include redundant attributes. Therefore, the attribute reduction based on the partition can be expressed as to strike a subset $C^{'}$ of the condition attribute set C. This subset $C^{'}$ is consistent with C for classification relative to the decision D.

4 The description of the reduction quality for data classification based on the equivalent division

In general case, the attribute information table containing more attributes of the universe is more detailed for object description. But too much attributes will influence the

comprehensive feature extraction for domain objects. Attribute reduction to facilitate the comprehensive knowledge extraction will lead to a quality decline on the object description in the domain. At present the property set of the absolute reduction forms smaller classification to the universe objects. This reduction is more detailed to the object description. In contrast the description to relative reduction on the domain object is rough. The data description quality for relative reduction is defined as follows.

Definition 12 The information system $S=(\tilde{X},R,V,f)$. $R^{'}$ is absolute reduction set and $C^{'}$ is relative reduction set, $\pi_{R^{'}}=\{X_1,X_2,\cdots,X_n\}$, $\pi_{C^{'}}=\{Y_1,Y_2,\cdots,Y_m\}$. The definition description of the quality of the classification to the domain as follows:

$$r_{C^{'}}(R^{'})=\left|\pi_{C^{'}}\right|/\left|\pi_{R^{'}}\right|. \tag{7}$$

It is shown in Table 2 by adding a row of decision attributes to Table 1.

TABLE 2 Information table of sample (2)

$\tilde{X}$	a_1	a_2	a_3	d
x_1	1	1	1	0
x_2	0	0	0	1
x_3	2	1	3	0
x_4	1	0	3	0
x_5	1	1	1	0
x_6	3	2	2	1
x_7	3	0	0	1
x_8	0	0	0	1

In Table 2, $R=\{a_1,a_2,a_3,d\}$. The division is constructed with the beginning of the property a_1 with respect to the decision attribute $\{d\}$. Part of the division is shown as following:

$$\tilde{X}/\{d\}=\{\{x_1,x_3,x_4,x_5\},\{x_2,x_6,x_7,x_8\}\},$$

$$\tilde{X}/\{a_1\}=\{\{x_1,x_4,x_5\},\{x_2,x_8\},\{x_3\},\{x_6,x_7\}\},$$

$$\tilde{X}/\{a_1,a_2\}=\{\{x_1,x_5\},\{x_4\},\{x_2,x_8\},\{x_3\},\{x_6\},\{x_7\}\},$$

$$\tilde{X}/\{a_1,a_3\}=\{\{x_1,x_5\},\{x_3\},\{x_4\},\{x_6\},\{x_7\},\{x_2,x_8\}\}.$$

The results can be drawn from the result. The absolute reduction generated by Table 2 is $\{a_1,a_2\}$ or $\{a_1,a_3\}$. The partition $\tilde{X}/\{a_1,a_2\}$ or $\tilde{X}/\{a_1,a_3\}$ is finer than $\tilde{X}/\{a_1\}$. But $\tilde{X}/\{a_1\}$ is subdivision of $\tilde{X}/\{d\}$ also and the classification can be done with respect to the decision attribute $\{d\}$. It is shown that the relative reduction results of Table 2 is $\{a_1\}$ by Theorem 2. The description classification quality of $\{a_1\}$ can be calculated by the Equation (7). $r_{\{a_1\}}(\{a_1,a_2\})=r_{\{a_1\}}(\{a_1,a_3\})=0.67$. Therefore striking minimum relative reduction of decision table is in the cost of the relative decline of description quality for classification. To improve the quality described on the domain objects some of the properties need to be increased on the results of the relative reduction for further refinement. The finer division are more specific for domain description.

Striking minimum attribute reduction and seeking higher classification description quality of the decision table cannot be taking into account. In the actual application this will be weighed in accordance with the characteristics of the field.

5 Conclusions

In classical rough set theory, attribute reduction is on the basis of positive region, distinguish matrix and information entropy. However calculating the distinguish matrix is excessive. And calculating equivalence classes is important basic steps for positive region or information entropy. The determination and proof for refinement reduction are given from absolute and relative reduction. It is shown that striking minimum relative reduction of decision table is in the cost of the relative decline of description quality for classification. These results must be processed according to the knowledge of specific fields.

Acknowledgments

This work is supported by the project of the people's security university of china about monitoring system research for internet public opinions.

References

[1] Tumasonis R, Rastenis R 2006 new statistical characteristics for mining frequent sequences in large databases *Computer Modelling and New Technologies* **10**(4) 46-52
[2] Shynybekov D, Bektemyssova G, Kuandykov A I 2012 decision making on the basis of group methods for systematization and project choice *Computer Modelling and New Technologies* 16(1) 35-40
[3] Pawlak Z 1982 Rough sets *International Journal Computer and Information Science* **11**(5) 341-56
[4] Wong S K M, Ziarko W 1986 a machine learning approach to information retrieval *Proceedings of the 9th annual international ACM SIGIR conference on Research and development in information retrieval* 228-33
[5] Pawlak Z, Skowron A 2007 Rudiments of rough sets *Information Sciences* **177**(1) 3-27
[6] Skowron A, Rauszer C 1992 The discernibility matrices and functions in information systems *Intelligent Decision Support Theory and Decision Library* **11** 331-62
[7] Zhou L, Jiang F 2011 a rough set approach to feature selection based on relative decision entropy Rough Sets and Knowledge Technology *Berlin Heidelberg:* Springer 110-9
[8] Zhang J B, Li T R, Ruan D 2012 neighborhood rough sets for dynamic data mining International *Journal of Intelligent Systems* **27**(4) 317-42
[9] Zhang W X, Wu Z W, Liang J Y 2001 *Rough set theory and methods Science Press Beijing* 12-22
[10] Wang G Y, Yu H, Yang D C 2002 decision table reduction based on conditional information entropy *Chinese Journal of Computers* **25** 759-66 *(in Chinese)*

Authors

Hui Wang, born in June, 1973, Inner Mongolia, China

Current position, grades: professor at People's Public Security University of China.
University studies: PhD degree in Control Theory and Control Engineering at University of Science and Technology Beijing of China, 2012
Scientific interest: data mining and analysing.
Publications: 15 papers.

Tao Zheng, born in December, 1974, Taiyuan, China

Current position, grades: professor at People's Public Security University Of China
University studies: bachelor's degree in Communication Engineering at the Chinese People's Police Officer University, 1997.
Scientific interests: wireless communication
Publications: 15papers

Weiwei Zhang, born in October, 1983, Tanjin, China

Current position, grades: senior engineer at Tanjin Design and Research Institute of Electric Drive.
University studies: PhD degrees in Control Theory and Control Engineering from University of Science and Technology Beijing of China, 2011.
Scientific interests: nonlinear control theory and applications in electric engineering.
Publications: 5 papers.

A path planning algorithm research for seeing eyes robot based on visibility graph algorithm

Chao Chen, Jian Tang, Zuguang Jin*

School of Mechanical Engineering, Jiangsu University of Science and Technology, Zhen Jiang 212003, China

Received 1 March 2014, www.cmnt.lv

Abstract

This paper presents a new indoor path planning algorithm for seeing eyes robot using the RFID system. Through combination of UHF radio frequency identification system and low radio frequency identification system to realize accurate positioning for robot. New algorithm combines the idea of visibility graph algorithm with A* search algorithm, which not only improves the efficiency of searching but also guarantees the feasibility of path planning at the same time. The simulation verify the effectiveness and feasibility of the method.

Keywords: path planning, seeing eyes robot, visibility graph algorithm

1 Introduction

Path planning is an important element of seeing eyes robot system, which can fall into two types: First, global path planning, information on the surrounding environment is completely known for path planning; Second, local path planning, information on the surrounding environment is completely unknown or part unknown for path planning. In this paper, we objective is to build an indoor robotic guide for the visually impaired, so the best choice for us is global path planning. Global path planning involves two basic problems are environment modelling and search strategy. For environment modelling the typically method is visibility graph, which need less modelling time and Smaller storage space. When accurate localization information is obtain hardly, this algorithm can made robot to achieve the target point. Visibility graph use similar polygon instead of obstacles which in surrounding environment to make a map, and then utilize straight lines that not intersection with obstacles polygon in map to connect initial point, target point and obstacles polygon vertex. These lines are "lines-of-visibility" and the endpoint of lines-of-visibility are "points-of-visibility". The visibility graph is consist of lines-of-visibility. Once the map is generated, the robot computes and follows the optimal path to destination autonomously with search strategy. But visibility graph algorithm lack of flexibility and have highly complexity. When initial point and target point changed, the robot need to reconfiguration the map [1]. In our strategy, the new algorithm combines the idea of visibility graph algorithm positioning with A*, which not only improves the efficiency of searching but also guarantees the feasibility of path planning at the same time. It is used heuristic means to enhance the environment adaptability and improve the real-time of visibility graph algorithm. The simulation verify the effectiveness and feasibility of the new algorithm.

2 Environment model

2.1 HARDWARE

In this paper, we use a independent design seeing eyes robot for the application object, which consist of four layers yakeli board. The robot has three wheels, one universal wheel in front and two driving wheel in the back, 8 ultrasonic sonars in the front, two infrared sensor on the both sides, a Siemens PLC, a voice module, a radio frequency identification module, and is equipped with two stepping motors that control by PLC. Figure 1 shows the integral structure of seeing eyes robot.

FIGURE 1 Integral structure of robot

* ***Corresponding author*** e-mail: snowden_chen@163.com

2.1 ENVIRONMENT MODEL CREATION

We use a radio frequency identification (RFID) module to obtain rang data of the surroundings. The RFID module consist of a radio frequency identification reader and several RFID tags. The radio frequency identification reader mounted on top of the robot. These RFID tags can be attached to obstacles in the environment and stored the information of the obstacles such as global coordinates, feature points and so on. They do not require any external power source or direct line of sight to be detected by the RFID reader. They are activated by the spherical electromagnetic field generated by the RFID reader with a radius of approximately 1.5 meters. Each tag is programmatically assigned a unique ID.

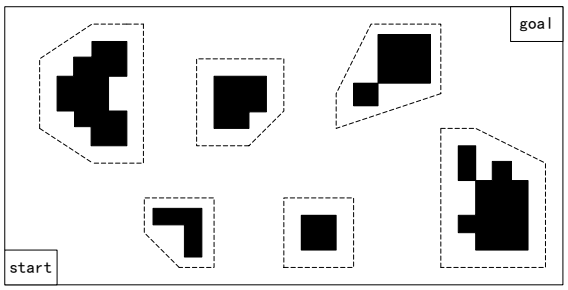


FIGURE 2 Mapping of environment model

As it can be seen in Figure 2, we use similar polygon instead of obstacles in the environment. Because of seeing eyes robot simplified to a point of 2D coordinates, obstacles polygon should be enlarge the object regions by $\omega/2 + a$, where ω is the robot width and a is the positional uncertainty of the robot (currently, 15cm) [2]. We only consider the obstacle polygon which is convex polygon and edge number less than seven. Those complex obstacles, such as concave polygon or roundness and so on, should be expansion to the convex polygon which accord with the requirements. Once the feature points of the obstacle polygon are known, we can record the information of the obstacle. During this process, we can record the information of the obstacles in environment to make a map. We should storage all points-of-visibility which in addition to itself of each feature point for path planning.

3 Indoor of robot localization

Because of the superiority to obtain the information of environment, the RFID system is very suitable for robot localization. But this system gained accurate localization information hardly, that it is easy by environmental influences [3]. In our method, we combine the uhf RFID module with low-frequency module, which reduce uncertainty of localization information due to effect of environment. We will introduces this method for the following aspects. First, the uhf RFID tags were attached to obstacles in the environment and stored the information of obstacles and the global coordinates of label. Second, the uhf RFID reader which mounted on the robot get the environment information, such as feature points of obstacles global coordinates of labels and received signal strength indicator (RSSI), by read the RFID tags in environment [4]. Because of the environmental influences is largely for RSSI, median strategy used to reduce this influences. By this strategy, the RFID reader collect n values of RSSI for a tag, and then these values be accorded to order of from big to small. The intermediate value is considered as the measurement result. Once the measurement result were obtained, we can make the RSSI value into distance value:

$$d = 10^{(ABS(RSSI)-A)/(10*n)}, \quad (1)$$

where A is the RSSI value when the distance from the reader to the tag one meter (currently, –45.8), and n is the attenuation coefficient of signal which because of the environmental influences (currently, 3). At a time, the reader receive many information of the tags in the environment. We choose three tags as nodes for localization which distance from the robot more recently, that the relative distance between tag and the robot is closer, the influence to robot localization is greater. Because of the complexity of environment, the estimated of distance between tag and the robot is greater than the actual one [5]. As can be seen in Figure 3, if three nodes' coordinates are $p_1(x_1, y_1)$, $p_2(x_2, y_2)$, $p_3(x_3, y_3)$, the estimated of distance values between tags and the robot are r_1, r_2, r_3, we used p_1, p_2, p_3 as the center of circle and used r_1, r_2, r_3 as the radius to painting circle.

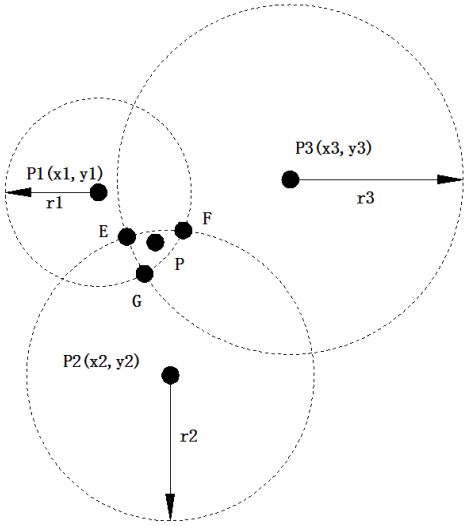


FIGURE 3 Location method

We were use triangle centroid localization algorithm to get the global coordinates of the robot in environment. In this method, the robot location is in two phases: First, we should calculation the three feature points coordinates which are the points of intersection of three circles and must in the overlap area of the three circles at the same time. And then, the position of robot is the centre of mass of tri-angular, which the three vertices of the triangular are the three feature points. To sum up, we can find out the

global coordinates of point $E(x_e, y_e)$ using the Equation (2). In a similar way, we can obtain the global coordinates of point $F(x_f, y_f)$ and point $G(x_g, y_g)$. As indicated above, we can get the global coordinates of the point $P\left((x_e + x_f + x_g)/3, (y_e + y_f + y_g)/3\right)$, which is the position of seeing eyes robot:

$$\begin{cases} \sqrt{(x_e - x_1)^2 + (y_e - y_1)^2} \le r_1 \\ \sqrt{(x_e - x_2)^2 + (y_e - y_2)^2} = r_2 \\ \sqrt{(x_e - x_3)^2 + (y_e - y_3)^2} = r_3 \end{cases}. \quad (2)$$

In the similar way, we can obtain the global coordinates of the point $P_1\left((x_{e+1} + x_{f+1} + x_{g+1})/3, (y_{e+1} + y_{f+1} + y_{g+1})/3\right)$ in other time. If the direction of the robot is θ. We can get this value as follows:

$$\theta = arctan\frac{(y_{e+1}+y_{f+1}+y_{g+1})/3-(y_e+y_f+y_g)/3}{(x_{e+1}+x_{f+1}+x_{g+1})/3-(x_e+x_f+x_g)/3}. \quad (3)$$

Then, we can layout low-frequency RFID tags in the environment which as a signpost. The low-frequency RFID reader were installed in the bottom of the robot for read the global coordinates which stored in the low-frequency RFID tags as the current coordinates of the robot, that reduce the location error which produced by the above localization algorithm and the robot move. Figure 4 show the layout of the low-frequency RFID tags in each intersection, that the read rang of the low-frequency RFID tags is small (currently, 10cm). In this manner, each tag apart to 20cm, and like this the robot can read the tags whether direction it passed through the intersection. If the robot read only to a tag, we used the global coordinates of this tag as the current coordinates of the robot. If the robot read two tags at the same time, we used the midpoint of the two tags as the current coordinates of the robot, for example, the robot read tag 1 and tag 2, the coordinates of the robot are $\left((x_1 + x_2)/2, (y_1 + y_2)/2\right)$. And if the robot read three tags, we used the center of mass of tri-angular, which the three vertices of the triangular are the three tags, as the current coordinates of the robot. For example, the robot read tag 2, tag 3 and tag 5, the coordinates of the robot are $\left((x_2 + x_3 + x_5)/3, (y_2 + y_3 + y_5)/3,\right)$. The location error of the robot were control within 10cm by this strategy.

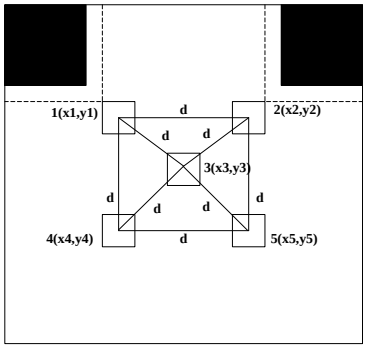


FIGURE 4 Mapping of low frequency tags

4 Description of the algorithm

4.1 THE RELATIVE POSITION OF LINE SEGMENT AND OBSTACLES

In the path planning process, we need to determine whether the desire line in conflict with obstacles in the environment [6]. In our strategy, as can be seen in Figure 5, if the desire line is the line segment SG which the starting point is S and the end point is G, we need to judgment whether S in conflict with the obstacle $ABCDE$. First, we should to work out the distances which between point S and the each feature of the obstacle are d_1, d_2, d_3, d_4, d_5 Second, we will make this distances to sorting by length and choose two feature points which distance from the S recently. Then, we need to judgment whether the line SG in conflict with the obstacle boundary line which the endpoint contain the two feature points. As shown in Figure 5, we need to judge whether the line SG in conflict with the lines AE, AB and ED. Obviously, if they are conflict that the line SG is in conflict with the obstacle; and if they are not intersect that the results is on the contrary. Therefore, the problem that whether the line SG in conflict with the obstacle is transformed into judge whether the line in conflict with the lines AE, AB and ED. As shown in Equation (4), the lines SG, AE, AB and ED were depicted as linear equation:

$$Ax + By + C = 0. \quad (4)$$

We can get the distance between points x, y and a line as follows:

$$d = (Ax + By + C)/\sqrt{A^2 + B^2}. \quad (5)$$

We can find it easily, in the coordinate system which as shown in Figure 5, that if the point is located in the top of the line, the value of d is less than zero, and if the point is located in the below of the line, the value of is greater than zero. Using the above relationship, if d_a is the distance between point A and line SG and d_e is the distance between point E and line SG, when $d_a * d_e < 0$ that the line SG is in conflict with the line AE; when $d_a * d_e = 0$ that the line SG is in conflict with the line AE and the node is the obstacles feature points, because of the line SG and AE only intersection one point and the obstacles were enlarged, so that the line AE not to interfere in the line SG; when $d_a * d_e > 0$ that the line SG is not in conflict with the line AE. In the similar way, we can judge whether the line SG in conflict with the line AB and line ED. Then we can judge whether the line SG in conflict with the obstacle.

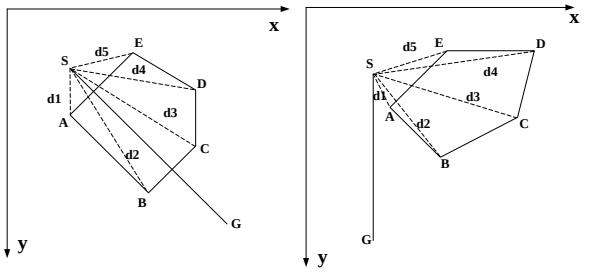


FIGURE 5 Verdict for relative of line SG and obstacle

4.2 VERDICT FOR POINTS-OF-VISIBILITY

In path planning process, when the change of the starting point and goal point, we need to search out new points-of-visibility for it. Therefore, we proposes a judgment method to judge whether a feature point of obstacles is the points of visibility for any point in space. We can find it easily that when obstacles polygon edge number less than seven, the number of the feature points of obstacle which can be the points of visibility for any point in space are no more than four, and the four points that distance from the point in space recently. Obviously, when without taking into consideration other obstacles in the environment, if the distance between the point in space and the feature point is shortest, and this feature point must be the points of visibility for the point in space [7]. In our strategy, the verdict for points of visibility in two phases: First, we only considerate one obstacle, if one feature point of the four points are not the points of visibility for the point in space, that the line of the feature point attach to the space point must be in conflict with the obstacle boundary line which the endpoints contain the nearest feature points. Therefore, we get the relative position which the line of the feature point attach to the space point and the obstacle boundary line which the endpoints contain the nearest feature points by the above method, and choose the feature points which are the points of visibility for the space point. Second, we need to judge whether the lines of this feature points attach to the space point in conflict with other obstacles in environment. If they are intersect, the feature point is not the points of visibility for the space point, and if they are not intersect that the results is on the contrary.

For example, as can be seen in Figure 6, the four points are point E, point F, point D and point A. Then, we need to judge whether the lines MF, MD and MA in conflict with the lines EF and ED. We can find that point E, point F and point D are the points of visibility for point M. And then, we should to judge whether the lines MF, MD and ME in conflict with other obstacles in the environment. Finally, we find that the points of E and D are the points of visibility for point M.

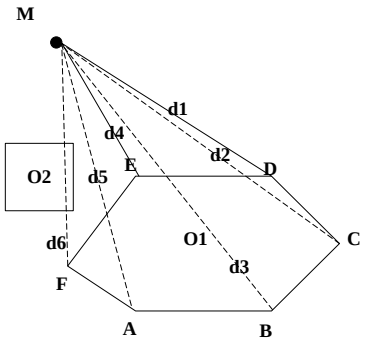


FIGURE 6 Verdict for points-of-visibility

4.3 THE ALGORITHM OF PATH PLANNING

A^* is a typical heuristic search algorithm in artificial intelligence, which uses a valuation function to estimate the weights of each node of the best, and work out the shortest path in the state-space [8]. In this paper, we should collect environment information by radio frequency identification module and voice module, and then combined with A^* algorithm for proposed a path planning algorithm. The algorithm processes as the following:

1) Create a two-dimensional environment model.

2) By UHF radio frequency identification module combined with low-frequency radio frequency identification module, the robot gets the current coordinates as starting point of S, and then obtain the information of the target point G by voice input functionality of voice module.

3) Search the feature points which can be the points of visibility for point S in the environment model.

4) Starting point of S joined the open list (open list is like a shopping list, and the path may through the point of it contain or may not be), as the current point. This article begins to search by formula 6 which as the valuation function.

$$f = g + h, \tag{6}$$

where the value of g is the actual value which is the travelled distance of from the starting point along the path to the current point. The value of h is the estimated value which is the euclidian distance of the current point to the target point. If the current point coordinates are (x_d, y_d) and target point coordinates are (x_m, y_m), you have:

$$h = \sqrt{(x_m - x_d)^2 + (y_m + y_d)^2}, \tag{7}$$

5) The current point was switched to the closed list (closed list saved all points that do not need to check again), and removed from the list of open. Then, we need to judge whether the feature points of obstacles which are the points of visibility for current point in the closed list. If in, we should skip over this point. Instead judge whether the feature points of obstacles which are the points of visibility for current point in the open list. If does not in, we are to the current point as its parent node for calculation

out its value of g, h and f and this feature point will joined in to the open list. If in, we should check new path whether more excellent by g value (g value is smaller, the path is better). If more excellent that we should change the parent node of the point to current point and update its value of g, h and f.

6) Judge whether the open list is empty, and if the open list is empty that the path does not exist, the algorithm ends.

7) Take the point which in the open list and the value of f is minimum as current point. And judge whether current point is the points of visibility for target point, if it is, we are to the current point as the parent node of target point for calculation out its value of g, h and f, and the target point joined in to the open list, the path is finding, end of the algorithm; otherwise go to step 5.

8) Save the path. Along every point of the parent point from the target point move until it is back to the starting point, and save the experienced points into the form of an array, reverse the order of the array, and then the array is the path.

4.4 PERFORMANCE ANALYSIS OF ALGORITHM

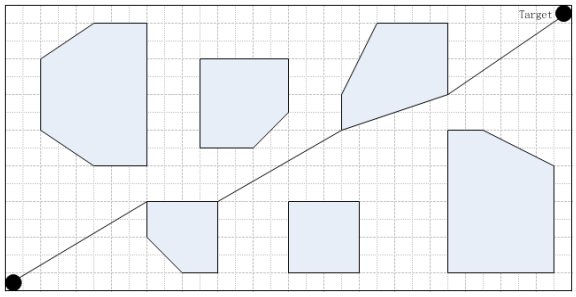
In this paper, we focus on the advantages and disadvantages of visibility graph algorithm, that proposed a path planning algorithm for the seeing eyes robot. This algorithm provides the method for judge the relative relationship of arbitrary points and the feature points of obstacles. The method ensure that path planning without need to refactor visual map when the starting point or target point is changed. Real time the path planning algorithm based on visibility graph are poor [9]. Such as the typical path planning algorithm of Dijkstra [10] based on visibility graph, it calculates the shortest path to each node to all other nodes, the main characteristics are use the starting point for the centre to the outer layers expand, until the expansion until the target point. Because of traverse a lot of nodes, its efficiency is very low. The new algorithm combines the idea of visibility graph algorithm with A^* search algorithm, and use of heuristic search approach greatly reduces the number of search nodes, that not only improves the efficiency of searching but also guarantees the feasibility of path planning at the same time.

5 Experimental debugging

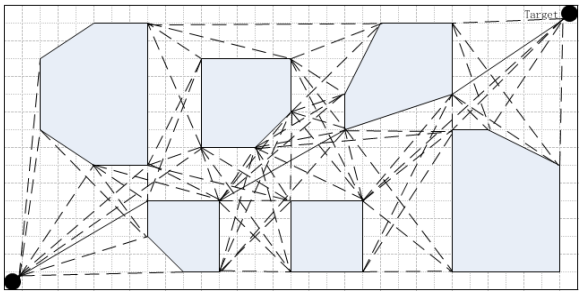
In the same of environment, we conducted simulation contrast test for the new algorithm and Dijkstra algorithm in MATLAB. We get the same path, but the new algorithm search time obvious is less than Dijkstra algorithm and the efficiency superior than the latter. And that the new algorithm can immediately again planning out effective path without need to refactor visual map when the starting point or target point is changed, but Dijkstra algorithm can't. As shown in Table 1 for the survey results of many experiments, Figure 7 are the find path of this article algorithm and Dijkstra algorithm, thus it can be seen that the new algorithm is better than Dijkstra algorithm.

TABLE 1 The time of search path

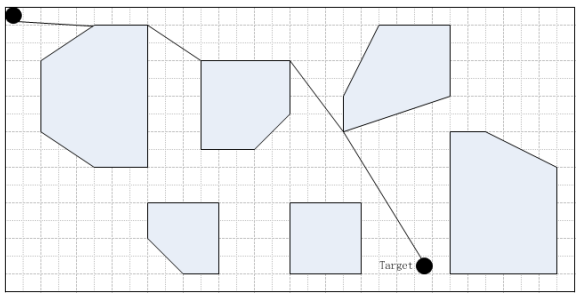
	minimum	**maximum**	**average**
New algorithm	0.4560s	0.8650s	0.6605s
Dijkstra algorithm	1.3550s	1.9460s	1.6505s



a) The path were found by the new algorithm



b) The path were found by Dijkstra algorithm



c) The path were found when the starting point or target point is changed

FIGURE 7 search path

Using the programming to the new algorithm to do the experiment, test sites and seeing eyes robot such as shown in Figure 8. In the process of moving, the robot getted the target point information by voice module, the system of robot will immediately planned an optimal or sub-optimal path which from the current position of the robot to the target point, and moving along the path until it reach the target point. Through the experiments that generate the executable path, meet the run-time requirements of robot, verify the viability and effectiveness of the algorithm.

FIGURE 8 Indoor environment experiment of robot

6 Conclusion

This article combines advantages of RFID module and voice module which on the data acquisition, first the robot gets the orientation of itself by UHF radio frequency identification module combined with low-frequency radio frequency identification module; then create the obstacles map by the idea of obstacles expansion and visibility graph; and then obtained the information of target point by voice module; finally use the search algorithm of *A** to search out the path which connection the starting point and target point. Experiments have shown that, the new algorithm which not only improves the efficiency of searching but also guarantees the feasibility of path planning at the same time.

References

[1] Zhang Y, Luo Y, Zheng T 2007 Mobile Robot Technology and It's Applications *Beijing Publishing House of Electronics Industry* 206-50 *(in Chinese)*

[2] Kodono K, Miura J, Shirai Y 2002 Autonomous visual navigation of a mobile robot using a human-guided experience *Robotics and Autonomous Systems* **40**(2-3) 121-30

[3] Hori T, Wada T, Ota Y, Uchitomi N, Mutsuura K, Okada H 2008 A Multi-sensing-Range Method for Position Estimation of passive RFID Tags *IEEE International Conference on Wireless & Mobile Computing, Networking & Communication* 208-38

[4] Liu J 2008 Application of RFID Technology in SLAM *Journal of South-Central University for Nationalities* **27**(3) 84-7

[5] Shi W, Xiong, Z, Xu L 2010 In-building RSSI-based user Localization Algorithm *Computer Engineering and Applications* **46**(17) 232-5

[6] Yang H, Xiao X, Yao D 2009 A V-graph based global path planning algorithm for mobile robot *Journal of Shenyang University of Technology* **31**(2) 225-9

[7] Hu G, Li X, Hu J 2007 A Jundge Algorithm for Polygon of The Visual Vertex *Gansu Science and Technology* **23**(10) 34-6 *(in Chinese)*

[8] Xu S, Cao Q 2011 A Visibility Graph Based Path Planning Algorithm for Mobile Robot *Computer Applications and Software* **3**(28) 220-36

[9] Qi Y, Yang Z, Huang Q 2009 Improved path planning algorithm based on *A** algorithm *Information and Electronic Engineering* **7**(4) 326-9

[10] Liu J, Ma S, Ma S 2011 Computation menthod of the dynamic shortest path based on improved-Dijkstra algorithm *Systems Engineering – Theory & Practice* (31) 1153-7 *(in Chinese)*

Authors

Chao Chen, born in August, 1974, JiangSu, China

Current position, grades: Doctor of Mechanical Engineering, associated professor in Jiangsu University of Science and Technology.
University studies: PhD degree in Mechanical Engineering at Shanghai Jiaotong University of China in 2005.
Scientific interests: robotics.
Publications: 4 patents, 31 papers.

Jian Tang, born in January, 1987, Jiangsu, China

University studies: master's degree in mechatronic engineering at Jiangsu university of science and technology of china in 2013.
Scientific interests: indoor path planning research for seeing eyes robot base on RFID.

Zuguang Jin, born in July, 1988, Beijing, China

Current position, grades: technician and quality inspector at Qingdao Haixi Heavy-duty Machinery CO., LTD, Qingdao, Shandong, China.
University studies: master's degree in Mechanical and Electronic Engineering at Jiangsu University of Science and Technology in 2013.
Scientific interest: autonomous navigation, RFID Recognition, voice interaction in intelligent service robots, wheeled auto-guide robots, industrial automation device process control.
Publications: 2 patents, 5 papers.

A framework to support flexible application collaboration in cloud computing

Meng Xu, Qingzhong Li*, Lizhen Cui

School of Computer Science and Technology, Shandong University, China

Shandong Provincial Key Laboratory of Software Engineering, China

Received 1 October 2014, www.cmnt.lv

Abstract

With the development of Internet, software has more and more collaborative trends. Collaboration application is becoming a new hotspot. Development of collaborative application process is complicated and need many technologies to realize it. It is difficult for an ISV (Independent Software Vendor) or a business department to provide all application modules of a whole collaboration application. Moreover it is difficult to integrate heterogeneous application module developed by different organizations (ISV or business department). According to the problems mentioned above, this paper proposes a PaaS platform framework to build collaboration application dynamically and flexibly. Such PaaS platform can provide the ability of application delivery deployment for organizations and provide the ability of collaboration application customization for end users.

Keywords: cloud computing, PaaS, collaboration application, process

1 Introduction

With the development of Internet, software has more and more collaborative trends. Collaboration application is becoming a new hotspot. Industry-specific functionality software has huge demand for complex collaborative applications. Collaboration is based on existing business organizations, through information exchange between applications and resource sharing, to achieve a unified business objective. For example, in the field of e-commerce, collaboration between different organizations is very obvious. For an e-commerce application collaboration system, it needs to achieve integration of many e-commerce applications, like online shop system, electronic payment systems, logistics systems. Resources and services that belong to different organizations and business execution is cross-organizational collaboration. It needs a lot of the underlying implementation technology to building such a huge complex process information systems. Currently it is lack of support for collaborative applications dynamically build support platform.

PaaS (Platform as a Service) [10] as a cloud computing [1, 2] service model that provides a new kind of software development, delivery, deployment and usage patterns, but also bring new technical support for designing, developing large-scale network collaborative applications. However, most studies on the PaaS focused on the delivery, deployment and on-demand of single application at present. In this paper, we propose a PaaS to support dynamically building collaboration applications and scheduling methods. The platform is notable feature is support dynamically building collaborative applications.

The key business capabilities of platform include:

1) Application delivery management. Platform supports rapid development, delivery and deployment of software applications with standard of platform. Software delivered by organizations will be resolved, reconstruction of cloud services which can be customized and deployed in the application runtime environment for the users by platform.

2) Collaboration applications customizing. Platform provides rapid collaboration applications customization, development capabilities. Users through leasing rather than invest to build collaboration applications. Moreover, collaboration applications can be customized to meet users' business needs.

3) Collaboration application scheduling. Platform can schedule customized collaborative applications.

4) Cross-cloud services called. Platform supports organization register local service to the platform to achieve cross-cloud service call.

By supporting dynamically building collaboration applications, users only need care about their own business logic implementation, without considering the complex collaboration in ecological changes. On the other hand, users can choose on-demand cloud services by dynamically building collaboration platform. Users do not need a separate investment to build its business collaboration applications. On the contrary users can quickly build the characteristics of the enterprise business processes only using a variety of software services provided by others. The platform can effectively use resources, reduce costs, and ultimately improve the enterprise's core competitiveness to meet the user demand

* ***Corresponding author*** e-mail: lqz@sdu.edu.cn

on business collaboration by supporting rapid collaboration applications building.

The following e-commerce example illustrates users to build collaboration application using PaaS. Entire e-commerce collaboration applications are so complex that no one organization can provide complete application. On the other hand, it is difficult for SMEs to have the economic strength to invest in construction of such a huge information system. Various types of e-commerce applications (such as the portal system, online payment systems, online e-commerce logistics system core applications), which are developed based PaaS can be delivered, deployed to the platform. E-commerce customers can browse, leasing, custom e-commerce software services, e-business collaboration applications quickly customize and development.

The rest of the paper is structured as follows: Related work is discussed in Section 2. Then section 3 describes the architecture of PaaS to support collaboration applications dynamically building. The method of dynamically building collaboration applications will be presented in Section 4. And Section 5 will show the customized approach of collaboration applications based PaaS. Finally Section 6 concludes.

2 Related works

PaaS is an Internet-based cloud computing service model that provides a novel software design, development, delivery, deployment and usage patterns. PaaS provides all the necessary resources, including virtual servers, storage, databases, and so on for software developers to build and deploy applications. PaaS hosted on web-based system through the application development platform, providing end to end development environment, or in some cases, all local online development environment [3].

Some PaaS currently exist, and different PaaS are technically differentiated:

1) One class is to provide application development and deployment container, which provide necessary resources to applications. A typical representative is AWS Elastic Beanstalk [7], on which developers can quickly deploy Java applications to Amazon's machine.

2) Another class is to provide an open API for third-party. A typical representative is Forec.com [5], which provides the necessary development tools and application services, so that software developers can use it to quickly build scalable applications and do not need to purchase, install, configure and manage hardware and software.

3) The last class is to provide the ability to generate the application's configuration. A typical representative is LongJump [9].

However, these PaaS less concerned about the collaboration of software services on the platform at present. Literature [11] presents a Service Delivery Platform for the telecommunications industry by using PaaS technology. It achieves develop value-added telecommunications services through the platform services and registered on the platform assembly of external services. Literature [12] studied how to improve information integration framework petrochemical execution efficiency through the cloud computing service model.

3 Architecture

The different software services are assembled together to form collaboration applications, in fact, is the process of software reuse. Researches show that software reuse in a specific domain is easier to achieve [13, 14]. With the same business objectives software delivered by different organizations typically form collaboration applications in specific domain. "Domain" refers to areas covered by a group application with similar functions [4]. For example, e-commerce or tax applications will be reflected in the relevant software applications and features in common. PaaS proposed in this paper can dynamically build collaboration applications in different fields. The application delivered to PaaS should be specified an area of software applications, so platform can dynamically build collaboration applications in specific areas of software services. During collaboration application development, the characteristics of the domain can be used, such as a common system architecture, which enhances the efficiency of collaboration application development.

PaaS which support dynamically building collaboration applications needs to provide application delivery, collaboration applications rapidly customize, instant deployment, external integration and other key operational capacity. By using the web-based development tools, dynamically build collaboration mechanism provided by the platform, in accordance with their specific business collaboration needs, users can quickly customize collaboration application system including personalized interface, business processes and data models. Moreover users do not need to focus on application deployment details and can ensure that applications can run stable.

The feature of PaaS proposed in this paper is the ability of dynamic building the collaboration of applications which is provided by different organizations within a specific domain and achieve personalized customization needs of users. Figure 1 shows the overall architecture of PaaS, including application delivery module, collaboration as a service, application customization module, resource management module, application runtime platforms, database as a service and PaaS operations management system. The following describes the role of each module.

3.1 APPLICATION DELIVERY MODULE

Application delivery module provides a Web-based application packaging, analytical, deployment, maintenance function and the necessary support of delivering the software for the ISV. It can achieve applications developed by ISV from the traditional

development environment to the platform. ISV can encapsulate the business logic, user interface, data model, and collaboration constraint in a compressed package for submission to the platform. The software developed by ISV can be expressed as App = <UI,BL,D,C>, where UI represents the application's user interface, including the interface of the function menu, the form layout, the optional background and Logo; BL is the core business logic of the application, which can be further divided into modules that can be customized by the user; D represents the data model of the application, including data objects and various data items of data objects; C represents the application collaboration constraints. These constraints can express which types of applications can be collaboration.

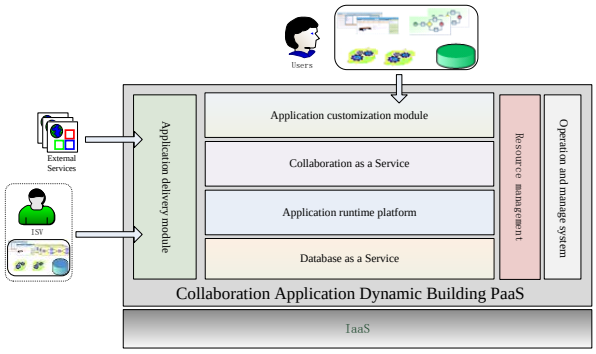


FIGURE 1 Overall architecture

PaaS will be responsible for the operation and maintenance work. Application Delivery module is responsible for parsing applications developed in accordance with the development standard and breaks it into the user interface, business logic, data models, and collaboration constraints. Then a service is built to enable users customizing collaboration applications. On the addition, PaaS can also accept registering local service on PaaS to enable cross-cloud service call capabilities. Use registration, organization needs maintenance its own software and infrastructure.

3.2 APPLICATION CUSTOMIZATION MODULE

Users can customize a single application on platform, including user interface, business logic, and data model and other aspects. Collaboration application customization module is shown in Figure 2. It includes UI customization module, collaboration application customization module, data customization module, cross-cloud service invocation interface module, custom verify module and custom isolation module. UI customization module supports the customization of menu, form layout, UI themes, interface colours, LOGO and other presentation layer.

Collaboration application customization module can implement customization on individual service and collaboration applications. Since collaboration constraints of each application are stored, platform will search, match, association, and recommend some abstract collaboration applications for users according to collaboration constraints.

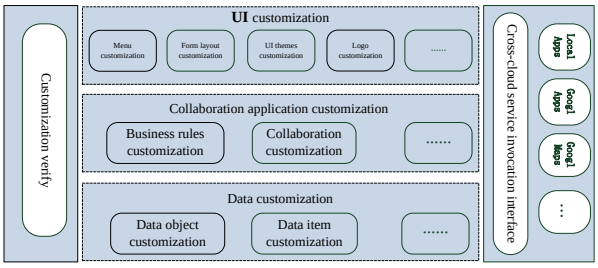


FIGURE 2 Application customization modules

Data customization module is responsible for the customization of data model, data object, and data item. It provides the ability to add, delete, edit and other customized functions of data objects and data items for tenants. Cross-cloud service call interface module is responsible for integrating a variety of cross-cloud applications. Custom verify module is responsible for verifying the user's customization results to ensure that the results did not violate the customization constraints.

3.3 COLLABORATION AS A SERVICE

The software delivered to PaaS is parsed and rebuilt as a standard cloud services by the application delivery module. There may be collaboration relationship between these services from different organizations. Collaboration as a service can discover the collaboration of services in services pool according to the collaboration constraints. Then it will link these services to form an abstract collaboration application. The abstract collaboration application refers each service node of abstract collaborative applications represents a class of services which have similar functions. For any service node, there may be many services from different organizations which can be chosen. The Collaboration as a Service is shown in Figure 3.

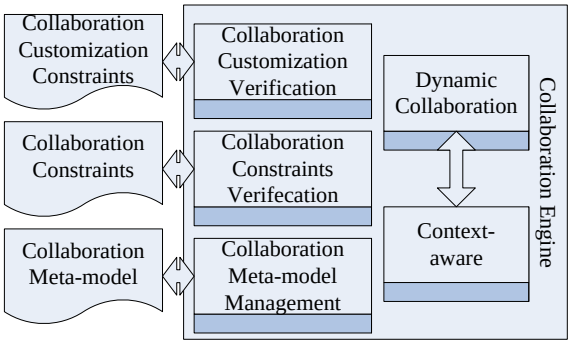


FIGURE 3 Collaboration as a service

3.4 APPLICATION RUNTIME PLATFORMS

Application runtime platform is the runtime environment, including UI engine, data engine and business process engine. These engines can interactive with application deliver module, application customization module, PaaS

operations management system. The collaboration applications generated by collaboration application customization module are deployed to the application running platform to provide services for users.

UI engine is responsible for UI presentation. It can rebuild, adjusted and displayed UI interface on the terminal. Data engine is used to parse data model of applications. Business process engine can parse and instantiate process of customized applications. It is responsible for creating and scheduling process instances, and process monitoring and exception handling. In addition, the application runtime platform can also call a Web service across cloud.

3.5 DATABASE AS A SERVICE

Database as a Service is responsible for managing data in the platform which include application data, meta-data of application, collaboration data and other data. Database as a Service can provide a uniform data model to all kinds of application which makes the manipulation of data more efficiently. On the other hand, applications can access their own data through different API. It is used as if only one was using the service. For example, application A can use MySQL database, while application B may use MongoDB.

3.6 RESOURCE MANAGEMENT MODULE

Resource management module can manage services according to its domain. The services which belonging to the same domain will be unified management, which will be beneficial for collaboration as a service dynamically building collaboration applications according to different domains. Services catalogue is also a part of it. The description of software delivered or registered to platform, including functional and non-functional information, will be in services catalogue. Services catalogue is available for user to find services satisfied his requirements. Besides resource management module is also responsible for data management and maintenance.

3.7 PAAS OPERATIONS MANAGEMENT SYSTEM

PaaS operations management system is used to manage the registration of tenants, as well as tenant directory structure, indexes, service access management. It is also used to coordinate between PaaS and tenants about QoS assurance. SLA requirements for different platforms QoS metrics to quantify dynamically adjusted based on quantitative indicators of shared resources, service monitoring, when the real understanding of custom applications running conditions and usage. Ensure that the platform's reliability, stability and robust to ensure tenants to meet SLA requirements. SLA-based resource dynamic adjustment strategy will meet two objectives: first, the SLA requirements for high-priority users will get the quality of service assurance; two SLA requirements is to ensure a high quality of service under the premise of the user, try to take care of low SLA requirements of users quality of service.

4 Building flexible application collaboration

Dynamically building collaboration applications rely on semantic rules. The rules of semantic description will be given in this section. Applications delivered by different organizations should follow these rules. When an application is delivered to PaaS, the application will be classified according to semantic information, and dynamically discover, associate, and build abstract collaboration applications according to the collaboration rules of the applications. The main roles of the semantic model are: first, it is the application development standards; second, it makes dynamically build collaboration applications possible.

4.1 BUILDING ABSTRACT COLLABORATION

An industry collaboration application will have its regular logic process models, such as e-commerce, collaboration applications in this area will involve electronic transactions, online payment, online logistics systems, etc. And the entire collaboration model is basically fixed. Differences between collaboration applications are also in framework of collaboration models. So an analysis can be made by domain experts to extract the business collaboration process to obtain collaboration application model for a specific domain. In addition, as more and more applications are delivered to platform, the potential collaboration between applications will be auto-discovery, association to form a new collaboration application model, or add new applications to the existing collaboration model. Users can customize the collaboration model according to their own needs.

Definition 1: Each service node of collaboration model represents one class of services but not a specific service instance in which all services have similar functions. Each service class contains many service instances which can achieve this functionality. The collaboration model is defined as BP = <SC,E>, where:

1) SC = {sc1, sc2, ..., scn} is set of the service classes, sc1 .... scn are the n classes of services that participate in abstract collaboration applications.

2) $E \subseteq SC \times SC$ is the set of connections between the service classes, in which $X \in E$ is the structure control operator. A number of basic services can be combined into a service workflow through some control structures. The control structure operators used by workflow are sequence, selection, cycle, and parallel.

Definition 2: The service class is a class services with same or similar function. Any service in the class which can meet the performance requirements of a specific application can be candidate service as a service node of abstract collaboration application.

Definition 3: Collaboration constraints C is defined as C = <Sd,F,Q,Dom>, where Sd represents the basic definition

and description of the application, including the service ID, a basic description of the application name. F represents the functional description of the application, including application interface parameters, preconditions and postconditions; Q is the non-functional attributes descriptions of the service, including time, cost, reputation, etc. Dom represents an application domain.

Users can cut, configure and build abstract collaboration applications which meet their own needs based on collaboration model. After building abstract collaboration applications, the next step is to bind service. The candidate services for each service node will be sorted according to their function by platform. Users can choose service to meet their non-functional requirements. Non-functional requirements include prices, response time, etc.

4.2 CUSTOMIZING COLLABORATION

Users can customize the application from the user interface, business logic, business process and data models four levels. This article focuses on the process logic customization of the collaboration application, i.e. users based on the abstract collaboration applications, according to their needs customize personalized collaboration applications. Platform provides the following custom methods:

Users select one collaboration application model, delete unnecessary path and change the service relationship based on this model. For example, for e-commerce collaborative applications, the user can adjust according to demand whether pay after shipping or shipping after pay. Eventually the abstract collaboration application is formed. Each service node of abstract collaboration application corresponds to many service instance, user can choose one service instance which meet his needs. For example, although many organizations have developed electronic payment applications, but there are some differences, user can choose on-demand according to their actual business conditions.

When there are a lot of collaboration models on platform, it is very difficult that users make choice manually. So a recommendation algorithm is proposed to quickly select the collaboration model.

Definition 4: Let ReqS is the services set of collaboration applications, Cm is a domain collaboration model, CmS is services included in the Cm , if ∀si∈ReqS, then si∈CmS, called ReqS match with Cms.

Definition 5: If ReqS match with CmS, and |ReqS| = |CmS|, called ReqS exactly match with CmS.

From the above definition, as long as ReqS match with CmS, the collaboration model will contain all functional requirements of user. But there may be some unwanted features. So the best case is the user needs and collaboration model exactly matches. If there is no exactly match, then find out the collaboration model with minimum difference.

Definition 6: Let C(ReqS, CmS) = $\frac{|\{si\,|\,(si \in \mathrm{Re}\,qS) \wedge (si \in CmS)\}|}{|\,\mathrm{Re}\,qS\,|}$ as the degree of collaboration model CmS cover user needs ReqS.

Definition 7: Let R(ReqS, CmS) = $\frac{|\,CmS - \{si\,|\,(si \in \mathrm{Re}\,qS) \wedge (si \in CmS)\}|}{|\,\mathrm{Re}\,qS\,|}$ as the degree of collaboration model CmS redundant of user needs ReqS.

Algorithm 1 recommendation algorithm:

```
Input: ReqS, userCov, userRed
Output: M

Ms= Ø //set candidate model null
for each m in M
  cov= C(ReqS, m)
  red =R(ReqS, m)
  If(cov>=userCov∧red<=userRed)
  M.add(m);
end for
Return M;
```

5 Scheduling Collaboration and Simulation

5.1 SCHEDULING COLLABORATION

Collaboration applications scheduling has two layers. The first layer maps the abstract collaboration application to the executable collaboration application. The second layer schedules the task to the appropriate computing nodes according to load balance.

Users customize the personalized collaboration applications based on the abstract collaboration applications. Then customized results are submitted, and will be scheduled by platform. The collaboration application scheduling model is shown in Figure 4.

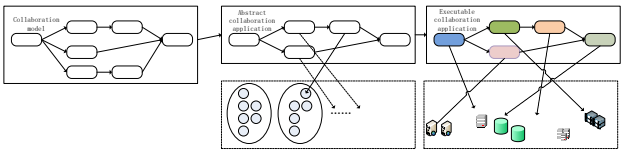


FIGURE 4 Collaboration application scheduling model

The first layer scheduling algorithm:

1) Process engine first reads the user's personalized needs, then the abstract collaboration is cut into customized collaboration application, i.e. removing the service node which user does not need.

2) The abstract customized collaboration applications bind services. The binding process is to find the corresponding services set (service instances set) based on the type of service. Then the services which are not satisfied QoS are deleted according to the user's QoS constraints.

3) After the above calculation, the service nodes of abstract collaboration application are mapped to specific service instance.

4) After executable collaboration applications generated, services is called and platform will monitor the execution of collaboration applications.

Taking into account the load balance, a service will not only deploy on one server, but have several copies on multiple servers. Then an executable collaboration application is scheduled, the appropriate server should be selected to make server load balancing. PaaS will receive a lot of request to execute applications the in a period of time, so multiple collaboration applications jointly scheduling should be considered. The strategy proposed in this paper is: first all executable services (execution condition is satisfied) is inserted into an execution queue according to the arrival time. Then try to do the following steps: a) remove the first task; b) schedule this task to a most appropriate server; c) If there is no appropriate server, this task is inserted into the tail; d) When the task is completed, notify scheduler task execution is completed.

Definition 8: The server load function is defined as $L(s) = \frac{C_{\max}(s) - C_{cur}(s)}{C_{\max}(s)}$, where $C_{\max}(s)$ is max load of server, i.e. the maximum execution server capacity, $C_{cur}(s)$ is the server's current load.

If the current server is idle, that is $C_{cur}(s) = 0$, then $L(s) = 1$. If the current server load reaches the maximum value, that is $C_{cur}(s) = C_{\max}(s)$, then $L(s) = 0$. So the second layer of collaboration applications scheduling algorithm is as follows:

1) For an executable collaboration application, all executable tasks will be inserted into a queue.

2) Take out each task, assign it to execution node. Execution node selection is based on its load function value. The value of the minimum execution node is selected. If all nodes load are full, i.e. L (s) <= 0, then this task is inserted into the tail.

3) Repeat steps 1 and 2 until the queue is empty.

5.2 SIMULATION

To verify the efficiency and scalability of the framework, an experiment is conducted. Five different test sets are used in the experiment. The collaboration time and result are recorded and shown in Table 1. If the collaboration which is generated dynamically by the framework in this paper can be executed and in accordance with the requirements, we call the collaboration is correct. And the accuracy in Table 1 is the correct rate of the collaborations. From Table 1, we can see that the execute time increases with the number of services linearly and the accuracy is always 100%.So we can conclude that the framework propose in this paper is effective and scalable.

TABLE 1 Collaboration time and accuracy

Test Set	Number of services	Time(ms)	Accuracy
Test Set 1	800	35	100%
Test Set 2	1000	52	100%
Test Set 3	400	116	100%
Test Set 4	1500	528	100%
Test Set 5	4000	1860	100%

6 Conclusion

Based on IT development and the demand of business collaboration, a PaaS for business collaboration is proposed that supports business collaboration applications quickly build and deployment. First, the platform provides a new application delivery model. The application can be parsed and rebuilt the services for the user customization. Secondly, platform provides automatic matching technology. The collaboration relationship between services can be matched and automatically generates business collaboration model. It simplifies the process of collaboration application customization. Moreover, the platform provides collaboration applications for deploying, runtime environment, including the UI engine, data engine and workflow engine.

Acknowledgments

This work has been supported by the National Key Technologies R&D Program under Grant No. 2012BAH54F01; the National Natural Science Foundation of China under Grant No. 61003253, No. 61272241, No. 61303085; the Natural Science Foundation of Shandong Province of China under Grant No. ZR2010FM031, ZR2010FQ010; Science and Technology Development Plan Project of Shandong Province under Grant No. 2012GGX10134; Independent Innovation Foundation of Shandong University under Grant No. 2012TS075, No. 2012TS074; Shandong Province Independent Innovation Major Special Project under Grant No.2013CXC30201.

References

[1] Buyya T, Yeo C S, Venugopal S, Broberg J, Brandic I 2009 Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility *Future Generation Computer Systems* Elsevier Science Amsterdam **25**(6) 599-616

[2] Armbrust M 2009 Above the Clouds: A Berkeley View of Cloud Computing *EECS Department University of California* Berkeley

[3] Lawton G 2008 Developing Software Online with Platform-as-a-Service Technology *Computer* **41**(6) 13–5

[4] Zhao J-F, Xie B, Zhang L, Yang F-Q 2005 A Web Services Composition Method Supporting Domain Feature *Chinese Journal of Computers* **28**(4) 732-8

[5] Salesforce *Force.com: The leading cloud platform for business apps* http://www.salesforce.com/platform/

[6] AppEngine http://code.google.com/intl/EN/appengine/

[7] Elasticbeanstalk http://aws.amazon.com/elasticbeanstalk/

[8] Anstett T, Leymann F, Mietzner R, Strauch S 2009 Towards BPEL in the Cloud: exploiting different delivery models for the execution of business processes *2009 World Conference on Services – I* 670-7

[9] Longjump, http://longjump.com/

[10] Zhou Y C, Liu X P, Wang X N, Xue L 2010 Business Process Centric Platform-as-a-Service Model and Technologies for Cloud Enabled Industry Solutions *2010 IEEE 3rd International Conference on*

Cloud Computing (CLOUD) 534-7
[11] Zhou Y C, Xue L, Liu X P, Wang X N, Liang X X, Sun C H 2010 Service Storm: A Self-Service Telecommunication Service Delivery Platform with Platform-as-a-Service Technology *2010 IEEE 6th World Congress on Services (Services – I)* 8-15
[12] Zhou Y C, Wang X N, Liu X P, Xue L, Liang S, Sun C H 2010 Enabled Integrated Information frameworks for chemical and petroleum industry *2010 IEEE 6th World Congress on Services (Services – I)* 1-7
[13] Prieto-Diaz R 1987 Domain analysis for reusability *Proceedings of COMPSAC'87* Tokyo Japan 23-9
[14] Mili H, Mili F, Mili A 1995 *IEEE Transactions on Software Engineering* **21**(6) 528-62

Authors

Meng Xu, born in October, 1978

Current position, grades: PhD Student, Shandong University, Jinan, Shandong, China.
University studies: master's degrees at Shandong University, 2010.
Scientific interests: cloud computing, service computing.
Publications: 10 papers.

Qingzhong Li, born in 1965

Current position, grades: PhD, professor and doctoral supervisor at Shandong University, Jinan, Shandong, China.
University studies: PhD degree at Institute of Computing Technology of Chinese Academy of Sciences, 2000.
Scientific interests: data management, cloud computing, data integration.
Publications: 60 Papers.

Lizhen Cui, born in 1976

Current position, grades: PhD, associate professor and master supervisor at Shandong University, Jinan, Shandong, China.
University studies: PhD degrees at Shandong University, 2006.
Scientific interests: cloud and service computing.
Publications: 30 papers.

Synchronization optimization under symmetry network

Shaohua Tao*, Meilian Li, Zhili Zhang

School of Information Engineering, Xuchang University, Xuchang City, 461000, PR. China

Received 1 October 2014, www.cmnt.lv

Abstract

Rich symmetries have been found in many real networks, which is an extensive structural of networks. We study the relationships between synchronization and symmetry of network. One of fascinating problems related to symmetry network is how to optimization phase synchronization by employing symmetric structure. For this purpose, we optimize the network structure by symmetry properties of network, which is statistics the orbits of network. Nodes in the same orbit have similar properties, and then we reduce the network size by nodes in same orbit. We simulate the BA and SW model. The simulations result show that symmetry structure can optimize the network topology and can enhance phase synchronization of network.

Keywords: synchronization, symmetry network, orbit

1 Introduction

The topologies of network have been extensively studied and some common architecture has been discovered [1-3]. The small-world property [4-6], the scale-free [7], communication structure [7-9], network motif [10], modular network [11] any symmetric network [12, 13]. From this viewpoint, systematically understanding the network structural effects on their dynamical processes is of both theoretical and practical importance. Synchronization behaviour, in particular, as a widely observed phenomenon in network systems, has received a great deal of attention in the past few decades [14, 15]. The effects of average distance [16, 17], heterogeneity [14, 18], clustering [19, 20] and weight distribution on network synchronizability have been extensively investigated [21, 22]. Since the synchronizability is correlated with many topological properties, but these properties affected the synchronization maybe is not a proper answer. Perhaps a good index, which can characterize the synchronizability, has not been found. In recent years, graph-theoretic were used to analyse the network synchronizability, e.g., degree sequences were discussed [23], complementary graph and subgraph techniques were used [24-26]. But using the automorphism of symmetry graph to study the synchronization has not been fully investigated. Motivated by the above of work, this paper focuses on the relationships between the network synchronization and the orbit of automorphism of symmetry network. We know that it is difficulty all nodes achieves the synchronization in whole network, the finding shows that the network size is smaller and achieving the synchronization is easier. From this, we know that we may optimize the network topology to enhance the network synchronizing through symmetric structure. In symmetry network nodes in same orbit are similar and we view the nodes in same orbit as one node and reduce the network size. We found that the nodes in same orbit are also very easy to achieve the synchronization. The results show that the optimized networks enhance the network synchronism.

2 Symmetry network and synchronization

2.1 SYMMETRY NETWORK

A graph or network is denoted by $H=(V,E)$, where V is the set of vertices and E is the set of edges. To any node $v\in V(G)$, $\varepsilon(v)=v$, obviously, there has a isomorphism to identity map $\varepsilon:V(H)\to V(H)$ [28]. We illustrate the concept through Figure 1. There has a map $\alpha_1:V(H)\to V(H)$ defined by $\alpha_1(v)=v$, which is a isomorphism from H to itself [27-29].

An automorphisms of a graph H is a isomorphism from H to itself. We denote the set of automorphisms of graph H under composition operation as $Aut(H)$. Figure 1 shows $Aut(H)=\{\alpha_1,\alpha_2,\alpha_3,\alpha_4\}$. The automorphisms group also is a permutation group. The set of all of permutation group base on n called symmetric group S_n. The network is considered as symmetry if its underlying graph contains not only an identity permutation.

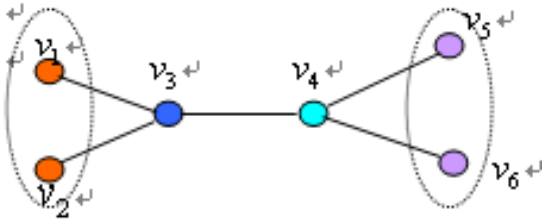


FIGURE 1 Symmetry network

*Corresponding author e-mail: shtao_2008@126.com

Next, we will introduce the concepts of automorphism partition and orbit. Given the automorphism group acting on vertex set V, we can get a partition $\pi = \{V_0, V_1, ..., V_k\}$, $x R y$ if and only if $\exists \alpha \in Aut(H)$, s.t. $\alpha(x) = y$. Such partition is called automorphism partition, each cell of the partition is called an orbit of $Aut(H)$, denote as $orbit(H)$. An orbit is trivial if it only contains a single vertex, otherwise, the orbit is non-trivial. The vertex u, v called similar if they are in same orbit. Figure 1 shows the automorphism partition $\pi = \{\{v_1, v_2\}, \{v_3\}, \{v_4\}, \{v_5, v_6\}\}$ and the orbits. Nodes in same orbit are marked with the same colour, shown as Figure 1. The nodes in same orbit have similar structure information [28]. We view them as one node, e.g.:

$$\{v_1, v_2\} \to v_1, \{v_3\} \to v_2, \{v_4\} \to v_3, \{v_5, v_6\} \to v_4 .$$

We reduce the network structure, shown as Figure 2. We call this reduction network as optimization network in this paper.

FIGURE 2 Optimization network.

2.2 SYNCHRONIZATION

To settled the natural frequencies of the oscillators, Kuramoto worked out a mathematically model to tackle this problem and each of the oscillators in the system are [30-31]:

$$\dot{\theta}_i = \omega_i + \frac{K}{N}\sum_{j=1}^{N} \sin\left(\theta_j - \theta_i\right), \ (i = 1, ..., N), \tag{1}$$

where N is the factor and $N \to \infty$, ω_i is the frequency of oscillator i, K is the coupling constant.

To deal with the synchronization on complex networks, the Kuramoto model can be reformulated to Equation (2): The synchronization of Kuramoto model [30, 31]:

$$\frac{d\phi_i}{dt} = \omega_i - \frac{\delta}{N}\sum_j a_{ij} \sin\left(\phi_i - \phi_j\right), \tag{2}$$

where ϕ_i and ω_i are the phase and the intrinsic frequency of vertex i, respectively. ω_i is chosen from the Gaussian distribution with unit variance. The summation runs over j, the nearest neighbours of vertex i. The coupling strength is used as δ. To study synchronization, each node in complex topologies is considered an oscillator, such as WS networks and BA graphs.

The order parameter M is defined as:

$$M = \left\langle \left| \frac{1}{N}\sum_{j=1}^{N} e^{i\phi j} \right| \right\rangle, \tag{3}$$

where $\langle ... \rangle$ is the ensemble average over different configurations and N is the total number of vertices. M is $0(1)$ in the fully incoherent (coherent) phase.

The dynamics of synchronization on symmetry networks through a modified Kuramoto model, where $orbit[i]$ is the vertex i in the same orbit.

$$\frac{d\phi_i}{dt} = \omega_i - \frac{\delta}{Orbit[i]}\sum_j a_{ij} \sin\left(\phi_i - \phi_j\right), \tag{4}$$

3 Scale-free network

To scale-free network our numerical simulations are based on the SF model. The model starts from m_0 nodes, at every time step adds a new node with $m(\le m_0)$ edges that link the odd node already existing in network. We assume that the probability connecting to an existing node i is proportional to $i's$ degree k_i. The algorithm of the BA model is [35]:

1) Growth: starting with a small number m_0 of nodes, each step adds a new node with $m(\le m_0)$ edges connect new node with different old nodes in network.

2) Preferential growth: when choosing old nodes which the new node connects to, and assuming that the probability depends on the degree k_i of node i. The link probability is:

$$\prod = \frac{k_i}{\sum_j k_j} . \tag{5}$$

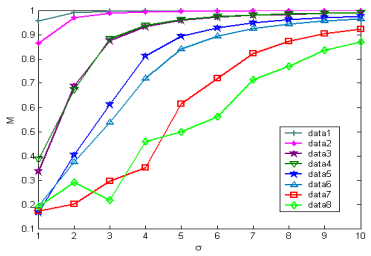
By using the rate equation, we can easily obtain the degree distribution of the whole network, $p(k) \sim k^{-2.8}$. In simulation, there are totally 200 nodes in the network. According to the new node m which added to the network, we assumed $m = 2, m = 3, m = 4, m = 5$, respectively. When $m = 2$, the $Aut(m_2) \approx 6.8 \times 10^6$, so the network is richly symmetric [10]. We only simulated the synchronization of non-trivial, because the nodes in non-trivial at lease are 2. In the 11th and 17th orbit, the node number is 14, respectively. In the 16th orbit, the node number is 17. In the 20th orbit, the node number is 8. In the 25th orbit, the node number is 9. In the 37th and 38th orbit, the node number is 3, respectively. In other orbit, the node numbers also are 2. To get the synchronization property of network, we investigate how the order parameters in different orbits, a network and quotient network change with the coupling strength. According to the Equation (2), Figure 3a shows the simulation results. From top to bottom, the node numbers in different orbits 2, 3, 8, 9, 14, and 17, corresponding to more and more node numbers in same orbit. In the case $orbit[i] = 2$, with the increasing of coupling strength, the order parameter of

$orbit[i]$ soon reaches 1, indicating the synchronized state within one orbit. With the further observed, an interesting phenomenon appears: for different $orbit[i] = 2,3,8,9,14,17$, order parameter for different $orbit[i]$ all reach 1 soon, which indicated oscillators belong to the different $orbit[i]$ are synchronized. We call this phenomenon orbit synchronization clustering.

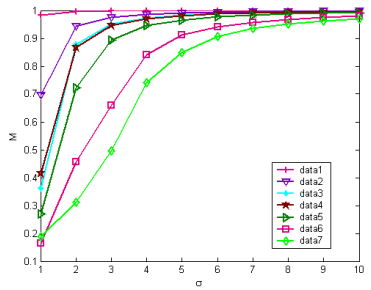
However, the phases of oscillators in different $orbit[i]$ are different, in the case $orbit[i] = 17$, the order parameter of $orbit[i]$ reaches 1 slower than $orbit[i] = 2$, which indicating more node within one orbit, slower reached synchronized state.

Similarly, to enhance the synchronization of the whole network, we defined the node in the same orbit as one node, like this may reduce the node number in the network greatly and enhances the network synchronism. Synchronization of the optimized network and original network shown as Figure 3a, the red square represents optimized network, indicated as $data8$, the green diamond represents the whole network, indicated as $data9$. In Figure 3a, obviously, the order parameter for the whole network is the lowest one. Optimization network reach 1 sooner than the whole network. That is, optimized network's topology may enhance the network synchronized performance.

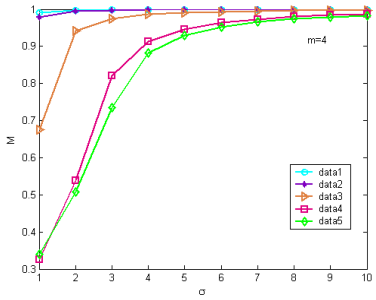
At the same time, we simulated the synchronism of $orbit[i]$, the optimized network and the whole network with $m = 3, m = 4, m = 5$, shown in Figures 3b-d. The simulation show that different m values form different network structure, but each network $orbit[i]$ may achieve the synchronization and forms the synchronized clustering. The synchronization of optimization network with different m values surpasses the original network.



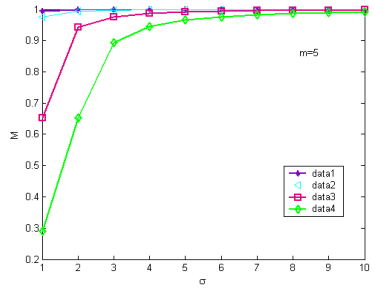
a) m=2



b) m=3



c) m=4



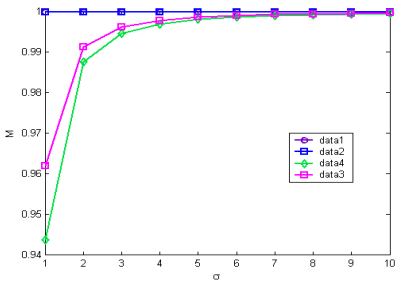
d) m=5

FIGURE 3 Synchronization of BA network

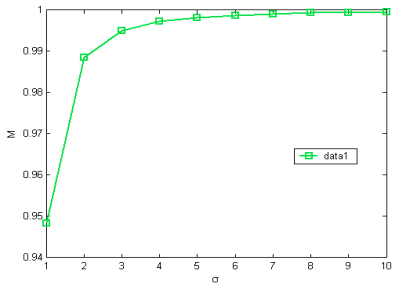
4 Small world network

The WS model was proposed by Watts and Strogatz. The WS model start from regular loop network with n nodes. Assuming that each node on loop links K edges with two sides nodes on it, then restart a new link with the probability p to each edge. This new network has small world character, which has smaller average path and bigger cluster. The degree distribution of small world obey the Possion distribution.

We made the synchronization simulation of small world network similar to BA network. Network has 200 nodes in simulation same as BA network size. When network connection probability smaller than $p = 0.49$, network has not non-trivial orbit, only trivial orbit. Hence, assuming that WS network connection probability is $p = 0.49$ and $p = 0.5$ separately. When the connection probability is bigger than $p = 0.5$, all of nodes in the same orbit. When $p = 0.49$, $Aut(H)$ is very small and has a few non-trivial orbit. We simulated the synchronization of non-trivial orbit and the optimization network when $p = 0.49$, shown as Figure 4a. When $p \geq 0.5$, $Aut(H) \approx 7 \times 10^5$ the network has richly symmetry and all of the node in same orbit. Although the topology of network has not optimized, the whole network reached 1 very soon, that is the network has better synchronization, shown as Figure 4b.



a)



b)

FIGURE 4 Synchronization of SW network

We found that the synchronization of SW reaches 1 sooner than BA model with the same scale. In BA, the network symmetry is richer, different non-trivial orbit are more and enhance the network synchronism. In SW, The network symmetry is richly, the node in the same orbit and the network is very completely easy to achieve the synchronization.

5 Conclusion

The paper captured the synchronization of symmetric network and introduced the concept of symmetry network. The more automorphism of a network have and the more symmetry of a network and the orbit constitutes the partition of automorphism. On the other hand, the nodes in orbit have similar properties, hence, we viewed the nodes in orbit as a node to optimize the network topology, and we called this network as optimization network. We simulated the synchronization of orbit clustering, the whole network and the optimization network using revised Kuramoto model. We discussed the relationship of synchronization and symmetry of BA network. In BA model, the different m values and the network topology are also different. According the different network topology formed by different m values, we simulated the synchronization of orbit clustering, the whole network and the optimization network of BA model. The results show that the optimized network topology can enhance the synchronization of network. Finally, we discussed the SW model, the synchronization of network is different according to different linking probability. All node in the orbit when $p \geq 0.5$, the network has the better synchronization. In BA model, the nodes in different orbit can reduce the network scale to optimize the network topology, and the optimization network has the better synchronization than the original network. To SW model, the nodes of the network in the same orbit when $p \geq 0.5$, and it can't optimize the network structure. But the results show that the SW model is more easy reach the synchronization than BA model to the same network scale if BA model is not optimization.

Acknowledgments

This work was supported by the Foundation Name under Grant No. (122102210544 and No.12A510022 and natural foundation fund (No. U1304403).

References

[1] Li H, Zhong L, Yao H 2013 Detecting Protein Complexes through Micro-Network Comparison in Protein-Protein Interaction Networks *Journal of Networks* **8**(3) 696-703
[2] Xu K, Cui W, Tie J, Zhang X 2012 An Algorithm for Detecting Group in Mobile Social Network *Journal of Networks* **7**(10) 1584-91
[3] Huang Z, Li W, Feng Z, Xing H 2013 On Nonnegative Signed Domination in Graphs and its Algorithmic Complexity *Journal of Networks* **8**(3) 365-72
[4] Amaral L A N, Scala A, Barthélémy M, Stanley H E 2000 Classes of small-world networks *Proceedings of the National Academy of Sciences* USA **97**(21) 11149-52.
[5] Kleinberg J M 2000 Navigation in a small world *Nature* **406** 845
[6] Watts D J, Strogatz S H 2000 Collective dynamics of small-world networks *Nature* **393** 440-2
[7] Barabási A L, Albert R 1999 Emergence of scaling in random networks *Science* 286(5439) 509-12
[8] Newman M E J, Girvan M 2003 Statistical Mechanics of Complex Networks *Springer* Berlin
[9] Newman M E J Mixing patterns in networks *Physical Review E* **67** 026126.
[10] Newman M E J 2004 Detecting community structure in networks *The European Physical Journal B – Condensed Matter and Complex Systems* **38**(2) 321-30
[11] Oh E, Rho K, Hong H, Kahng B 2008 Modular synchronization in complex networks *Complex Sciences Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* **4** 429-34
[12] MacArthur B D, Sánchez-García R J, Anderson J W 2008 Symmetry in complex networks *Discrete Applied Mathematics* **156**(18) 3525-31
[13] Xiao Y, Xiong M, Wang W, Wang H 2008 Emergence of Symmetry in complex networks *Physical Review E* **77** 066108.
[14] Ji P, Peron T K DM, Menck P J, Rodriguez F A, Kurths J 2013 Cluster Explosive Synchronization in Complex Networks *Physical Review Letters* **110** 218701.
[15] Nicosia V, Valencia M, Chavez M, Díaz-Guilera A, Latora V 2013 Remote Synchronization Reveals Network Symmetries and Functional Models *Physical Review Letters* **110** 174102
[16] Zhou T, Zhao M, Chen G, Yan G, Wang B H 2007 Phase Synchronization on scale-free networks with community structure *Physics letters A* **368**(6) 431-4
[17] Hasegawa H 2004 Dynamical mean-filed approximation to small-world networks of spiking neurons: From local to global, and/or from regular to random couplings *Physical Review E* **70** 066107
[18] Zhao M, Zhou T, Yang H J, Yan G, Wang B H 2009 Synchronization on complex networks with different sorts of communities *Complex*

Sciences Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering **4** 924-33
[19] McGraw P N, Menzinger M 2005 Clustering and the synchronization of oscillator networks *Physical Review E* **72** 015101.
[20] Zhao M, ZhaoT, Wang B H, Yan G 2006 Relations between average distance, heterogeneity and network synchronizability Physica A **371** 773-80
[21] Chavez M, Hwang D U, Amann A, Boccaletti S 2006 Synchronizing weighted complex networks *Chaos* **16**(1) 015106
[22] Arenas A, Diaz-Guilera A, Perez-Vicente C J 2006 Synchronization processes in complex networks *Physica D* **224**(1-2) 27-34
[23] Do A L, Boccaletti S, Gross T 2012 Graphical Notation Reveals Topological Stability Criteria for Collective Dynamics in Complex Networks *Physical Review Letters* **108** 194102
[24] Liu C, Duan Z, Chen G, Huan L 2007 Analyzing and controlling the network synchronization regions Physica A **386**(1) 531-42
[25] Duan Z-S, Wang W-X, Liu C, Chen G-R 2009Are networks with more edges easier to synchronize, or not? *Chinese Physics B* **18**(8) 3122-30
[26] Godsil C, Royle G F 2001 Algebraic Graph Theory *Graduate Texts in Mathematics* Springer **207**
[27] Biggs N 1974 Algebraic graph theory *Cambridge University Press* London
[28] Beineke L W, Wilson R J, Cameron P J 2004 Topics in algebraic graph theory *Encyclopedia of Mathematics and its Applications* Cambridge University Press Cambridge **102** 250-66
[29] Kuramoto Y 1984 Chemical Oscillations, Wave and Turbulence *Springer-Verlag* Berlin
[30] Pikovsky A 2001 Synchronization *Cambridge University Press* Cambridge
[31] Acebrón J A, Bonilla L L, Vicente C J P, Ritort F, Spigler R 2005 The Kuramoto model: A simple paradigm for synchronization phenomena Reviews of Modern Physics **77** 137

Authors

Shaohua Tao, born in March, 1978, Shangqiu, China

Current position, grades: lecturer at Xuchang University.
University studies: Computer Technologies at Normal Central University.
Scientific interests: data analysis, complex network, communication network, and online network.
Publications: 15 papers.

Meilian Li, born in March, 1974, Anyang, China

Current position, grades: assist professor at Xuchang University
University studies: Computer Technologies.
Scientific interests: computer control, network communication.
Publications: 10 papers.

Zhili Zhang, born in October, 1964, Xuchang, China

Current position, grades: professor at Xuchang University.
University studies: PhD degree on Computer Communication at Institutes of Technology of South China in 2003.
Scientific interests: computer communication.
Publications: 35 papers.

Solution of compressed sensing for wide angle EM scattering analysis based on MFIE

Xinyuan Cao*, Mingsheng Chen, Bingbing Chen, Liangliang Cheng, Qi Qi

School of Electronics and Information Engineering, Hefei Normal University, Hefei, Anhui, China

Received 1 June 2014, www.cmnt.lv

Abstract

Fast analysis of electromagnetic scattering problems over a wide incident angle is always a difficult problem in computational electromagnetics. Up to the present, almost all of the traditional numerical methods need to solve one discrete angle after another to finish calculating this kind of problem. In this paper, we propose a new method, which can fix it effectively by applying compressed sensing into method of moments for magnetic field integral equation. The theory and calculation process of the solution are described in detail in the paper, and by numerical experiments of different three dimensional objects, the accuracy and the efficiency of the algorithm are also discussed.

Keywords: compressed sensing, method of moments, magnetic field integral equation

1 Introduction

As an important research field of computational electromagnetics (CEM), up to the present, electromagnetic (EM) scattering problems can already be computed by a lot of solutions [1-3]. However, fast analysis of wide angle EM scattering problems is still a difficult task for all traditional numerical methods, almost all of them need to solve one discrete angle after another to finish calculating this kind of problem [4-6]. It means iterative operations are adopted as incident angle changes, which must result in low efficiency. Aiming at this difficulty, we propose a new method, which can solve it effectively by applying compressed sensing (CS) into method of moments (MoM) for magnetic field integral equation (MFIE).

CS is called as 'a big idea' in the field of signal processing [7]. One of the most interesting advantages in CS is that it breaks the restriction of Nyquist-Shannon sampling theorem [8] - it can capture and represent compressible signals at a rate significantly below the Nyquist rate. MoM is a classical numerical method of electromagnetic field integral equation, which is applied extensively in solving EM scattering problems [9]. MFIE, as is well known to all, is considered to have some advantages by its smaller condition number and faster iterative convergence speed [10], so we choose it as the basic integral equation. In this paper, the theory and implementation of the new solution are elaborated, and numerical simulation for different three dimensional objects is presented and discussed - it is shown that the new method can obtain accurate results by only several measurements and the efficiency can be improved greatly.

2 Traditional MoM for MFIE

For a closure perfect electric conductor (PEC) whose surface is S radiated by electromagnetic waves, MFIE could be represented as:

$$\mathbf{J}(\mathbf{r}) = 2\hat{\mathbf{n}} \times \mathbf{H}(\mathbf{r}) + 2\hat{\mathbf{n}} \times P.V. \int_{S'} \mathbf{J}(\mathbf{r}') \times [\nabla' G(\mathbf{r}, \mathbf{r}')] \mathrm{d}S' , \quad (1)$$

where $\mathbf{J}$ represents induced current density on S, $\hat{\mathbf{n}}$ represents unit normal vector of the surface of object, $\mathbf{H}(\mathbf{r})$ represents incident magnetic field, $G(\mathbf{r}, \mathbf{r}') = \mathrm{e}^{-\mathrm{j}kR}/4\pi R$, $P.V.\int$ means principal value integral.

Calculation process of traditional MoM for MFIE is as follows:

Step 1: Use RWG vector basis functions $\mathbf{f}_n^S$

$$\mathbf{f}_n^S(\mathbf{r}) = \begin{cases} \dfrac{l_n}{2A_n^{\pm}} \boldsymbol{\rho}_n^{\pm S} & \mathbf{r} \in \mathrm{T}_n^{\pm} \\ 0 & \text{else} \end{cases} , \quad (2)$$

to expand $\mathbf{J}(\mathbf{r})$.

Step 2: Make inner product operation by Galerkin's method:

$$< \mathbf{f}_m^S(\mathbf{r}), \hat{\mathbf{n}} \times \mathbf{H} >= \sum_{n=1}^{N} I_n [< \mathbf{f}_m^S(\mathbf{r}), \frac{\mathbf{f}_n^S(\mathbf{r}')}{2} > - < \mathbf{f}_m^S(\mathbf{r}), \hat{\mathbf{n}} \times \int_{S'} (\mathbf{f}_n^S(\mathbf{r}') \times \nabla' G(\mathbf{r}, \mathbf{r}')) \mathrm{d}S' >], \quad (3)$$

that is:

$$\mathbf{Z}_{n \times n} \mathbf{I}_n = \mathbf{V}_n , \quad (4)$$

* ***Corresponding author*** e-mail: xycaoBL@163.com

where:

$$Z_{mn} = \frac{1}{2}\int_S \mathbf{f}_m^S(\mathbf{r})\cdot\mathbf{f}_n^S(\mathbf{r}')\mathrm{d}S - \int_S \mathbf{f}_m^S(\mathbf{r})\cdot\hat{\mathbf{n}}\times \int_{S'}\mathbf{f}_n^S(\mathbf{r}')\times\nabla' G(\mathbf{r},\mathbf{r}')\mathrm{d}S'\mathrm{d}S, \quad (5)$$

$$V_m = \int_S \mathbf{f}_m^S(\mathbf{r})\cdot[\hat{\mathbf{n}}\times\mathbf{H}(\mathbf{r})]\mathrm{d}S. \quad (6)$$

Step 3: Solve the matrix equation with gauss integration, mean value theorem and treatment of singularities.
Step 4: Numerical result of **J**(**r**) will be finally calculated as:

$$\mathbf{J}(\mathbf{r}) = \sum_n I_n \mathbf{f}_n^S(\mathbf{r}). \quad (7)$$

While the angle of incident wave is not certain but in a wide range, traditional MoM for MFIE has to compute the matrix equations at each small discrete angle repetitively, so the efficiency is low.

3 Basic CS theory

CS is based on sparse representation of signals, its mathematical model [11-13] can be formulated as follows:

$$\mathbf{\Phi}_{M\times N}\mathbf{X}_{N\times 1} = \mathbf{\Phi}_{M\times N}\mathbf{\Psi}_{N\times N}\boldsymbol{\alpha}_{N\times 1} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{pmatrix} = \mathbf{s}_{M\times 1} \quad (M \ll N), \quad (8)$$

where **X** stands for original signal, **α** is the sparse projection of **X**, **Ψ** stands for sparse basis, **Φ** stands for measurement matrix and it is incoherent with **Ψ**. From the *M*-dimensional measurement **s**, the approximation of **α** can be calculated from a *L*-minimization problem as:

$$\hat{\boldsymbol{\alpha}} = \min\|\boldsymbol{\alpha}\|_L \quad s.t.\ \mathbf{\Phi\Psi}\boldsymbol{\alpha} = \mathbf{s}, \quad (9)$$

finally, the original signal can be approximated as:

$$\hat{\mathbf{X}} = \mathbf{\Psi}\hat{\boldsymbol{\alpha}}. \quad (10)$$

4 Solution of CS introduced to MFIE

We introduce CS into traditional MoM for MFIE to fasten calculating wide angle EM scattering problems, the procedure is as follows:
Step 1: Assume the discrete angles of incident waves are $\theta_1,\theta_2...\theta_n$, accordingly, the excitations could be denoted as $\mathbf{V}(\theta_1)$, $\mathbf{V}(\theta_2)$, ..., $\mathbf{V}(\theta_n)$, construct a new group of excitations as:

$$\mathbf{V}_i^{\mathrm{CS}} = \alpha_{i1}\mathbf{V}(\theta_1) + \alpha_{i2}\mathbf{V}(\theta_2) + \cdots + \alpha_{in}\mathbf{V}(\theta_n), \quad (11)$$

where $i = 1, 2, ...m$ and $m \ll n$.

Step 2: Substitute these new excitations into the matrix equation of MFIE, that is,

$$\mathbf{Z}\mathbf{I}_i^{\mathrm{CS}} = \mathbf{V}_i^{\mathrm{CS}}, \quad (12)$$

solve these matrix equations as the way of traditional MoM, then $\mathbf{I}_1^{\mathrm{CS}}, \mathbf{I}_2^{\mathrm{CS}}, ..., \mathbf{I}_m^{\mathrm{CS}}$ can be obtained.
Step 3: Since the impedance matrix **Z** does not vary with the angle of incidence, $\mathbf{I}_i^{\mathrm{CS}}$ can be expanded as

$$\mathbf{I}_i^{\mathrm{CS}} = \alpha_{i1}\mathbf{I}(\theta_1) + \alpha_{i2}\mathbf{I}(\theta_2) + \cdots + \alpha_{in}\mathbf{I}(\theta_n) \ (i = 1, 2, ...m), \quad (13)$$

thus $\mathbf{I}_1^{\mathrm{CS}}, \mathbf{I}_2^{\mathrm{CS}} ... \mathbf{I}_m^{\mathrm{CS}}$ are *m* measured values of $\mathbf{I}(\theta_1), \mathbf{I}(\theta_2) ... \mathbf{I}(\theta_n)$.
Step 4: According to the theory of CS, from these measured values, current coefficient vectors over the wide angle can be reconstructed accurately by the recovery algorithm and sparse basis, so we get the relation:

$$\begin{bmatrix} \mathbf{I}_1^{\mathrm{CS}} \\ \vdots \\ \mathbf{I}_m^{\mathrm{CS}} \end{bmatrix} = \mathbf{\Phi}_{m\times n}\begin{bmatrix} \mathbf{I}(\theta_1) \\ \mathbf{I}(\theta_2) \\ \vdots \\ \mathbf{I}(\theta_n) \end{bmatrix} = \mathbf{\Phi}_{m\times n}\mathbf{\Psi}_{n\times n}\begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_n \end{bmatrix}, \quad (14)$$

where $\mathbf{\Phi}_{m\times n}$ represents measurement matrix, $\mathbf{\Psi}_{n\times n}$ represents sparse basis, $[\mathbf{s}_1\ \mathbf{s}_2\ ...\ \mathbf{s}_n]^{\mathrm{T}}$ is sparse projection of $[\mathbf{I}(\theta_1)\ \mathbf{I}(\theta_2)\ ...\ \mathbf{I}(\theta_n)]^{\mathrm{T}}$.
Step 5: Calculate the optimization problem:

$$\begin{bmatrix} \hat{\mathbf{s}}_1 \\ \hat{\mathbf{s}}_2 \\ \vdots \\ \hat{\mathbf{s}}_n \end{bmatrix} = \min\left\|\begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_n \end{bmatrix}\right\|_L \quad s.t.\ (\mathbf{\Phi\Psi})\begin{bmatrix} \hat{\mathbf{s}}_1 \\ \hat{\mathbf{s}}_2 \\ \vdots \\ \hat{\mathbf{s}}_n \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1^{\mathrm{CS}} \\ \vdots \\ \mathbf{I}_m^{\mathrm{CS}} \end{bmatrix}. \quad (15)$$

Step 6: Finally, current coefficient vectors over the wide angle can be reconstructed as:

$$\begin{bmatrix} \hat{\mathbf{I}}(\theta_1) \\ \hat{\mathbf{I}}(\theta_2) \\ \vdots \\ \hat{\mathbf{I}}(\theta_n) \end{bmatrix} = \mathbf{\Psi}_{n\times n}\begin{bmatrix} \hat{\mathbf{s}}_1 \\ \hat{\mathbf{s}}_2 \\ \vdots \\ \hat{\mathbf{s}}_n \end{bmatrix} \quad (16)$$

and the induced current density **J** will be solved by Equation (7).

Above all, based on $m(m \ll n)$ times of calculation of MoM, $\mathbf{I}(\theta_1), \mathbf{I}(\theta_2) ... \mathbf{I}(\theta_n)$ are solved. Compared with *n* times of calculation of matrix equations needed by traditional MoM, the amount of computation is reduced.

5 Numerical results

Take different three dimensional objects as numerical examples. Experimental electromagnetic parameters are set as follows: the frequency of incident wave f=1.35×10^{9}Hz, permittivity ε=1/(4π×9×10^9)F/m, permeability μ=4π×10^{-7}H/m. Consider 360 angles of incidence (from 1° to 360°, the angle of the wave which is propagating along y axis and E-polarized in the x direction is defined as 0°). Choose Gauss random matrix [14] as the measurement matrix, Hermite basis [15] as the sparse basis and orthogonal matching pursuit (OMP) [16,17] as the recovery algorithm, take 35 times of measurement as example. Compare the calculation results of our method with the ones of traditional MoM.

5.1 NUMERICAL EXAMPLE 1

Consider a PEC cuboid with size of 0.2×0.1×0.05m, as shown in Figure 1.

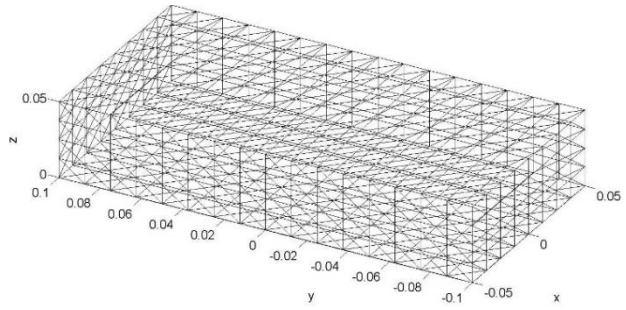


FIGURE 1 The PEC cuboid model

Calculate the current coefficients over 1°, 2°, ..., 360° by both traditional MoM and our solution, and compare the results of two methods. Figure 2 shows the comparison of the current coefficients over the wide angle on an arbitrary RWG basis (take basis number 327 which is centered at (0.025,-0.056,0) as example).

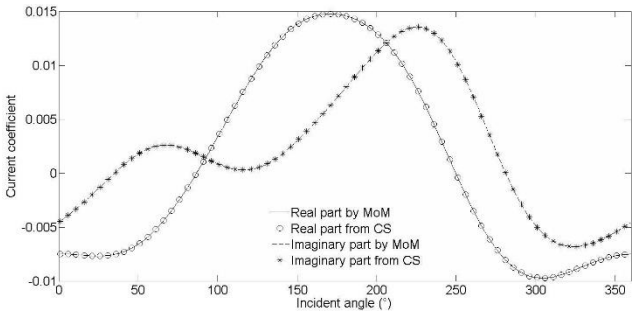


FIGURE 2 Comparison of the calculation results of the current coefficients over the wide angle on basis 327 of the cuboid

From Figure 2, we can see that the results of our method are completely consistent with the ones of traditional MoM.

5.2 NUMERICAL EXAMPLE 2

Consider a multi-objective model which consists of a PEC sphere, a PEC cube and a PEC rectangular pyramid, as shown in Figure 3 (assume the radius of the sphere is 0.05m, the edge-length of the cube is 0.1m and the size of the rectangular pyramid is 0.1×0.1×0.1m.).

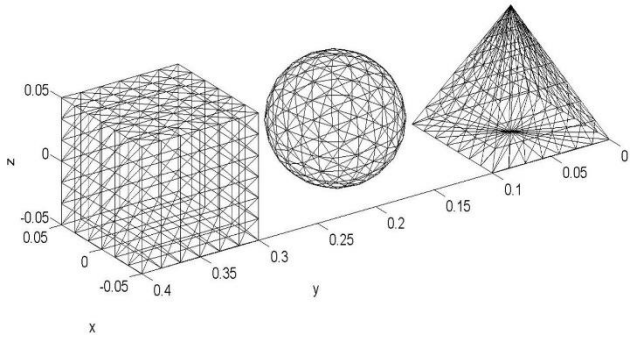
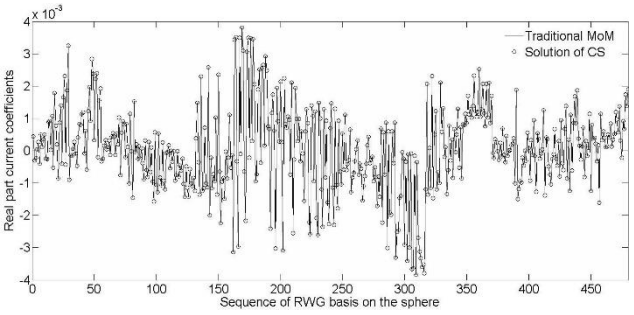
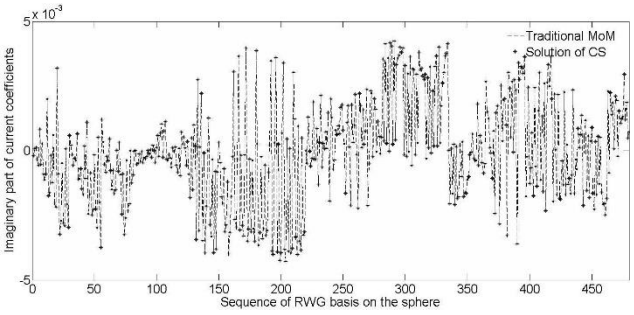


FIGURE 3 The multi-objective model

Figures 4-6 show the comparison between the current coefficients of all RWG basis on the sphere, the cube and the rectangular pyramid over an arbitrary incident angle (take 100°, 200°, 300° as examples) of the wide range calculated by our method and the results of traditional MoM respectively.

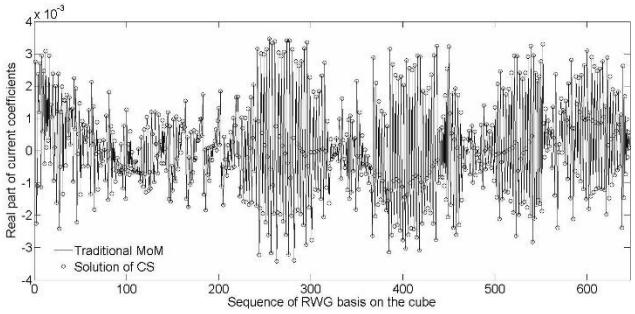


a)

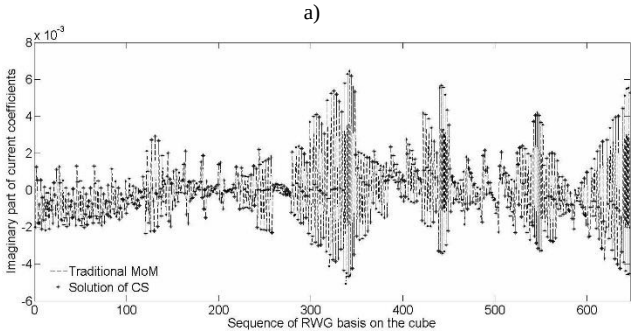


b)

FIGURE 4 Comparison of the calculation results of the current coefficients on the sphere (incident angle=100°): a) Real part, b) Imaginary part

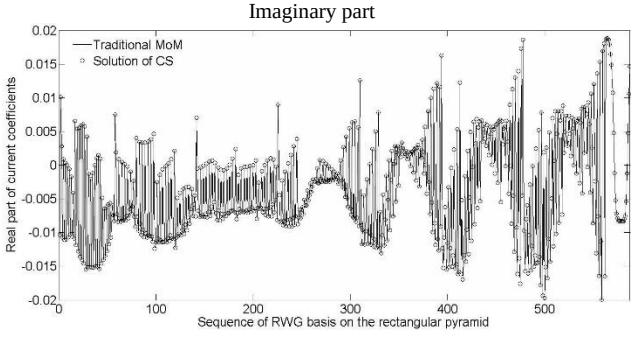


a)

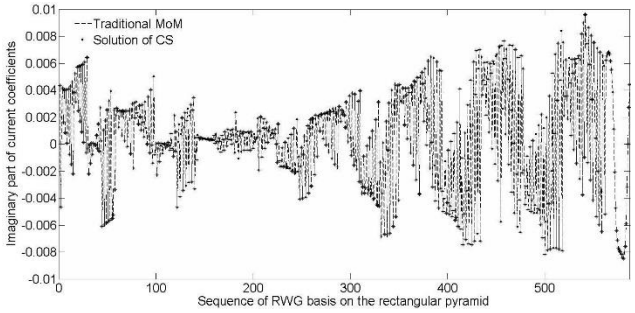


b)

FIGURE 5 Comparison of the calculation results of the current coefficients on the cube (incident angle=200°): a) Real part, b) Imaginary part



a)



b)

FIGURE 6 Comparison of the calculation results of the current coefficients on the rectangular pyramid (incident angle=300°): a) Real part, b) Imaginary part

From Figures 4-6, one can see that the results of our method are also accurate for the multi-objective model.

5.3 NUMERICAL EXAMPLE 3

Consider a simple missile-like model, as shown in Figure 7.

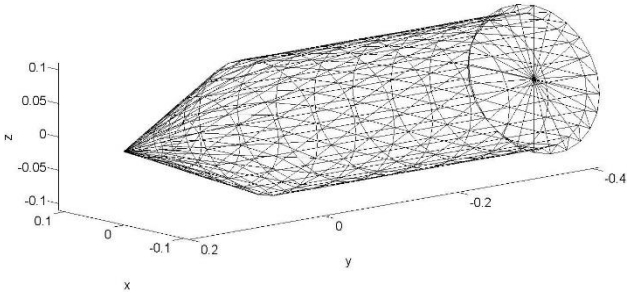
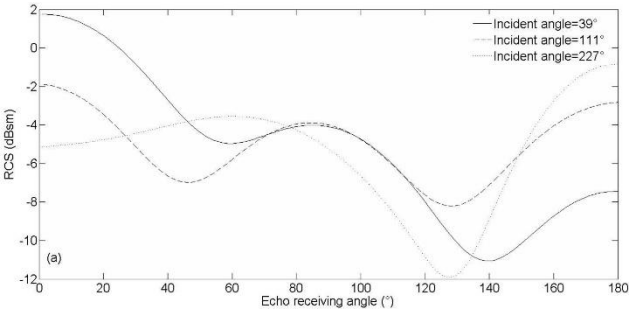
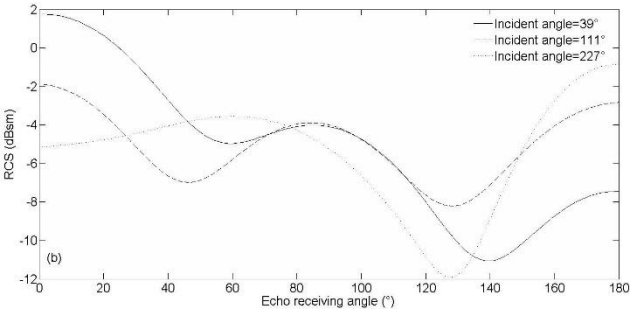


FIGURE 7 The simple missile-like model

Calculate the values of RCS (take *E* surface as example) of this simple missile-like model over some discrete angles among the wide range (take 39°, 111°, 227° as examples) based on both our method and traditional MoM, comparisons of the results of the two solutions are shown in Figure 8.



a)



b)

FIGURE 8 Comparison of the RCS results of the simple missile-like model by a) traditional MoM and b) our method

From Figure 8, one can see that the numerical results calculated by our method are still accurate.

5.4 ERROR STATISTICS

Define the calculation error as:

$$\Delta = \frac{\left\| \hat{\mathbf{I}}_\theta - \mathbf{I}_\theta \right\|_2}{\left\| \mathbf{I}_\theta \right\|_2} \times 100\% , \tag{17}$$

where $\mathbf{I}_\theta$ stands for the current coefficient matrix over the wide angle calculated by traditional MoM, $\hat{\mathbf{I}}_\theta$ stands for the one calculated by our method. The calculation error statistics for the cuboid, the multi-objective model and the simple missile-like model are shown in Table 1.

TABLE 1 The calculation error statistics

Object	Example 1	Example 2	Example 3
Calculation Error	2.7374×10^{-5}	6.1254×10^{-5}	1.8901×10^{-6}

5.5 COMPARISON OF OPERATION TIME

Table 2 shows the comparison between the computational time of our method and the one of traditional MoM. (The operation environment of the programs is Mathworks Matlab7.0, Pentium(R) Dual-Core CPU at 2.10GHz and an internal memory with capacity of 2GB.)

TABLE 2 Comparison of the computational time (s)

Object	Example 1	Example 2	Example 3
Traditional MoM	402.5571	431.6213	510.1818
Solution of CS	93.6437	99.1613	111.7152

From the tables above, we can see that the calculation results of our method relative to the ones of traditional MoM are still accurate highly but the operation time is reduced a lot.

6 Conclusion

Overall, aiming at fast analyzing wide angle EM scattering problems, this paper proposes a new solution by applying CS into MoM for MFIE, and the feasibility of the method is verified by numerical experiments -- the expected effect of fastening calculating EM scattering problems over a wide angle is achieved.

Acknowledgments

This research was partially supported by Anhui Provincial Natural Science Foundation of China under Grant No. 1408085QF104, the Key Project of Provincial Natural Science Research of University of Anhui Province of China under Grant No. KJ2014A206, the Key Project of the Ministry of Education of China under Grant No. 212081 and the National Natural Science Foundation of China under Grant No. 61301062.

References

[1] Christopoulos C A 2011 *IET 8th International Conference on Computation in Electromagnetics (CEM 2011)* **1**(1) 85
[2] Song J M, Lu C C, Chew W C 1997 *IEEE Transactions on Antennas Propagation* **45**(10) 1488-93
[3] Li J, Guo L X, He Q, Wei B 2011 *Chin Phys Lett* **28**(10) 104101
[4] Moss C D, Teixeira F L, Yang Y E, Kong J A 2002 *IEEE Transactions on Geoscience and Remote Sensing* **40**(1) 178-86
[5] Zhao K Z, Vouvakis M N, Lee J F 2006 *IEEE Transactions on Magnetics* **42**(4) 583-6
[6] Fasenfest B, Rockway J D, Champagne N J, Sharpe R M 2004 *IEEE Antennas and Propagation Society International Symposium* **4**(1) 3944-3947
[7] Baraniuk R 2007 *IEEE Signal Proc Mag* **24**(4) 118-21
[8] Tropp J A, Laska J N, Duarte M F, Romberg J K, Baraniuk R 2010 *IEEE Trans. Inform. Theory* **56**(1) 520-544
[9] Shafieipour M, Jeffrey I, Aronsson J, Okhmatovski V I 2014 *IEEE Transactions on Antennas Propagation* **62**(2) 772-82
[10] Chew W C, Davis C P, Warnick K F, Nie Z P, Hu J, Yan S, Gurel L 2008 *IEEE Antennas and Propagation Society International Symposium (AP-S)* **1**(1) 3221-2
[11] Zhang J D, Zhu D Y, Zhang G 2012 *IEEE Transactions Signal Processing* **60**(4) 1718-29
[12] Yu G S, Sapiro G 2011 *IEEE Trans Signal Proces* **59**(12) 5842-58
[13] Baraniuk R G, Candes E, Nowak R, Vetterli M 2008 *IEEE Signal Processing Magazine* **25**(2) 12-3
[14] Candes E J, Romberg J, Tao T 2006 *IEEE Transactions on Information Theory* **52**(2) 489-509
[15] Cao X Y, Chen M S, Wu X L 2013 *Chin Phys Lett* **30**(2), 028401
[16] Hussain Z, Shawe-Taylor J, Hardoon D R, Dhanjal C 2011 *IEEE Transactions on Information Theory* **57**(8) 5326-41
[17] Needell D, Vershynin R 2007 *IEEE Journal of Selected Topics in Signal Processing* **4**(2) 310-6

Authors

Xinyuan Cao, born in February, 1983, Hefei, Anhui, China

Current position, grades: Doctor of Electronic Information Engineering, lecturer at Hefei Normal University.
University studies: PhD degree in Electronic Information Engineering at Anhui University of China in 2013.
Scientific interests: electromagnetic scattering and signal processing in electronic information engineering.
Publications: 5 papers.

Mingsheng Chen, born in January, 1981, Wuhu, Anhui, China

Current position, grades: professor at Hefei Normal University and doctoral supervisor at Anhui University.
University studies: PhD degree in Electronic Information Engineering at Anhui University of China in 2008.
Scientific interests: theory of electromagnetic field integral equations, microwave imaging, wavelet signal processing.
Publications: 60 papers.

Bingbing Chen, born in December, 1986, Anqing, Anhui, China

Current position, grades: assistant in Hefei Normal University.
University studies: master's degree on Electronic Information Engineering at Zhejiang Sci-Tech University of China, 2012.
Scientific interests: signal processing and motor control in electronic information engineering.
Publications: 3 papers

Liangliang Cheng, born in July, 1987, Chizhou, Anhui, China

Current position, grades: assistant in Hefei Normal University.
University studies: master's degree on Signal and Information Processing at Zhejiang Sci-Tech University of China in 2012.
Scientific interests: computer network, smart home and the internet of things in electronic information engineering.
Publications: 3 papers

Qi Qi, born in December, 1988, Anqing, Anhui, China

Current position, grades: assistant in Hefei Normal University
University studies: master's degree in Electronic Information Engineering at Anhui University of China in 2013.
Scientific interests: electromagnetic scattering and signal processing in electronic information engineering, electromagnetic field integral equations, sparse transform.
Publications: 3 papers.

The influence and improvement of scanning speed to the line width in fused deposition manufacturing

Zili He, Tingchun Shi*, Biaobiao Gao

School of Life information and Instrument Engineering, Hang Zhou Dianzi University, Hangzhou, Zhejiang, China

Received 1 October 2014, www.cmnt.lv

Abstract

During the process of fused deposition manufacturing, the extruding speed and the scanning speed of nozzle don't match in some path of large change in direction may lead the extrusion of material more or less than demand too much which causes accumulation of material or the lack of line width. The method of controlling FDM's forming accuracy which is under the restrictions of the mechanical structure of the work platform is studied. By adjusting the extruding speed and measuring the line width on the work platform the width of forming line in process can be controlled in allowed range. The experiment results indicate that through controlling the speed of extruding motor the extrusion of material is controlled efficiently and the accuracy of molding is improved.

Keywords: movement theory of nozzle, scanning speed, extruding motor controlling, measure of line-width

1 Introduction

The Fused deposition manufacturing (FDM) is one of the rapid prototyping manufacturing technology and is today the second most common commercial layered manufacturing system [1]. The FDM without the need of laser is lower cost and has a simpler structure compared to SLA, LOM, SLS and other rapid prototyping manufacturing technology. So the maintenance is more convenient, too. But the accuracy of line width is the constraint to the FDM's further improvement [2].

2 Theory of scanning speed

The manufacturing accuracy of the FDM is influenced by many factors including the temperature of the nozzle, material viscosity, the contraction size of material's solidification [3], the accuracy of the equipment, the performance of software system and so on. Besides, the manufacturing is also restricted by working platform. The movement of the FDM mainly depends on the three motors of *X* axis, *Y* axis and *Z* axis. The *Z* axis motor controls the height while the *X* and *Y* axes motors controls the nozzle's movement in horizontal direction. That means the nozzle's horizontal scanning speed is composed of motors speed of *X* and *Y* axes. As shown in Figure 1.

It's hard to form when the scanning speed is faster than the extruding speed which causes the line broken due to the lack of the material. Instead, when the scanning speed is slower than the extruding speed will lead to the extrusion of the material too much and it causes the material on the working platform spreading non-uniform, which will affect the model's quality [4]. The set scanning speed of the nozzle matches the appropriate extruding speed can make the line width of the material on the working platform in a stable range.

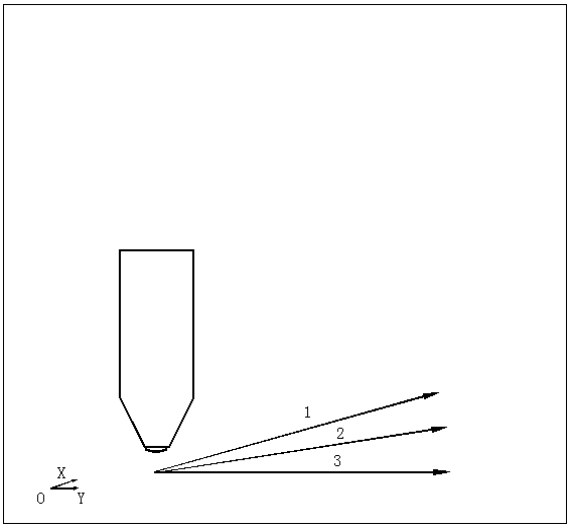


FIGURE 1 The velocity of the nozzle 1. The speed of the *X* axis 2. The actual speed of the nozzle 3. The speed of the *Y* axis

However, in some path, such as right-angled path, the scanning speed will change substantially while the extruding speed doesn't change which may cause accumulation of the material on the working platform. Although the scanning speed that is set by PC in different paths, which the PC cannot tell the difference, the motor speed of X and Y axes will change anyway. The most paths during FDM's working process are the straight line path, the right-angled path and the curved path. Here is an experiment between the line width and them. The right-angled path takes the bevel width, which is the critical point of the velocity's change, as the line width. The motor

* ***Corresponding author*** e-mail: stc@hdu.edu.cn

speed of *X* and *Y* axes don't change when it comes the straight line path. The Table 1 shows the line width of straight line path doesn't change much. The line width in curved path is wider than it in straight line path but the difference between them is not big in generally even the motor speed of X and Y axes are changing all the time. The scanning speed couldn't keep the same when its direction is about to change 90 degree. Because the scanning speed is composed of 2 phase motors' speed and the change in direction of the nozzle need both change of the two motors substantially in a very short time.

TABLE 1 The line-width in different trajectory (mm)

	straight line path	the curved path	right-angled path
d	0.65	0.7	1.10
	0.62	0.69	1.05
	0.61	0.72	1.07
	0.59	0.71	0.99
	0.62	0.75	1.08
	0.63	0.7	1.09
$\bar{d}$	0.62	0.71	1.06

Nozzle's temperature: 200°C the scanning speed: 10mm/s material: Polyurethane

The change of the shaft's speed is accomplished by changing the speed of the servo motor. In normal case, the starting of the no-load servo motor is very quick and the time of it can be ignored compared to the while working process. Even it may cost longer when it is loaded. The right-angled path is done by the nozzle instantly which makes time of the servo motor's acceleration not be ignored and it can be proved by Table 1.The acceleration of the servo motor in the starting and stopping stays the same, which is showed by Figure 2, thus, the relation between the velocity of servo motor and the velocity of the shaft can be calculated. The rotation of the shaft driven by servo motor makes the nozzle move in the direction of X(Y) axis. According to relation between the spiral and the helical pitch in mechanical transmission, the equation can be got as follows:

$$L = M \times P_h . \tag{1}$$

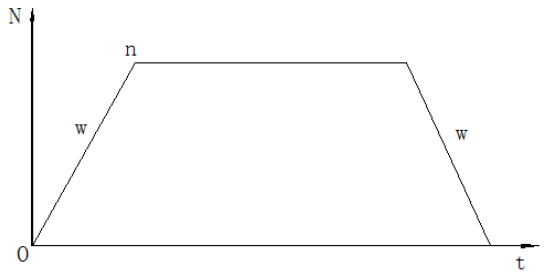


FIGURE 2 The process of motor's starting and stopping. *n* is the rated speed of the motor, *w* is the acceleration of the motor

The *L* represents the distance moved by the nozzle in the direction of *X*(*Y*) axis, the *M* represents the number of rotations of the shaft driven by servo motor, the P_h represents the helical pitch. Both sides of the Equation (1) are divided by the time, the relation between the velocity in axial of the nozzle and the velocity will be:

$$v = \frac{L}{t} = \frac{M}{t} P_h = n \times P_h , \tag{2}$$

and both sides of Equation (2) are differentiated, the relation is as follows:

$$a = \frac{dv}{dt} = \frac{dn}{dt} P_h = w \times P_h , \tag{3}$$

a represents the axial acceleration, the *w* represents the servo motor's acceleration.

Establish a model and analysis it for the process when the nozzle is going through a righted-angled path (Figure 3). The direction of the scanning speed is assumed to parallel the *X* and *Y* axes (For the convenience of analysis, the same conclusion can be got in other angles).

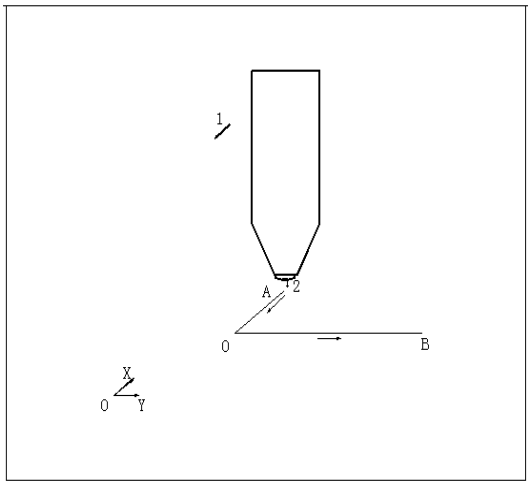


FIGURE 3 The scanning path 1. The scanning speed v_2. 2. The extruding speed v_1

As the scanning speed of the nozzle is composed of the motors' speed in the *X* and *Y* axes, The scanning speed in path AO means the speed of the motor in *X* axis. The nozzle decelerates in path AO and the deceleration can be deduced by Equation (3).The scanning speed of the nozzle is v_1,The extruding speed is v_2,the distance of decelerating is *S*, the time of nozzle spending in the decelerating distance is t_1.

The time spent in path AO with decelerating speed:

$$t_1 = v_1 / a . \tag{4}$$

The time spent in path AO with uniform speed:

$$S = v_1 \times t_2 ,$$

$$S = v_1^2 / 2a ,$$

$$t_2 = v_1 / 2a . \tag{5}$$

The different time between them:

$$\Delta t = t_1 - t_2 = v_1 / 2a . \tag{6}$$

Analysing the process of extruding material and the condition of the material on the working platform, as shown in Figure 4.In the circumstances of uniform speed, the quantity *m* of the material comes out of the nozzle in unit time *t*:

$$m = v_2 \times t \,. \tag{7}$$

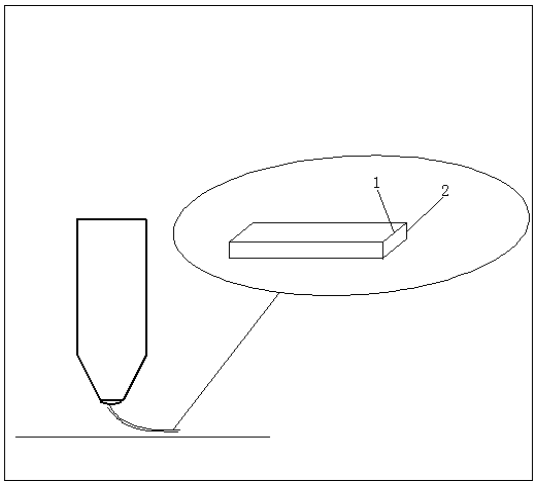


FIGURE 4 the amplification of line-width 1. Line width *d*. 2. Height of layer *h*

The quantity of the material extruded by nozzle on the working platform (ρ is the density of the material) [5]:

$$m = v_1 \times d \times h \times t \times \rho \,. \tag{8}$$

The quantity of the material stays the same from the nozzle to the working platform because there is no loss, the following formula can be got combining Equations (7) and (8) [6]:

$$v_2 \times t = v_1 \times d \times h \times t \times \rho \,. \tag{9}$$

If the line width and line height are wanted to keep the same when scanning speed v_1 is decelerating, the v_2 should be decelerated along with the v_1, which can be known from Equation (6).

$$\Delta m = \rho \times v_2 \times \Delta t = (v_1 \times v_2 \times \rho) / 2a \,. \tag{10}$$

Therefore, there is 2 times of quantity extruded when the nozzle goes through a righted-angle path combining the path AO and path OB. Substituting Equations (2) and (3) into Equation (10), the quantity extruded more than the standard is $(\mathrm{n} \times v_2 \times \rho) / w$.

There are two mainly filling ways in rapid prototyping technology presently which are straight filling and offsetting filling. The way of the straight filling (Figure 5a) working is moving back and forth in a straight direction and has nothing to do with the outline of the model to be done. The outline is added after each layer is done. But the offsetting filling (Figure 5b) is working in a way, which is close to the outline of each layer. The scanning path from inside to outside parallels the outline related to each layer of the model [7].

It fills perpendicularly to the working path between the two back and forth in order to keep the material continuous although the straight filling is not complex. The nozzle always meets the condition of big change in direction which causes the accumulation of the material no matter the simple straight filling or the complex offsetting filling. There are more than dozens and hundreds of layers in a model and each layer's error will be accumulated which may affect the accuracy of the model even destroy the structure.

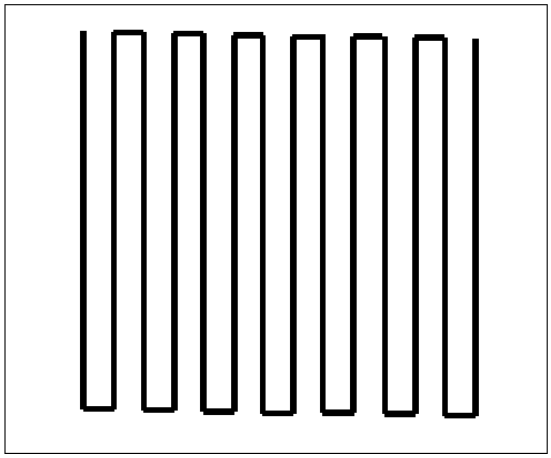


FIGURE 5a straight filling

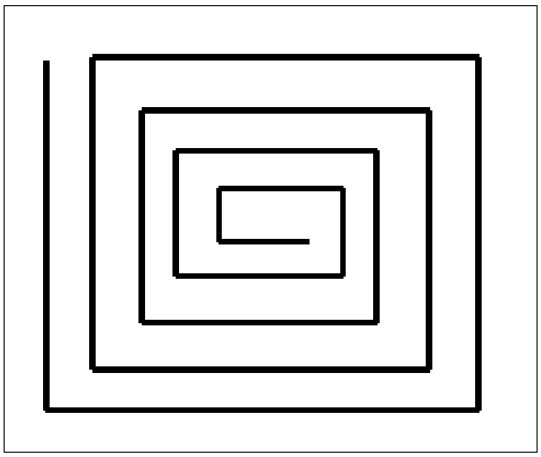


FIGURE 5b offsetting filling

3 The measures of improvement

The paper takes the nozzle of fused deposition molding machine the MEM-300 (Figure 6) of Tsinghua University as the experimental object and bring a method to improve the controlling. The screw extruder which simplifies the process and reduces the cost of the trial [8] is one of the main extruder in the field of FDM presently. The main factor affecting the extruding speed is the balance between the flowing resistance and driving force of the extruding motor in liquefaction pipe and nozzle where the liquid material goes through [9]. The extruding motor is the only problem to be considered to the extruding speed because the model is made by one material. The motor driver transmits the fixed frequency of pulse to the extruding

motor in order to make the material out of the nozzle with an unchanged speed in the whole process once the machine starts working. According the above analysis, the reason to the error is that the extruding speed does not match the scanning speed when the nozzle is going to have a big change in direction.

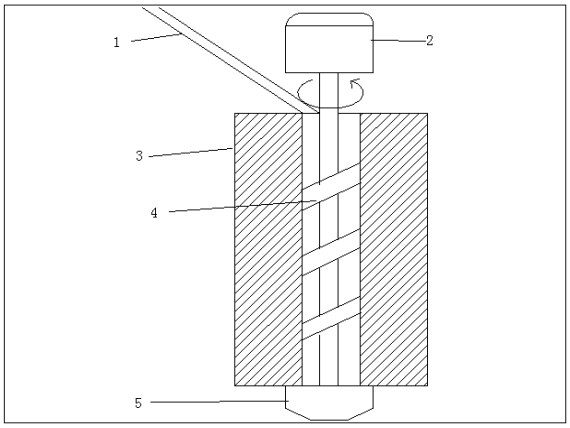


FIGURE 6 The structure of the nozzle 1. Feeding channel. 2. The extruding motor. 3. The heater. 4. Screw arbour. 5. The nozzle

If the extruding speed is reduced during the changing through controlling the extruding motor, the error will be decreased. The purpose of the study is to measure the line width in the different working frequency of the extruding motor. The reasonable speed could be chose when the nozzle is going to have a big change in direction. It may lay the foundation of the whole process's automation. The extruding motor is controlled by the PC software in normal condition and the extruding motor works with a set frequency once the software is started. The set speed is suitable which is based on a large number of experiments and the experiment is much significate based on the speed above.

After the oscilloscope detects the signal which is sent out by the PC to control the stepper motor driver, it could found the frequency of the signal is 1000Hz.

The external equipment of extruding motor is composed of step motor drive controller YJ01, stepper motor driver DM422C and the DC power to the driver.

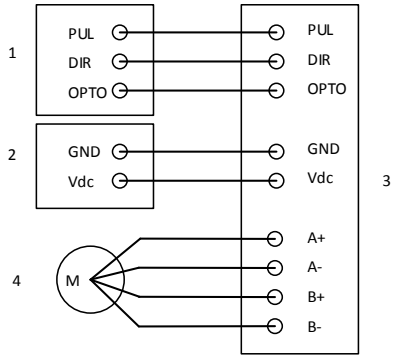


FIGURE 7 The connection of equipment 1. Stepper motor drive controller. 2. DC power. 3. Stepper motor driver. 4. Stepper motor

The Figure 7 is the connection of the equipment. The Figure 8 is the analysis of the righted-angle path. The line width D is the right line to be measured because the line width D is the critical place of the velocity's change and the maximum variation of the line width.

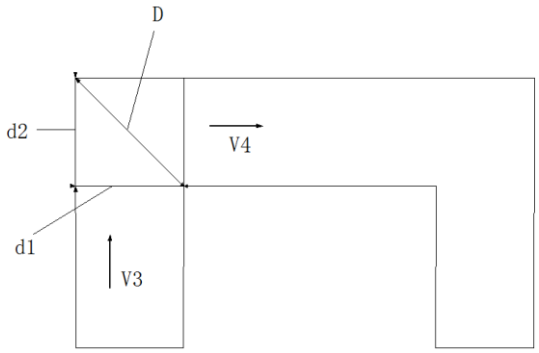


FIGURE 8 The amplification of the right-angled path

The scanning speed of the nozzle is changeless in the ideal condition, $v_3 = v_4$, according to the Equation (8):

$$v_3 \times d_1 \times h \times t \times \rho = v_4 \times d_2 \times h \times t \times \rho , \quad (11)$$

$$d_1 = d_2 = d . \quad (12)$$

So, $D = \sqrt{2}d$.

TABLE 2 The line-width under different frequency of the motor (mm)

Motor's frequency(Hz)	500	550	600	650	700
Line-width in right-angled d	0.85	0.82	0.84	0.89	0.89
	0.86	0.81	0.85	0.83	0.93
	0.82	0.83	0.89	0.86	0.90
	0.81	0.87	0.86	0.90	0.95
	0.78	0.86	0.86	0.92	0.95
	0.78	0.88	0.91	0.91	0.94
	0.75	0.88	0.93	0.88	0.93
	0.87	0.83	0.84	0.89	0.91
$\overline{d}$	0.815	0.8475	0.8725	0.885	0.925

Nozzle's temperature: 200°C the scanning speed: 10mm/s material: Polyurethane

According to the Table 1, the average of the straight line width d is 0.62mm and the bevel width D of righted-angle should be 0.87mm. The accuracy of the model is only controlled in the range of 0.1~0.2mm due to many factors' influence in the FDM process at present [10]. It is known from Table 1 the range of straight line width d is from 0.59mm to 0.65mm and the range of bevel width D should be from 0.83mm to 0.91mm.

TABLE 3 The percentage of the different frequency

motor's frequency (Hz)	500	550	600	650	700
percentage	37.5%	75%	87.5%	87.5%	50%

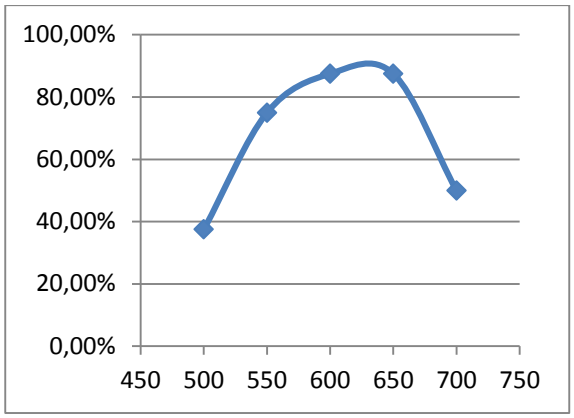


FIGURE 9 Scatter diagram of Table 2

Table 3 is percentage of Table 2 and Figure 9 is scatter diagram of Table 3.The data of line width is most in range of 0.83~0.91mm when the extruding motor's frequency is 550Hz,600Hz and 650Hz after analysing the above charts. The line width meets the requirement when the frequency of the extruding motor is between 550Hz and 650Hz during a righted-angle path according to Figure 9. In order to get rid of the limitation of the machine's structure, the frequency of the extruding motor should be changing along with the different path in the whole working process.

4 Conclusions

This paper has deeply analysed the accuracy of the FDM from the mechanical structure of the nozzle and brought a new method to improve the accuracy with the proof by experiment. The model's accuracy is still restricted by the structure of the working platform even the nozzle and other factors which affect the FDM's accuracy have been improved and it still may not meet the requirement. The accuracy of FDM will be much improved if the intelligent control system which makes the controlling of extruding motor automatic rapidly was taken.

Acknowledgments

This study is supported by the National Natural Science Foundation of China, No.61272389, 2012.

References

[1] Tyberg J, Bøhn J H 1999 FDM systems and local adaptive slicing *Materials & Design* **20**(2-3) 77-82
[2] Huo X-Y, Tao M-Y, Ye C-S 2004 Analysis for Factors that Affect the Precision of the Shaped Pieces on Fused Deposion Modeling Process *Machinery & Electronics* (9) 77-8 *(in Chinese)*
[3] Jia Z-Y, Zou G-L, Guo D-M, Wang D-X 2002 Study on spinning line model and method for compensation in FDM *China Mechanical Engineering* **13**(23) 1997-2000
[4] Zhang Y, Zhou T-R, Xu C-H 2007 Factors and Countermeasures Influencing the Precision in FDM *Journal of Nanchang University* **29**(3) 252-5 *(in Chinese)*
[5] Wang Y-Q, Fang Y, Le G et al. 2009 Finite Element Aided Design of Nozzle Head in FDM *Aviation Precision Manufacturing Technology* **45**(3) 32-6 *(in Chinese)*
[6] Li Y-Q, Shi T-C 2010 Accuracy Control Studyof Rapid Prototyping Machine by FDM *Machine Tool & Hydraulics* **38**(23) 65-9 *(in Chinese)*
[7] Chen J 2009 Analysis of scanning ways in FDM *Huazhong University of Electronic Science and Technology Thesis (in Chinese)*
[8] Guo Y-B, Fan Z, Dong XZ 2007 Discussion of screw extruding and plasticizing fuses deposition advanced rapid molding technology *Modern Manufacturing Engineering* **12** 76-8 *(in Chinese)*
[9] Gu YH, Xiao Q 2003 Analysis on the Extrusion Rate Control of Molten Polymer in Fused Deposition Modeling *New Technology& New Process* (6) 25-6 *(in Chinese)*
[10] Xiao Q, Gu Y-H, Jiang K-Y 2004 Optimal Control of Nozzle Extrusion Parameters for Fused Deposition Modeling Processes *China Mechanical Engineering* **15**(1) 17-9 *(in Chinese)*

Authors

Zili He, born in May, 1989, Wenzhou, China

Current position, grades: Master's degree.
University studies: Bachelor degree at Hangzhou Dianzi University in August 2012.
Scientific interests: additive manufacturing, mechanical manufacturing and automation.

Tingchun Shi, born in May, 1966, Tangshan, China

Current position, grades: PhD, Professor.
University studies: PhD degree at Tsinghua university in August 2003.
Scientific interests: computer application, biofabrication, additive manufacturing.
Publications: more than 40 papers.

Biaobiao Gao, born in October, 1989, Jiaozuo, China

University studies: bachelor of science in engineering in Nanyang Institute of Technology in June 2012.
Scientific interests: biofabrication.

Query semantic data from relational database: an on-demand mapping approach

Haifei Zhang*

School of Mechanical and Electrical Engineering, Nantong Textile Vocational Technology College, CN-226007 Nantong, China

Received 1 March 2014, www.cmnt.lv

Abstract

One of the tasks for semantic web is to integrate large amounts of current information in relational database, which behind Web into machine-understandable RDF data model to form a "web of data". So relational database semantic query namely RDF access to relational database is an important issue in semantic web research. To realize the query is to build mapping relation between relational database schema and ontology. However, there is natural isomerism between them. The traditional method to eliminate the isomerism is to convert relational database schema into a similar ontology form and then to build all concepts and attributes mappings between conversion ontology and input ontology. This paper realized an on-demanding mapping method when users request query, avoided building all concepts and attributes mappings between conversion ontology and input ontology and improved mapping efficiency.

Keywords: relational database, semantic data, conversion ontology, input ontology, on-demanding mapping

1 Introduction

Relational database semantic query namely RDF [1] access to relational database is an important issue in semantic Web research. There are two access approaches: one is to build mapping between relational database schema and known ontology [2], the other is to issue or transform existing relation schema into RDF ontology [3, 4]. With the development of semantic Web, the number of ontology increases, thus we can make full use of existing known ontology to realize the RDF access to relational database that is the first access method is discussed in above. However, there is natural isomerism between relational database schema and ontology. The isomerism can be eliminated by converting relational database schema into similar ontology (we call the conversion ontology) and then building mapping between conversion ontology and input ontology (that is known ontology mentioned above). The basic graph pattern of input ontology SPARQL query reflects ontology classes and attributes that users need to query. This paper realized an on–demanding mapping method when users request query, avoided building all concepts and attributes mapping between conversion ontology and input ontology and improved mapping efficiency.

2 Relational database schema

Definition 1: A relational database schema can be described as $D=(N,attr,pk,fk,datatype)$, and is pentad, where name set $N=ET\cup RT\cup DT$ is a finite set composed of pairwise disjoint sets, including set *ET* of entity relation name, set *RT* of relationship relation name, and set *DT* of data type name, each data type name is data type name predefined by RDBMS. For $\forall T\in ET\cup RT$, the $Rel(T)$ denotes the *T* is a relational table.

In $\forall T\in ET\cup RT$, T has a non-empty set of attributes $attr(T)$ and each attribute $A\in attr(T)$ has a relevant predefined data type $datatype(A)\in DT$ as its range, $datatype(*)$ denotes the predefined data type of "$*$".

In $\forall T\in ET\cup RT$, T has only a single attribute or attribute set determining its tuple, called primary key $pk(T)$ of *T*. If $pk(T)\in attr(T)$, $pk(T)$ is called as single primary key, while if $pk(T)\subseteq attr(T)$, $pk(T)$ is composite primary key.

In $\forall T\in ET\cup RT$, if there is relation attribute quotes the primary key $pk(G)$ of other entity relation $G\in ET$ in T, then this attribute is called foreign key, satisfying $fk(T,G)\subseteq attr(T)$ and:

$$value(fk(T,G))\subseteq value(pk(G))\cup\{null\},$$

$value(*)$ denotes the range of "*", and "*null*" denotes null value. *T* may have $n\,(n\geq 0)$ foreign keys.

3 Ontology and semantic modelling

An ontology is an explicit specification of a conceptualization [5], which can be formalized as O(C, R, F, A, I), where C is the set of concepts/classes; R is the set of relations over elements of C; F is the set of functions; A is a set of axioms; and I is a set of instances.

* ***Corresponding author*** e-mail: zhanghaifei@nttec.edu.cn

Concepts can be used to describe anything, concrete or abstract. The ontology engineer analyzes relevant entities and organizes them into concepts. The backbone of ontology consists of a generalization/specialization hierarchy of concepts, i.e., taxonomy. Relations between concepts of a domain are a subset of the Cartesian product: C×C, which describe the relationships between two entities of the entities in C; Functions are special relations, which presents a functional property on pair entities. For example, mother(x, y) presents that x has a mother y, so, x determines y, like a function: y=mother(x); A is a set of axioms which are tautology assertions; and I is a set if instances of the concepts in C.

In the perspective of Semantic Web, the ontology can be used to model the knowledge about a specific domain, which requires a terminology to represent the concepts and the relations between these concepts. What's more, these concepts and the relations are organized as a hierarchy form. Ontology provides the Semantic Web a formal model of domain knowledge, on which we can use some logic to inference. In Semantic Web, the logical foundation of ontology is the Description Logic (DL) [6].

Based on the concept of ontology, W3C recommends a set of specifications to model the information on the Web of data. RDF (Resource Description Framework) [1] is a lightweight ontology language, which originally designed as a metadata data model. RDF is based upon the idea of making statements about Web resources in the form of triple (subject, predicate, object) expressions. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. Therefore, RDF can be used to describe the entities and the relationships between these entities. However, RDF has limited weak semantics, and we need a rich-semantic ontology language. OWL (Web Ontology Language) [7] is another proposal which is recommended by W3C. OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans. It can facilitate greater machine interpretability of Web content than that supported by RDF and RDFS by providing additional vocabulary along with a formal semantics. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full.

Semantic data modelled with OWL described in RDF is the foundation of Semantic Web applications, how to get the RDF data is an emergency question for many environments. So, lots of researchers have focused on the fields of mapping the existing traditional data to the RDF data [8]. W3C also recommends a language to map the relational data to RDF data, the R2RML (RDB to RDF Mapping Language) [9]. It can be used to express customized mappings from relational databases to RDF datasets. With R2RML, we can view existing relational data in the RDF data model. The target RDF data is modelled in an existing ontology and different distributed relational data can be mapped to the unified semantic data which implements a data integration application.

4 Relational database semantic query

Traditional database semantic query is as follows: First generating conversion ontology from relational database schema by semantic wrapper, then building mapping relation from input ontology to conversion ontology. So SPARQL query on input ontology can be transformed into conversion ontology according to the mapping relation. Lastly submitting conversion ontology to semantic wrapper for transforming into SQL query of relational database. The construction of semantic wrapper refers to the author's previous work [10]. And the key function is to generate conversion ontology and rewrite from conversion query to SQL query and SQL query results transform into RDF data. Relational database semantic query is shown as Figure 1. This paper mainly describes mapping on query, so query results are not discussed here and there is no hint of query results in the graph. All these are realized in the author's previous work [10].

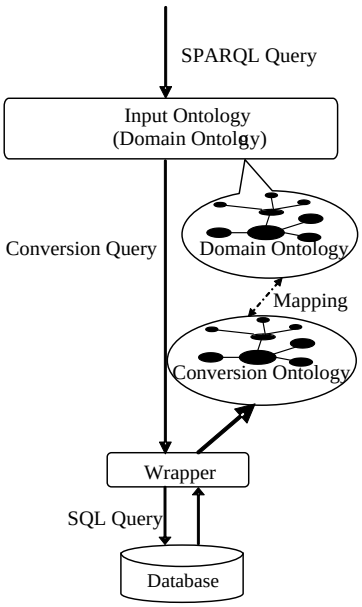


FIGURE 1 The structural diagram of relational database semantic query system

5 SPARQL basic graph pattern

SPARQL is the query language of semantic Web and is matched on graph patterns. Complex graph patterns can be obtained by combining simple patterns. Combining methods include: 1. Basic graph pattern; 2. Group graph pattern; 3. Optional graph pattern; 4. Alternative graph pattern.

Definition 2: A SPARQL query graph pattern *GP* is defined as the following expression:

$GP \to TP \mid GP\ AND\ GP \mid GP\ OPTIONAL\ GP \mid GP\ UNION\ GP \mid GP\ FILTER\ \exp$, where

$GP\ AND\ GP$ is namely group graph pattern, which could be the combination of any other graph patterns.

TP namely basic triple graph pattern, which could be the set consists of many RDF triples.

GP OPTIONAL GP stands for alternative graph pattern, the latter *GP* of which is the range of optional graph pattern.

GP UNION GP means that query graph pattern contains alternative graph pattern.

GP FILTER exp stands for value constraint, in which exp is the expression of value constraint.

Basic graph patterns (abbreviated to BGP) are sets of triple patterns. SPARQL graph pattern matching is defined in terms of combining the results from matching basic graph patterns. A sequence of triple patterns interrupted by a filter comprises a single basic graph pattern. Any graph pattern terminates a basic graph pattern [11].

Figure 2 shows the SPARQL query statement pattern, in which ?x, ?y and ?z respectively stand for instances of Class A, Class B and Class C. Class A has value property A1 and relationship property has B; Class B has value property B1 and B2 and relationship property has A and has C. Among them, A1 is value property variable with value constraint Exp(?A1); B2 is value property variable under OPTIONAL pattern; B1 is value property variable and takes the Value(B1); C1 is value property variable. From this SPARQL query statement pattern, it is known that this query covers basic graph pattern, optional graph pattern, value constraint and so on; for alternative graph pattern (UNION), its query rewriting result can be regarded as union of two SQL statements; SPARQL query statement of UNION graph pattern is not given.

```
PREFIX ex:<http://example.org/schemas/university#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?A1 ?B2 ?C1
WHERE{
    ?x rdf:type ex:ClassA.        }
    ?x ex:A1 ?A1.                 }
    ?x ex:hasB ?y.                }  Basic Graph
    ?y rdf:type ex:ClassB.        }    Pattern
    ?y ex:hasA ?x.                }
    ?y ex:B1 Value(B1).           }
    OPTIONAL{                     }  Optional Graph
    ?y ex:B2 ?B2. }               }    Pattern
    ?y ex :hasC ?z.               }
    ?z rdf:type ex :ClassC.       }  Basic Graph
    ?z ex :C1 ?C1.                }    Pattern
    FILTER(?A1=constant)          }  Value Constraint
}                                     Graph Pattern
```

FIGURE 2 SPARQL query statement pattern

6 On-demanding mapping on query request

In addition to the above described SPARQL basic graph patterns that are composed of triples, ontology is also composed of triples. Inputting the basic graph patterns of ontology SPARQL query in relational database semantic query reflects ontology classes and attributes that users need to query. This chapter presents how to realize the on-demanding mapping method when users request query by examples to illustrate how to build the mapping between the basic graph patterns in SPARQL query and conversion ontology.

Assuming conversion ontology generated from relational database schema is shown as Figure 3.

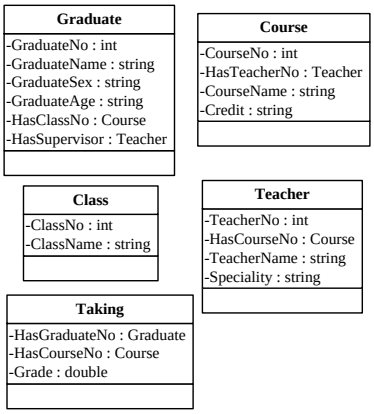


FIGURE 3 Graduate ontology

Figure 5 shows SPARQL query that users input into input ontology shown as Figure 4.

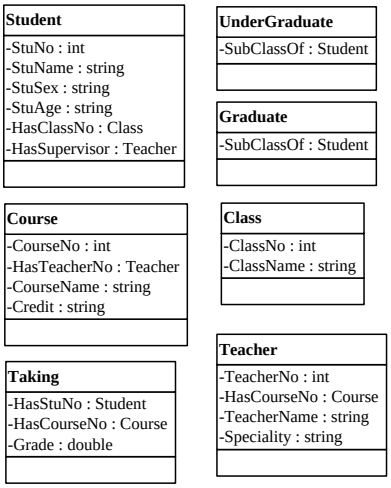


FIGURE 4 Student ontology

```
PREFIX ex:<http://example.org/schemas/university#>
PREFIX  rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
  SELECT ?cid ?cname ?tid ?specia
   WHERE {?x rdf:type ex:Course.
          ?x ex:CourseNo ?cid.
          ?x ex: CourseName ?cname.
          ?x ex: Credit ?Credit.
          ?x ex: hasTeacherNo ?y.
          ?y rdf:type  ex:Teacher.
          ?y ex:TeacherNo ?tid.
          ?y ex:hasCourseNo ?x.
          OPTIONAL{?y ex:Speciality ?specia.}
          FILTER(?credit=3)}
```

FIGURE 5 SPARQL query

The mapping relation between triples of basic and optional graph patterns and conversion ontology Graduate Ontology (as shown in Figure 3) is shown as Figure 6.

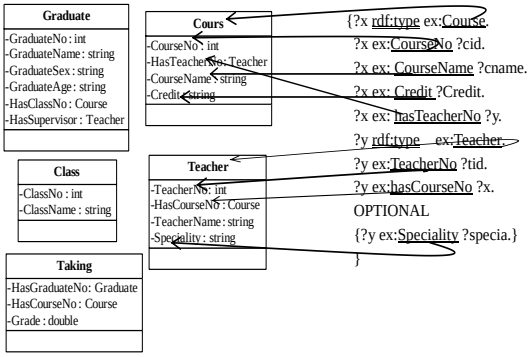


FIGURE 6 Mappings from BGP of SPARQL to conversion ontology

ALGORITHM 1 (the mapping algorithm): Triples to conversion ontology mapping algorithm TP2OMap

Algorithm TP2OMap (TP_{BGP}, TP_{OPT})
Input: TP_{BGP}, TP_{OPT} Output: *Match*(C_{ti}, C_{oi}), *Match*(P_{ti}, P_{oi}) Steps: 1. FOR each triple T_i in TP_{BGP} and TP_{OPT} 2. String S_i ←obtain the subject in T_i; 3. String P_i ←obtain the predicate in T_i; 4. String O_i ←obtain the object in T_i; 5. IF "type" EQUALS P_i THEN 6. BEGIN 7. Obtain corresponding(matching) concept C_{oi} of O_i from conversion ontology; 8. $C_{ti} \leftarrow O_i$; 9. Return *Match*(C_{ti}, C_{oi}) 10. END 11. ELSE 12. BEGIN 13. Obtain corresponding(matching) concept P_{oi} of O_i from conversion ontology; 14. $P_{ti} \leftarrow P_i$; 15. Return *Match*(P_{ti}, P_{oi}) 16. END 17. END FOR

7 Conclusion

This paper describes an on-demanding method when users request query by building mapping between basic graph patterns in SPARQL query and conversion ontology. To compared with the method described in classic literature [2] of the first access approach in section 1, this method avoid constructing all concepts and properties mapping between input ontology and conversion ontology, thus the efficiency is obviously.

The match between basic graph pattern and conversion ontology is similar to directly regarding input ontology SPARQL query as query on conversion ontology. But this query term (class and attribute) is not an exact match and needs to establish matching relation by using lots of similarity calculation algorithm in existing mapping of basic graph patterns and terms on conversion ontology. The on-demanding mapping method using basic graph patterns of SPARQL query statement in relational database semantic query avoided building all concepts and attributes mapping between conversion ontology and input ontology and undoubtedly improved query efficiency.

Acknowledgment

This work is sponsored by the following grants: (i) a grant No. 61163057 from the National Natural Science Foundation of China, (ii) a grant No. BK2010520 from the Natural Science Foundation of Jiangsu Province of China, (iii) Excellent Youth Scholar of "Qing Lan Project" of Jiangsu Province in China (2012), and (iv) two grants No. FYKY/2012/2 and No. FYKY/2013/8 from Science and Technology Program of Nantong Textile Vocational Technology College.

References

[1] Klyne G, Carroll J eds 2004 Resource Description Framework (RDF): Concepts and Abstract Syntax *W3C Recommendation* 10 Feb 2004 http://www.w3.org/TR/rdf-concepts/
[2] Hu W, Qu Y 2007 Discovering simple mappings between relational database schemas and ontologies *Proceedings of ISWC/ASWC2007* Busan, South Korea **4825** 225-38
[3] Xu Z, Dong Y, Lu Y 2006 Semantic-preserving translation from ER schema to OWL DL ontology *Chinese Journal of Computers* **29**(10) 1786-96 *(in Chinese)*
[4] Bizer C, Cyganiak R 2006 D2R Server-Publishing Relational Databases on the Semantic Web *Poster at the 5th International Semantic Web Conference (ISWC2006)* Athens GA USA
[5] Perez A G, Benjamins V R 1999 Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods *Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods* Stockholm Sweden 1-15
[6] Baader D, Calvanese D, McGuinness D L, Nardi D, Patel-Schneider P F 2003 The Description Logic Handbook: Theory, Implementation, and Applications *Cambridge Cambridge University Press* 1-3
[7] McGuinness D L, van Harmelen F 2004 OWL Web Ontology Language Overview W3C Recommendation 10 February 2004 http://www.w3.org/TR/owl-features/
[8] Sequeda J F, Tirmizi S H, Corcho O, Miranker D P 2011 Survey of Directly Mapping SQL Databases to the Semantic Web *The Knowledge Engineering Review* **26** 445-86
[9] Das S, Sundara S, Cyganiak R 2012 R2RML: RDB to RDF Mapping Language *W3C Recommendation* 27 September 2012 http://www.w3.org/TR/r2rml/
[10] Zhang H, Lin L, Qiang B 2012 A Semantic Wrapper of Deep Web Data Source Based on RDF View *Journal of Computational Information Systems* **8**(9) 3527-39
[11] Prud'hommeaux E, Seaborne A 2008 SPARQL Query Language for *RDF W3C Recommendation* 15 January 2008 http://www.w3.org/TR/rdf-sparql-query/

Author

Haifei Zhang, born in November, 1980, Nantong, China

Current position, grades: lecturer of Mechanical and Electrical Engineering School of Nantong Textile Vocational Technology College. PhD candidate at Hohai University.
University studies: MS degree in Software Engineering from Soochow University in 2005.
Scientific interests: databases, GIS, ontological engineering, semantic web and web intelligence.
Publications: 1 patent, 2 software copyrights, 18 papers.

The study and design on the campus network surveillance

Xu Bing*, Yizhi Zhang

Chongqing Three Gorges University, Wanzhou Chongqing, China

Received 1 August 2014, www.cmnt.lv

Abstract

In order of improving the efficiency of network administration, one of the most useful methods is to inspect and measure of the network traffic. By introducing technology relevant to network flow monitor and analysing monitor system of campus network flow, this paper brings forth collection of Campus network flow and statistic of network flow, adopting Visual C++ 6.0 technology to design structure of this plan. This paper also points the key technology and means to realize monitor system of campus network flow and monitor system of campus network flow. It is reliable and extensible to for realization of system to improve administrative function. It is easy to realize the function of collection of campus network flow and flow statistic.

Keywords: network monitor, network flow, collection of flow, data statistic, VC++ 6.0

1 Introduction

With the rapid development and wide application of computer network, network has been interfered into various fields of social life. Similarly, as an important infrastructure establishment so as to innovate education in high colleges, improve managing level and educational quality, campus area network has been elevated correspondingly. However, the secure problems consisted in its operation and management has become prominent. Therefore how to establish sophisticated secure system of campus area network has become a crucial problem to be faced and tacked by network managers.

Network will determine our life style; the quality of network will have direct influence on various respects of social and economic life. With the increase of users' demand on network property, stable operation and efficient development of network cannot be guaranteed without efficient managing system and network system. With the rapid development of infrastructure technology of constructive network and network application and increase of demand on network property, network management has become an issue to be solved immediately. Efficient network management can guarantee stable operation and lasting development of network. What is more important, due to enlargement of network scale and development of hacker technology, the number of cases about invasion and attack become large, which takes challenge to stable network service and information security and internet rule, so internet security has become an important role in the whole system of internet management. At present, P2P, mainstream media and network game and some new application has occupied above 60 percent of network flow [1], and hostile attack on network has become more and more. Thus identification of network flow is quite important, which is helpful for network manager to timing supervise and manage various business flows and for supplier of network service to understand condition of each business flow of network when lay outing and constructing and for internet researchers to understand character of each flow of network and correspondent users' action. Therefore supervision on network flow has attached wider attention from academic and applicative fields.

2 Adopting relevant technology

2.1 NETFLOW

NETFOLW is a network exchanging technology stipulated by CISCO Company [2], which classify and handle network flow by concept of data flow. The definition of data flow given by NETFPLOW is consisted with seven key words marked by one- way network data assemblage of two communication terminals. The seven keywords are original IP, target IP, original terminal, target terminal, negotiation type, service type and interface search. In the case of traditional network interchange, each input subgroup is handled separately, then router will take series of independent consultation for each subgroup and check visiting list, obtain charging data and exchange subgroup by resort of series of functions, after that send {exchange}it to destination. These checks include whether to adopt secure visit to infiltrate and renewing network charging record. On contrary, in terms of NETFLOW exchange, in the process of consultation, it only works on the first subgroup of the first data flow. When a data flow is identified and determined its relevant service, the later subgroups will be treated as a part of this data flow and handle it on the base of linkage. In this case, it can avoid check on visiting list and can exchange subgroups in turn.

* ***Corresponding author*** e-mail:xb_2316@163.com

Besides exchange of data, important applications of NETFLOW are count information of data flow and supervise network flow. The information of data flow counted by NETFLOW is sent to address of preset information collector in the form of UDP data packet. Information collector will receive NETFLOW data packet constantly, then analyze various flow information.

The applicative model of NETFLOW (illustrated by Figure 1), NETFLOW exchange or route equipment analyze and count flow, then send information to information collector. NETFLOW information of various information collectors will be assembled to NETFLOW SERVER, after that it will handle flow information in accordance with need of application.

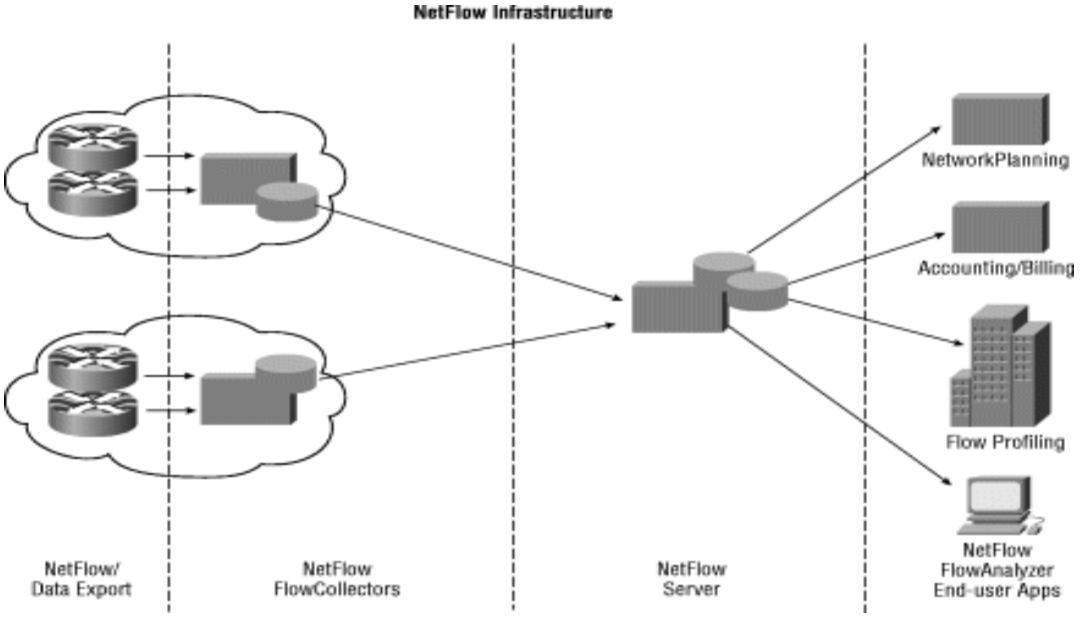


FIGURE 1 Netflow application structure diagram

2.2 PROGRAMMING INTERFACE OF TRANSPORT LAYER (WINDOWS SOCKET PROGRAMMING TECHNOLOGY)

2.2.1 The concept of Window socket

Window socket - SOCKET, as a foundation of network communication, which is suit of standard API supporting network negotiation data communication for exploration of Windows system. That is TCP/IP network programming interface. In 1994 after set as network programming standard, it experienced Winsock 1.1 and Winsock 2.0, with the final aim to adapt to the demand of explorer of applicative programmer and network servicer, and the process of communication is completed by socket communication domain. Socket is charge of input and output of data in transport layer and network layer so that shield data link layer and physic layer.

2.2.2 Type of socket

According to the nature of communication, the socket can be divided into stream socket and data gram socket. Their differences mainly are that stream socket offers bidirectional, regular, unrepeated data flow service. By contrast with data gram socket, the expense of its system is high. Data gram socket also supports bidirectional data flow ignoring reliability of transportation, unrepeated and regulation. However, it retains record delimitation. Owing to high efficient transportation of data gram, it is still used widely.

2.2.3 Sequence of byte, block and unblock

In terms of sequence of byte, different computer will adopt their own sequence to deposit data, thus when these data are used to communicated, sequence of byte need to be transferred. IP addresses and terminal numbers of network sockets functions should be arranged according to network sequence constructed by sockaddr_in. To be more cautious, before applicative procedure constructs applicative procedure, users need to transfer system- unit series number to network sequence (use htons function, on the contrary ntohs function).

Block and unblock, socket possesses synchronization block and a synchronization; in the case of block model, socket only to be finished when operation is accomplished; when socket are in unblock model, socket is marked by whether new data is up to block. Block model socket is simple, convenient, but is inefficient, by contrast, unblock model is complicated by efficient. To be mentioned, Winsock provides several I/O model to solve a synchronization issue, such as “selection”, “overlap”, “event selection”, “async selection” and so on.

3 Designing goal

Supervision of network flow and analysis tool mainly aims to timing collect network data constantly and do statistic

to obtain nature index of main opponents. Combined with theory of network flow, we can observe network situation by counted nature index and analyse changing tendency of network and find factors affecting network nature. It will realize following functions:

1) Adopting Winsock to compile original socket, socket- Raw collects data packet.

2) Analysing obtained data packets

3) Visiting parameter interface of network nature offered by operative system, obtaining total amount of flow of network card, input flow and output flow;

4) System will provide various ways to show result, such as curve graph, list and so on; it mainly adopt curve to show.

5) Adopting IP to help API obtain statistic information of network.

6) Realizing part of common alarms to continue exploring function of alarm.

4 Realization of system of network flow supervision

4.1 REALIZATION OF OBTAINING ORIGINAL DATA PACKETS

Obtaining system of data packet in network mainly relies on operative system; there are different realizing ways in different operative systems. In Window environment, we can realize the obtaining function of network data packet by network driver interface specification (NDIS), SOCK-RAW of WinSock or driving technology of virtual equipment and so on.

According to above account, original socket can avoid functions suffered by Socket, use and explore bottom negotiation, and generate needed data sheet in accordance with their own desire. I start to introduce some technical knowledge relevant to exploring catch data packet by use of original socket.

4.1.1 Before using socket, we need to understand working principle of receiving data of network card

In normal condition, network interface only correspond two kinds of data frame, one kind is just match their hardware, the other one is applied to all computer broadcast. In system, the send and reception of data frame is completed by network card. Network procedure receives data packet sent by network, and judges whether to match with hardware by hardware address. If it works, it will take notice CPU to stop to respond, then it calls network card set by driver procedure to stop procedure address to call driver procedure to receive data, after all put it into stack to do systematic handle, if not, this data packet will be discarded directly.

In the case of network interface, it possesses four receiving models of data that are broadcast, multicast, and direct and mix. Only when we set interface as mixed model, network interface can receive all data, ignoring whether addresses match, thus only we set mixed model when designing, and we can realize collection of data.

4.1.2 Working procedure and usage of socket

Generally speaking, exploring network procedure by adopting socket will be experienced by following steps: start, establishment, binding, monitor (receiving connection), connection, sending/receiving data, and close, unload and so on.

4.1.3 Designing how to obtain network data by means of original socket in windows

1) Starting socket;
2) Establishing a original socket;
3) Binding socket to local address
4) Setting operative parameter
5) Setting network interface as mixed model
6) Starting monitoring thread to receiving data
7) Quitting and closing socket.

4.2 KEY FUNCTIONS OBTAINED BY ORIGINAL DATA PACKET

4.2.1 Start function WSAS startup

int PASCAL FAR WSAStartup (DWORD wVersionRequested, LPWSADATA lpWSAData).

Each socket applicative procedure must do series of initialization work by calling this function, and this socket only can be used when call is accomplished and returned. Parameter wVersionRequested among them is the version number, high byte is the minor version, low byte is the major version, and parameter lpWSAData is the guiding principle of WSADATA structure.

4.2.2 Socket establishes function socket

SOCKET socket (int af, int type, int protocol).

All communications must be established on starting up a socket, the function of socket function is establishing socket, parameter af among them refers to address family. When socket established based on UDP or TCP, we need set af as AF_ INET and adopt IP negotiation. Function type is one type of negotiation socket. When we adopt stream socket, we use SOCK_ STREAM. When we adopt data gram socket, we use SOCK_DGRAM. When adopting original socket, we use SOCK_ RAW. Function protocol can set as 0 in the situation of default.

4.2.3 binding function bind

int bind (SOCKET s, struct sockaddr_in name, int namelen).*

The next step is bind local network interface with socket after establishing socket, function s is the established socket, parameter name is the guide of information structure of communicative object needed to be bound, namelen is the length of this structure. We should pay more attention to structure of sockaddr_in:

```
struct sockaddr_in{
short sin_family;    / address family, set as AF_INET
unsigned short  sin_port;      //pointed terminal number
struct in_addr   sin_addr;     //IPaddress
```

char sin_zero
};[8].

Due to relationship between System- unit serial number and network number, in the procedure we need to use htons and some functions to transfer.

4.2.4 Setting function WSAIoctl of interface model

int WSAAPI WSAIoctl(SOCKET s, DWORD dwIoControlCode, LPVOID lpvInBuffer, DWORD cbInBuffer, LPVOID lpvOutBuffer, DWORD cbOutBuffer, LPDWORD lpcbBytesReturned, LPWSAOVERLAPPED lpOverlapped, LPWSAOVERLAPPED_COMPLETION_ROUTINE lpCompletionRoutine).

s is a handle of socket, dwIoControlCode is the code of operative control, lpvInBuffer is the address of input block, cbInBuffer is the size of input block, lpvOutBufferis the address of output block, cbOutBuffer is the size of output block, lpcbBytesReturned is the address of the number of practical output byte, lpOverlapped is the address of structure of WSAOVERLAPPED, lpCompletionRoutine is a called routine guide after finishing guiding operation.

After finishing call, WSAIoctl return to 0, or will return to INVALID_SOCKET, applicative procedure can obtain wrong code by WSAGetLastError [5]

4.2.5 Recv

int recv (SOCKET s, char buf, int len, int flags);*

4.3 DATA STATISTIC

By use of API function in IP assistant of software, network manager can, to some extent, find bottom neck of network nature by statistic data. Main relevant functions are:

GetUdpStatistic, GetTcpStatistic, GetIcmpStatistic, GetIStatistic, we should pay more attention to IPHelpapi.lib needed to load in the project. The result of function calling can show directly by list that network managers can supervise network nature by observing change of statistic, the design of statistic interface is following.

5 Conclusions

On the basis of network flow supervision system, this essay introduces in detail main technology of network flow supervision, collection of campus network flow and statistic of network flow. This system can improve reliability and flexibility of flow supervision by measure of VC++ 6.0. In the light of operation test of campus network shows that system works stably to basically satisfy designing requirement, setting foundation of research and realization of exploring further campus network flow supervision system.

Acknowledgment

It is a project supported by the scientific fund assistance project from Chongqing Education Committee (NO: KJ1401027).

References

[1] Sen S, Wang J 2002 Analyzing P2P traffic across large networks *Proceedings of the Second SIGCOMM Internet Measurement Workshop (IMW2002)* Marseille France
[2] White Paper NetFlow Services and Applications http://www.cisco.com
[3] Liang W, Li H 2008 Research on Ways of Identifying Network Flow *Communication Technology* **41**(11) 88-90 *(in Chinese)*
[4] Yao Q, Ma H, Zhang G 2006 Flow Forecast based on self-adaptation a network managing software *Computer Measure and Control* **14**(1) 45-6
[5] Li F, Li A, Wu C 2008 Application of Advanced Arithmetic of Network to Supervision of Ultrasonic flow *Computer Measure and Control* **16**(2) 163-5
[6] Zhao G, Ji C, Xu C 2010 Research on Identified Technology of Internet Flow *Microsoft Computer system* **31**(8) 1514-20 *(in Chinese)*

Authors

Xu Bing, born in February, 1976, Chongqing, China

Current position, grades: School of Computer Science and Engineering, Associate professor and master supervisor in Chongqing Three Gorges University.
University studies: Computer Science and Engineering in Chongqing University.
Publications: 5 patents, 45 papers.
Scientific interests: computer networks and applications.

Yizhi Zhang, born in March, 1963, Dazhou, Sichuan, China

Current position, grades: Network information Center. Professor and master supervisor in Chongqing Three Gorges University.
University studies: Computer Science and Engineering in Zhejiang University.
Publications: 5 patents, 23 papers
Scientific interests: computer networks and applications.

A self-aware strategy for virtual machines placement on clouds

Fen Guo[1*], Huaqing Min[1], Ming Yin[2]

[1]*School of Software Engineering, South China University of Technology, Guangzhou Higher Education mega center, 510006, Guangzhou, China*

[2]*School of Automation, Guangdong University of technology, Guangzhou Higher Education mega center, 510006, Guangzhou, China*

Received 1 August 2014, www.cmnt.lv

Abstract

Cloud computing is a new computing service mode, and virtualization is a key technology of it. A self-aware strategy (SAST) for Virtual machines (VMs) management on clouds is proposed which is multi-attributed weighted on the resources. It manages the virtual resource basing on the requests of users and the real-time state of the system dynamically. It consists of three phases: (1) monitoring the cloud performance including VMs and Physical Machines (PMs), with the data standardized; (2) measuring the cloud load balance value with the attribute weighted measurement model; (3) using the placement algorithm to choose the best appropriate PM to place the VM requested. The main contribution of the paper is that a cloud load balance measurement model is introduced and a VM scheduling strategy is proposed which includes the VMs placement optimization algorithm and the VMs dynamic migration algorithm. The SAST is tested on the simulation platform comparing with other traditional ones. As a result, we concluded that it guaranteed the SLA and achieved better load balance of cloud. And at the same time, it minimized the number of the started PMs on clouds to reduce energy consumption.

Keywords: cloud computing, virtualization, placement, scheduling

1 Introduction

Cloud computing [1] is considered a new computing service model and has got a lot of attentions. Cloud computing provides the infinite shared resources to the customers via the Internet. The customers would get them with not understanding the fundamental technology and the application running environment. The more quickly the demand of cloud computing grow, the larger data centre is needed. As most of all know, the user requirement for cloud platform is often heterogeneous and irrelevant, so the unreasonable resource distribution will lead to the waste of resources inevitably. In order to improve the utilization rate of resources, cloud computing platform also needs the dynamically balanced load of various kinds of services. On the other hand, large-scale computing infrastructure consumes a lot of power resources, and the power consumption increased year by year. FORREST W [2] has predicted that the data centre power consumption would reach 2% of the world's total energy consumption in 2020. So, how to dynamically and effectively manage the cloud computing platform resource becomes a key problem.

Virtualization technology [3, 4] provides an effective method to manage the resources of the cloud computing platform dynamically, and it has broken the tight coupling between the computing and the hardware. Customers use Virtual Machines (VMs), based on SLA; cloud providers take advantages of VM's flexible management on PMs to optimize resources allocation so as to meet customers' requests. Server virtualization technology enables multiple VMs running on a physical node at the same time. It has greatly improved the utilization of computing resources and implemented the on-demand deployment [5, 6]. In addition, the VM migration technology makes that cloud computing platform can be dynamically adjusted to deploy the VM to the less physical machines according to the changes of the service load, and as a result, it will achieve energy saving.

Since different resource utilization is caused by different mapping between VMs and PMs, for cloud providers, the key issue is how to effectively manage and schedule VMs to meet the customer's requests, and at the same time reduce the energy consumptions to minimize the cost. Nowadays, the algorithm for building the mapping from the customer's requesting VMs sets to the servers in the resource pool is becoming a hot issue. The algorithm will choose the most appropriate PM as the host for the requested VM and establish the specific mapping.

The number of the VMs requested on cloud is increasing with the development of cloud computing, and then the deployment of VMs becomes more important. A VM will be bound a PM for a lifelong under simple managing strategy, which will lead to a load imbalance. In the same way, the managing and scheduling strategy based on single attribute will also cause uneven load, such as a VM with demand for network could be deployed in a PM with sufficient CPU residue but poor network resource, resulting the VMs on the same PM competing for bandwidth. At the same time, the quantity and load of VMs and PMs will vary over time with the demand of the customer and application, static management of VMs will cause waste or shortage of resources, and the artificial scheduling is an obvious lag one.

[*]***Corresponding author*** e-mail: csguofen@scut.edu.cn

For VM management, one major direction is to minimize the number of the PMs for reducing energy consumption. Such problems are usually to be interpreted as packing problem [7-9]. These studies are usually only consider how to reduce the number of physical, but does not take into account the needs of customers (for example, now more and more customers have specified requirements about the network).What's more, it will lead SLA violation. Some studies [10-22] considered the needs of customers and applications. However, the optimization of load balancing is less concerned, and some studies do not consider the dynamic performance of the cloud. In view of this, the VMs management strategy should not only consider the utilization rate of resources but also the overall load balancing combining with the dynamic demand of customers.

Therefore, our paper presents a dynamic managing strategy for VMs on cloud. It is concerned chiefly with meeting the customer requirement, keeping the system load balancing and saving power. With meeting the constraints of PM resources(CPU, network, etc.), the strategy managed the VMs on PMs to achieve cloud's load balancing for reducing SLA violation, which allows the idle PMs in a sleeping state so as to reduce energy consumption on cloud finally.

Based on the above, we proposed a self-aware strategy (SAST) for VMs management on clouds which is called as SAST. The strategy is based on the user requests of virtual resources and the real-time system status. And it is composed of three phases: Firstly, to monitor the cloud load performance including VMs and physical machines, with the data standardized; secondly, to measure the cloud load balancing value with the proposed measurement model. Thirdly, to use the managing and deploying algorithm to choose the best appropriate PM to deploy the VM requested. In addition, the batch requests are treated in accordance with the single application process in order in SAST.

Our paper is divided into six major sections as follows. Section 1 introduces the related work. Section 2 opens with the description of the SAST. Section 3 shows the load balance measurement method. Section 4 introduces the main scheduling algorithm, while section 5 provides experimental result and analysis, and section 6 draws a conclusion.

2 Related work

There are two focuses on VM managing strategy for reducing energy consumptions. One is to consider how to place VM to the PMs in cloud. Eucalyptus [10, 11] has proposed round robin and greed algorithm to deploy the VMs. While OpenStack [12] takes a random scheduling strategy as the default one. However, the match-making scheduler [13] of OpenNebula achieved the ranking algorithm. Varma et al. [14] dynamically readjust server's location and consider the cost of migration and energy, with a simple algorithm; it shows that dynamic migration technology realizes low energy cost. Norman Bobroff et al. in [15] put forward a VM deployment algorithm using the forecasting techniques and heuristic algorithm, and ensured the SLA with minimizing the physical machine number. Singh et al. formed the question into a multi-dimensional knapsack problem, and treated the constraints of the deployment as a separate dimension [16]. Tsakalozos [17] used a two-phase mechanism for mapping problem of VM for large heterogeneous infrastructure: First, to synthesize the physical machine set which can be used at this stage; second, to determine the approximate optimal VM-to-PM mapping with satisfying constraints. Breitgrand et al. formalized the problem to be a multi-unit combinatorial auction model in [18]. Zhang Wei et al. of [19] proposed a kind of strategy for deploying and scheduling VMs based on multiple attributes analysis for the uneven loads between physical machines in the cloud computing.

The strategies or algorithm above consider the energy consumption, and they neglect the user requests and the load balancing of the cloud. The other type [20-22] is to consider the dynamic factor of the cloud. Nik Bessis [20] explored two configurations, the static case in which VMs are generated according to the cloud orchestration, and the dynamic case in which VMs are reactively adapted according to the job submissions, using migration, for optimizing performance time metrics. Yang Xing [21] proposed the Performance Matching-Load Balancing (PM-LB) algorithm of VM deployment. In [22] Zhao Hongwei put forward a kind of efficient resource management strategy based on the domain. Studies above covered various aspects, and they emphasized with the merchant's profits or only one single aspects of the optimization. But the study didn't consider how to achieve best load balancing. So they are also different from our objective.

3 Description of the SAST

3.1. SYSTEM ARCHITECTURE

The framework of the SAST is shown in Figure 1. It includes: central controller; VMs scheduling strategy generator, cloud monitor, VMs placement module.

The central controller is responsible for the overall system, including receiving the requests from the client layer, sending alarm information from the monitor and sending command to the VMs scheduling generator, etc.

The scheduling generator would build the load balance model of the cloud, and judge whether the node is overloaded or low load by analyzing the real-time data from the monitor. And then, it will calculate the best VMs placement scheme using the proposed algorithm in SAST combining with the requests of the alarm queue and the user's optimizing request queue.

The monitor will receive the load performance of each VM and PM periodically, the requests of the migration, the usage of the storage and the network of the cloud. If it found a PM load had exceeded the specified threshold, it would alarm. And at the same time, it would add this PM

into the alarm queue, and informed the central controller triggering the VMs scheduling generator to deal with it.

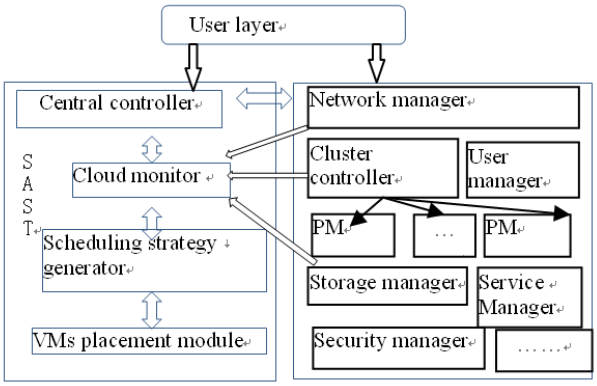


FIGURE 1 The system architecture

3.2 THE OBJECTIVE

The details of our objective are as follows:

Meet the existing customer's request, and maximize the load balancing degree on each physical server scheduling domain.

Minimize the number of PMs to start on IaaS on the basis of 2 to save energy.

Control the load rate of each PM in the scheduling domain not higher than the threshold

3.3 MONITORING STRATEGY

The traditional distributed systems often use the passive monitoring strategy. The node does not only need to exchange information between each other, but also report the status to the information centre regularly. Consequently, it will affect the overall performance of the platform.

1. The monitoring contents are shown in Table 1.

TABLE 1 The monitoring contents

CPU	Memory	Hard disk	Network
1-usage	1-usage	1-usage	1-usage

2. The sampling method.

The average sampling method is used to collect the monitoring data. The monitor collects every node's (including PMs and VMs) load data from the system for every *T* seconds, and defines the average value of the last N data as the load value of the node, which is denoted as $X = \sum_{i=1}^{N} X_i / N$. The value of *T* and *N* can be dynamically set by the monitor according to the individual request, the default value of *T* is 10 seconds, and the *N* is 5.

3. The monitor saves the latest load data vector $\boldsymbol{S}_x$ as the real-time load information of each *X* as follows(*X* can be a VM or a PM).

$$\boldsymbol{S}_x = (C_X, M_X, H_X, N_X). \quad (1)$$

C_x, M_x, H_x, N_x, respectively represents the monitoring data of CPU usage, memory usage, hard disk usage and bandwidth usage of the *X*.

3.4 SCHEDULING STRATEGY

The scheduling strategy is the core of the SAST, and it will be detailed in the below. This scheduling strategy combined with users' requirements to overcome the shortcomings of existing technology:

1) A cloud platform load balance measurement model is introduced with setting the weigh vector and matrix referring to the customer requirements and the monitoring data.

2) A VM scheduling algorithm is proposed, which includes the VM deployment optimization strategy and the VM dynamic migration strategy.

4 Load balance value measurement method

4.1 DEFINITION AND LEMMA

Suppose the number of the PMs in the scheduling domain is *m*, and the one of the VM types is *t.*

Definition 1. Let *P* stands for the PMs Set. *P* is defined as follows.

$$\mathrm{P} = \left\{ \mathrm{G}_i | \mathrm{i} \in (0, n) \, and \, \mathrm{G}_i P \right\} ,$$

$$\mathrm{G}_i = \left\{ p_{i1}, p_{i2}, ..., p_{ij} | \mathrm{j} = \mathrm{Num}(\mathrm{G}_i) \, \mathrm{and} \sum_{i=1}^{n} Num(G_i) = \mathrm{m} \right\},$$

$$p_i = \{SC_i, SM_i, SH_i, SN_i\} .$$

G_i stands for the i^{th} group on cloud, and p_i stands for the configure vector of the PM *i*. SC_i, SM_i, SH_i, SN_i respectively stands for the CPU size, the memory size, the disk size and the bandwidth size of P_i.

Definition 2. Let VT stands for the VM type's set. VT is defined as follows.

$$\mathrm{VT} = \{vt_1, vt_2, vt_3, \ldots \ldots, vt_t\}),$$

$$vt_i = \{sc_i, sm_i, sh_i, sn_i).$$

vt_i stands for the configure vector of the VM of type *i*, and sc_i, sm_i, sh_i, sn_i respectively stands for the CPU size, the memory size, the disk size and the bandwidth size of vt_i.

Definition 3. $v_i = \{sc_i, sm_i, sh_i, sn_i)$ can be deployed on $p_i = \{SC_i, SM_i, SH_i, SN_i\}$ only when

$$sc_i \le SC_i \ ; sm_i \le SM_i; sh_i \le SH_i; sn_i \le SN_i.$$

Definition 4. $p_j \le p_i \ (i, j \in (0, m])$

$$\text{s.t. } SC_j \le SC_i \ ; SM_j \le SM_i; SH_j \le SH_i; SN_j \le SN_i$$

Definition 5. Let V be the set of the VMs vectors which are deployed in cloud.

$$\mathrm{V}=\left\{\begin{matrix}V_1,V_2,\ldots\ldots,V_m|V_j=\{v_{j1},v_{j2},\ldots\ldots,v_{jk}\},\\ k\in N\ and\ \ j\in[1,\ \ m]\end{matrix}\right\}$$

Definition 6. V_j represents the set of the VMs on the PM p_j, the v_{ji} represents the i^{th} VM on the PM p_j.

Let W_v be the pre-supposed weighted matrix of the VMs, W_p be the pre-supposed weighted matrix of the PMs. The W_v and W_p are stored in the VM scheduler as follows.

$$\boldsymbol{W}_v=\begin{bmatrix}wv_{11} & wv_{12} & \ldots & wv_{14}\\ wv_{21} & wv_{22} & \ldots & wv_{24}\\ \ldots & \ldots & \ldots & \ldots\\ wv_{t1} & wv_{t2} & \ldots & wv_{t4}\end{bmatrix}$$

$$\boldsymbol{W}_p=\begin{bmatrix}wp_{11} & wp_{12} & \ldots & wp_{14}\\ wp_{21} & wp_{22} & \ldots & wp_{24}\\ \ldots & \ldots & \ldots & \ldots\\ wp_{m1} & wp_{m2} & \ldots & wp_{m4}\end{bmatrix}$$

$$\mathrm{W_{v_s}}=(\mathrm{wv_{s1}},\mathrm{wv_{s2}},\mathrm{wv_{s3}},\mathrm{wv_{s4}})$$

$$\mathrm{W_{p_j}}=(\mathrm{wp_{j1}},\mathrm{wp_{j2}},\mathrm{wp_{j3}},\mathrm{wp_{j4}})$$

$$\mathrm{wv_{s1}}+\mathrm{wv_{s2}}+\mathrm{wv_{s3}}+\mathrm{wv_{s4}}=1$$

$$\mathrm{wp_{j1}}+\mathrm{wp_{j2}}+\mathrm{wp_{j3}}+\mathrm{wp_{j4}}=1$$

$\boldsymbol{W}_{v_s}$ stands for the weighted vector of v_s, while the $\boldsymbol{W}_{p_j}$ stands for the weighted vector of p_j. And $wv_{s1},wv_{s2},wv_{s3},wv_{s4}$ respectively represents the pre-supposed weight of CPU, memory, hard disk, network for the v_s, while $wp_{j1},wp_{j2},wp_{j3},wp_{j4}$ has the similar meaning. So we can use $\boldsymbol{W}_x$ (X can be a VM or a PM) to represent the performance weight vector of X.

4.2 LOAD VALUE MEASUREMENT MODEL

Let $bl(j)$(j∈[1,m]) be the load value of the PM p_j, O as the load balance value of the scheduling domain. Function $f_n(j)$ returns the number of the VMs deployed on the PM p_j, while $H_p(p_j)$ represents the load value of PM p_j, and $H_v(j,i)$ stands for the load value of the VM v_{ji}. The performance weighted strategy calculates the load value according to the following Equations. σ is a systematic parameter which can be set by the system administrator.

$$H(x)=\boldsymbol{W}_x\times\boldsymbol{S}_x, \tag{2}$$

$$bl(j)=\sigma\times H_p(p_j)+(1-\sigma)\times\sum_{i=1}^{f_n(j)}H_v(j,i)/\ f_n(j). \tag{3}$$

$$O=\sqrt{\sum_{j=1}^{m}\left(bl(j)-|bl|\right)^2}/m\ ,$$

$$\sigma\in(0,1)\ ,\qquad\left(|bl|=\sum_{i=1}^{m}bl(i)/m\right). \tag{4}$$

Lemma 1. Based on the definitions above, we can deduce that the smaller the O is, the higher the cloud load balance degree is.

Lemma 2. Based on the configure set of PMs and VMs, we can get the VM number matrix $\boldsymbol{R}$, while the r_{ij} represents the max number that the vt_i can be deployed on the p_j. The detail is as follows:

$$\boldsymbol{R}=\begin{bmatrix}r_{11} & r_{12} & \ldots & r_{1m}\\ r_{21} & r_{22} & \ldots & r_{2m}\\ \ldots & \ldots & \ldots & \ldots\\ r_{t1} & r_{t2} & \ldots & r_{tm}\end{bmatrix}.$$

$$r_{ij}=min\{l_1,\ l_2,\ l_3,\ l_4\}\ (\ i\in[1,\ t],\ j\in[1,\ m]).$$

$$l_1=\ SC_{p_j}/sc_{v_i},l_2=\ SM_{p_j}/sm_{v_i}.$$

$$l_3=\ SH_{p_j}/sh_{v_i}\ ,l_4=\ SN_{p_j}/sn_{v_i}.$$

Lemma 3. When a new VM q_x is requested whose type number is be k, we can find the corresponding weighted vector in W_v. Suppose $H'_v(v_x)$ returns the presupposed load value for q_x, while $H'_p(p_j)$ returns the pre-supposed load value of the PM p_j if the v_x is deployed on it, and $\boldsymbol{S'}_{p_j}$ returns the pre-supposed load data vector of the PM p_j. Then we can deduce the following Equation.

$$\boldsymbol{S'}_{p_j}\approx\boldsymbol{S}_{p_j}+1/r_{kj}\ ,$$

$$H'_p(p_j)=\boldsymbol{W}_{p_j}\times\boldsymbol{S'}_{p_j}\approx H_p(p_j)+\boldsymbol{W}_{p_j}\times100/r_{jk}, \tag{5}$$

$$H'_v(v_x)=\sum_{i=1}^{f_n(j)}H_v(j,i)/f_n(j). \tag{6}$$

Lemma 4. Given $bl'(j)$ returns the pre-supposed load value of the PM p_j. Then we can deduce it as follows basing on Equation (5) and (6).

$$bl'(j)=\sigma\times H'_p(p_j)+$$
$$1-\sigma\cdot\left(\sum_{i=1}^{f_n(j)}H_v(j,i)+H'_v(v_x)\right)/\left(f_n(j)+1\right)\ . \tag{7}$$

Lemma 5. Suppose $\Delta bl(j)$ as the changes in the load values of p_j if the v_x is deployed on itself, then we can deduce as follows.

$$\Delta bl(j)=bl'(j)-bl(j)$$
$$=\sigma\times H_p(p_j)+\sigma\times100\times\overline{W_{p_j}}/r_{jk}+\ . \tag{8}$$
$$(1-\sigma)\times\left(\sum_{i=1}^{f_n(j)}H_v(j,i)+H'_v(v_x)\right)/\left(f_n(j)+1\right)$$

When $f_n(j)$ is very great, we can deduced the below Equation.

$$\Delta bl(j) \approx \sigma W_{p_j} / r_{jk} + (1-\sigma) H'_v(v_x) / f_n(j). \quad (9)$$

Lemma 6. Suppose $\Delta O(j)$ as the changing value in the load balance value of the cloud scheduling domain if the v_x is deployed on itself, then we can deduce as follows.

$$|O'| = \left(\sum_{i=1 \text{ and } i \neq j}^{m} bl(i) + bl'(j) \right) / m, \quad (10)$$

$$\Delta O(j) = \sqrt{\sum_{i=1 \text{ and } i \neq j}^{m} \left(bl(i) - |O'|\right)^2 + \left(bl'(j) - |O'|\right)^2 / m - O}. \quad (11)$$

5 Scheduling algorithm

5.1 ALGORITHM SUMMARY

Suppose there are n VMs which the user layer has requested, while the request set is called Q:

$$Q = \{q_1, q_2, q_3, \ldots\ldots, q_n\} | qi \in VT.$$

Then the main work of the deploying and scheduling for VMs is to find the mapping f:

$$f: Q \xrightarrow[q_i \longrightarrow p_j]{deployed} P \quad (q_i \text{ is deployed on } p_j).$$

The details are as follows.

Adding all the alarm tasks into the alarm queue Q_R, and invoking the alert processing module to process each PM in queue sequentially.

Adding the user-specified optimization tasks into the optimization queue Q_O, and calling the optimization module to optimize each PM in queue.

Adding new requested task into the new task queue Q_N, and calling the VM deployment scheduling module to treat each request in order.

5.2 PREREQUISITES AND RESTRICTIONS

Never deny the user's request ($v = \{sc, sm, sh, sn\}$) except when:

$c \leq \sum_{i=1}^{m} SC_i$ or $sm \leq \sum_{i=1}^{m} SM_i$

or $sh \leq \sum_{i=1}^{m} SH_i$ or $sn \leq \sum_{i=1}^{m} SN_i$.

Only allow the user to request one of the VM types among the given VT.

Each PM and VM can be monitored, and when there was a PM or VM load had exceeded the specified threshold, the monitor would alarm.

The first group G_1 is the default initial scheduling domain on the cloud.

5.3 THE DETAILS OF THE ALGORITHM

5.3.1The main process

Input:

$P = \{G_i \mid i \in (0, n) \text{ and } G_i P$

$P = \{p_1, p_2, \ldots, p_m \mid m \in N\}$

1) V, W_v, R, W_p;
2) $S = \{S_x \mid x \in P \text{ or } x \in V\}$;
3) The queues Q_R, Q_O, Q_N.

Output: $f: Q_N \xrightarrow[q_i \longrightarrow p_j]{deployed} P$; finish status.

Steps

```
1)  while(QR!=NULL)
    {
    q=QR->head;
     if (dynamicsheduler (q->id, q->threshold, (P-QR))= = SUCCESS)//
process PM q
     {QR->head=q->head; delete (q);}
     else
     go to 4);
     }
2)  while(QO !=NULL)
    {     q=Qo->head;
      if (dynamicsheduler (q->id, S, (P-Qo))= =SUCCESS) // process
      PM q
        {Qo->head=q->head;
         delete(q);
         }
      else
      go to 4);
    }
3)  while(QN !=NULL)
    {q= QN->head;
     if (staticsheduler (q->id, P)= =0)
        // process PM q
       go to 4);
     else {QN->head=q->head;
        new p;
        p->id= staticsheduler(q-> id, P);
        insert(q ⟶ p) to f;
        Update(V, q ⟶ p);
        deploy(q,p);//deploy q on p
        delete(q);
        }
    }
4)  if (Num(the current Gi)= =m)
    //if all the PMs in the current scheduler domain
         return FAIL;
    else {
    start Gi+1;merge(Gi, Gi+1);
    }
```

5.3.2 The VM placement module

staticscheduler

Input: the requested new VM q; ()

Output: the id of the target PM (0 stands for failure)

Steps:

```
1)  for(i=1;i<=num(G);i++)
    {p=G(i);
     if(S(p) ≈0) return i;
    }
```

2) for(i=1;i<=num(G);i++)
if (G(i)>q) insert (G(i),Q_L);
{if(Q_L!=NULL)
{Scan Q_L orderly;
if (m> system threshold number)
//if m is very large
return the first j in $\{j|j \in \min_{j\in[1,num(Q_L)]} \Delta bl(j)\}$;
else
return the first j in $\{j|j \in \min_{j\in[1,num(Q_L)]} \Delta O(j)\}$;
}
else
return dynamicsheduler (q->id, the specific threshold, G);
}

5.3.3 The dynamic migration scheduling module

dynamicsheduler

Input: ID (PM id or new VM type number);
The default threshold of PM load; the current schedule domain G;
Output: FAIL or SUCCESS

Steps:

1) Make *P* to *P'*(*P'* is in descending order according to the available size of resources);
2) if the module is invoked by Q_O or Q_R
{delete p whose id = =ID from *P'*;
v=Max (V_{id}); //v is the max VM on p;
}
else v=vt_{id}; //v is the requested VM;
3) for(x=1;x<num(*P'*),x++)
{
if (there is *V'* on V_x which meet the following condition:
a); (V_x- *V'*)>*v*
b) for each $v' \in V'$
staticsheduler (*v'*->id, P')! =0
)
{
for each $v' \in V'$ deploy (*v'*, *P'*);
return SUCCESS;
}
else return FAIL;
}

6 Experimental result and analysis

This section shows that the SAST is feasible and exact. This experiment is simulated with Java programming language, using a set of simulated experiment set to compare the performance of the round robin algorithm, the best fit heuristic packing algorithm and the SAST.

6.1 CONFIGURATION

1. The configuration of the PMs and VMs is described in Table 2. The PMs consists of 10 nodes and 2 kinds of configuration. There are 6 PMs about node1, and 4 PMs about node 2; The VMs has two types, one is the high computing (HC), another is the high storage (HS). Each type is divided into big, medium and small class, so there are six different VM configurations. Clusters are divided into 10 groups, each group of only one PM. Each CPU has a core (Intel(R) Xeon (R) E5606). There are 10 PMs with two types of configuration.

TABLE 2 Experiment configuration

Type	CPU (G)	Memory (G)	Storage (GB)	Network (MB)
node1	16	96	20*1024	1000
node2	16	96	10*1024	1000
VM_A	8	8	200	100
VM_B	8	4	100	100
VM_C	8	2	50	100
VM_D	4	8	200	100
VM_E	4	4	100	100
VM_F	4	2	50	100

2. All strategies using the same set of monitoring data.

3. In each strategy the VMs are requested at random for 15 times.

4. When talking about round robin strategy, let's suppose that Wp=(0.25,0.25,0.25,0.25), and for SAST, σ=0.6,and the setting of Wv is shown as Table 3.

TABLE 3 W for six VMs

w	VM_A	VM_B	VM_C	VM_D	VM_E	VM_F
w_1	0.3	0.2	0.4	0.2	0.5	0.2
w_2	0.25	0.25	0.2	0.2	0.15	0.15
w_3	0.2	0.3	0.2	0.4	0.2	0.5
w_4	0.25	0.25	0.2	0.2	0.15	0.15

6.2 RESULTS AND ANALYSIS

The number of PMs started of the SAST is the same with the best fit heuristic packing algorithm at each application as shown in Figure 2, and the early number were less than the largest PM number. But when the system entered into a stable state, the load balance degree of cloud of the SAST is relatively high, particularly as shown in Figure 3. The abscissa represents the number of VMs, and the ordinate represents the value of load balance value of the cloud.

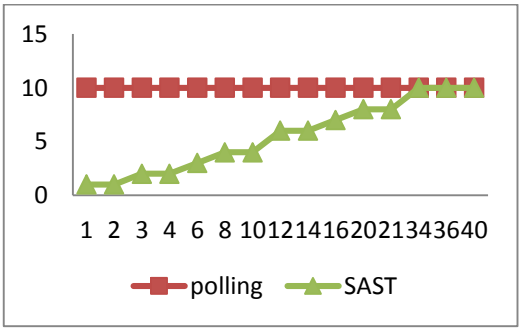


FIGURE 2 The PMs started

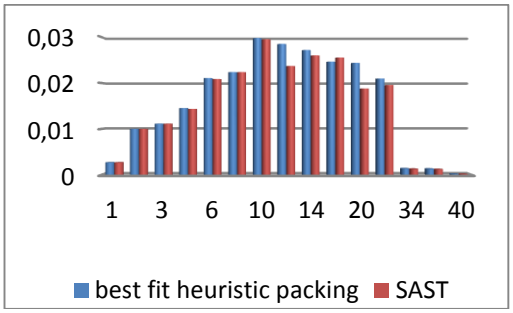


FIGURE 3 The load balance value

We let two strategies applied 20 VMs at random for 4 times, the total number of migration at the steady state of the cloud platform was compared and shown in Table 4.

TABLE 4 The total number of migration

Strategy	Random Equilibrium	SAST
the number of test 1	4	2
the number of test 2	4	3
the number of test 3	3	0
the number of test 4	5	2

7 Conclusions

Considering the user requests of virtual resources and setting the weight of performance parameters to calculate the load balance value of the cloud, we present a self-aware strategy for VMs placement on clouds. It achieved better load balance of cloud, and at the same time, it minimized the number of the started PMs on clouds to reduce energy consumption. However, the time complexity of the SAST is relatively high when it is compared with the round robin algorithm and other traditional algorithms. We will emphasize with the security on cloud in the future work.

Acknowledgments

This work is supported by the national science of nature fund project: The Computing Plat tern based on Formal Area (x2jsb55101680) and the Guangdong province science of nature fund research group project: Grid Theory Model and Analysis (x2jsb6110010).

References

[1] The NIST Definition of Cloud Computing *National Institute of Standards and Technology* 2011
[2] FORREST W 2008 How to cut data centre carbon emissions http://www.computerweekly.com/feature/How-to-cut-data-centre-carbon-emissions
[3] Goldberg R P 1974 *IEEE Computer* **7**(6) 34-45
[4] Barham P, Dragovic B, Fraser K, Hand S, Harris T, Ho A, Neugebauer R, Pratt I, Warfield A 2003 Xen and the art of virtualization *ACM SIGOPS Operating Systems Review* **37** 164-77
[5] Tian W, Zhao Y 2011 Cloud Computing Resource Scheduling Management *National defense science and technology university press*
[6] Hai J 2009 Computing System Virtualization - the Principle and Application *Tsinghua university press* (*in Chinese*)
[7] Verma A, Ahuja P, Neogi A 2008 Mapper: power and migration cost aware application placement in virtualized systems *Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware* 243-64
[8] Bobroff N, Kochut A, Beaty K 2007 Dynamic placement of virtual machines for managing sla violations *Integrated Network Management IM'07 &10th IFIP/IEEE International Symposium* 119-28
[9] Cardosa M, Korupolu M R, Singh A 2009 Shares and utilities based power consolidation in virtualized server environments *Integrated Network Management IM'09 & IFIP/IEEE International Symposium Papers* **31** 327-34
[10] Tan T X, Cameron K 2009 An Assessment of Eucalyptus Version 1.4 2009-929-07 Calgary *Department of Computer Science University of Calgary*
[11] Wikipedia. Eucalyptus. http://en.wikipedia.org/wiki/Eucalyptus
[12] Corradi A, Fanelli M, Foschini L VM consolidation: A real case based on OpenStack Cloud *Dipartimento di Elettronica, Informatica e Sistemistica (DEIS) University of Bologna Italy*
[13] Opennebula 2012 *Opennebula Scheduling Policies 2.0* http:// www.opennebula.org
[14] Aiash M, Mapp G, Gemikonakli O 2014 Secure live virtual machines migration: issues and solutions *Advanced Information Networking and Applications Workshops (WAINA) 28th International Conference* 160-5
[15] Bobroff N, Kochut A, Beaty K 2007 Dynamic Placement of VMs for Managing SLA Violations *Proceedings of 10th IFIP/IEEE International Symposium on Integrated Network Management papers* **118** 119-28
[16] Singh A, Korupolu M, Mohapatra D 2008 Server-Storage Virtualization: Integration and LoadBalancing in Data Centers *Proceeding of the 2008 ACM/IEEE conference on Supercomputing (SC'08) papers* **49** 1-12
[17] Tsakalozos K, Roussopoulos M, Delis A 2011 VM: placement in non-Homogeneous Iaas-clouds *Proceedings of the 9th international conference on Service-Oriented Computing* 172-87
[18] Breitgand D, Epstein A 2011 SLA-aware placement of multi-VM elastic services in compute clouds *Proceeding of IFIP/IEEE International Symposium on Integrated Network Management* 161-8
[19] Zhuang Wei, Gui Xiaolin, Lin Jiancai1, Wang Gang, Dai Min 2013 Deployment and scheduling of vms in cloud computing: an "AHP" approach *Journal Of Xi'An Jiaotong University* **47**(2) y1-y7
[20] Bessis N, Sotiriadis S, Xhafa F, Asimakopoulou E 2013 Cloud scheduling optimization: a reactive model to enable dynamic deployment of virtual machines instantiations *Journal of High Speed Network* **24**(3) 357-80
[21] Yang X, Ma Z, Sun L 2012 Performance vector-based algorithm for VM deployment in infrastructure clouds *Journal of Computer Applications* **32**(1) 16-9
[22] Zhao H, Song B, Shao Y 2012 High-effect resource management strategy in cloud computing environment *Computer Science* **39**(2) 212-5

Authors

Fen Guo, China

Present position, grades: teacher in South China University of Technology, China.
University study: Doctoral students in School of Computer Science & Engineering, South China University of Technology, Guangzhou, China, 2009-present.
Research activities: cloud computing, database, pattern recognition, artificial intelligence.

Huaqing Min, China

Present position, grades: professor in South China University of Technology, China.
University study: PhD in School of Computer Science & Engineering, Huazhong University of Science and Technology, Wuhan, China, in 1998.
Research activities: artificial intelligence, cloud computing, database, pattern recognition.

Ming Yin, China

Present position, grades assistant professor with the School of automation in Guangdong University of Technology, China.
University study: PhD in information & communication engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2006.
Research activities: image/videocoding, image deblurring, sparse representation, unsupervised/semi-supervised data cluster/classification.

Searching security policy with acyclic directed graphs

Xiaorong Cheng[1*], Sizu Hou[2]

[1]*Department of Computer Science, North China Electric Power University, Baoding 071000, China*

[2]*Department of Electronic & Communication Engineering, North China Electric Power University, Baoding 071000, China*

Received 1 October 2014, www.cmnt.lv

Abstract

To improve the efficiency of policy searching, a method based on the use of a weighted directed graph is studied. Regarding security states as vertices and trigger conditions as edges, the security policy knowledge base can be described as an acyclic weighted directed graph. Firstly, the graph is divided into some areas with just has an initial state node and a termination state node. Secondly, weights for each edge are set according to trigger condition frequencies, and then the optimal path from the initial state node to the termination state node is found using the A* algorithm. Finally, all state nodes are reordered on the basis of their optimal path to build an adjacency matrix and conduct depth-first traversal to search policies. Experiments showed that this method improved policy search efficiency.

Keywords: security policy search, acyclic directed graph, A* algorithm, adjacency matrix

1 Introduction

With the expansion of computer networks and increases in their complexity, information systems face serious security threats. Reasonable and effective security policy can ensure that network devices work together, so that an information system can deal with a series of security incidents. The security policy is event-driven and enforced by changing the configuration of security devices. Based on the model of a state machine, security devices or their related functions can be regarded as security states, and the security policy can be described as changes in these security states under special trigger conditions. Combined with the frequency of these trigger conditions, the set of security policies can be created as an acyclic weighted directed graph. On the basis, this paper discusses improvements to the efficiency of search policy rules on a directed graph.

The complexity of searching a security policy depends on how its description. A security policy can be expressed as a Datalog program [1], by using the Datalog semantic query algorithm to implement searched thereon. Based on first-order logic, a security policy can also be expressed by well-founded semantics [2], by introducing the SLG algorithm [3, 4] to realise policy searching. In this research, a direct graph was used to express the policy knowledge base instead of creating a logical system through well-founded semantics. The method of searching policies by traversing the graph was simpler and more efficient.

2 Definition of the security policy and system operation principle

A security policy is a set of rules, which control the security management of people and resources [5]. A security policy expresses multiple requirements related to system security while not all these requirements can be characterised by security models [6]. This paper describes a security policy with $P =< D, C, R >$, where D represents the security domain. C represents the trigger condition which consists of a security incident and a system incident (only security incidents are shown in the acyclic directed graph). R represents the rule set (each rule represents a configuration action described by $r =< i, o, s_{0,} l >$). i represents the former state. ^{o} represents the output state (state transition can be described as $i(C) \to o$). s_o represents the initial state of the security domain, and l represents whether the output state of policy is a final state or not. All state transition rules in the system constitute the policy knowledge base, which can be described as an acyclic directed graph containing one initial state node and some termination state nodes: it is shown in Figure 1, which shows part of the rules as based on policy within the policy knowledge base.

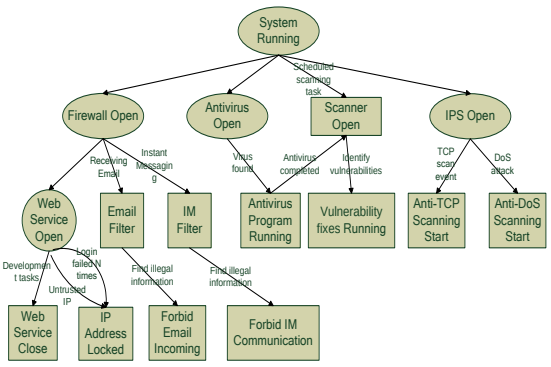


FIGURE 1 Part of the rules in the knowledge base described as a directed graph

[*]***Corresponding author*** e-mail: cheng3100@sohu.com

There is a current state set in the system, which is used to save the active states, and a state will be deleted from the current state set when it changes under the conditions imposed by the security incident. The system operation principle is shown in Figure 2. When the system receives notification of a security incident, it searched the former state and determined whether the former state belonged to the current state set. If the current state set contained the former state, then system distributed the security policy to related devices and enforced it, otherwise no action was taken. When policy enforcement was complete, the output state of the policy was merged into the current state set. Sometimes the trigger condition would be a system incident such as "system running" provided that the trigger condition continued in the system, then its output state would always remain in the current state set.

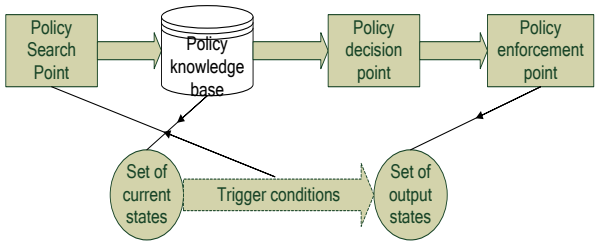


FIGURE 2 System operation principle

As shown in Figure 2, policy searching plays an important role in system operation: it is the premise of policy enforcement. In the direct graph, depth-first traversal or breadth-first traversal can be used when searching policies: both methods required an adjacency matrix. A path from the initial state node to the termination state node forms a workflow, considering the association among states in the same workflow; depth-first traversal was a better choice. When building the adjacency matrix, the sequence of nodes determined the nodal access order in depth-first traversal. To give priority to nodes, which stand for high frequency security incidents, the weights of the graph and the method of arranging the nodes were combined: this was feasible when building the adjacency matrix as proposed.

3 Method of searching security policy based on weighted direct graph

3.1 INSTANCE OF WEIGHTED DIRECTED GRAPH

There is no loop in the directed graph, which is created according to the knowledge base. The graph only has an initial state node (the in-degree of the node is zero) and some termination state node (the out-degree of the node is zero). With a small-scale knowledge base (e.g. Figure 3), V1 represents the initial state node, V10, V11, and V12 are termination state nodes, c1 to c16 are trigger conditions, the frequency of trigger conditions is described by an integer from 1 to 10, where 10 stands for the highest occurrence probability and 1 the lowest.

Using an adjacency matrix to describe this directed graph, each item in the matrix was represented by the name of its trigger condition and event frequency. In accordance with the sequence from V1 to V12, the weighted direct graph was transformed into a matrix as follows:

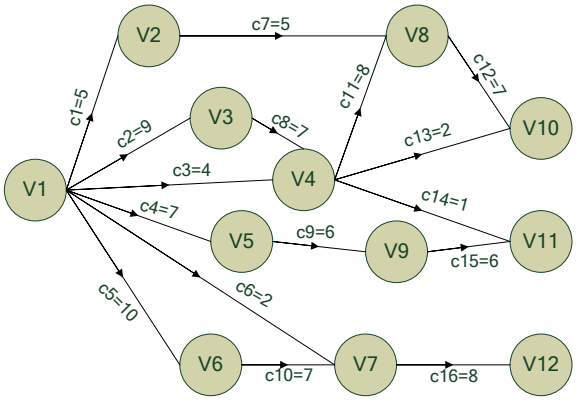


FIGURE 3 A small-scale policy knowledge base represented by a weighted direct graph

$$A=\begin{bmatrix}
null & c1,5 & c2,9 & c3,4 & c4,7 & c5,10 & c6,2 & null & null & null & null & null \\
null & null & null & null & null & null & null & c7,5 & null & null & null & null \\
null & null & null & c8,7 & null & null & null & null & null & null & null & null \\
null & null & null & null & null & null & null & c11,8 & null & null & c13,2 & c14,1 \\
null & null & null & null & null & null & null & null & c9,6 & null & null & null \\
null & null & null & null & null & null & c10,7 & null & null & null & null & null \\
null & null & null & null & null & null & null & null & null & null & null & c16,8 \\
null & null & null & null & null & null & null & null & null & c12,7 & null & null \\
null & null & null & null & null & null & null & null & null & null & c15,6 & null \\
null & null & null & null & null & null & null & null & null & null & null & null \\
null & null & null & null & null & null & null & null & null & null & null & null \\
null & null & null & null & null & null & null & null & null & null & null & null
\end{bmatrix}$$

If the frequency of the trigger condition was not taken into account when building the adjacency matrix, then depth-first traversal would not be conducive to a timely response to high frequency events, especially when the scale of the directed graph were large. As shown in Figure 3, the frequency of c16 was relatively high in all trigger conditions while almost all state nodes needed to be traversed to find the former state (V7) of c16. So the sequence of nodes when building the adjacency matrix was important when improving the efficiency of a policy search. Early detection, early implementation methods of arranging the nodes entailed two key steps: division of the graph into sub-areas, each containing one initial state node and one termination state node and discovery of the optimal path for each area using the A* algorithm. On the basis of each optimal path, all state nodes were reordered.

3.2 DIVIDE THE DIRECT GRAPH

Starting with one of the termination state nodes, the previous state of that node was successively searched until reversion to its initial state node. All state nodes were visited from the termination state to initial state node to

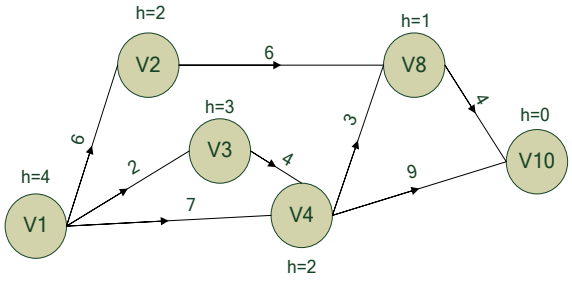
form a division of the directed graph. The number of termination state nodes on a graph was the number of areas after sub-division. This was realised by a recursion method.

For a direct graph with n nodes and e edges, the time complexity of the algorithm was $\Omega(n+e)$. By executing the algorithm, all relevant state nodes from the initial state node to the termination state nodes can be found. The graph was sub-divided as shown in Figure 3 into three areas, and the node sequence was output in reverse order as follows:

{{V1,V2,V3,V4,V8,V10},{V1,V3,V4,V5,V9,V11},{V1, V6,V7,V12}}

3.3 FIND THE OPTIMAL PATH OF EACH AREA AND REORDER ALL STATE NODES

The optimal path was that most likely to have been chosen from the initial node to the termination node corresponding to a series of higher frequency security events. Using the A* algorithm to find the optimal path required alteration of the optimal path to the minimum dissipation path. In the weighted direct graph, value of arcs represented the frequency of security events, the higher the frequency, the smaller the dissipation. The frequency values ranged from 1 to 10, so the conversion formula from frequency to dissipation value was set to $g_i = 10 - c_i + 1$ where g_i represents the dissipation value and c_i represents the frequency. Meanwhile, to reduce the number of nodes in this extended search, a heuristic function was required. The value of this heuristic function h was set to encompass the maximum number of steps to a given termination node. One of the sub-areas of the graph shown in Figure 3 was chosen, marking the heuristic function value of each node and the dissipation value of each arc, as shown in Figure 4.



FIUGRE 4 One graph sub-division showing its heuristic function value and dissipation value

The A* algorithm [7] was used to find the optimal path from initial node V1 to termination node V10: Table 1 shows such the search table thus created.

TABLE 1 A* algorithm search table

OPEN`	CLOSED
Initialisation(V1(4))	()
1 (V2(8) V3(5) V4(9))	(V1(4))
2 (V2(8) V4(8))	(V1(4) V3(5))
3 (V4(8) V8(13))	(V1(4) V2(8) V3(5))
4 (V8(10) V10(15))	(V1(4) V2(8) V3(5) V4(8))
5 (V10(13))	(V1(4) V2(8) V3(5) V4(8) V8(10))
End of search	

According to the search result, the optimal path was V1 →V3 → V4 →V8 → V10, and the dissipation of the path was 13. Similarly, for the division from V1 to V11, the optimal path was V1 → V5 → V9 → V11, the dissipation value was 14, and for the division from V1 to V12, the optimal path was V1 → V6 → V7 → V12 and the dissipation value was 8. According to the dissipation value of each optimal path, the division of nodes was arranged in ascending order:

{{V1,V6,V7,V12}, {V1,V3,V4,V8,V2,V10}, and {V1,V5,V9,V3,V4,V11}}.

The first node was removed from each set to form the final node arrangement, if the node was already in the final arrangement, then the original node was overwritten to give the termination nodes for each division. After nodal arrangement the set was:

{V1,V6,V3,V5,V7,V4,V9,V8,V2,V12,V10,V11}.

According to this nodal arrangement, the adjacency matrix was rebuilt as follows:

$$A' = \begin{bmatrix}
null & c5,10 & c2,9 & c4,7 & c6,2 & c3,4 & null & null & c1,5 & null & null & null \\
null & null & null & null & c10,7 & null & null & null & null & null & null & null \\
null & null & null & null & null & c8,7 & null & null & null & null & null & null \\
null & null & null & null & null & null & c9,6 & null & null & null & null & null \\
null & null & null & null & null & null & null & null & null & c16,8 & null & null \\
null & null & null & null & null & null & null & c11,8 & null & null & c13,2 & c14,1 \\
null & null & null & null & null & null & null & null & null & null & null & c15,6 \\
null & null & null & null & null & null & null & null & null & null & c12,7 & null \\
null & null & null & null & null & null & null & c7,5 & null & null & null & null \\
null & null & null & null & null & null & null & null & null & null & null & null \\
null & null & null & null & null & null & null & null & null & null & null & null \\
null & null & null & null & null & null & null & null & null & null & null & null
\end{bmatrix}$$

A depth-first traversal was run using the adjacency matrix above: nodes representing high frequency events were always searched first so the method could improve the overall efficiency of a policy search.

4 Experimental verification

An experiment was designed to verify the advantages of this policy search method. The experiment was run with the Java programming language on a computer with a Core 2 Duo processor and 2 GB of random access memory. Some 26 sample of security incidents were generated according to frequency. The 26 trigger conditions were searched in the original adjacency matrix of the directed

graph and in the adjacency matrix after nodal reordering. The time required under both cases was recorded (see Figure 5). It was seen from the result of this experiment that, with the expansion of the direct graph, the advantages of the proposed policy searching method were obvious.

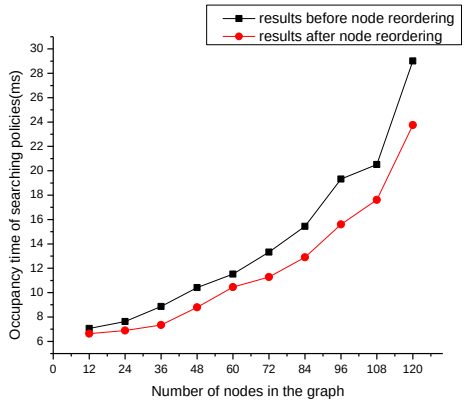


FIGURE 5 Comparison of the times required for searching the security policy: the case for nodal reordering

5 Conclusions

This research into methods of improving the efficiency of searches in a security policy based on direct graph depth-first traversal was successful. By means of a state machine, the policy knowledge base could be described as an acyclic weighted direct graph with a single initial state node and some termination state nodes.

Combined with the reverse topological sorting algorithm and the A* algorithm, and according to the frequency of each trigger condition, all state nodes were reordered before creating the adjacency matrix for a given graph. As a result, nodes associated with high frequency security events could be searched first. The experimental result showed that the proposed method could improve search efficiency within a security policy.

References

[1] Bonatti P A, Shahmehri N, Duma C, Olmedilla D, Nejdl W, Baldoni M, Baroglio C, Martelli A, Patti V, Coraggio P, Antoniou G, Peer J, Fuchs N E 2004 Rule-Based policy specification: State of the art and future work *Technical Report IST506779 Project Deliverable D1 Working Group I2*

[2] Bao Y, Yin L, Fang B, Guo L 2012 Approach of security policy expression and verification based on well founded semantic *Journal of Software* 2012 **23**(4) 912-27

[3] Chen W D, Warren D S 1996 Tabled evaluation with delaying for general logic programs *Journal of ACM* **43**(1) 20-74

[4] Chen W D, Warren D S 1993 Query evaluation under the well-founded semantics *Proc 12th ACM SIGACT-SIGMOD-SIGART Symp. On Principles of Database Systems Washington*

[5] Tang C, Yu S 2009 Verifying Network Security policy based on features *Journal of Computer Research and Development* **46**(11) 1854-61

[6] Wen H, Zhou Y, Qing S 2005 A formal commercial security policy model based on framework *Chinese Journal of Electronics* 2005 **33**(2) 222-6

[7] Ma S, Zhu X 2004 Artificial Intelligence *Tsinghua University Press the first edition* 34-46

Authors

Xiaorong Cheng, born on April 25, 1963, Handan, Hebei province, China

Current position, grades: North China Electric Power University, professor.
University study: Master of Engineering of North China Electric Power University, Power Systems and Automation, 1994.
Research activities: computer network technology, network information security.
Professional Activities and Memberships: Network Security Research, GIS-based Ethernet management software, Panjin Power Administration Cable Management Software, Concentrate on the practice teaching reform and research.

Sizu Hou, born on May 23, 1962, Yuncheng, Shanxi province, China

Current position, grades: North China Electric Power University, professor.
University study: Master of Engineering of North Jiaotong University, specializing in communication and electronic systems, 1988.
Research activities: power systems communication technology, information and communication engineering.
Professional Activities and Memberships: Broadband Power Line Communication Technology Research, Communications network monitoring and management system, Integrated Network Management System.

Design and implementation of an event correlation model in a network security linkage system

Sizu Hou[1*], Xiaorong Cheng[2]

[1]*Department of Electronic & Communication Engineering, North China Electric Power University, Baoding 071000, China*

[2]*Department of Computer Science, North China Electric Power University, Baoding 071000, China*

Received 1 October 2014, www.cmnt.lv

Abstract

Based on the shortcomings of poor compatibility, weak practicality, and low accuracy in current linkage systems, this paper designed a gradation of event correlation model with real-time response mechanism. It will be analyzing technology in data mining association rules is introduced to analyse the processing of security incidents. Then the system through the analysis of a large number of real-time data collecting all kinds of security devices found hidden in the data and related information to improve detection precision and safety accident treatment work. At last, apply this model into the system, to demonstrate the effectiveness of the model and priority.

Keywords: linkage system, event correlation model, correlation rule, security event

1 Introduction

With the rapid development of the Internet and the widespread use of e-commerce, the importance of computer communication networks in all aspects of society is increasing; computer and Internet technologies are changing the face of society. Security and reliability issues [1, 2] have become the main problems facing network security operations. To ensure that critical network information is always complete and confidential, to protect against attacks from external and internal networks, security products, such as firewalls [3], intrusion prevention systems [4], antivirus systems for all aspects of internal and external networking were developed. On the one hand, they guarantee the safety of system, on the other hand they generate many security events. These events are multifaceted in nature, vary considerably, take different formats, are sometimes false alarms, or are not be able to be used fully.

How to effectively manage all kinds of heterogeneous security devices, extract important information from massive information sets (and do so timeously), forecast system attacks that may occur accurately and efficiently, has become the urgent problem facing network security linkage systems.

By associating various security events, an event correlation [5] model matches security events with other of the same type, large quantities, different formats, and false alarms mixed with genuine cases, can not only be greatly reduced in number, and unified in format, improved in readability, and ultimately produce more real event reports, but also allow a system administrator to discover intrusive behaviours timeously and take effective measures to reduce loss.

2 Design of the event correlation model

2.1 OVERALL DESIGN OF THE MODEL

The overall structure of the event correlation model is shown in Figure 1.

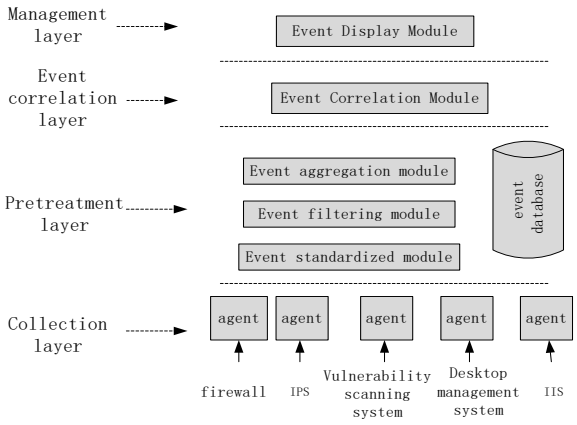


FIGURE 1 Structure of the event correlation model

The basic composition of the event correlation model included: a security event information collection module, an event standardised module, an event filtering module, an event aggregation module, an event correlation module, an event display module, and a security event database. The entire model was divided into four levels: the event collection layer, the pre-treatment layer, the event correlation layer, and the event management layer.

The lowermost layer was the collection layer, which comprised various types of information collection agent. These agents can run on a variety of operating systems and

[*]***Corresponding author*** e-mail: 867038858@qq.com

are responsible for collecting all kinds of security event information, such as IDS alarm logs, firewall logs, host logs, vulnerability information, *etc*., from different security devices.

The event pre-treatment layer included the event standardised module, event filtering module, event aggregation module, and the event database. The event standardised module unified the format of the information collected for subsequent processing. The event filtering and event aggregation modules were used to remove useless information for data mining, reducing the number of original data items. The security event database was mainly used to store all kinds of security event information produced by pre-treatment.

The event correlation layer was to uncover specific, potentially threatening, security event information through associating/matching events: this was a core part of the network security event correlation system and a critical section of the system.

The uppermost layer was the management layer, which was responsible for displaying security events, so that administrators can query, delete, or browse current network conditions.

2.2 DESIGN OF EACH MODULE IN THE MODEL

In this section, the design of the function of each module is considered at four levels, and the corresponding implementation method is introduced.

2.2.1 Event collection layer

This layer was mainly used to collect event information generated by various types of safety equipment. Here Agent is used to collect information. Agent can be installed on all kinds of safety equipment or on the router in the network where the devices sit. Agent can fully capture event information generated by the security equipment which it has been monitoring. The flowchart through the event collection layer is shown in Figure 2.

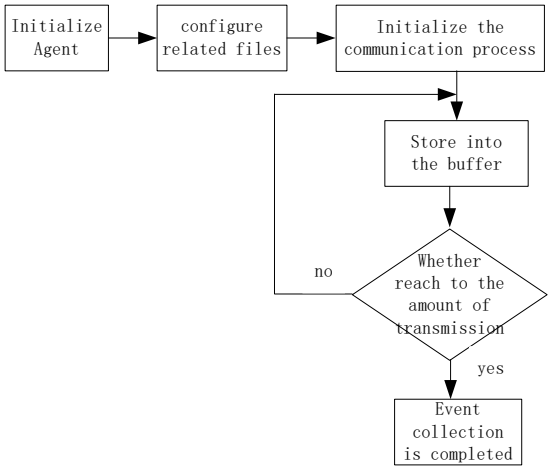


FIGURE 2 Flowchart for security event information collection

As seen in Figure 2, before collecting a certain type of security event, the information collection agent had to be initialised; secondly, we configured related files, specifying the server's IP address, port number, file location including security event/information, and the quantity of data sent each time. Then the communication process was initiated and a communication connection with the server-side established; thirdly, the specified information was collected, and finally sent from agent to server.

The event collection process mainly depended on each information collection agent which was developed independently in this research and cross-platformed, including host log collection agents, IIS service log collection agents, intrusion detection alarm collection agents, firewall log collection agents, vulnerability scan results collection agents, *etc*.

2.2.2 Pre-treatment layer

The flowchart through the event pre-treatment process is shown in Figure 3. In the pre-treatment process, firstly, we normalised and unified the formatting of various events. A security event correlation process may receive security events of different types or of one type but from different devices: there are certain differences between the formats of these events, such as security events from different types of firewall which can be compiled into different formats. A standardised module aimed to put all security events into a unified format, filter the events, removing those with nothing to do with system security, to get events that are real threats to system security. Finally, aggregation and incorporation of repeat events to reduce the number of data were undertaken and the output stored in the event database.

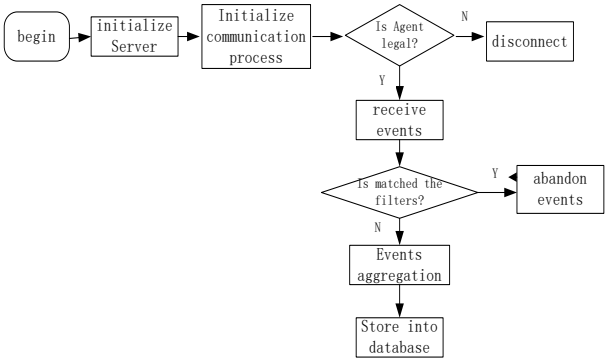


FIGURE 3 Flowchart through the pre-treatment process

2.2.2.1 Event standardised module

Security events refer to the original security event information produced by all kinds of security equipment, including log events produced by the firewall, intrusion alarm events generated by intrusion detection systems, system vulnerability events produced by vulnerability scanners, host system log events, router log information,

IIS Service log events, etc. These network security events are the main analysis objects in this research, they not only have different characteristics, but contain different content and record formats. The following is a simple introduction to some type of events:

1) Properties included in firewall logs.

int log_id
time createtime
string type
string src
string dst
string protocol
string service
string operation
int severity
int probability
int priority

2) Properties included in intrusion alarm events.

int analyzeid
time createtime
string classification
string src
string dst
string protocol
string suggestion
int probability
int priority

3) Properties included in vulnerability scanning events.

string name
string owner
time release_time
int cve_id
string request
string consequence
int severity
string solution

From the above, it can be seen that the formats of security events differed: therefore to facilitate the unified analysis of all kinds of security incidents, it was necessary to carry out the standardisation, to derive a unified format capable of expressing diversified event information from all relevant security devices.

To facilitate event correlation, events will be expressed in the following octuple form: (ID, Time, Host name, Type, Src, Dest, PRI, MSG). Among them, ID is a unique identifier of the information; Time is the production time; Host name is the device name of the recorded information; Type describes the information type; Src is the description of the source of invasion; Dest describes the purpose of the invasion; PRI refers to the priority of the event; and MSG is a description of the log.

2.2.2.2 Event filtering module

Network security events which occur in large numbers and are of low quality, contain both real events corresponding to attacker intrusion behaviour and false events with little or nothing to do with system security. The purpose of filtering for all kinds of network security events is to filter out false information as well as event information that had no relationship with the system security, at the same time, obtain only those events which posed a real threat to system security. The work included: excluding false information from detection and revealing event information corresponding to attacker intrusion behaviour from a large event log dataset.

Filtering rules mainly includes two aspects:

1) Filtering based on priority: we read PRI values of the event, and compared them with the pre-set threshold *k*. If PRI > *k*, this data had to be dealt with instantly, then sent to the next process. Otherwise it was abandoned with further processing.

2) Filtering based on key field: we included two methods:

a) Ignorance. Treat security events lacking key parameters as illegal event information, and delete it directly. For example, event information with empty log types had no analytical value and was filtered out. In addition, we deleted events with obvious error messages. For example, each part number of an IP address is an integer between 0 and 255, if the attribute value of a certain security event lay outwith that range it was clearly erroneous and was filtered out.

b) Filling. When missing only a few attributes of a record event, linear prediction, a global constant, or the average of the local properties were used to provide the missing attributes. A flowchart through the filtering process is shown in Figure 4.

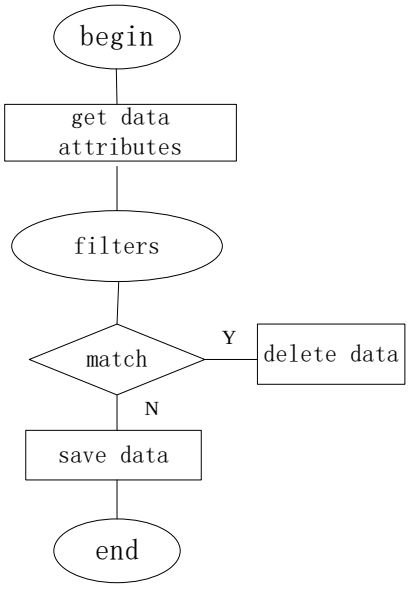


FIGURE 4 Filtering flowchart

2.2.2.3 Event aggregation module

Event aggregation aimed to merge a large number of duplicate events and reduce data redundancy. The main causes of a large number of repeats in security events were

that: the same attack, which is from the same place at the same time, often triggered different security equipment items to generate security incident records respectively; attackers often test different attack methods against specific targets or run the same attack multiple times to evaluate some parameters (such as, the offset and memory address of cache attack). Generally, all attributes or key attributes of these events are the same, and if we stored all of them in the database, they would not only occupy a large amount of storage space, but also the database grew too large, which was not conducive to the follow-up work and affected system efficiency: merging was therefore necessary.

2.2.3 Event correlation layer

After pre-treatment, the number of security events was reduced significantly. Analysing the processed results with the event correlation analysis algorithm and finding the association rules that may exist behind these data, while predicting potential dangerous events in the network, provided decision-making information for system managers.

2.2.4 Event management layer

The event management layer was responsible for displaying the data in the event database to the administrator in a pre-set way: at this time, the alarm information presented to the administrator was concise and clear, which made it convenient for the system administrator to respond to the current display and take appropriate precautionary measures. The main function of this layer included: security event querying: according to specific requirements, administrators can search associated event information by typing the appropriate query condition in the query box; and statistical analysis, for example the statistical analysis tables of various security events over a certain period (flow statistics or attack type statistics) making it easier for administrators to have an overall grasp of the current security situation.

3 Implementation of the event correlation model

To evaluate this correlation model, a strategy-based network security equipment linkage platform was used as the experimental platform and logs generated by the Venus firewall and user were taken as the data source for this experimental trial.

3.1 EXPERIMENTAL PROCEDURE

Specific steps were as follows:

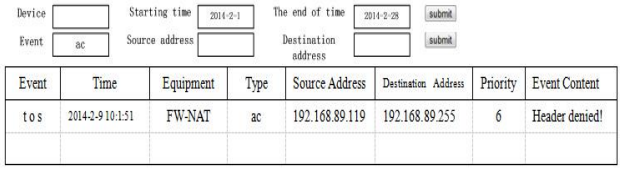
1) Pre-process the data collected by Agent to unify formats and streamline the data, store the results were then stored in the event database;

2) Discretise the data (correlation rules only deal with logical data, while these data had both numeric and class properties), thus necessitating their translation into logical data;

3) Analyse the data in the event database with an improved FUP [6] algorithm and automatically mine the dangerous events that may exist behind these data. Mining results were as follows:

TCP SYN scan => SYN flood attacks 0.8
Ftp, Telnet password detection => Trojan 0.6

Network administrators searched correlated event information through the management interface, part of which looked as follows:



Event	Time	Equipment	Type	Source Address	Destination Address	Priority	Event Content
tos	2014-2-9 10:1:51	FW-NAT	ac	192.168.89.119	192.168.89.255	6	Header denied!

FIGURE 5 Information query interface

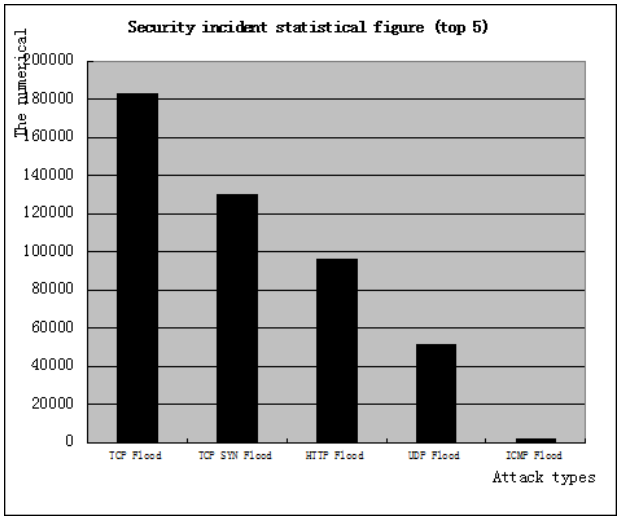


FIGURE 6 Security event statistics chart

3.2 EXPERIMENTAL RESULTS ANALYSIS

As the policy-based network security equipment linkage platform was still in its research stage, we just tested it simplistically: from the test results, this event correlation model was seen to effectively analyse alarm events, reducing the number of alarms significantly. Compared with traditional correlated systems, this event correlation model had the following advantages:

1) As it was the core module of a real-time network security equipment management system, it could process security events immediately for a timeous response.

2) The associated security events came from various heterogeneous security devices; therefore it could also improve the accuracy of correlation operations.

3) It undertook a series of processes to deal with the original security event(s) before executing correlation operations, including event normalisation, event filtering

and event consolidation, which improved the quality of the security event management and reduced the number thereof significantly. It was therefore deemed to have improved the efficiency of correlated events.

4 Conclusions

This research first introduced the overall design of the model, and then elaborated each module hierarchically. This model collected event information from agents, formatted the data into a unified event form; then eliminated redundant duplicate data with filtering and aggregation rules, stored the processed security event into an event database, and uncovered association rules hidden behind data by mining security events with an improved FUP algorithm. It then notified the system administrator thus allowing timeous preventative measures to be taken to prevent possible dangerous events. The system administrator can query the specific security event, current network status, *etc*. Finally, this model was applied to a network security equipment linkage system, using a firewall log and a user log as its data sources, and, according to the data processing flow of the model, realised the prediction of abnormal behaviour and attack behaviour, verified the effectiveness of the model, and so further improved overall system performance.

References

[1] Xin L 2009 Security Event Extraction Methods Based on Multiple Log Sources *Harbin Engineering University*
[2] Li H 2008 Research of Network Security Events Correlation and Design of System *PLA Information Engineering University*
[3] Wu Q 2006 Research on Integration of Intrusion Detection System and Firewall *Chongqing University*
[4] Luo Z 2012 Analysis network intrusion detection system *Network Security Technology and Application* **18** 54-5
[5] Agrawal R, Imielinski T, Swami A 1993 Mining association rules between sets of items in large database *Proceedings of the ACM SIGMOD Conference on Management of Data* 207-16
[6] Zhang S, Liang Z, Hu L 2012 Research and Application of Improved Multidimensional Association Rule Mining Algorithm *Engineering and Computer Science* **34** 174-9

Authors

Sizu Hou, born on May 23, 1962, Yuncheng, Shanxi province, China

Current position, grades: North China Electric Power University, professor.
University study: Master of Engineering of North Jiaotong University, specializing in communication and electronic systems, 1988.
Research activities: power systems communication technology, information and communication engineering.
Professional Activities and Memberships: Broadband Power Line Communication Technology Research, Communications network monitoring and management system, Integrated Network Management System.

Xiaorong Cheng, born on April 25, 1963, Handan, Hebei province, China

Current position, grades: North China Electric Power University, professor.
University study: Master of Engineering of North China Electric Power University, Power Systems and Automation, 1994.
Research activities: computer network technology, network information security.
Professional Activities and Memberships: Network Security Research, GIS-based Ethernet management software, Panjin Power Administration Cable Management Software, Concentrate on the practice teaching reform and research.

Research on key technologies of medicine grain defect detection system based on machine vision

Yueqiu Jiang*, Yujun Wang, Hongwei Gao, Shuang Ma

School of Information Science and Engineering, Shenyang Ligong University, Nanping Str. 6, 110159 Shenyang, China

Received 1 March 2014, www.cmnt.lv

Abstract

In the process of medicine grain production may generate many kinds of defects. If these unqualified medicine granules are not timely detected, it will not only affect the company's reputation but also the health of the patient. This paper mainly studied how to detect the unqualified medicine grain base on machine vision. It mainly consists of three kinds of common defects segmentation and defect area calculation. Firstly, preprocess the medicine grain image for the following procedures. Secondly, obtain the defect region by improved segmentation algorithm, in order to deal with three different drugs grain defects this paper improved three segmentation algorithms, for the damaged tablets propose a local edge detection algorithm that based on grey level difference, for the irregular-shaped tablets adopts the ellipse detection algorithm, which based on Hough transform technology, for the Capsule has air uses the multi-scale Canny edge detection operator. Finally, adopt Chain code contour tracking algorithm and Three-point method calculate the area of the damaged tablets, determine whether the tablets meet the requirements. Experiments show that the system can detect unqualified medicine granule quickly and accurately, it is of great practical value.

Keywords: image processing, median filtering, defect detection, region segmentation, area calculation

1 Introduction

In recent years, many developed countries in Europe and America have carried out deeply research on product quality real time detection, and have made a great progress in the online quality detection and production control [1, 2]. But the technical details strictly confidential to the outside world. In china, due to the research starts late, the domestic pharmaceutical testing is mainly relay on manual that production efficiency relatively low, so there often comes out error detection or misses detection and other undesirable phenomena. It does not be able to satisfy the efficient industrial automation requirements. So study on the drug quality inspection has a very good social significance [3-5].

Three common unqualified medicine granules are: damaged tablets, irregular-shaped tablets, Capsule has air bubbles. The medicine grain defect detection is to detect the unqualified medicine grains on the production line. This task includes two main parts: the granule region segmentation and the damaged areas calculation. Image segmentation is dividing image into non-overlapping regions which based on the image grey, texture, shape and colour. According to the prior knowledge separate the target area and background. Through analysis three types of medicine grain image find that each type has their own characteristics. Because of the tablets interior region high connectivity features, it can use the SUSAN edge detection algorithm get edge information. But SUSAN operator is sensitive with threshold lead to edge detection unstable. Therefore, based on the grey level difference propose a local edge detection algorithm. This algorithm can effectively solve the problem. For the irregular-shaped tablets detection, it adopted the ellipse detection based on Hough transform technology. For the bubble capsule, this paper mainly uses the multi-scale Canny edge detection operator [6]. After obtained the defect region, in order to quantify the damage rate [7], this paper mainly uses a method based on Freeman chain code [8]. Because the medicine grains image has the connectivity characteristics, so adopt the algorithm that chain code combined with pixel count to calculate the area. In order to prevent misjudge, in this paper for first time proposed the “three point method” to calculate the area.

2 Image preprocessing

2.1 IMAGE GREY PROCESSING

Dealing with colour images is complicated and needs a long operation time. RGB image does not reflect the morphological characteristics. It is better to convert the colour image into grey image, which only contains brightness information with 256-level grey scales. The conversion Equation for grey image and colour image is:

$$Y = 0.299R + 0.587G + 0.114B, \quad (1)$$

**Corresponding author* e-mail: missjiangyueqiu@sina.com

2.2 MEDIAN FILTER

In the process of collection and transportation, the image may be added noise, so there will be influence the image quality [9]. Median filter can smooth the image to reduce the noise effects.

When the noise is rarely and evenly distributed, median filter can better reduce noise in the image [10]. But the noise distribution in the image is relatively concentrated, traditional median filtering effect obviously cannot meet the requirements of the system [11, 12].

In order to meet the requirements that the median filter not only effectively remove the concentrated distribution noise, but also preferably preserve medicine grain image detail [13]. We propose an improved median filter algorithm. The main idea is to get rid of the maximum and minimum values in the sliding window, get the median of the two pixels, denoted m_p . Then get the difference between m_p and the value of the corresponding pixel within the window, values Δ_p . If Δ_p is within a predetermined range of threshold ΔTH , use the mean value m_p replace the grey value of the pixel point.

Let $W_n(\mathrm{i,j})$ window size is $n\times n$ its centre $(\mathrm{i,j})$. Improved filtering algorithm detailed steps are as follows:

Step 1: First assign $n=3$, find out the maximum and minimum grey value $p_{\max}$ and $p_{\min}$ in window $W_n(\mathrm{i,j})$.

Step 2: Statistics the number of pixels value is not equal to $p_{\max}$ or $p_{\min}$ in the window and make it as c_n.

Step 3: If c_3 less than or equal to 1, then let $n=5$ execution Step 2.

Step 4: If c_5 equal to 0, then let $n=7$ execution Step 2.

Step 5: Within the sliding window, calculate the median of all elements pixel values:

m_{p3} = median (the number of pixels in the window $W_3(\mathrm{i,j})$ and the grey does not equal $p_{\max}$ or $p_{\min}$;

m_{p5} = median (the number of pixels in the window $W_5(\mathrm{i,j})$ and the grey does not equal $p_{\max}$ or $p_{\min}$);

m_{p7} = median (the number of pixels in the window $W_7(\mathrm{i,j})$ and the grey does not equal $p_{\max}$ or $p_{\min}$).

Step 6: Calculate the overall average m_p .

$$m_p=\begin{cases} m_{p3} & c_3\ge 2 \\ m_{p5} & c_5>0, c_3<2 \\ m_{p7} & c_5=0 \end{cases}. \tag{2}$$

Step 7: Filtering result $F(\mathrm{i,j})$ is calculated by the original pixel values and median values according to the Equation (3).

$$F(\mathrm{i,j})=\mathrm{a}_1\cdot E(\mathrm{i,j})+\mathrm{a}_2\cdot m_p, \tag{3}$$

Where a_1 and a_2 are the weighting coefficients satisfy the condition $\mathrm{a}_1+\mathrm{a}_2=1$, wherein a_2 is determined by the Equation (4).

$$a_2=\begin{cases} 0 & \left|E(\mathrm{i,j})-\mathrm{m}_p<TH_1\right| \\ \dfrac{E(\mathrm{i,j})-m_p-TH_1}{TH_2-TH_1} & TH_1\le\left|E(\mathrm{i,j})-\mathrm{m}_p\right|\le TH_2 \\ 1 & \left|E(\mathrm{i,j})-\mathrm{m}_p\right|>TH_2 \end{cases}, \tag{4}$$

where parameters TH_1 and TH_2 are preset threshold.

Step 8: Go Step 1, processing the next target pixel.

3 Medicine grain of region segmentation algorithm research

Because there is no segmentation algorithm can adapt to all situations [14, 15]. According to the characteristics of each medicine grain image it proposes three different image defect detection segmentation algorithm based on machine vision.

3.1 REGIONAL SEGMENTATION FOR DEFECT TABLETS

Image edge is often accompanied by grey mutation that is to say the grey value of edge pixels satisfy discrete criteria [16, 17]. According to the principle, this paper put forward the improved edge detection algorithm based on local greyscale difference. The core of the algorithm includes computing greyscale differences, selecting candidate point, determining threshold value and refining edge. Details as follows:

Firstly, obtain candidate edge pixels. It can use the formula to quantify the grey step.

$$\delta=A-B. \tag{5}$$

Before determine *A* and *B*, first define a constant metric *n*. When use a 3×3 template traverse the image, *n* can measure the number of valid point in the template. According to the actual circumstance of medicine grain of image detection, here $n=3$. Sort the pixels within the coverage area of the template, taking the top *n* larger pixels summation assigned to *A*, taking *n* smaller pixels summation assigned to *B*. For each pixel x_i, difference in local greyscale δ constantly changes with different x_i position. When x_i is located in the edge, δ is larger. So edge candidate can be selected according to the size of the value of δ , the greater of the value, this point the more likely is edge pixel.

After traversing the whole image, local differences in per-pixel can be respectively calculated. According to the ideas of the SUSAN algorithm, here need to set a threshold to determine the candidate edge points. In order to avoid the shortcoming that SUSAN algorithm is affected by the threshold is too big. It can get the adaptive threshold according to the Equation (6):

$$T_{hr} = \frac{1}{M}\sum_{i=1}^{M}\delta_i \,. \tag{6}$$

If a pixel is the centre of 3×3 area and its greyscale difference δ is greater than T_{hr}, and then the pixel is the candidate edge point, otherwise not edge point.

Secondly, get the ultimate edge.

While get a set of candidate edge points, it still need to further refine the edge. Refinement idea as follows:

1) Use a sliding window size 3×3 traverse the whole image.

2) If the pixels is the candidate edge points, Calculate as Equation (7):

$$t = \delta\frac{x_i}{x_M} \tag{7}$$

where x_M is the average of the three largest grey value in the 3×3 neighbourhood which centre is x_i.

(3) Set the mean of t in the whole image as the threshold, if t is greater than the threshold, this point as edge points.

3.2 REGIONAL SEGMENTATION FOR IRREGULAR-SHAPED MEDICINE TABLETS

According to the detection index of a pharmaceutical company, irregular drug testing need not quantify different degrees, only need to detect whether medicine grain of aliens.

This section is mainly to detect the ellipsoid medicine granule whether irregular, so it can use Hough transform to determine the centre coordinate. According to the results of the Hough transform locate the tablets centre. When we got ellipse parameters, plug the tablet edge point coordinates into elliptic equations to calculate its value, if get the value greater than 1 means that the oval medicine grain is qualified, or medicine granule irregular.

The main idea of Hough transform is to map the point of original image space to the line of parameter space, or do the opposite process. If directly using the Hough principle testing ellipse centre, in the image space $(\mathrm{x,y})$, elliptical Equation (8):

$$\frac{(\mathrm{x}-\mathrm{a})^2}{m^2}+\frac{(\mathrm{y}-\mathrm{b})^2}{n^2}=1, \tag{8}$$

where m, n, a and b can be obtained by the production design specification. There are four parameters in the Equation above: (a,b) is the centre coordinates of the ellipse, oval semi-major axis m semi-minor axis n. According to the idea of Hough Transform, there need to create a cumulative four-dimensional array $A(\mathrm{a,b,m,n})$. For each a must traverse all the possible values of b, m and n, this calculation is too large to detect in real time. So in this paper, we propose an improved method.

Using gradient information, when $dy/dx = \tan\theta$, for x do derivate, get the following equation:

$$\frac{(\mathrm{x}-\mathrm{a})^2}{m^2}+\frac{(\mathrm{y}-\mathrm{b})^2}{n^2}\tan\theta=0, \tag{9}$$

$$a = x \pm \frac{\sqrt{m^2\tan\theta}}{\sqrt{m^2\tan^2\theta+n^2}}, \tag{10}$$

$$b = y \pm \frac{n^2}{\sqrt{m^2\tan^2\theta+n^2}}. \tag{11}$$

Through comparing the Equations (10), (11) with (8) can found that the number of parameters reduced from the original four to three, so just set up two three-dimensional accumulation array $A_x(\mathrm{a,m,n})$ and $A_y(\mathrm{b,m,n})$ can meet the requirements. Every point in the image, in turn, change m and n calculate the corresponding a and b, then respectively for m, n accumulation, after the improved algorithm it obviously reduce amount of calculation.

3.3 THE SEGMENTATION ALGORITHM OF BUBBLES INSIDE THE CAPSULE

This section is mainly to detect capsules whether containing bubbles. Because of the bubbles within the capsule have a clear boundary and its colour significantly different with the surrounding, so the bubbles boundary can be clear detected. Canny operator detects the edge of the more complete and connection degree is also very good, but the result is sensitive to noise, so this paper proposed an adaptive multi-scale Canny edge detection algorithm. Specific steps are as follows:

1) Combined with capsule medicine grain images to determine the scale collection.

If σ represent one scale of the image, then $0\le\sigma\le\sigma_{\max}$, the scale of the image edge is always in a range, so assume medicine granule image edge contains K scale. It can use $\{\sigma_k\}_{k=0,1,\ldots,K-1}$ represent the scale collection, the collection can cover all of the image edges scale, it's element is monotonic distribution and scale difference between the two adjacent elements within the set should be less than the scale interval $\Delta\sigma$, generally $0<\Delta\sigma<0.5$. $\sigma_{\max}$ is the maximum edge scale. $\sigma_{\max}=2^{n-1}\sigma$, After sampling at equal intervals on the scale space $0\sim\sigma_{\max}$ can get the scale set.

2) Use Canny edge detection operator [11] to detect the capsules image edge information on multiple scales.

In the course of capsules edge detection, first with a Gaussian function for image smoothing, then derivative of capsules image get the original image edge information on various scales. The edge information at the scale σ_k can be expressed as

$$f_k'(r,c), k=0,1,\ldots,k-1.$$

3) Through fusion the information of scale edge to determine the edge.

After obtaining the information of each scale on the capsule image edge, complete the synthesis of various scale edge information to get the new multi-scale image edge information. In this paper, use the weighted summation algorithm.

$$f_k'(r,c) = \sum_{k=0}^{k-1} w_k f_k'(r,c), \quad (12)$$

where w_k represents the weight factor of each scale edge information. If ignore the noise it can use mean-weighted factor. However, due to uneven illumination bring significant noise effect to the capsule image, so the mean-weighted does not meet the requirements. The weights of different scales case determined by the following steps:

Step 1: On the scale σ_k use a Gaussian function which standard deviation is σ smooth the capsules image to reduce noise, $f_k = f \cdot G(x,y,\sigma_k)$.

Step 2: In various scales of the capsule image, calculate the difference $\Delta_k = |f - f_k|$ and variance $\Delta_k^2 = |f - f_k|^2$.

Step 3: Calculates the difference and variance of image and the weight can be determined by the ratio of the two parameters.

4) The Non-maxima Suppression method to extract the edge points.

When using non-maxima suppression to get edge, mainly in order to find the ridge of the original image from the gradient image, that is to say our target point is maximum gradient, so discarded the edge is not a maximum, set $a_1,...,a_8$ are the 3×3 neighbourhood pixels of a:

$$\begin{matrix} a_1 & a_2 & a_3 \\ a_4 & a & a_5 \\ a_6 & a_7 & a_8 \end{matrix}$$

If a satisfy one of the following conditions it could as the edge points, otherwise use $\min(a,a_1,...,a_8)$ instead of a.

1) $a_1 + a_4 + a_6 < a_2 + a + a_7$ and $a_2 + a + a_7 > a_3 + a_5 + a_8$, a is vertical ridge.

2) $a_1 + a_2 + a_3 < a_4 + a + a_5$, and $a_4 + a + a_5 > a_6 + a_7 + a_8$, a is horizontal ridge.

3) $a_1 + a_2 + a_4 < a_3 + a + a_6$, and $a_3 + a + a_6 > a_5 + a_7 + a_8$, a is 45-degree angle ridge.

4) $a_2 + a_3 + a_5 < a_1 + a + a_8$, and $a_1 + a + a_8 > a_4 + a_6 + a_7$, a is 135-degree angle ridge.

4 The area calculation of the damaged tablets

Calculate the area of the damaged tablets is mainly used it quantify the damage rate. In the process of drug quality inspection need to pick out the tablet damage rate more than 3%. In this paper, the basic unit of area is pixel, by calculating the number of valid pixels in the target area to obtain the damaged area and other geometric characteristics. Since the area calculation methods all have their advantages and disadvantages and application scope. We combine with the actual situation of medicine grain of defects, proposed two methods to calculate area.

The first method: Combined chain code contour tracking algorithm [18], devised an effective multi- region contour area calculation method.

Chain code element of the contour is determined by a line segment which has both the length and direction, and each direction has a sequential number, shown in Figure 1.

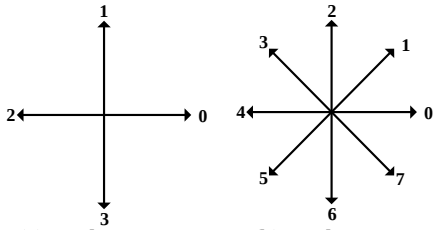


FIGURE 1 Chain code

After medicine grain image segmentation and binarization, the target region and the background region were respectively marked as 0 and 1. According to the idea of chain code contour tracing, use eight to the chain code tracking medicine grain of image boundary specific steps are as follows [19]:

1) Set the starting point b_0 as the first marked black dots of picture. Use c_0 represents b_0 neighbouring point on the left side. Figure 2b, c_0 always background points. Starting from the c_0 along the same sequence inspect b_0 eight adjacent points, find the first black adjacent points b_1 and set c_1 (background) as the point in front of the sequence of b_1. Save the location of the b_0 and c_0, in order to use in step 4.

2) Set $b = b_1$ and $c = c_1$, Figure 2c.

3) Starting from the c travels clockwise, set b eight adjacent points as $n_1, n_2, ..., n_8$ respectively. Find the first n_k marked 1, set $b = n_k$ and $c = n_{k-1}$.

4) Repeat step 3, until $b = b_0$ and find the next boundary point b_1.

In Figure 2, the next point to be processed is marked in black, has been processed pixels marked grey, c is always background points in step 3, because n_k is the first point value is 1 when clockwise scanning. But the algorithm also can further improve in boundary description. After finding target area, the area is marked as grey, follow-up point may also be judged, which will be duplicate detection. Mark the point has been detected as a boundary pixel can reduce

unwanted detection can further accelerate the processing speed.

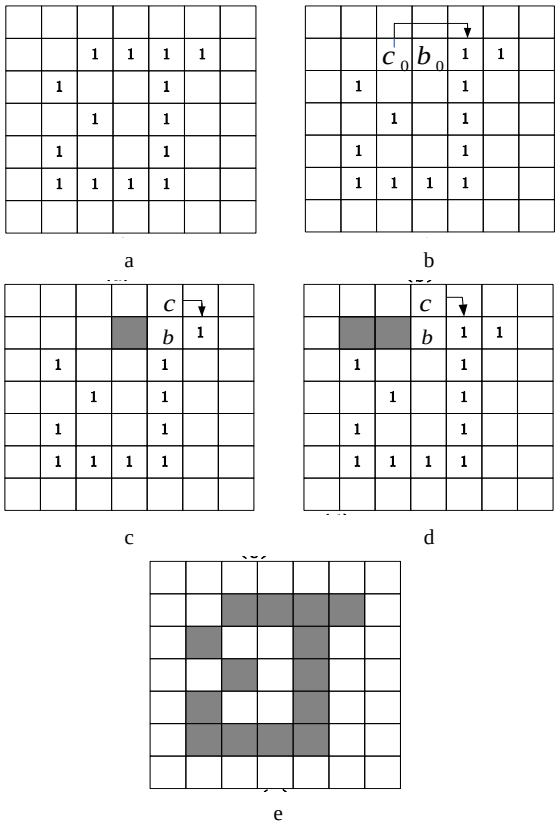


FIGURE 2 Boundary tracing algorithm description

The second method: "three point method"

"Three point method" is mainly for round tablets, because its approximate circle area after image segmentation, regional boundary has good connectivity, traverse the medicine image, and select the outermost (uppermost, lowermost, leftmost and rightmost) "point". Make straight line in the direction of maximum gradient points, it is vertical line. Calculate the intersection point of straight line, the intersection is the circular medicine circle, finally statistical the number of pixels that centre to each point (up, down, left and right edge point), calculated their average that is the drug tablet radius r, according to the area Equation $S = \pi \cdot r^2$, we can get the medicine grain area. Set four points $p_1(x_1, y_1)$, $p_2(x_2, y_2)$, $p_3(x_3, y_3)$, $p_4(x_4, y_4)$ as shown in Figure 3, after detection $p_1(x_1, y_1)$ is the rightmost edge point, but in theory, the rightmost point should be $p_1'(x_1, y_1)$. So according to the criterion, the point of $p_1(x_1, y_1)$ should be discarded. Choose the remaining three points for subsequent calculation. Determining whether the detection point is an edge point by the following criterion as in Equation (13):

$$\begin{cases} |x_2 - x_4| < c \\ |y_1 - y_3| < c \\ |y_2 - y_4 + x_3 - x_1| < 2c \\ |y_1 - y_3 + x_4 - x_2| < 2c \end{cases}, \tag{13}$$

where c is the preset threshold, it is determined whether the point is a valid edge point.

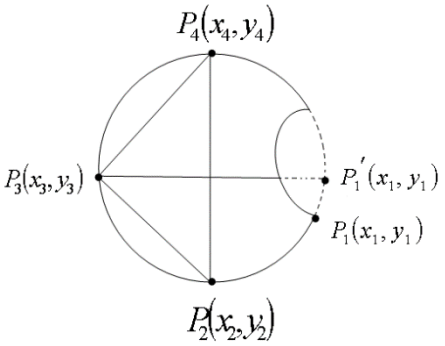


FIGURE 3 Three-point method schematic

Based on the above ideas and combined with the specific circumstances of this topic, set target pixel is white, background pixels is black. Three point method measuring area details in the following five steps:

Set P_1, P_2 are two one-dimensional arrays.

Step 1: Detecting started from the top and left of the image and gradually went on towards the down and the right, until detect the first white pixels; this point is edge point record the pixel coordinates $p_1(x, y)$.

Step 2: After obtain the first white pixel, scanning a new line and put the first met white point (the leftmost edge points) in the array P_1. At the same line when a black pixel is scanned again to put its front white pixel coordinates into the array P_2.

Step 3: When there is no white pixel to be detected in the whole row, means the end of detecting. Then the last element of the array P_2 is the most under the lowermost edge point.

Step 4: Extract the horizontal ordinate component of the each object in the array P_1, P_2 to sort. Determining the minimum and maximum of the horizontal component, thus the left-most, rightmost, uppermost, lowermost edge point coordinates are obtained.

Step 5: According to the Equation (13), determine whether the detection point is required edge point.

5 Experimental results and analysis

5.1 REGIONAL SEGMENTATION RESULTS

The improved segmentation algorithm was respectively applied to the medicine grain image, to test the effect of image segmentation and edge detection.

Figure 4 shows the damaged medicine grain region segmentation results obtained by the improved edge detection algorithm based on local greyscale difference.

Figure 4a shows the noise free image segmentation Figure 4b the image adding 3% of Gaussian noise and 2% salt and pepper noise.

Through the experimental results it can be seen that in the case of less noise, the algorithm can segment the region of medicine grain clearly. The second group is the result of edge detection after adding noise, common medicine grain image quality is much better than that situation. The improved algorithm detects the edge more complete and clear than SUSAN algorithm detected.

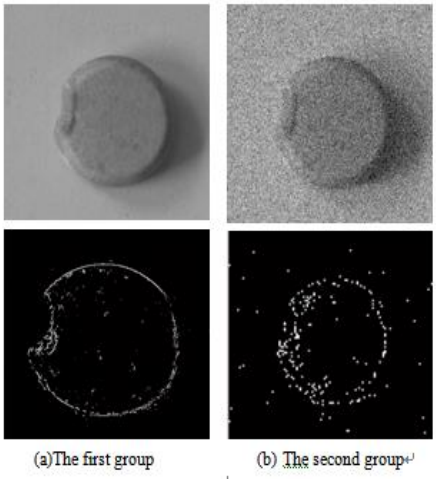


FIGURE 4 The original and add noise image region segmentation results

Figure 5 shows the irregular-shaped medicine tablets region segmentation results obtained by modified ellipse detection algorithm based on Hough transform.

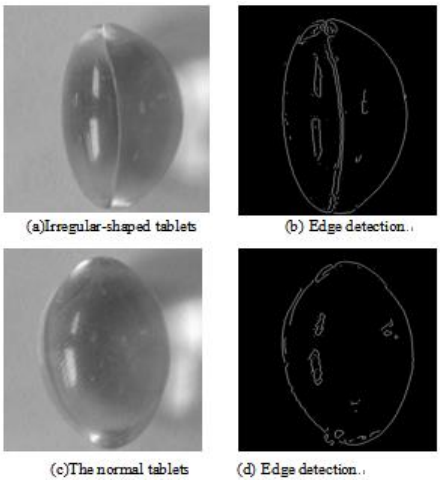


FIGURE 5 Irregular-shaped and normal tablets region segmentation results

Medicine grain qualified analysis: after dealing with the various steps above a single medicine area has to be designated. At the same time the elliptic equation is obtained by Equations (9-11), so it only needs to traverse each element of the $A_x(a,m,n)$ and $A_y(b,m,n)$, if get the value greater than 1 means that the oval medicine grain is qualified, or medicine granule irregular.

Figure 6 shows the bubble region segmentation in capsules results obtained by Multi-scale adaptive Canny edge detection algorithm.

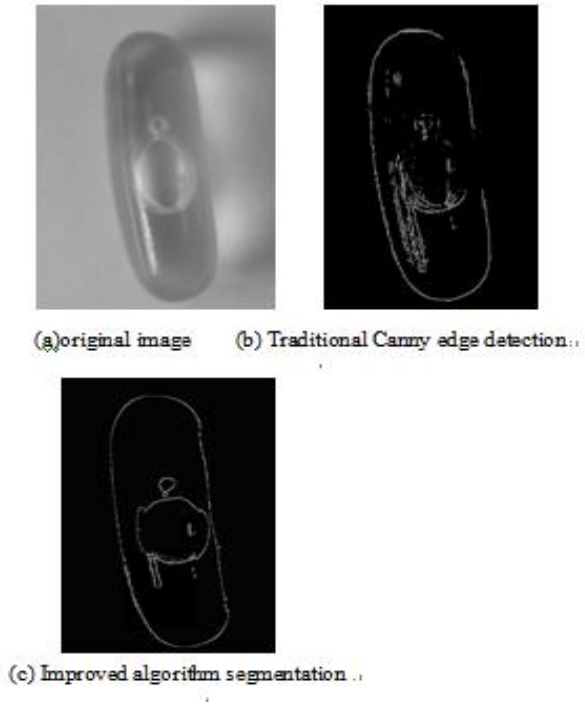


FIGURE 6 Multi-scale adaptive Canny edge detection

5.2 MEDICINE GRAIN OF DEFECT QUANTITATIVE RESULTS

First, use the Freeman chain code area algorithm processing the image, for the damage medicine grain of defect areas can be detected, quantification of the defect region, obtained theoretical defect area and actual detection area. According to the drug quality specification, damage rate more than 3% the tablet is unqualified. So that can further calculate damage rate and judge whether the drug tablets are qualified. The experimental results are shown in Table 1.

TABLE 1 Freeman chain code area algorithm of quantitative results (pixels)

No.	Theoretical area	Actual area	Damage rate	Qualified
1.	147730	137813	6.71%	no
2.	151327	147476	2.54%	yes
3.	150968	146682	2.84%	yes
4.	149978	136597	8.92%	no
5.	161468	154567	4.27%	no
6.	146798	136892	6.75%	no
…	…	…	…	…

In order to better compare the reliability and accuracy of the two algorithms, both methods deal with the same group of tablet images. The quantitative results are shown in Table 2.

TABLE 2 Three point method of quantitative results (pixels)

No.	Theoretical area	Actual area	Damage rate	Qualified
1.	153863	139867	9.09%	no
2.	157638	153433	2.67%	yes
3.	149863	145658	2.81%	yes
4.	151267	139685	7.66%	no
5.	161468	153534	3.85%	no
6.	160357	147586	7.96%	no
…	…	…	…	…

Quantitative results obtained from two algorithms can be seen: although the algorithm implementation have differences lead to concrete numerical value different, it has same results.

6 Conclusions

The experimental results show that the segmentation result satisfies the requirement of experiment, in order to quantify the damage of drug, we use the Freeman chain code area algorithm and "three point method" respectively calculated the defect area. Although the algorithm implementation have differences lead to concrete numerical value different, it has same results. But this article is just study the area of damaged medicine grains, other types without quantitative treatment, might exist misjudgments, so the algorithms also need to be improved. The system cannot achieve complete automation, there still needs a further improvement and enhancement.

References

[1] Zhu M, Zou W 2010 Application of image processing in tablets defect inspection *Computer Engineering and Design* **31**(23) 5151-4
[2] Wu H, Ji J, Zhu Y, Xi X, Fu B 2010 The design and research of capsule defect detection system *Computer Knowledge and Technology* **6**(28) 8093-4
[3] Li Z, Li H, Sun J 2012 Detection of PCB based on digital image processing *Instrument Technique and Sensor* **8** 87-9
[4] Dong W, Sun Z, Tan T 2010 Iris recognition system at a distance based on stereo vision and pan-tilt-zoom camera *Science & Technology Review* **28**(5) 35
[5] Jin L, Xia L, Yang S 2000 Histogram adaptive thresholding using index of fuzziness *Journal of Image and Graphics* **5**(5) 390-5
[6] Lixia X, Li T, Wang Z 2010 Adaptive canny edge detection algorithm *Application Research of Computers* **27**(9) 3588-90
[7] Li B, Liu D, Liang G 2002 A new algorithm for calculating the area of arbitrary enclosed shape *Journal of National University of Defense Technology* **24**(4) 61-3
[8] Yang W, Ren M, Yang J 2008 Object area algorithms based on chain code in digital image *Computer Engineering* **34**(1) 30-3
[9] Liu S, Wu G 2010 A novel method for image enhancement based on generalized histogram *Electronics Optics & Control* **17**(3) 12-4
[10] Ko S J 1991 *IEEE Transactions on Circuits and Systems* **38**(9) 984-93
[11] Lang L, He W, Lei L, Zhang W 2010 Survey on enhancement methods for non-uniform illumination image *Application Research of Computers* **27**(5) 1625-8
[12] Sun H, Shi W, Ju Y 2003 Image processing with medium value filter *Journal of Changan University* **23**(2) 104-6 (*in Chinese*)
[13] Qu C, Li H, Zhang C 2006 Wavelet package denoising method based on multiseale edge detection *Information Security and Communications Privacy* 77-8
[14] Lang Y, Zheng D 2012 An improved sobel edge detection operator *Proceedings of the 2012 Third International Conference on Mechanic Automation and Control Engineering IEEE Computer Society* 841-3
[15] Chen Q 2011 Research of image segmentation algorithm based on mathematic Morphology *Dissertation for the Master Degree in Engineering of Harbin University of Science and Technology*
[16] Lim Y W, Lee S U 1990 On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques *Pattern Recognition* **9**(23) 935-52
[17] Zhang C 2011 Study on edge detection method *Dissertation for the Master Degree of Jiangnan University* (*in Chinese*)
[18] Chochia P A 2010 Image segmentation via contour tracking in application to the analysis of the photographs of electronic microcircuits *Journal of Communications Technology and Electronics* **55**(12) 1466-73
[19] Du Y, Shi W, Liu C 2012 Research on an efficient method of license plate location *Physics Procedia* **24**(Part C) 1990-5 (*in Chinese*)

Authors

Yueqiu Jiang, China

Current position, grades: professor in Shenyang Ligong University, China.
University study: PhD degree in Computer application technology, Northeastern University, Shenyang, China, 2004.
Research activities: network management, image processing, wireless network, pattern recognition, signal and information process.
Professional Activities and Memberships: a leader of subject direction for signal and information process.

Yujun Wang, China

Current position, grades: postgraduate student in Shenyang Ligong University, China.
University study: Graduate degree in Signal and information processing in Shenyang Ligong University in 2012.
Research activities: network management, image processing, pattern recognition, signal and information process.

Hongwei Gao, China

Current position, grades: associate professor in Shenyang Ligong University, China.
University study: PhD degree in Pattern recognition and intelligent system, Chinese Academy of Sciences (CAS), Shenyang, China, 2007.
Research activities: network management, image processing. wireless network, pattern recognition, signal and information process.
Professional Activities and Memberships: a leader of academic direction for optical and electrical detect technology and system.

Shuang Ma, China

Current position, grades: postgraduate student in Shenyang Ligong University, China.
University study: Graduate degree in Signal and information processing, Shenyang Ligong University in 2012.
Research activities: network management, image processing. pattern recognition, signal and information process.

Research on a regional innovation system: viewed from the degree distribution of complex networks theory

Jiangbo Zheng*

School of Management, Jinan University, Guangzhou, China

Received 1 March 2014, www.cmnt.lv

Abstract

A Regional Innovation Systems is viewed as a special type of complex network in this paper, and the complex features of the entities in the system as well as the interactions between the entities are discussed in detail. Based on the Degree Distribution of Complex Networks theory, firstly this paper proves the feasible degree distributions in practical networks through mathematical reasoning and analysing. Then, an empirical case of degree distribution characteristics for the Regional Innovation System is tested by mean of calculating the Correlation Coefficient with the statistical data of Guangdong Province. Coupling Complex Networks Theory with these data and the practical conditions of Guangdong province, this paper sheds light on the insight of the calculating results. Such research shows that the methodologies and conclusions of this paper are proper.

Keywords: regional innovation system, degree distribution, complex networks, data calculation, Pearson correlation coefficient

1. Introduction

Since the early 1990s, the concept of regional innovation systems has obtained considerable attention for advancing the understanding of the innovation process in regional economies. The increasing popularity of this concept has been impelled partly by the intense international competition in globalizing environment, as well as the deficiencies of traditional regional development models and policies in many regions around the world. Although it is typically understood to be a set of interacting private and public institutions or organizations that function according to organizational arrangements and relationships conducive to the generation, use, and dissemination of knowledge, the concept of regional innovation systems has no generally accepted definitions. The basic argument is that such a set of actors produce pervasive and systemic effects that encourage firms within the region to develop specific forms of capital that are derived from social relations, norms, values, and interactions within the community in order to reinforce regional innovative capability and competitiveness [1].

The original concept can be found in two main bodies of relative research. The first is systems of innovation. Built on evolutionary theories of economic and technological change, the literature [2] conceptualizes innovation as an evolutionary and social process. In the literature [3], innovation is stimulated and influenced by many actors and factors, both internal and external to the firm. In the literature [4], the social aspect of innovation refers to the collective learning process between several departments of a company (for example, R&D, production, marketing, commercialization, etc.) as well as to external collaborations with other firms, knowledge providers, financing, training, etc. The second body is regional science and its explanations about the social environment from which innovations emerge. From this point of view, innovation is localized and a locally embedded, not placeless, process [5, 6]. Namely, a regional innovation system is characterized by cooperative innovation activities for knowledge creation and diffusion between firms and organizations, such as universities, training organizations, R&D institutes, technology transfer agencies, and so forth.

A fundamental problem in all studies of regional innovation systems is that we cannot yet determine how a regional innovation system might appear in reality [7]. For instance, how much and what type of innovation must occur within a region for it to be considered a regional innovation system? Do all regions that aspire to take a lead in organizing and innovating become regional innovation systems by default? If something like a regional innovation system already exists, but the extant literatures on the subject are not clear. We contend that the interactions between actors in regional innovation systems have not been sufficiently explored, while the institutional context of these interactions has been largely overlooked. As a result, the validity of recommendations for innovation policy making based on the current analyses of regional innovation systems is somewhat questionable. Therefore, it is very necessary to analyse a regional innovation system from new viewpoint to obtain substantial insight.

* ***Corresponding author*** e-mail: zhengjbjnu@126.com

2 The definition of a regional innovation system in the sight of network

The innovation theory initially emphasizes on the technological innovation within corporations. For instance, American economist Joseph Alois Schumpeter's book The Theory of Economic Development: An inquiry into profits, capital, credit, interest and the business cycle pointed out that technological innovation is a linear process including developing, designing, manufacturing and selling [8]. However, as researches deepened, innovation was no longer confined to a linear model within corporations or between corporations. People generally consider it in a higher level--regional innovation system. In 1987, British economist Christopher Freeman introduced the concept of "National Innovation System" [9]: the network of institutions in the public and private sectors whose activities and interactions initiate, import, modify and diffuse new technologies. The OECD report "Managing National Innovation Systems" [10] in 1999 pointed out: "Innovation performance depends on the way in which the different components of the 'innovation system' -- businesses, universities and other research bodies - interact with one another at the local, national and international levels, and concludes that the public authorities must change their approach to the promotion of innovation." Throughout the researches published, although regional innovation systems have drawn great attention in the academic fields, there is no plausible explanation about it given by mainstream Economics, Industrial Economics, or Evolutionary Economics [11]. The researches in managerial fields are even in a worse situation.

The world-famed physicist Stephen Hawking once said that the 21st century would be the Century of Complexity. This statement is widely recognized, and the researches on complexity have been carried out in many fields in recent years. Among many new researches on complexity science, one of the most representative theories is Complex Networks Theory. According to this theory, a complex network is a highly abstract complex system. In fact, complex networks can be found in natural, engineering, and even our social fields [12,13]. For instance, the metabolism in cells, neural networks in brains, food chain networks forming the ecological system, social relationship networks, science research cooperation networks, trade networks, internet, and electric power networks, etc [14-19]. Recent years, with the rapid development of network and computer technologies, researchers are able to obtain abundant data from major real networks to conduct statistical analysis. Such results show that complex networks are not homogeneous; on the contrary, they are heterogeneous. Their degree distribution does not follow Poisson distribution, but power-law distribution. A network that follows power-law distribution is called a scale-free network [20]. Because of the scale-free feature possessed by complex networks, its degree distribution is completely determined by its power-law index (degree distribution index). For instance, the degree distribution indexes of most artificial networks are between 2 and 3. And some key natures (like transmission threshold) of network dynamics (like transmission of infectious disease or virus) are directly related with degree distribution index.

All of the achievements have shown the successful application of complex networks theory. However, there are not yet specific researches utilizing this theory into the important academic field-regional innovation. Based on the previous researches, this paper defines a regional innovation system as a complex network constituted by corporations, research institutions, universities and colleges, government and institutions with intermediary functions within a given region, which develop the ability of regional innovation by interacting with each other. Based on the definition, this paper first analyses the connotation of the degree distribution based on complex networks theory. Then, taking Guangdong Province of China as the object of demonstration, this paper calculates the degree distribution with the related data and tries to explain the calculation results. It is believed that this paper will offer important reference for future research related to regional innovation.

3 The complex network's features of a regional innovation network

Based on the complex networks theory, both the characteristics of the subjects in regional innovation networks and interactions between subjects have embodied complex network features.

3.1 FEATURES OF THE SUBJECTS

There are many categories of subjects building a regional innovation system. These subjects have different forms and abilities. Furthermore, each type subject can be described in different accurate degrees. Such subjects have different layers and scales, and play different roles in the innovation systems. This paper has classified subjects into 4 following categories.

Creative Subjects: this paper considers that the sources of innovation are ideas, which produce new knowledge. In a regional innovation network the subjects that produce new knowledge are research institutions, universities, R&D departments of enterprises. They are in the one edge of the frontier in the innovation network. And the aggregates of these subjects are creative ones. These subjects should be depicted according to the layer of the network studied or the particular characteristics of data.

Supportive Subjects: if creative, informational new knowledge can be transformed into economic interests, it has to gain support and coordination from material elements. That is why we think governments (political supports), banks (financial credits), venture capitals, etc. are on the other edge of the frontier in the innovation network. They are supportive subjects.

Flowing Subjects: in an innovation network system, the combinations and flows from source boundary (inputs of economic interests) to department boundary (outputs of economic interests) of every elements related to innovation cannot fulfil the transmission function. They need corresponding channels, patterns and drivers as inter-mediums. We see intermediaries in technical markets and related government sectors that perform communicative functions and resource allocation are in the middle area of the innovation network. They are flowing subjects.

Productive Subjects: in an innovation network, many innovation elements finally come together and transform into economic interests. This process need to be performed by a specific category of subjects. In this paper I view corporations as the elements that produce outputs of innovation networks. They are in the last area of an innovation network and can be defined as productive subjects.

As *we* can see, different kinds of subjects in a regional innovation network not only have important effects on themselves, but also have corresponding attributes and features that can directly affect the whole innovation network. Based on the description of different accurate degree, therefore, if we see the subjects in an innovation network as nodes in the network, they have obviously embodied the features of nodes in the complex network.

3.2 FEATURES OF THE INTERACTIONS BETWEEN SUBJECTS

In a regional innovation network, interactions between subjects substantially reflect the interactive relations between material, energy and information. These interactions are non-linear synergistic actions. With these interactions, subjects can modify their behaviors to response to the changes of environment. In other words, every subject can gain innovative ability through information sharing, mutual complementarities of abilities, diffusion effect of knowledge and cumulative effect of innovation accumulation. Therefore, theoretically speaking, the interactions between subjects have features like direction and intensity (or weight). In practice, of course, they can be dealt as undirected and weightless interactions.

The Directions of the Interactions: almost every interaction between subjects specifically directs from one to another. For example, an action which directs to a corporation from a research institute means that the researches flow from the latter to the former, while an action directing to universities from government financial sectors means the flows of financial expenses to the universities. Therefore, the in-degree and out-degree of every subject should be considered respectively. And also we can study the ratio of input to output (like innovation effectiveness), the proportion of patent, etc. about the corresponding innovative behavior.

The Multiple-characteristics of the Interactions: the interactions between subjects in an innovation network have different characteristics. Take universities as an example, an action directing the university from government means the government supports the university financially; an action from a corporation means it supports the university financially as well. The former is much more policy behavior, while the latter is more business cooperation.

The Weights of the Interactions: in innovation networks, the interactions not only differ in quality, but also in quantity - it means that the interactions have weights. Furthermore, in this paper weight means the intensity of *interactions* between subjects.

The Dynamics of the Interactions: the interactions between subjects in innovation networks are not static, but dynamic. In other words, every action happens only in a specific time, being a function of time.

Therefore, if we see the interactions between different subjects in an innovation network as the "edges", they will obviously reflect the features of edges in complex networks.

4 Theoretical analyses on degree distribution

Based on the principles of complex networks, the degree of a node refers to the number of the edges to which link the node. Suppose that there is no isolated node in a network, no self-circled phenomenon, and there is at most one edge between any two given nodes, then the definition of degree distribution should be:

$$P(k) = \frac{\text{the number of nodes whose degrees are k}}{\text{the total number of nodes}}$$

($\forall$ positive integer k).

Set the total number of nodes as N, and the total number of edges as W. Then the minimum degree of every node is 1, the maximum is N-1. So the relationship of degree distribution can be expressed as Equation (1) which is called completeness:

$$\sum_{k=1}^{N-1} P(k) = 1. \tag{1}$$

For a scale-free network [12~13], $P(k)$ is a power function, namely $\gamma > 0$ and $C_N > 0$, making:

$$P(k) = C_N \cdot k^{-\gamma}. \tag{2}$$

In the equation above, γ is called degree exponent; C_N is a Power law coefficient, whose value can be calculated according to the completeness of degree distribution in Equation (3).

$$C_N = 1 \Big/ \sum_{k=1}^{N-1} k^{-\gamma}. \tag{3}$$

In order to further discuss degree distribution, set the fist moment and second moment of degree distribution as d_{M1} and d_{M2} respectively, expressed as followed:

$$d_{M1} = \sum_{k=1}^{N-1} k \cdot P(k) = C_N \cdot \sum_{k=1}^{N-1} k^{1-\gamma}, \quad (4)$$

$$d_{M2} = \sum_{k=1}^{N-1} k^2 \cdot P(k) = C_N \cdot \sum_{k=1}^{N-1} k^{2-\gamma}. \quad (5)$$

Set the average value and standard deviation as μ and σ respectively, let $\mu = d_{M1}$, then:

$$\sigma^2 = d_{M2} - \mu^2 = d_{M2} - d_{M1}^2. \quad (6)$$

Obviously, with γ increasing, both d_{M1} and d_{M2} will decrease. When reflecting on network topology, that means the network transits from heterogeneity to homogeneity. When $\gamma \to +\infty$, there would be $\mu \to 1$, $\sigma \to 0$, which means the network is completely homogeneous. Since the length of this paper is limited, only one situation is discussed to demonstrate the relations between the range of γ and network topology.

When $0 \le \gamma < 1$, there is:

$$\sum_{k=1}^{N-1} k^{-\gamma} = O\left(N^{1-\gamma}\right), \sum_{k=1}^{N-1} k^{1-\gamma} = O\left(N^{2-\gamma}\right), \quad (7)$$

$$d_{M1} = O\left(N^{2-\gamma} / N^{1-\gamma}\right) = O(N). \quad (8)$$

Obviously, when N is abundantly large, d_{M1} is divergent. The same:

$$d_{M1} = O\left(N^{3-\gamma} / N^{1-\gamma}\right) = O\left(N^2\right) \quad (9)$$

So d_{M2} is also divergent. Therefore based on the Equation (6), it can be proved that the average and variance of degree distribution are divergent. The total number of edges in the network is:

$$W = N \cdot \mu = N \cdot d_{M1} = O\left(N^2\right). \quad (10)$$

The Equation (10) shows that the number of edges in the network is of the same magnitude with $N \cdot (N-1)/2$, the number of edges in a complete network. The calculation does not quite fit the real evidence obtained by other researchers since many researches with real evidence show that large networks are almost sparse networks, therefore the network where γ is in the range of [0,1] does not exist in the real world.

Specially, when $\gamma = 1$, there will be:

$$\sum_{k=1}^{N-1} k^{-\gamma} = O[\ln(N)], \sum_{k=1}^{N-1} k^{1-\gamma} = O(N), \sum_{k=1}^{N-1} k^{2-\gamma} = O\left(N^2\right)$$

and we can yield:

$$d_{M1} = O\left[N / \ln(N)\right], \; d_{M2} = O\left[N^2 / \ln(N)\right], \quad (11)$$

$$W = N \cdot \mu = N \cdot d_{M1} = O\left[N^2 / \ln(N)\right]. \quad (12)$$

Based on the principles above, a large network with $\gamma = 1$ does not exist in the real world either.

5 Demonstration and analyses based on data calculation

According to the complex networks theory, different types of degree distribution of networks can reflect the nature of the whole structure. Moreover, nodes with different degrees have different status and functions within the network. Besides, the degree distribution of complex network theory is built on the hypothesis that edges and weights are equal. But in practice, what matters is the overall weight. In examining the degree distribution, we can ignore the source of the edges (in the case of in-degree) and the direction of the edges (in the case of out-degree), just focus on the calculation of input and output distribution of nodes.

5.1 DEFINITION OF NODES AND EDGES IN THE NETWORK

Ideally speaking, universities, corporations and government sectors can be identified as the innovation subjects in a region because these subjects meet with the requirements of complex network theory well. But since data that completely fits theoretical requirements is unavailable, we instead employ rough method to define nodes. That is, to view Guangdong Province as an innovation network and the 21 cities of Guangdong as nodes.

Some other measuring methods are often subjective when defining whether there are relationships between nodes or not. In general, those arbitrary weights will affect the result of the measurement. To eliminate the subjectivity in defining the edges, this paper applies Pearson Correlation Coefficient as an objective indicator to measure the relationship between nodes. Pearson correlation coefficient is a statistical measure of the degree of linear dependence between two variables, explaining the strength of linear relationship between two phenomena (X and Y). Sample correlation coefficient is commonly represented by the letter *R*, giving a value between -1 to 1 inclusive, where *R*>0 indicates positive correlation, *R* <0 indicates negative correlation, and *R*=0 is no correlation. The Equation is:

$$R = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 \sum (Y - \overline{Y})^2}} = \frac{\sum XY - \frac{\sum X \cdot \sum Y}{n}}{\sqrt{[\sum X^2 - \frac{(\sum X)^2}{n}][\sum Y^2 - \frac{(\sum Y)^2}{n}]}} . \quad (13)$$

Then we choose some statistics data with temporal continuity of nodes, and calculate its relevance with time, to gauge whether and at what degree they are related. The method is to subjectively define the critical value of the correlation coefficient between any two cities as R_0, then to analyze the correlation between every two different nodes. When the correlation coefficient is larger than R_0, we consider there is a edge between the two nodes. This assumption is built on such theoretical basis: if two cities, at a particular time period, have a relatively consistent development trend, it means that they have similar policies and other internal and external elements. The reason they have such similarities is generally due to communications and interactions between cities (for instance, one city often learns some patents from other cities to promote the experience of development). Therefore, this method avoids subjectivity, as well as better reflecting the connections between cities.

5.2 DATA SELECTION AND CALCULATION RESULTS

Table 1 shows the statistics of the number of patents of 21 cities of Guangdong Province (data from statistics yearbook, only part of the data is shown below), which is calculated to see if an edge is formed between any 2 cities. When the given R_0 =0.97, the degree distribution (by computing programming) is shown in Figure 1.

Make sure that you have used Type 1 or True Type Fonts (check with the Acrobat Reader or Acrobat Writer by clicking on File>Document Properties>Fonts to see the list of fonts and their type used in the PDF document).

In Figure 1, the network built by numbers of patents in Guangdong cities is scale-free. It is found that the degree distribution is negative power law distribution with $2 < \gamma \leq 3$. Based on the complex network theory, its network architecture is highly self-similar and divergent, meaning that most nodes' degree have low values, but a minority of nodes are central nodes with pretty high degree values. Therefore, these central nodes with high degree values are obviously of great importance, like Guagzhou, Shenzhen, Zhuhai and Foshan. The results fit the actual situation of Guangdong Province quite well, in a way demonstrating the correctness of our research.

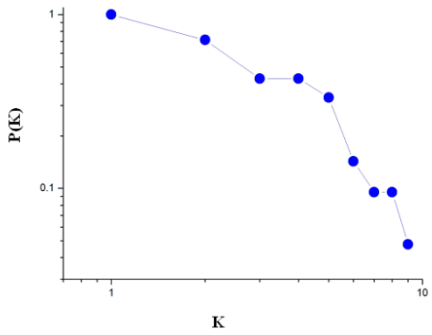


FIGURE 1 Degree distribution of domestic patents application

TABLE 1 Domestic patent application of the main cities in Guangdong province

City/year	2007	2008	2009	2010	2012	2013
Guangzhou	4410	4998	6288	8206	8230	11012
Shenzhen	4447	6177	8337	12344	15088	20943
Zhuhai	594	759	825	989	1220	1830
Shantou	1061	1285	1849	2270	2586	2715
Foshan	4238	5490	6897	7388	10809	17248
Shaoguan	98	106	114	129	169	303
Heyuan	51	28	39	65	133	83
Meizhou	72	87	69	123	148	190
Huizhou	378	331	708	854	1109	1041
Shanwei	51	104	169	182	188	160
Dongguan	1653	2914	3100	3865	4325	6694
Zhongshan	1372	1829	2115	2159	2545	3399
Jiangmen	695	767	1097	1848	2116	2787
Yangjiang	195	283	425	369	499	699
Zhanjiang	217	195	248	235	406	475
Maoming	94	84	157	205	172	314
Zhaoqing	118	179	190	255	343	303
Qingyuan	53	108	169	119	218	167
Chaozhou	746	997	737	805	879	1155
Jieyang	498	346	331	504	539	523
Yunfu	28	43	84	94	152	148

6 Conclusions and future research

The following conclusions can be drawn from the research:

1) A regional innovation system can be considered as a complex network. The definitions of nodes and sides reflect its features as a complex network.

2) The calculations show that the Guangdong Province regional innovation network follows the negative power law distribution and the degree index is between 2 and 3.

3) The calculation results show that degree distribution theory is able to reflect the status of subjects and their relations with one other.

4) The complex networks theory brings new view, new tools and new methodologies for regional innovation researches. Relative researches worth going even further.

Acknowledgment

This work is sponsored by: the Philosophy and Social Science Fund of Guangzhou, foundation item 13G52.

References

[1] Doloreux D 2003 Regional innovation systems in the periphery: the case of the Beauce in Canada *International Journal of Innovation Management* **7**(4) 67-94
[2] Edquist C 2004 Systems of innovation--a critical review of the state of the art *In: Fagerberg J, Mowery D, Nelson R editors Handbook of innovation Oxford: Oxford University Press*
[3] Dosi G 1988 The nature of innovation process *In: Dosi G. editor Technical change and economic theory London Pinter*
[4] Cooke P, Boekholt P, Dtling F 2000 The governance of innovation in Europe *London Pinter*
[5] Storper M *1997* The regional world *New York Guilford Press*
[6] Malmberg A, Maskell P 1997 Toward an explanation of regional specialization and industrial agglomeration *European Planning Study* **5**(1) 25-41
[7] Markusen A 1999 Fuzzy concepts, scanty evidence, policy distance: the case for rigour and policy relevance in critical regional studies *Regional Study* **33**(8) 869-84
[8] Wasserman S, Faust K 1994 Social Network Analysis *Cambridge: Cambridge University Press*
[9] Watts D J, Strogatz S H 1998 Collective Dynamics of Small-World Networks *Nature* **393** 440-2
[10] Amaral L A N, Scala A, Barthelemy M J, Stanley H E 2000 Classes of Small-World Networks *Proceedings of Natl Acad Sci USA 97* 11149-52
[11] Newman M E J 2000 The Structure of Scientific Collaboration Networks *Proceedings of Natl Acad Sci USA 98* 404-9
[12] Scott J 2000 Social Network Analysis: A Handbook *London: Sage Publications*
[13] Barabdsi A L, Albert R 1999 Emergence of scaling in random networks *Science* **286**(10) 509-12
[14] Albert R, Barabási A L 2009 Statistical Mechanics of Complex Networks *Review of Modern Physics* **74**(1) 47-97
[15] Schumpeter J A 1961 The Theory of Economic Development: An inquiry into profits, capital, credit, interest and the business cycle *Oxford: Oxford University Press*
[16] Reeman C 1987 Technology Policy and Economic Performance: Lessons from Japan *London Pinter*
[17] OECD Managing National Innovation Systems *Paris: OECD* 1999.
[18] Cooke P 1993 Regional innovation system: an evaluation of six European cases *Urban and Regional Development in the New Europe* **74**(1) 133-54
[19] Diez M A 2011 The evaluation of regional innovation and cluster policies: towards a participatory approach *European Planning Studies* **9**(7) 907-23
[20] Carlsson B, Jacobsson S, Holmen M, Rickne A 2012 An innovation system: analytical and methodological issues *Research Policy* **31**(1) 233-45

Author

Jiangbo Zheng, China

Current position, grades: associate professor of School of Management, Jinan University, China.
University studies: PhD in Management Science and Technology, School of Management, Tianjin University in 2005.
Research activities: operations research, supply chain alliances, products' development, logistics

A GPU-based simulation system for infrared images of deep space targets

Weiqing Li*, Ranran Xu, Mengyu Yuan

School of Computer Science & Engineering, Nanjing University of Science & Technology, Nanjing, China

Received 1 October 2014, www.cmnt.lv

Abstract

In study of deep space targets recognition, infrared images of deep space targets are needed for repeat testing and evaluating. Since the limitation of deep space flight experiments, it is difficult to obtain sufficient infrared images under different conditions. Infrared image simulation technology is brought up to solve this problem efficiently. The principle of deep space targets infrared imaging was studied. Based on the infrared sensor's optical properties, a hierarchical imaging model was built. The infrared camera and all the effects were simulated respectively, including motion trail of target and space objects, blurring, dispersion, blind elements, and noise. A mixed noise model was introduced by combining the random noise and Perling noise model. In the image simulating process, Graphic Processing Unit was used to produce noise image in real time. According to the reference photo of infrared sensors, infrared simulated images were evaluated using histogram distribution, the trend of intensity, and Signal to Noise Ratio, and the results show these images satisfied targets recognition algorithm.

Keywords: deep space targets, infrared optical properties, infrared imaging, graphic processing unit, Perling noise

1 Introduction

According to the infrared radiation characteristics of the target and the background, infrared image simulation technology produces infrared thermal image, this is very significant for the target tracking and recognition algorithm. The first principle model, proposed in 1988, considers infrared features of objects in different conditions, by solving equations to get radiation intensity distribution on the surface of the object [1]. But as solving the first principle model is very complicated and unpractical, a semi-empirical first principle model is proposed.

With the development of the infrared image simulation technology, the reality of infrared image becomes more and more strong. In 1996, in order to overcome the problem in lack of reality sense and measured infrared data, people rendered the scene and identified the target by the synthesis of real images and artificial images. In 2005, Gaussian proposed an infrared imaging simulator IRISIM, which can simulate the complete imaging process of the broadband and multispectral infrared imaging system, and evaluate the reliability of the simulation image [2].

Currently, the domestic research adopts the mature semi-empirical first principle model. In studying the effect of the infrared detector on imaging, domestic researchers have obtained certain achievements, including optical dispersion, detector noise, etc, and set up some models and methods [3]. Based on these models and methods, fully considering the influence of various infrared detectors imaging effect, this paper proposes the hierarchical model of imaging, and mixed noise model is established. The whole process implemented with the GPU, provides an effective way for the study of deep space target detection recognition algorithm.

2 Layered model for infrared detector imaging

Different from the traditional scanning detectors, the infrared staring detectors' imaging lack some factors such as blur, which is essential in the real world. To generate the final infrared image, the dispersion effect, blur, noise and blind pixel need to be added to the image.

Therefore, imaging can be layered in modelling, to simplify the modelling process. A complete infrared image can be divided into six layers. As shown in Figure 1.

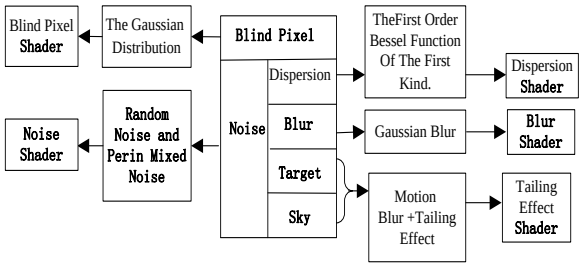


FIGURE 1 Layered model of infrared detector

The top is blind pixel, which will cover all the lower objects. The blind pixel conforms to the Gaussian distribution. The closer to the neutral, the lower the risk of producing blind pixel is. The blind pixel is achieved by blind pixel Shader. The blind pixel is divided into two branches, the relationship between which is not covered

*Corresponding author-e-mail: li_weiqing@139.com

with each other, but for a certain pixel, the larger brightness value of two branches is obtained as the final brightness value. The left branch is noise. The noise of infrared image is simulated in the way of mixing random noise and Perling noise, and produced by the noise Shader. The right branch is top-down with four levels: blur, dispersion, target and stars. The dispersion phenomenon conforms to the trend of the first-order the first kind of Bessel function, which is implemented by blur Shader; blur phenomenon is simulated with Gaussian blur function, and implemented by blur Shader; Target and stars produce tailing effect, which is simulated by the motion blur algorithm and OpenGL camera Settings for the calculation of the interaction, and implemented by the tailing effect Shader.

3 Simulation of infrared characteristics for deep space target

Due to the difficulty of the deep space target intercept test in the reality, which spends high cost and a lot of manpower, the way of simulation experiment is adopted.

3.1 THE TARGET TEMPERATURE FIELD CALCULATION WITH FINITE VOLUME METHOD

Deep space target temperature directly determines the target radiation condition. Target surface temperature is affected by many factors. To study the influencing factors, a deep space target surface thermal equilibrium relation is established and the equation is solved. Therefore, the target surface element temperature and the target surface temperature field distribution is obtained.

The system mainly uses the finite volume method [4] for target temperature field calculation. The basic idea of finite volume method is a series of grid in the calculation area, which does not overlap with each other, is divided, and each grid point is surrounded by a control volume. Each control volume of the differential equation is integrated, and then a set of discrete equations are obtained. The biggest advantage of the finite volume method is that even in the condition of coarse grid; the accurate integral conservation is also shown.

The main source of external thermal radiation that the target in the deep space environment received is as follows:

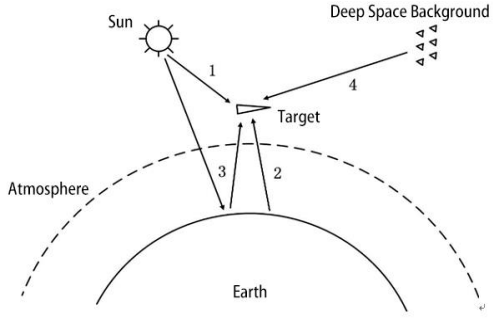


FIGURE 2 Thermal environment of the deep-space targets among them

Among them:

1 is the direct radiation from sun to target Q_{sun}, 2 is the radiation from earth to target Q_{earth}, 3is the reflection of the sun from the earth radiation Q_{ErefS}, 4 is the deep space background radiation Q_{space}.

When a target surface unit receives the radiation and is regarded as the research object, it also exchanges the heat source $Q_{exchange}$ with the nearby unit and internal heat source Q_{inside} of heat. Target surface unit, at the same time, under the influence of various thermal itself, also radiate heat energy, remember to Q_{self}. All of them are factors that make effects on the change of the target surface internal energy Q_{tempr}. The thermal balance equation is established by selecting a binning on targets.

Because deep space background in addition to the sun and the earth's makes little effect on the target radiation Q_{space}, it will be ignored and a balance Equation is obtained:

$$Q_{self}+Q_{tempr}=Q_{sun}+Q_{earth}+Q_{ErefS}+Q_{exchange}+Q_{inside}. \quad (1)$$

Then, balance Equation is established for each binning. As there is T4 factors in the thermal factors of binning, the difference method [5] is used in the linearization of equations, and solution is obtained according to the linear equations. There are many methods of solving linear equations. The methods often used are: Jacobi iteration method, the fourth-order Runge-Kutta method and Gauss-Seidel iterative method. Gauss-Seidel iterative method is used in the system, to calculate the temperature field distribution on the surface of the target on every moment.

3.2 TARGET RADIATION CALCULATION

From the perspective of deep space detector on target, the radiation it receives is related with energy and radiation on the surface of the target, the target reflection, and the target distance, detecting angle and so on. Thus these factors will be analysed.

3.2.1 The radiation factor of itself

The radiation of target unit is related with the temperature of the unit. When the temperature of the target is high, the radiation increases. When the temperature reduces, the radiation will be reduced accordingly. Since infrared radiation of the target is distributed at different wavelengths, it only needs to study the radiation of observed wavelengths. With an optical variable to describe the target, using Planck's law to describe the spectral distribution of blackbody radiation E_λ:

$$E_\lambda=\frac{c_i\lambda^{-5}}{\exp(\frac{c_2}{\lambda T})-1}, \quad (2)$$

among them: T is temperature; λ is wavelength; c_1 is the first radiation constant, value is 3.742×10-6 W m2; c_2 is the second radiation constant, value is 1.4388×10-2 m K.

With the integration of the infrared sensors from floor to ceiling, effective radiation flux for sensor of the target unit itself outward radiation can be obtained:

$$E_{Self} = \int_{\lambda_1}^{\lambda_2} \varepsilon_\lambda E_\lambda d_\lambda , \tag{3}$$

where λ_1 is infrared wavelength floor; λ_2 is infrared wavelengths ceiling; ε_λ is the surface spectral emissivity of target unit.

3.2.2 Reflection radiation factors

The calculation of target unit reflection from the external radiation is complicated. The direct radiation of the sun on the target Q_{sun}, the radiation of the earth on the target Q_{earth}, the sun radiation from the reflection of the earth Q_{ErefS}, and the radiate on from the other units Q_{others}.At the same time the effective receiving band range of the detector is also taken into account. The target unit calculating formula for the reflection of radiation is:

$$E_{reflect} = \rho_s K_s \left(Q_{sun} + Q_{ErefS}\right) + \rho_l K_e (Q_{ErefS} + Q_{others}) , \tag{4}$$

among them: ρ_s is the reflectivity of the target unit to the infrared solar radiation; ρ_l is the reflectivity of the target unit to the infrared earth radiation; K_s is the proportion of the total radiant energy from solar radiation in the infrared radiation of the detector; K_e is the proportion of the total radiant energy from earth radiation in the infrared radiation of the detector.

3.2.3 Detector factor

Detector receives the radiation flux associated with the target and the radiation flux, also with the environment, and the target distance and Angle. That is:

$$H = \sum_{i=1}^{N} \mu_i A_i E_{Mi} , \tag{5}$$

among them: N is the total number of factors, μ_i is the pupil coefficient of unit normal and the angel of the detector; E_{Mi} is the radiation flux of target unit, the value is:

$$E_{Mi} = E_{self} + E_{reflect} \tag{6}$$

Using optical device to focus infrared energy of objects in the scene's on the infrared detector, and then transforming infrared data from each detector element into standard video formats, they can be displayed on the standard video detector, or recorded on videotape. As a result that the infrared detector is heat rather than light, they can be used all-day long.

4 Infrared image simulation based on GPU

4.1 INFRARED DETECTOR IMAGING

Temperature change rate of the object is reflected in the image grayscale change rate. When the temperature of the object changes, the imaging grayscale of warhead target and decoy change over time, then we can obtain the grayscale time series and grayscale change rate.

4.2 INFRARED IMAGING EFFECTS

4.2.1 Tailing

In the real world, the infrared detector exposure is not instantaneous, this will result in that images it captured have tailing phenomenon. In addition, as interceptors constantly do orbit adjustment and attitude adjustment, Astronomical do back and forth motion relative to the infrared detector, this also produces tailing phenomenon.

To simulate tailing phenomenon, there are two approaches as follows:

1) Directly tailing phenomenon is seen as Motion blur to deal with. This method is simple, in each frame captured image, apply motion blur algorithm to target point to achieve the trailing effect. As shown below:

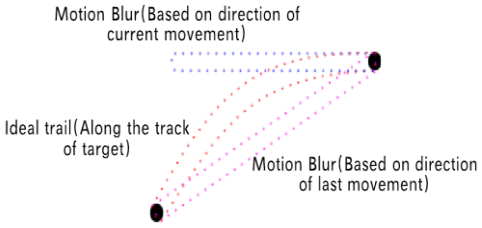


FIGURE 3a Tailing simulated by Motion blur

2) Set the camera so as not to remove the previous frame image when drawing the next frame image. This approach is slightly more complicated. Firstly several intermediate frames need to be inserted between each two frames. Provided frame A_1, A_2, A_3 are inserted between the original frame A and frame B. When rendering frame A_1, frame is not output to the screen, instead is output to a separate frame buffer to store until A_2, A_3 and B rendering are completed, then the image in the frame buffer is displayed on the screen. As shown below:

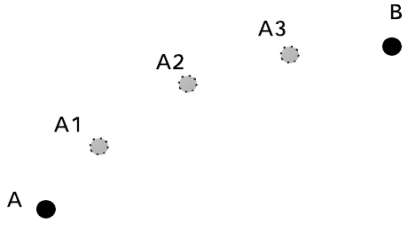


FIGURE 3b Tailing simulated by camera features

We combine the above two methods, namely firstly use Method 2 to insert intermediate frames, and each frame image output to the frame buffer, while use Method 1 to do motion blur for each frame image. This guarantees that the images were real enough.

4.2.2 Blur

Blur mainly refers to the Gaussian blur. Noticing that images outputted by the computer simulation are too sharp, but in real life there is always some blur exciting in the edges of the image captured by camera. In order to simulate this phenomenon, we need to apply a slight Gaussian blur to the image, so that the edge of objects in the image is slightly soft, close to the effect of the real world.

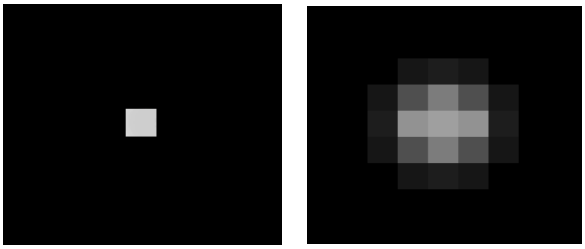
Gaussian blur uses probability density function of normal distribution to calculate the weight of the surrounding pixels, which is defined as follows:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-u)^2}{2\sigma^2}). \tag{7}$$

We use the standard normal function to generate natural effect, when $\mu=0$, $\sigma=1$ this is probability density function of standard normal distribution as follows:

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2}). \tag{8}$$

Point target in infrared image before and after joining blur is displayed as follows:



a) before adding blur b) after adding blur

FIGURE 4 Blur effect

4.2.3 Dispersion

In the background of deep space, the main impact on the size of dispersion circle is the diffraction and aberrations. Aberrations can be weakened as much as possible by improving the optical system. But the diffraction phenomenon is the inevitable result of the spread of radiation, cannot eliminate.

Optical instruments generally use circular aperture [6], for dispersion energy of circular aperture diffraction calculation Equation is:

$$I(\omega)=|U(\omega)|=I_0\left[\frac{2J_1(k\alpha\omega)}{k\alpha\omega}\right]^2, \tag{9}$$

where I_0 is the centre intensity of dispersion circle, $k=2\pi/\lambda$ is wavelength constant, α is the diameter of incident aperture, ω is the angle, $J_1(\ )$ is the first order the first kind of Bessel function.

Assuming the radius of incident aperture is α , the radius of reflector is $\varepsilon\alpha$, the diffraction energy distribution is:

$$I(\omega)=|U(\omega)|=\frac{I_0}{(1-\varepsilon^2)^2}\times \left\{\frac{2J_1(k\alpha\omega)}{k\alpha\omega}-\varepsilon^2\left[\frac{2J_1(k\in\alpha\omega)}{k\in\alpha\omega}\right]\right\}^2. \tag{10}$$

The second factor impacting dispersion circle is aberration. In the simulation, we can simply use uniform optical defocus to simulate the impact caused by aberration by adding defocus coefficient μ in the above function. For reflection system can be written as:

$$I(\omega)=|U(\omega)|=\frac{I_0\mu^2}{(1-\varepsilon^2)^2}\times \left\{\frac{2J_1(k\alpha\mu\omega)}{k\alpha\mu\omega}-\varepsilon^2\left[\frac{2J_1(k\in\alpha\mu\omega)}{k\in\alpha\mu\omega}\right]\right\}^2. \tag{11}$$

This can reflect the diffusion phenomena caused by aberration without changing the basic distribution of the diffraction. The implementation of dispersion phenomenon is estimated and simplified in the system, Dispersion effect as shown below:

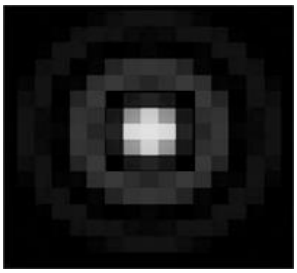


FIGURE 5 Dispersion effect

4.2.4 Blind pixel

Blind pixel is also known as failure, including death pixel and overheating pixel. Performance on the infrared image, death pixel is a black spot for all time, while overheating pixel is white dot forever. According to the project demand, system just simulate white dot.

Central area of infrared detectors undertakes main task of recording the target position. Therefore, blind pixels of central area are less and obey the Gaussian distribution, namely the closer from the centre, the smaller odds of blind pixel. However, the peripheral blind pixels are more and obey uniform distribution. Blind dollar rate for the entire focal plane is about 0.1%.

4.2.5 Noise

Noise can be divided into two parts to simulate. One part is the white noise shown in the uniform distribution. The other part use Perling to simulate the lower image noise of "piece" light and dark areas. Then the brightness values of the upper is remapped around 1.0, and superpose the brightness values of the lower image again to get close to the true noise pattern.

1) Random Noise

The upper noise is easier, regarded as the effects of white noise. Then the white noise has nothing to do with the frequency, and is even distributed on the whole infrared image. The brightness range of your white noise is set as the maximum and the minimum brightness, and the two values are between 0 and 1. The implementation only needs to produce the brightness value of the image between the maximum value and the minimum value with the random number generator.

Because generation of the noise is accomplished in the GPU Shader language and Shader does not directly provide random noise function, this system uses the following hash function to generate pseudo random numbers:

```
float rand(vec2 seed){
return fract(sin(dot(seed.xy, vec2(12.9898,78.233))) * 43758.5453);
}
```

When called, the *x*, *y* coordinates of points on the image are introduced to the function as a seed vector, to get the brightness values of point. Then we obtain the more-random image.

2) Perling noise

It needs to use Perling noise function to generate the lower noise. Perling noise is mainly used to simulate real objects, to replace the general random number generator to produce hard and rough result.

Ken Perling proposed a simple method of one-dimension noise function in 1985, called as classical Perling noise. It can be generalized to the two-dimensional case. The implementation steps are as follows:

1) A value and a gradient is specified for all integers (x, y) coordinates, and the space is divided into many squares;

2) In the plane, to coordinate for integer points, the value specified for it is set as a value of a point; for points inside square, we make the interpolation with four vertices values and gradient in squares.

For example, for point (x, y), i and j are integers which x and y are respectively made down. The square four vertices are: (i, j), $(i+1, j)$, $(i+1, j+1)$, $(i, j+1)$ .Make $u = x - i$, $v = y - j$, and then the contribution of the four vertices to the point (x, y) can be achieved with the dot product in the gradient $(g_{00}, g_{10}, g_{11}, g_{01})$ and (x, y) point and the direction of the four vertices $((u, v), (u-1, v), (u-1, v-1), (u, v-1))$. But in the case of two-dimension interpolation is more complex, and the number of interpolation grows as the growth of the dimension exponential growth.

In order to simplify the calculation and use the hardware implementation easier, Ken Perling improved his method in 2001 and the improved noise is called Simplex noise. The algorithm of Simplex noise effectively reduces the number of interpolation.

As mentioned above, the two-dimensional classic Perling noise makes that the two-dimensional space is filled with a square, using four vertices used with three interpolations, and Simplex noise makes that the two-dimensional space is filled with an equilateral triangle, using three vertices with interpolation.

4.3 INFRARED IMAGE GENERATION BASED ON GPU

We use GPU programming, specifically Shader programming for image simulation. The Shader can't save intermediate results in memory, we can use a technology known as the "render to texture" (RTT), to save the intermediate results rendered in a certain extent, and used to image transmission in the multiple Shader.

Concrete implementation steps are shown in the Figure 6 below:

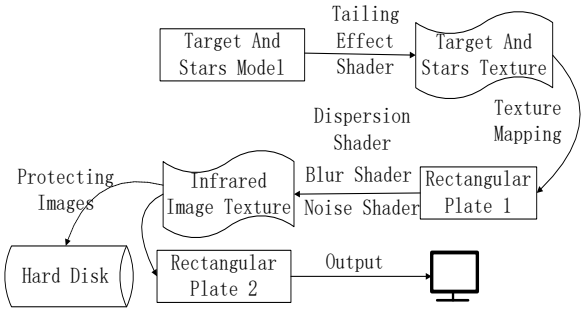


FIGURE 6 Implementation steps based on Shader

Because in the middle dispersion and blur effect needs to be based on the current coordinate view around the point, and the dispersion is to check a certain distance from the current point on the existence of a target. If any, the current point assignment is set; Blur is according to the brightness of the points around and reset the brightness of the current point. So we need to save target and star to join the tail effect to a texture, that it is convenience to check dispersion and blur query Shader. And noise Shader does not need to use the result of a step, so the noise Shader does not introduce another texture.

We can notice that a system is a requirement to save and read infrared image, or store it in hard disk, or sent to the network, so the above is the two nodes "infrared image texture" and "rectangle plate 2". If there are not the two nodes, the infrared image directly output to the screen, and cannot be read.

5 Simulation results evaluation

5.1 SIMULATED PICTURES OF INFRARED TARGETS

The following figures show the real-time infrared simulation images (the original image is too big, the picture is cropped):

a) The 100th frame b) The 800th frame c) The 900th frame

FIGURE 7 Simulation of infrared images

These pictures show the change of the target in infrared image after the start of the simulation. Target is getting closer, and the target point is also growing. Noticed that the tiny white spot in the background is blind pixel, the noise also can be clearly seen.

5.2 THE EVALUATION OF INFRARED IMAGE

In the literature [7], it gives a deep space infrared background image by a certain type of infrared detector:

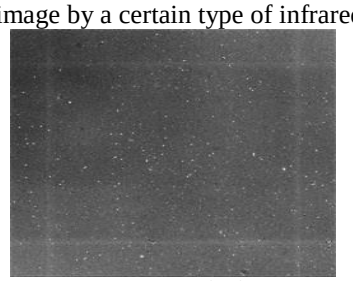


FIGURE 8 Real images of infrared detector

Considering the final goal of the system is target recognition, the infrared image generated by our system should meet an indicator, which is suitable for an ideal target recognition algorithm. We can assess simulated images based on it. As shown in Figure 8, using infrared background image as a reference image, compare it with the simulation image, thereby assess the simulation results.

1) The assessment about characteristics of histogram distribution.

What target recognition algorithm first need to do is eliminating noise in the background. However, filtering background noise is related to the histogram of images. The histogram of referential image as shown below:

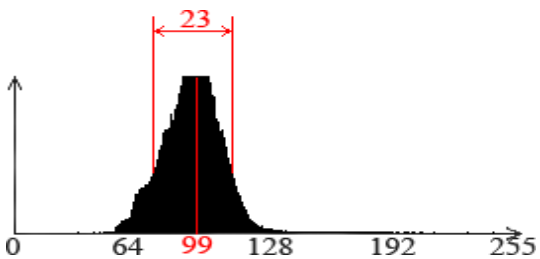


FIGURE 9 Histogram of real deep space infrared background image

The luminance values of reference image is trapezoidal distribution, middle value is 99. Among the pixels, whose luminance value is more than the half of middle value, the difference between max luminance value and min luminance value (hereinafter referred to as "the width of the ladder") is 23.

Histogram distribution of simulation image is as follows:

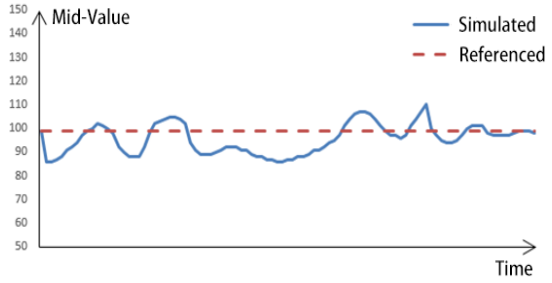


FIGURE 10 Comparison of middle value

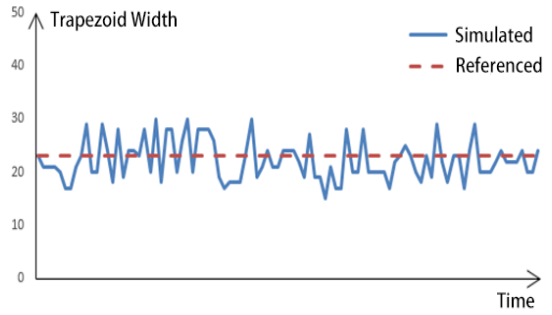


FIGURE 11 Comparison of trapezoidal width

In the simulation image, the histogram changes along with the noise. But the change is always kept within a certain range. The middle value of the simulation image has always been in the vicinity of 99, the maximum deviation is -13.13%, average deviation is -3.43%. The standard deviation of the difference was 0.05; Keystone width of simulated image is always around 23. The maximum deviation is -34.78%, average deviation is -3.52.%, the standard deviation of the difference was 0.13. The histogram distribution of simulation image is close to reference image, we can consider that the simulation image truthfully reflects the infrared image.

2) SNR evaluation of infrared images

In this experiment, we assume the point target is a pixel, and the luminance value is 255. We have a random sample of ten in a series of simulation images to compare with the real image, SNR of real images is 27.74 dB and SNR of simulation image is shown as below:

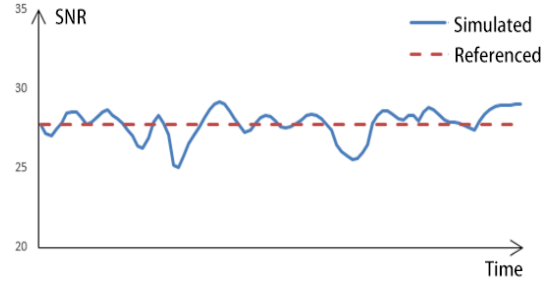


FIGURE 12 Comparison of SNR

The SNR of simulated image fluctuated around 27.74dB, the maximum deviation is -9.73%, average deviation is 0.27%, the standard deviation of the difference was 0.02. The SNR of simulated image is close to the reference image, and we can consider that the simulated image is close to real infrared image.

6 Conclusions

In this paper, according to the imaging principle of infrared detectors, we set up a hierarchical imaging model and give a method for simulation of targets, stars, blur, dispersion, blind pixel, as well as noise. The modelling of noise is discussed in detail. We implementation this method based on GPU to generate more realistic infrared images. According to reference image of an infrared detector, the histogram trend of infrared simulation images, and the signal to noise ratio were evaluated. Simulation results verify the modelling method and show these images satisfied targets recognition algorithm.

References

[1] Cathcart J M 1988 Target and background infrared signature modeling for complex sythetic scenes *SPIE Infrared Systems and Components* 1988 **890**
[2] Guissin R, Lavi E, Palatnik A, Gronau Y, Repasi E, Wittenstein W, Gal R, Ben-Ezra M 2005 IRSIM: infrared imaging simulator *Proceedings of SPIE* **5784** 190-200
[3] Zhang Z, Guo T, Lu H 2007 Simulation study of inherent effects of infrared sensor *Journal of System Simulation* **19**(3) 601-4
[4] Liu J, Shang H M, Chen Y S 1999 Parallel simulation of radiative heat transfer using an unstructured finite-volume method *Numerical Heat Transfer Part B Fundamentals* **36**(2) 115-37
[5] Cheng Jinfa The solution of fractional difference equations of order (k, q), Xiamen: *Acta Mathematicae Application Sinica* **34**(2) 314-30
[6] Mai W 1979 Optical Transfers Function and Mathematics Foundations *Beijing National Defense Industry Press*
[7] Herring J, Bollengier L, Madajian D 1998 Staring 256 x 256 LWIR Focal Plane Array Performance of the Raytheon Exoatmospheric Kill Vehicle *Alexandria USA ASTIA ADA 400061* 10-6

Authors

Weiqing Li, China

Current position, grades: assistant professor in School of Computer Sci. & Eng., Nanjing University of Sci. & Tech, China.
University studies: Doctor's degree on Computer Application from Nanjing University of Sci. & Tech in 2007.
Scientific interest: virtual reality, human machine interface, computer graphics

Ranran Xu, China

Current position, grades: academic master candidate in Nanjing University of Sci. & Tech, China.
University studies: bachelor's degree on Software Engineering from QUFU Normal University in 2013.
Scientific interest: virtual reality and simulation system.

Mengyu Yuan, China

Current position, grades: academic master candidate in Nanjing University of Sci. & Tech, China.
University studies: bachelor's degree on Software Engineering from Nanjing University of Sci. & Tech in 2011.
Scientific interest: virtual reality and simulation system.

Analysis based on computer graphic design and visual communication design

Nan Yao[1*], Kaisheng Wang[2], Jin Yu[3]

[1]*Jiangsu Electric Power Company Research Institute, China*

[2]*Yangzhou Power Supply Company, China*

[3]*Computer Science and Technology, Nanjing University, China*

Received 1 October 2014, www.cmnt.lv

Abstract

The development and application of computer graphic design and visual communication design greatly changes people's life. Designers can use design tools, which are highly-advanced and professional to create a broader scope of design and design theme works. In our daily life, computer graphics technology is widely applied into various fields, such as the military, medical, communication, art etc. Computer graphics and image technology has been emphasized in the visual field, which improves the overall effect and the level of visual communication design to a large extent. This article mentioned the definition of the technology of computer graphics and the visual communication, then described the significance of realizing the visual communication design elements, and hoped the results will be helpful to relative fields.

Keywords: computer graphics, visual communication, the overall effect, the significance

1 Introduction

Since the computer generated, the software technology has gradually penetrated into every corner of people's life with further extension and development of modern technology. From the visual communication way, the whole concept of visual communication design elements appeared for a long time, but the theory of visual communication design elements and computer graphics technology proposed in recent decades. If the technology of computer graphics and visual communication design elements are well-linked, it will be more convenient and easily for the viewer to understand the concept and the intrinsic meaning. Therefore, we can say that the computer graphics technology has important significance for the design of the visual communication elements.

2 The definition of computer technology graphics and visual communication design

Technology of computer graphics and visual communication design refers to use a sort of computer technology as a means to do design data processing in its dram, and show the design drawings in the display or in the drawing software. From the perspective of the angle of visual communication design, it refers to the basic elements of text, colour, graphics, layout and other information through the computer technology to create, and we usually call it the computer graphics technology [1]. Visual communication refers to the visual symbols as the channel of information transfer, use visual language as expression and transmission way, which means people use "view" as a means to communicate with others. Visual communication design aims at processing the design content and let it have the visual function. Visual communication design mostly refers to the graphic design, and printed matter as the media, which include book design, corporate image design, logo design, packaging design, advertising design etc. Visual communication design is mainly aimed at colour, graphics, layout, text and other elements of artistic creation, which plays an extremely important role in people's life. Graphics is a kind of visual symbol, and the intonation is intuitive and conscious. Computer graphics and image processing technology have been fully applied in the visual communication fields, and its prosperous development has brought a breakthrough in the visual communication design field.

The rapid development in computer technology today, which can not only enrich people's cultural and spiritual world with applying the computer graphics and image technology in visual communication design, but also play an important role which should not be ignored in the development of the art design [2, 3].

[*]***Corresponding author*** e-mail: nanynyao@163.com

3 The significance of computer graphics technology to the visual communication design elements

3.1 THE SIGNIFICANCE OF COMPUTER GRAPHICS TO REALIZE THE LAYOUT ELEMENTS

It is of great significance of computer graphics technology in the layout elements of visual communication, and the layout elements includes a wide range of fields, such as advertising design, creative imagery which has a close contact with it. Perfect layout on the advertising and image picture can bring extraordinary creative effects. One of the main branches of the art design is the creative visual design, which has obvious business purpose; but it provides people who appreciate design with a broad design space, consequently, a growing number of designers become interested in designing the computer graphics, and their creativity and innovation make the visual communication more creative and has obvious personal characteristics. It is a particular demanding job in advertising photography and hand-painted for the visual communication arrangement elements. The purpose is to use the visual communication design to make the company's packaging style, the behaviour of enterprise idea, management idea, management characteristics, marketing strategy and guidelines into a complete and unified image, and it hopes to make the product layout satisfactory through computer graphics technology [4]. In addition, the imagination and creativity of computer image design is very rich, and it can make the totally unrelated things united by using the powerful graphics processing software, the image can be totally changed by using a variety of replacement and modification technology. For example, we can achieve the effect which exceed our expectations by designing the layout of fruit juice beverage bottles, and we often see the pattern design character with tick juice fall into the mouth, thus it is not only enhance the realistic visual effect, but also make people feel fresh and lovely. There are a number of such graphic design and bring great economic benefits for manufacturers.

3.2 THE SIGNIFICANCE TO REALIZE COMPUTER GRAPHICS AND IMAGE PROCESSING TECHNOLOGY ON COLOR ELEMENTS

Colour is the first feeling for eyes when receive the visible light, so in visual communication, colour element is one important element among them. Computer can adjust and grasp the colour, brightness and purity, then reproduce the three dimensional effect so as to let people observe a thing more intuitively. Consequently, computer designers can make use of the computer graphics and image technology to transmit the rich and colourful life to the viewers, then viewers can know more about the world by the information created by the computer, and applied this information into all works of life. For example, the packaging design is always the first thing to attract consumer's eyes, so we would see the novel and unique packaging design in our daily life. The garment packaging, and food packaging are closely related to communication design of computer graphics [5]. The packaging of goods is the means to realize the value and the use value of goods, and whether we see posters, brochures, or the book cover we are reading, these graphic printings have rich visual effect. To arouse people's interest, the first thing to do is to design the novel colour through computer graphics design software. Only by doing this can arouse more interest, and enhance the understanding of the cultural, intellectual and ideological aspects. As long as we set the product different transparency and colour, different visual effects will be created.

3.3 THE SIGNIFICANCE OF REALIZING GRAPHIC ELEMENTS AND TEXT ELEMENTS

The adjustments and division of the graphics and image are two main parts in terms of the significance of computer graphics and image technology. The acquisition images can be adjusted and processed by computer graphics technology in two aspects. One is good for the design need, which can do effects of filters, colour adjustment and image synthesis processing. The second is to draw with the computer drawing technology, for example, a graphic or image can achieve large number of scattered in vertical and horizontal by technical parameter settings, meanwhile, more changes in colour, transparency, shape ,and these changes are finished in a few seconds. It is impossible to achieve it in such a short time if hand- painted is used. The graphics and image are clear divided into vector graphics and bitmap by using computer graphics and image technology. The bitmap is bit by bit to draw Icon, but vector record and plot the various elements of the geometric figure painting in the digital form. The former can design the rich colour and brightness of the image, while the latter can show things the original contours and lines. In addition, the computer graphic design can also be applied to character design, and it is easy to operate. In the commonly used software, and in the toolbar, as long as you choose to design "front colour" "typeface" "font size", then input the text, then you can do the art-processing. The text will experience a sort of changes after processed by the computer graphic design technology, and it is not only enhancing the visual effect, but also improve the work demand rate. For instance, a variety of text design is not only enhance the advertising effect but also transmit the happiness atmosphere; therefore, computer graphics design can make any form of words realize its potential image value.

4 The field of computer graphic design with visual communication application

4.1 TEXT DESIGN

Character design is the most basic application of graphic design, software operation is simple/ Such as Photoshop, as long as one choose to design "typeface" "font size" "font colour", then select areas on the selected input text box, you can input the text and word [6]. Afterwards, you can do the art processing. When the text encounter image design processed, everything will be changed. It is stressed that the hotkey settings is very skilled in software tools, and some sets of data should be noted. Set the word adjustment ratio of 0，and select the pixels "Ctrl+ Shift+ Q". When we want to trim the Kerning decrease or increase to 20 -1000 ems, please use "Alt+ /".However, we should use "Ctrl+ Alt+ /"to trim the kerning decrease or increase to 100-1000 ems. Computer graphics design can let the text or words have a variety of changes, and improve the work demand, and in addition, it can also enhance the communication visual effect. For instance, a variety of text design is not only enhancing the advertising effect but also transmit the happiness atmosphere.

4.2 PACKAGING AND ADVERTISING DESIGN

When we walk in the market, a shopping canter or a small supermarket, all kinds of packaging of various commodities, and its novel and unique design can often attract the attention of consumers. The packaging of food or clothing cannot exist without the visual communication design of computer graphic image. Packaging is a means to realize the value of the commodity. Whether we are reading the book cover, or see posters, brochures in the street, these graphic paintings have rich visual communication effects. Most of them need computer graphic design to do art processing to let more people interested in it, then achieve a deep understanding of the ideological, cultural and intellectual nature. Different colours and different opacity can convey different visual effect. Creative advertising design and image cannot do without a computer graphic image processing, and it will bring out different creative effect if we do reasonable and exaggerated treatment on the picture. Visual creative design is a branch of the art design, such design usually does not have obvious business purpose, but because it provides a broad space for the general design enthusiasts, so more and more designers began to learn computer image design, creative visual communication and add the individual characteristic and style [7]. Advertising photography and hand-painted are very strict work for visual communication. It is a means that make enterprise business philosophy, behaviour concept, management characteristics, packaging, marketing style guidelines and strategies reach an overall image, the final product can get a satisfied effect after image design modification. In addition, image innovation is the specialty of computer image design, we can make the objects which have no relevance together through the powerful computer graphic image software, moreover, the replacement technology can make the image totally changed. The design of tick coffee falling into the mouth of the characters that enhances coffee cup visual vivid effect, and make people feel fresh and lovely, so graphic design will bring great advertising benefits and economic benefits to the coffee cup manufacturers.

4.3 PAINTING INTERFACE DESIGN

Many illustration design producers often use the pencil to draw a draft, then use the colouring method of graphic design software to draw illustrations, since the software has good function of computer image design and colour painting [8]. In addition, pixel art is very popular in recent years also for designers to use image design works, these works have strong visual impact. The success of IPhone, which is largely attributed to the exquisite interface design, and show the most incisive visual performance, details decide the excellence! The interface of IPhone design is solemn and elegant, and let a person feel it with quality. Computer graphics interface design is no longer an emerging field, which attracts more and more attention, and most designers use the computer graphic software to design the visual communication.

5 Conclusions

This research made a brief analysis and research mainly in computer graphics and image processing technology to realize the significance of visual communication design elements and other aspects. The graphic design and visual communication design can be effectively improved with the help of computer image processing software and provide sufficient design content to the designers, and allow designers to use more processing techniques and processing method on the design of the works to adjust or perfect. In addition, the combination of graphic images and visual communication design is able to create a more vibrant, more appealing works of art, and constantly promote people to use modern tools to show their own thinking and innovative ability

References

[1] Huang L, He T 2010 The Specialty Construction of Computer Aided Design in Visual communication of College *Art Education* (12) 45-6 (*in Chinese*)
[2] Tong J 2011 The Communication Visual Analysis on Teaching Courseware *Art Panorama* (06) 07-08 (*in Chinese*)
[3] Lu Q 2011 Discussion on the Combination of Teaching and Graphic Design Software *CD-ROM Computer Applications and Software* (06) 7-8 (*in Chinese*)
[4] Hou T 2011 Photoshop Auxiliary Colour Composition Teaching Exploration *Art and design (Theory)* (04) 28-9 (*in Chinese*)

[5] Cui J 2012 3D Compute Technology and Application to Modern Film & entertainment Media (24) 56-7 (*in Chinese*)
[6] Jiao X, Duan A, Sun S 2009 Research on the Design of the User Interface *Technology and Digital Application* (10) 160 (*in Chinese*)
[7] Gong L 2012 Computer Graphic Design and Visual Communication Design *CD-ROM Computer Software and Applications* (20) 181 (*in Chinese*)
[8] Nan C 2011 Image, Graphics, Text-From the Visual Angle of the Connotation of Design Thinking *Science & Technology Information* (06) 10013 (*in Chinese*)

Authors

Nan Yao, Jiangsu Province of China

University studies: Master's degree of computer science and technology, Nanjing University in 2005.
Scientific interest: software engineering, power and application technology of video, electric power information and communication technology.

Kaisheng Wang, Jiangsu Province of China

University studies: Bachelor's degree of applied mathematics, Nanjing University in 1997.
Scientific interest: digital music teaching and theoretical research.

Jin Yu, Jiangsu Province of China

University studies: Master's degree of computer science and technology, Nanjing University in 2005.
Scientific interest: information security.

Content of smart wireless sensor network security and its network security policy

Xiehua Yu*

Minnan Science and Technology Institute, Fujian Normal University, Fujian, 362332, China

Received 1 October 2014, www.cmnt.lv

Abstract

Wireless sensor network is generally composed by plenty of micro-sensors that arranged on designated area. The supervision of these sensor nodes is used to finish the collecting, disposing and uploading of vast information. However, the security of wireless sensor network has many problems, since the sensor node itself exists plenty of limitation. Aiming at the problems of smart wireless sensor network security as well as the analysis of wireless sensor network node easy been attacked, wiretapped and forged without safeguard, this paper put forward a security policy of smart wireless sensor network. Simple and useful intrusion detection policy was realized from a series of improvement of LEACH protocol of low-energy self-adaptation cluster routing protocol.

Keywords: Smart wireless sensor network security, secure routing, key management, intrusion detection

1 Introduction

According to characteristics of limited self resource, poor computing power and small storage space of wireless sensor network, this paper put forward a security policy of smart wireless sensor network. Through the improvement of LEACH protocol of low-energy self-adaptation cluster routing protocol, and the requirement of low-energy and real-time needed for intrusion detection of wireless sensor network, it also put forward whole network cooperated intrusion detection policy. Low–energy and real-time intrusion detection policy is realized by effectively using base station energy and setting up dynamic parameter.

2 Introduction of Wireless Sensor Network

Wireless sensor network technology is an inter-discipline, which involves in many fields, like computer, micro-electronics, sensor, network, communication, signal processing, etc. With the development of many relevant new technologies, this technology is also rapidly booming [1]. Sensor node is made up of data acquisition module, data processing module, data communication module and energy supply module [2], as shown in Figure 1. Data acquisition module is used for monitoring the collection of information within formulation range and the transformation of information data. Data processing module is used for controlling the disposal operation, routing protocol, synchronization, location, energy management, task management and data infusion of all nodes. Data communication module is used for node to collect data and transmit data. Energy supply module provides the needed energy for the above three modules.

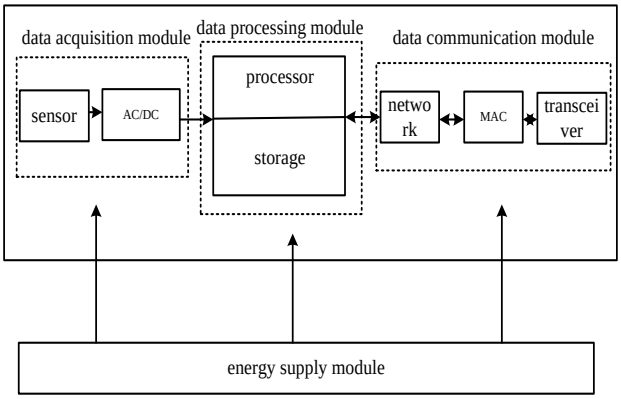


FIGURE 1 Architecture of sensor node

Structure of wireless sensor network is shown in Figure 2 [3]. Wireless sensor network is the newly developing network, which consists of many sensors, reactors and base stations. It finishes the corresponding reaction task and distribution induction with the cooperation of infinite medium. Such kind of wireless sensing actor network must be extensively used in society [4].

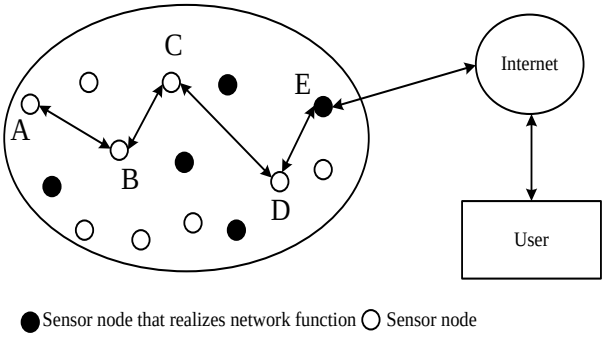


FIGURE 2 Architecture of wireless sensor network system

* ***Corresponding author*** e-mail: yxhxieh@163.com

3 Contents of wireless sensor network security and its technologies

3.1 SECURE ROUTING

The routing algorithms of wireless sensor network are mainly put forward based on characteristics of network itself and energy problems. However, these routing algorithms are not taking security issue into consideration. In order to solve problems like network failure caused by misuse of routing protocol, insecure information transmission, etc, this paper will put forward a safe and reliable routing protocol. Generally, routing security protocol is designed from two large aspects: aspects of message encryption, intrusion detection, identity authentication, etc; the application of multiple path transmissions for providing reliable transmission path.

3.2 KEY MANAGEMENT

Based on wireless sensor key management, the authentication mechanism, confidentiality mechanism, integrity, usability, secure routing, secure localization, etc of wireless sensor network are guaranteed. Key managements based on deployment knowledge, multipath reinforce key, random key and non-symmetric cryptography algorithm are all currently common key management plans [5].

3.3 CRYPTOLOGY

The characteristic of limited energy of wireless sensor network itself will be the main problem of cryptology of wireless sensor network. Low-energy and light weight key algorithm is the characteristic of self-organizing network. More security mechanisms of wireless sensor network are designed based on the key algorithm of wireless sensor network. Many scholars are trying to effectively use public key algorithm on wireless sensor node. With the improvement of technology and the development of wireless sensor technology, the previously not utilized key algorithms begin to be widely accepted and used in wireless sensor network.

3.4 AUTHENTICATION TECHNOLOGY

The authentication technology of wireless sensor network is made up of entity authentication and message authentication. Entity authentication refers to identify user status through method of key management. E-G algorithm and LeaP algorithm are the main representatives of authentication protocol of symmetric key, which are based on the authentication of symmetric cryptography and authentication of identity key mechanism. TinyPk authentication plan is the main representative of the authentication of identity key mechanism. μTESLA protocol is the main protocol of message authentication, of which the function is the guarantee that message is not been forged or falsified [6].

3.5 INTRUSION DETECTION TECHNOLOGY

Intrusion detection includes two kinds of detection models: anomaly detection and misuse detection. The process of intrusion detection is divided into three parts: information collection, information analysis and result processing. Wireless sensor network is subject to be larger intruded, since wireless sensor is generally located in the easily intrusive environment and resource is limited. Intrusion detection is also just emerging. The current intrusion detection technology based on active defense cannot be realized because of the characteristics of wireless sensor itself. Currently, traditional intrusion detection technology cannot apply to wireless sensor network [7].

4 Security Policy of Smart Wireless Sensor Network

4.1 SECURITY POLICY MODEL OF SMART WIRELESS SENSOR NETWORK

Based on the relay node routing protocol of smart energy detection, this model put forward efficient group secrete protocol encryption communication of smart identity authentication matched with it based on cluster structure. In addition, it also added the intrusion detection plan based on smart intrusion detection recognition. This security policy model of wireless sensor network integrated the network security technology of these three aspects together, thus to make this network security policy model more comprehensive, safe and reliable. The model structure is shown in Figure 3.

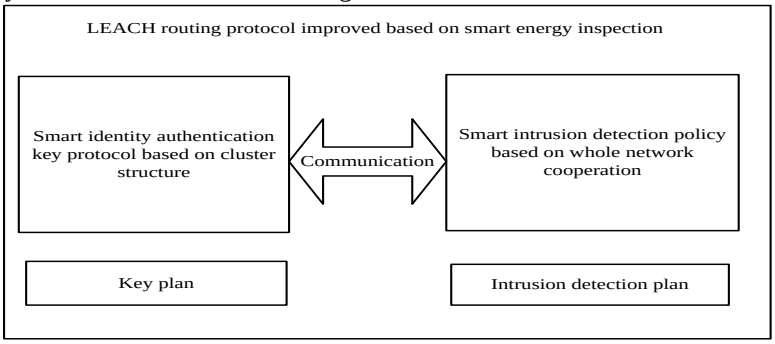


FIGURE 3 Security policy model of smart wireless sensor network

4.2 LEACH ROUTING PROTOCOL IMPROVED BASED ON SMART ENERGY INSPECTION

Routing protocol of wireless sensor network is the carrier of whole network communication, which determines the energy consumption of wireless sensor network and security of network. The current common sensor network routing protocols are all wireless routing protocol. The characteristics of wireless sensor network are greatly different from those of wireless network, thus a more appropriate routing protocol is needed.

The application of this protocol to wireless sensor network may cause the network paralysis with exhausted energy, since LEACH routing protocol consumes large energy on cluster head node. The primarily solved problem is energy consumption that the application of this protocol is needed. This paper put forward the improvement plan based on the routing protocol of smart energy inspection transmission relay mode. Its main content is to set up a relay node and sensor node, which locate on designated detection area. This relay node serves as transit of communication between cluster head node that far away from base station and base station, and it also reveal itself independently in the whole network. This protocol has a kind of smart detection mechanism, and all nodes are equipped with voltage and current detection ability (including relay node). As shown in Figure 4.

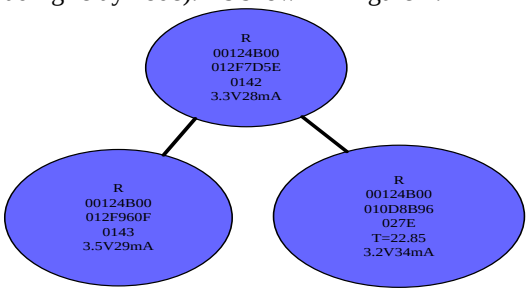


FIGURE 4 Communication mode of wireless sensor

Relay node will detect and record data of cluster head at fixed time. When it detects that cluster head node has little energy, then it will select another node with more energy as cluster head, which will effectively solve problem of energy exhaust of cluster head node. In addition, an independent relay node is established to decrease the intrusion degree of node after been intruded.

4.3 SMART IDENTITY AUTHENTICATION EFFICIENT GROUP KEY AGREEMENT PROTOCOL MODEL BASED ON CLUSTER STRUCTURE

Key management protocol plays an important role in secure communication of sensor. It is responsible for the management task of key, for the development of wireless sensor network also needs larger guarantee for network security. In addition, the application of group key protocol in wireless sensor network has strong reliability. The calculated amount, communication traffic and security of group key agreement protocol also meet the characteristics of wireless sensor network. However, the group key agreement protocol proposed by Burmester and Desmedt (or BD protocol) needs only two round of communication, and its calculated amount is small. BD protocol can resist passive attack brought by the wiretap of external node, but it cannot resist attack from internal node. Based on the above analysis, the application of BD protocol in wireless sensor network is feasible. As for the characteristics of not able to resist internal node attack, this paper put forward a kind of group key agreement protocol of logical key hierarchy in hierarchical structure through the improvement of BD protocol [8].

4.4 SMART INTRUSION DETECTION POLICY MODEL BASED ON WHOLE NETWORK COOPERATION

General intrusion detection model cannot adapt to wireless sensor network, since wireless sensor network have the characteristics of small calculated amount. A da Silva, et al put forward distributed wireless sensor network intrusion detection system, which has synthesized anomaly detection and misuse detection. These two models conducted feature library contrast on attack in network and analysed abnormal data through cooperative work, then found new intrusion feature [9]. These models provided integrated intrusion detection plan, but the plan consumed more energy. Based on the improvement of distributed wireless sensor network intrusion detection model, this paper put forward the intrusion detection plan that suit to cluster wireless sensor network model. Base station implemented complex analysis and detection algorithm through each cluster head node and terminal node. The communication method among nodes was provided through the secure routing protocol of wireless sensor network. At the same time, a simple and lo-energy algorithm was also set up based on intrusion feature comparison for cluster head node as small-scale communication network rendezvous point.

4.5 OVERALL WORKING MODE OF POLICY MODEL

The security policy of wireless sensor network is integrated security architecture. Taking the improved LEACH routing protocol as carrier, an information transmission path is provided for data that need to be send through the networking and information transmission protocol provided for routing protocol, thus to transmit information. In the process of information transmission for routing protocol, key agreement protocol of smart identity authentication and smart intrusion detection policy of whole network cooperation were added to guarantee the security of information transmission, which provided a low-energy security plan of secure transmission for whole information transmission based on smart identity authentication, encryption and decryption of information and smart intrusion detection policy of nodes. The main procedure is shown in Figure 5.

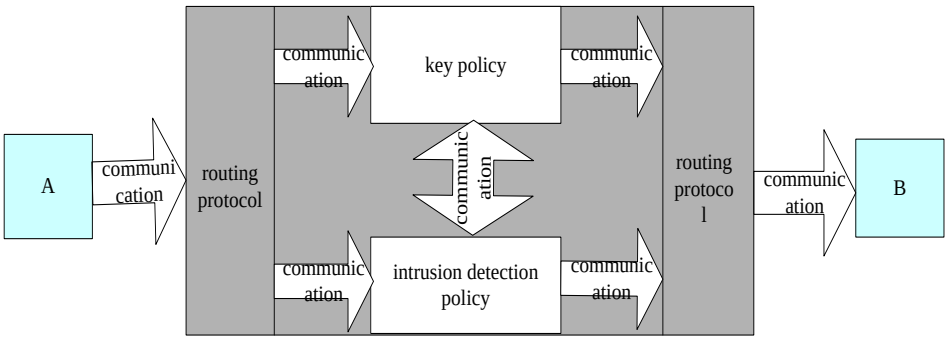


FIGURE 5 Diagrammatic figure of the overall working mode of policy model

5 Implementation of smart wireless sensor network security policy

5.1 IMPLEMENTATION OF LEACH ROUTING PROTOCOL IMPROVED BASED ON SMART ENERGY INSPECTION

According to the analysis and thought of routing protocol based on the relay mode of smart energy inspection, the idea in this paper was realized and verified through experiment and design. Therefore, node was considered based on routing protocol of smart detection relay node. In order to realize the protocol, we also need to provide feasible hardware and interface design plan. The component of wireless sensor are mostly made up of hardware, thus the most of application protocols also need the support of bottom hardware. The analysis was on the need of application layer of network, other communication network layer and physical layer and data link layer that supply hardware support. In addition, communication requirement was also analysed, since communication of protocol has many modules. The primary programming mode of design aid software based on Zigbee protocol was C procedure. Each communication module defines the corresponding function.

5.2 SMART IDENTITY AUTHENTICATION GROUP KEY AGREEMENT PROTOCOL BASED ON CLUSTER STRUCTURE

Burmester and Desmedt put forward the efficient key agreement protocol that only needs two wheel communication processes (BD protocol) [10]. However, this protocol is a non-authenticated group key agreement protocol, which cannot provide authentication function. Zheng Minghui [8] put forward the improved function that added message authentication, which proved the security of protocol under ROM. The identity authentication group key agreement protocol based on cluster structure applied BD protocol that has authentication function to wireless sensor network, which has improved the existing problem of protocol itself. At the same time, it also met the low-energy requirement of wireless sensor network. The requirement of small calculated amount and storage space has met the requirement of wireless sensor network security on data security and node security.

5.3 IMPLEMENTATION OF SMART INTRUSION DETECTION POLICY BASED ON WHOLE NETWORK COOPERATION

Based on smart intrusion detection model, the provided methods for model was analysed and solved according to the detection requirement of wireless sensor network, which verified the idea proposed by the model. According to the characteristics of wireless sensor network, the wireless sensor network security policy that integrated active intrusion detection policy and passive intrusion prevention key policy together was realized with secure routing policy as carrier. It put forward the corresponding implementation method and process of secure routing, key policy and intrusion detection plan.

6 Conclusions

Wireless sensor network is extensively used in modern society. However, such kind of network has many security problems because of the characteristics of wireless sensor network. In order to solve these problems, this paper put forward a security policy of wireless sensor network that integrated three aspects together. Each policy of this module cooperates mutually, and through a set of smart analysis, recognition, authentication and detection mechanism to form a comprehensive security defense and problem solving model for the overall wireless sensor network.

7 Acknowledgements

Supported by Foundation of Fujian Educational Committee (Grant No: JB12280); Supported by Training the Key Members of the Outstanding Young Teacher of Minnan Science and Technology Institute, FuJian Normal University (Grant No: mkq201008)

References

[1] Shao J 2010 Technology and Applications of Wireless Sensor Networks *China Electric Power Press* 3-4 (*in Chinese*)
[2] Akyildiz I F, Su W, Sankarasubramaniam Y, Cayirci E 2002 *IEEE Communications Magazine* **40**(8) 102-14
[3] Yao J, Yang X, Yi J, Han J 2012 Principle and Application of Wireless Sensor Network *Higher Education Press* 51-5
[4] Li J, Gao H 2008 Survey on Sensor Network Research *Journal of Computer Research and Development* **45**(1) 1-15
[5] Levis P, Lee N, Welsh M, Culler D 2005 TOSSIM: Accurate and Scalable Simulation of Entire TinyOS Applications *The 1st International Conference on Embedded Networked Sensor Systems Los Angeles CA USA New York ACM* **30**(1) 122-73
[6] Enge A 2007 Elliptic Curve and Its Application and Guidance in Cryptology *Beijing Science Press* (*in Chinese*)
[7] Li X 2009 Intrusion Detection Method of Wireless Sensor Network. *Beijing University of Technology* (*in Chinese*)
[8] Zheng M 2008 Key Agreement Protocol of Provable Security *Wuhan Huazhong University of Science and Technology* (*in Chinese*)
[9] DaSilva A, Martins M, Rocha B, Loureiro A, Ruiz L, Wong H 2005 Decentralized Intrusion Detection in Wireless Sensor Networks *Proceedings of the 1st ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks Montreal Quebec Canada*
[10] Chen J, Cheng L, Si T, et al 2007 Intranet Security Strategy Based on a Monitor System *Journal of Tsinghua University (Science and Technology)* **47**(4) 606-9 (*in Chinese*)

Author

Xiehua Yu, born in 1982, Hunan Province of China

Current position, grades: associate professor.
University studies: Master's degree of computer science and technology, Central South University in 2004.
Scientific interest: wireless sensor network, artificial intelligence, information and network security.

Duality for multi-objective semi-infinite programming with $K-(F_b,\rho)-$ convexity

Hong Yang*

School of Mathematics and Statistics, Yulin College, Yulin719000, Shanxi, China

Received 1 June 2014, www.cmnt.lv

Abstract

In this paper, some nonsmooth generalized convex functions called uniform $K-(F_b,\rho)-$ convex function, uniform $K-(F_b,\rho)-$ pseudoconvex function, uniform $K-(F_b,\rho)-$ quasiconvex function are defined using $K-$ directional derivative and $K-$ subdifferential. Nonsmooth multi-objective semi-infinite programming involving these generalized convex functions is researched, some Mond-Weir type duality results are obtained.

Keywords: Nonsmooth, Multi-objective Semi-infinite Programming, Mond-Weir type duality, Uniform $K-(F_b,\rho)-$ Convex Function

1 Introduction

The convexity theory plays an important role in many aspects in mathematical programming. In recent years, to relax convexity assumption involved in sufficient conditions for optimality or duality theorems, various generalizations of convex functions have appeared in the literature. Hanson and Mond introduced type I and type II function [1]. Reuda and Hanson extended type I function and obtained pseudo type I and quasi type I function [2]. Bector and Singh introduced $b-$ convex function [3]. Bector, Suneja and Cupta extended $b-$ convex function and defined univex function [4]. Mishra discussed the optimality and duality for multi-objective programming with generalized univexity [5]. Preda introduced $(F_b,\rho)-$ convex function as extension of $F-$ convex function and $\rho-$ convex function [6, 7, 8].

In this paper, we introduce a new classes of generalized convex functions, that is, uniform $K-(F_b,\rho)-$ convex function, uniform $K-(F_b,\rho)-$ pseudoconvex function, uniform $K-(F_b,\rho)-$ quasiconvex functioned. Then we consider nonsmooth multi-objective semi-infinite programming involving these generalized convex functions and obtain some Mond-Weir type duality results.

2 Definitions

Throughout this paper, let R^n be the $n-$ dimensional Euclidean space and R^n be its non-negative orthat. Now we consider the following multi-objective semi-infinite programming problem:

$$(VP)\begin{cases}\min f(x)=(f_1(x),f_2(x),\cdots,f_p(x))\\ s.t.g(x,u)\le 0, x\in X, u\in U\end{cases},$$

where X is an open subset of R^n, $f:X\to R^p$, $g:X\times U\to R^n$, $U\subset R$ is an infinite parameter set.

Let

$A=\{x\mid g(x,u)\le 0, x\in X, u\in U\}, \Delta=\{i\mid g(x,u^i)\le 0, x\in X, u^i\in U\}$, $I(x^*)=\{i\mid g(x^*,u^i)\le 0, x^*\in X, u^i\in U\}$, $U^*=\{u^i\in U\mid g(x,u^i)\le 0, x\in X, i\in\Delta\}$ is any countable subset of U, $\Lambda=\{\mu_j\mid \mu_j\ge 0|, j\in\Delta,$ there is only finite μ_j such that $\mu_j\ne 0\}$.

Notations. If $x,y\in R^n$, then $x\le y\Leftrightarrow x_i\le y_i$, $i=1,2,\cdots,n$, and there exists at least one $i_0\in\{1,2,\cdots,n\}$ such that $x_{i_0}<y_{i_0}$.

Definition 2.1 [9] Let $K(\cdot,\cdot)$ is a local cone approximation, the function $f^K(x,\cdot):X\to R$ with $f^K(x;y):=\inf\{\xi\in R\mid (y,\xi)\in K(epif,(x,f(x)),y\in R^n)\}$ is called $K-$directional derivative of f at x.

Definition 2.2 [9] A function $f:X\to R$ is called $K-$subdifferentiable at x if there exists a convex compact set $\partial^K f(x)$ such that $f^K(x,y)=\max\limits_{\xi\in\partial^K f(x)}<\xi,y>,\forall y\in R^n,$ where $\partial^K f(x):=\{x^*\in X^*\mid <y,x^*>\le f^K(x;y),\forall y\in R^n\}$ is called $K-$subdifferential of f at x.

Definition 2.3 A functional $F:X\times X\times R^n\to R$ ($X\subset R^n$) is called sublinear with respect to the third variable, if for any $x_1,x_2\in X$,

* ***Corresponding author*** e-mail: yhxy888@sina.com

(i) $F(x_1,x_2;a_1+a_2)\le F(x_1,x_2;a_1)+F(x_1,x_2;a_2),\forall a_1,a_2\in R^n$;

(ii) $F(x_1,x_2;ra)=rF(x_1,x_2;a),\forall r\in R, r\ge 0, a\in R^n$.

Definition 2.4 $x^*\subset X^0$ is called an efficient solution for (VP) if and only if there exists no $x\subset X^0$ such that $f(x^*)\le f(x)$.

In the following definitions, we suppose $C\subset R^n$ is a nonempty set, $x_0\in C$, $f:C\to R$ is a local Lipschitz function at x_0, $F:C\times C\times R^n\to R$ is sublinear with respect to the third variable, $\phi:R\to R$, $b:C\times C\times[0,1]\to R_+$, $\lim_{\lambda\to 0^+} b(x,x_0;\lambda)=b(x,x_0)$, $d(\cdot,\cdot)$ is a pseudo-metric in R^n. In [9], Elster and Thierfelder defined $K-$ directional derivative and $K-$ subdifferential and pointed out that $K-$ subdifferential is most generalized. Now we will define some new generalized convex functions using $K-$ directional derivative and $K-$ subdifferential.

Definition 2.5 A function $f:C\to R$ is said to be uniform $K-(F_b,\rho)-$ convex at x_0 with respect to F,ϕ,b,d, if for all $x\in C$, there exists $\rho\in R$ such that

$b(x,x_0)\phi[f(x)-f(x_0)]\ge F(x,x_0;\xi)$
$+\rho d^2(x,y),\forall\xi\in\partial^K f(x_0)$.

Definition 2.6 A function $f:C\to R$ is said to be strictly uniform $K-(F_b,\rho)-$ convex at x_0 with respect to F,ϕ,b,d, if for all $x\in C, x\ne x_0$, there exists $\rho\in R$ such that

$b(x,x_0)\phi[f(x)-f(x_0)]>F(x,x_0;\xi)$
$+\rho d^2(x,x_0),\forall\xi\in\partial^K f(x_0)$.

Definition 2.7 A function $f:C\to R$ is said to be uniform $K-(F_b,\rho)-$ pseudoconvex at x_0 with respect to F,ϕ,b,d, if for all $x\in C$, there exists $\rho\in R$ such that

$b(x,x_0)\phi[f(x)-f(x_0)]<0\Rightarrow$
$F(x,x_0;\xi)+\rho d^2(x,y)<0,\forall\xi\in\partial^K f(x_0)$.

Definition 2.8 A function $f:C\to R$ is said to be strictly uniform $K-(F_b,\rho)-$ pseudoconvex at x_0 with respect to F,ϕ,b,d, if for all $x\in C, x\ne x_0$, there exists $\rho\in R$ such that

$b(x,x_0)\phi[f(x)-f(x_0)]\le 0\Rightarrow$
$F(x,x_0;\xi)+\rho d^2(x,y)<0,\forall\xi\in\partial^K f(x_0)$.

Definition 2.9 A function $f:C\to R$ is said to be uniform $K-(F_b,\rho)-$ quasiconvex at x_0 with respect to F,ϕ,b,d, if for all $x\in C$, there exists $\rho\in R$ such that

$b(x,x_0)\phi[f(x)-f(x_0)]\le 0\Rightarrow$
$F(x,x_0;\xi)+\rho d^2(x,y)\le 0,\forall\xi\in\partial^K f(x_0)$.

Definition 2.10 A function $f:C\to R$ is said to be weak uniform $K-(F_b,\rho)-$ quasiconvex at x_0 with respect to F,ϕ,b,d, if for all $x\in C$, there exists $\rho\in R$ such that

$b(x,x_0)\phi[f(x)-f(x_0)]<0\Rightarrow$
$F(x,x_0;\xi)+\rho d^2(x,y)\le 0,\forall\xi\in\partial^K f(x_0)$.

3 Mond-Weir type duality

We consider the following Mond-Weir type dual programming for (VP):

$$(VD)\begin{cases}\max f(y)\\ s.t.0\in\partial^K(\sum_{i=1}^p\lambda_i f_i)(y)+\sum_{j\in\Delta}\mu_j\partial^K g(y,u^j),\\ \sum_{j\in\Delta}\mu_j g(y,u^j)\ge 0,\\ \lambda_i\ge 0, i=1,2,\cdots,p,\sum_{i=1}^p\lambda_i=1,\\ \mu_j\in\Lambda.\end{cases}$$

Let $W=\Big\{(y,u^j,\lambda,\mu)\,|\,0\in\partial^K(\sum_{i=1}^p\lambda_i f_i)(y)$
$+\sum_{j\in\Delta}\mu_j\partial^K g(y,u^j),\sum_{j\in\Delta}\mu_j g(y,u^j)\ge 0,\ \lambda_i\ge 0,\ i=1,2,\cdots$
$,p,\sum_{i=1}^p\lambda_i=1,\mu_j\in\Lambda,u^j\in U^*,U^*\subset U\Big\}$.

Theorem 3.1 (Weak duality) Assume that $x\in A$, $(y,u^j,\lambda,\mu)\in W$, if for any $\lambda_i>0,i=1,2,\cdots,p$, $\mu_j\in\Lambda, j\in\Delta$, there exist $F,\phi_1,\phi_2,b_1,b_2,\rho_1\in R^1$, $\rho_2^j\in R^{|I|}$ such that

(i) $\sum_{i=1}^p\lambda_i f_i$ is uniform $K-(F_b,\rho_1)-$ convex or strictly uniform $K-(F_b,\rho_1)-$ convex at y;

(ii) $-b_2(x,y)\phi_2[\sum_{j\in I(y)}\mu_j g(y,u^j)]\ge$
$F(x,y;\sum_{j\in I(y)}\mu_j\eta_j)+\sum_{j\in I(y)}\mu_j\rho_2^j d^2(x,y)$,
$\forall\eta_j\in\partial^K g(y,u^j),u^j\in U^*,j\in I(y)$;

(iii) $\alpha<0\Rightarrow\phi_1(\alpha)<0,\alpha\ge 0\Rightarrow\phi_2(\alpha)\ge 0$,
$b_1(x,y)>0,b_2(x,y)\ge 0$;

(iv) $\rho_1+\sum_{j\in\Delta}\mu_j\rho_2^j\ge 0$.

Then the following inequality cannot hold: $f(x)\le f(y)$.

Proof

Suppose that $f(x)\le f(y)$, then there exists i_0 such that $f_{i_0}(x)<f_{i_0}(y)$, $f_i(x)<f_i(y),\forall i\ne i_0$.

Since $\lambda_i>0,i=1,2,\cdots,p$, we have $\sum_{i=1}^p\lambda_i f_i(x)<\sum_{i=1}^p\lambda_i f_i(y)$.

By hypothesis (iii), we have

$b_1(x,y)\phi_1[\sum_{i=1}^{p}\lambda_i f_i(x)-\sum_{i=1}^{p}\lambda_i f_i(y)]<0.$

By hypothesis (i), we get

$$F(x,y;\xi)+\rho_1 d^2(x,y)<0,\forall\xi\in\partial^K(\sum_{i=1}^{p}\lambda_i f_i)(y), \qquad (1)$$

observe that $\mu_j\in\Lambda$ and $g(y,u^j)=0, j\in I(y)$, we have $\sum_{j\in I(y)}\mu_j g(y,u^j)\geq 0.$

By hypothesis (iii), we have $-b_2(x,y)\phi_2[\sum_{j\in I(y)}\mu_j g(y,u^j)]\leq 0.$

By hypothesis (ii), we get $F(x,y;\sum_{j\in I(y)}\mu_j\eta_j)+\sum_{j\in I(y)}\mu_j\rho_2^j d^2(x,y)\leq 0,$

$\forall\eta_j\in\partial^K g(y,u^j),u^j\in U^*, j\in I(y).$

Let $\mu_j=0, \quad j\notin I(y),$ we have

$$F(x,y;\sum_{j\in\Delta}\mu_j\eta_j)+\sum_{j\in\Delta}\mu_j\rho_2^j d^2(x,y)\leq 0,$$

$$\forall\eta_j\in\partial^K g(y,u^j),u^j\in U^*, j\in\Delta. \qquad (2)$$

adding (1) and (2) ,using the sublinearity of F, we can obtain $F(x,y;\xi+\sum_{j\in\Delta}\mu_j\eta_j)+(\rho_1+\sum_{j\in\Delta}\mu_j\rho_2^j)d^2(x,y)<0.$

By hypothesis (iv), we have $F(x,y;\xi+\sum_{j\in\Delta}\mu_j\eta_j)<0.$

So $\xi+\sum_{j\in\Delta}\mu_j\eta_j\neq 0$, which contradicts the first subjective condition of (VD).

Theorem 3.2 (Weak duality) Assume that $x\in A$, $(y,u^j,\lambda,\mu)\in W$, if for any $\lambda_i>0, i=1,2,\cdots,p$, $\mu_j\in\Lambda, j\in\Delta$, there exist $F,\phi_1,\phi_2,b_1,b_2,\rho_1\in R^1$, $\rho_2^j\in R^{|I|}$ such that

(i) $\sum_{i=1}^{p}\lambda_i f_i$ is uniform $K-(F_b,\rho_1)-$ pseudoconvex at y;

(ii) $-b_2(x,y)\phi_2[\sum_{j\in I(y)}\mu_j g(y,u^j)]\leq 0\Rightarrow$

$F(x,y;\sum_{j\in I(y)}\mu_j\eta_j)+\sum_{j\in I(y)}\mu_j\rho_2^j d^2(x,y)\leq 0,$

$\forall\eta_j\in\partial^K g(y,u^j),u^j\in U^*, j\in I(y);$

(iii) $\alpha<0\Rightarrow\phi_1(\alpha)<0,\alpha\geq 0\Rightarrow\phi_2(\alpha)\geq 0,$

$b_1(x,y)>0,b_2(x,y)\geq 0;$

(iv) $\rho_1+\sum_{j\in\Delta}\mu_j\rho_2^j\geq 0.$

Then the following inequality cannot hold: $f(x)\leq f(y).$

Theorem 3.3 (Weak duality) Assume that $x\in A$, $(y,u^j,\lambda,\mu)\in W$, if for any $\lambda_i\geq 0, \quad i=1,2,\cdots,p$, $\mu_j\in\Lambda, j\in\Delta,$ there exist $F,\phi_1,\phi_2,b_1,b_2,\rho_1\in R^1$, $\rho_2^j\in R^{|I|}$ such that

(i) $\sum_{i=1}^{p}\lambda_i f_i$ is strictly uniform $K-(F_b,\rho_1)-$ pseudo-convex at y;

(ii) $-b_2(x,y)\phi_2[\sum_{j\in I(y)}\mu_j g(y,u^j)]\leq 0\Rightarrow$

$F(x,y;\sum_{j\in I(y)}\mu_j\eta_j)+\sum_{j\in I(y)}\mu_j\rho_2^j d^2(x,y)\leq 0,$

$\forall\eta_j\in\partial^K g(y,u^j),u^j\in U^*, j\in I(y);$

(iii) $\alpha\leq 0\Rightarrow\phi_1(\alpha)\leq 0,\alpha\geq 0\Rightarrow\phi_2(\alpha)\geq 0,$

$b_1(x,y)\geq 0,b_2(x,y)\geq 0;$

(iv) $\rho_1+\sum_{j\in\Delta}\mu_j\rho_2^j\geq 0.$

Then the following inequality cannot hold: $f(x)\leq f(y).$

Theorem 3.4 (Weak duality) Assume that $x\in A$, $(y,u^j,\lambda,\mu)\in W$, if for any $\lambda_i\geq 0, \quad i=1,2,\cdots,p$, $\mu_j\in\Lambda, j\in\Delta,$ there exist $F,\phi_1,\phi_2,b_1,b_2,\rho_1\in R^1$, $\rho_2^j\in R^{|I|}$ such that

(i) $\sum_{i=1}^{p}\lambda_i f_i$ is uniform $K-(F_b,\rho_1)-$ quasiconvex at y;

(ii) $-b_2(x,y)\phi_2[\sum_{j\in I(y)}\mu_j g(y,u^j)]\leq 0\Rightarrow$

$F(x,y;\sum_{j\in I(y)}\mu_j\eta_j)+\sum_{j\in I(y)}\mu_j\rho_2^j d^2(x,y)<0,$

$\forall\eta_j\in\partial^K g(y,u^j),u^j\in U^*, j\in I(y);$

(iii) $\alpha\leq 0\Rightarrow\phi_1(\alpha)\leq 0,\alpha\geq 0\Rightarrow\phi_2(\alpha)\geq 0,$

$b_1(x,y)>0,b_2(x,y)\geq 0;$

(iv) $\rho_1+\sum_{j\in\Delta}\mu_j\rho_2^j\geq 0.$

Then the following inequality cannot hold: $f(x)\leq f(y).$

Theorem 3.5 (Weak duality) Assume that $x\in A$, $(y,u^j,\lambda,\mu)\in W$, if for any $\lambda_i>0, \quad i=1,2,\cdots,p$, $\mu_j\in\Lambda, j\in\Delta,$ there exist $F,\phi_1,\phi_2,b_1,b_2,\rho_1\in R^1$, $\rho_2^j\in R^{|I|}$ such that

(i) $\sum_{i=1}^{p}\lambda_i f_i$ is weak uniform $K-(F_b,\rho_1)-$ quasiconvex at y;

(ii) $-b_2(x,y)\phi_2[\sum_{j\in I(y)}\mu_j g(y,u^j)]\leq 0\Rightarrow$

$F(x,y;\sum_{j\in I(y)}\mu_j\eta_j)+\sum_{j\in I(y)}\mu_j\rho_2^j d^2(x,y)<0,$

$\forall\eta_j\in\partial^K g(y,u^j),u^j\in U^*, j\in I(y);$

(iii) $\alpha<0\Rightarrow\phi_1(\alpha)<0,\alpha\geq 0\Rightarrow\phi_2(\alpha)\geq 0,$

$b_1(x,y)>0, b_2(x,y)\geq 0$;

(iv) $\rho_1+\sum_{j\in\Delta}\mu_j\rho_2^j\geq 0.$

Then the following inequality cannot hold: $f(x)\leq f(y)$. The proofs of Theorem 3.2 —Theorem 3.5 are similar to Theorem 3.1.

Theorem 3.6 (Strong duality) Assume that x^* is an efficient solution for (VP), if for any $\lambda_i>0,\ i=1,2,\cdots,p,$ $\mu_j\in\Lambda,\ j\in\Delta,$ there exist $F,\phi_1,\phi_2,b_1,b_2,\rho_1\in R^1,$ $\rho_2^j\in R^{|I|}$ such that

(i) $\sum_{i=1}^{p}\lambda_i f_i$ is uniform $K-(F_b,\rho_1)-$ pseudoconvex at y;

(ii) $-b_2(x,y)\phi_2[\sum_{j\in I(y)}\mu_j g(y,u^j)]\leq 0\Rightarrow$

$F(x,y;\sum_{j\in I(y)}\mu_j\eta_j)+\sum_{j\in I(y)}\mu_j\rho_2^j d^2(x,y)\leq 0,$

$\forall\eta_j\in\partial^K g(y,u^j), u^j\in U^*, j\in I(y)$;

(iii) $\alpha<0\Rightarrow\phi_1(\alpha)<0, \alpha\geq 0\Rightarrow\phi_2(\alpha)\geq 0,$

$b_1(x,y)>0, b_2(x,y)\geq 0$;

(iv) $\rho_1+\sum_{j\in\Delta}\mu_j\rho_2^j\geq 0.$;

(v) The *Kuhn-Tuker* condition is hold at x^*, that is, there exist $\lambda^*\in R_+^p,\mu^*\in R_+^{|I|}$ such that

$$0\in\partial^K(\sum_{i=1}^{p}\lambda_i^* f_i)(x^*)+\sum_{j\in I(x^*)}\mu_j^*\partial^K g(x^*,u^j),$$

$$\mu_j^* g(x^*,u^j)=0,\ j\in\Delta,$$

$$(\lambda^*,\mu^*)\geq 0,\sum_{i=1}^{p}\lambda_i^*=1.$$

Then there exist (λ^*,μ^*) such that $(x^*,u^j,\lambda^*,\mu^*)$ is an efficient solution for (VD), and the optimal values of (VP) and (VD) are equal.

Proof

Since x^* is an efficient solution for problem (VP) and the *Kuhn-Tuker* condition is hold at x^*, therefore, for any $u^j\in U^*$, there exist $\lambda_i^*>0,\ i=1,2,\cdots,p,$ $\sum_{i=1}^{p}\lambda_i^*=1$, $\mu_j^*\in\Lambda, j\in\Delta$ such that

$$0\in\partial^K(\sum_{i=1}^{p}\lambda_i^* f_i)(x^*)+\sum_{j\in\Delta}\mu_j^*\partial^K g(x^*,u^j),$$

$$\sum_{j\in\Delta}\mu_j^* g(x^*,u^j)\geq 0.$$

So $(x^*,u^j,\lambda^*,\mu^*)$ is a feasible solution for (VD).

Suppose that $(x^*,u^j,\lambda^*,\mu^*)$ is not an efficient solution for (VD), then there exist (y,u^j,λ,μ) and i_0 such that $f_{i_0}(x^*)<f_{i_0}(y),\ f_i(x^*)\leq f_i(y),\forall i\neq i_0.$ that is,

$$f(x^*)\leq f(y),\qquad (3)$$

which contradicts the conclusion of Theorem 3.2. Therefore, $(x^*,u^j,\lambda^*,\mu^*)$ is an efficient solution for (VD). Obviously, the optimal values of (VP) and (VD) are equal.

4 Conclusions

In this paper, we introduce a new classes of generalized convex functions, that is, uniform $K-(F_b,\rho)-$ convex function, uniform $K-(F_b,\rho)-$ pseudoconvex function, uniform $K-(F_b,\rho)-$ quasiconvex function, etc. Then we consider nonsmooth multi-objective semi-infinite programming involving these generalized convex functions and proved some weak and strong duality theorems of Mond-Weir type duality.

References

[1] Hanson M A, Mond B 1987 *Mathematical Programming* **37**(3) 51-8

[2] Rueda N G, Hanson M A 1998 Optimality *J.Math.Anal.Appl.* **130**(6) 375-85

[3] Bector C R, Singh C 1991 *J. Optim. Theory Appl.* **71**(5) 237-53

[4] Bector C R, Suneja S K, Gupta S 1992 Univex Functions and Univex Nonlinear Programming *Proceedings of the Administrative Science Association of Canada* Vancouver 115-24

[5] Mishra S K 1998 *J. Math.Anal.Appl.* **224**(11) 131-48

[6] Preda V 1992 *J. Math.Anal.Appl.* **166**(8) 365-77

[7] Hanson M A, Mond B 1986 *J. Inform.Optim.Sci.* **22**(3) 25-32

[8] Vial J P 1989 *Bull.Austral.Math.Soc.* **39**(2) 287-99

[9] Elster K H, Thierfelder J 1988 *Optimization* **19**(3) 315-41

Author

Hong Yang, born on October 19, 1979, China

University study: Hong Yang received the master degree in basic mathematics from Yan'an University, China in 2004. Currently, he is a teacher at Yulin College, China.

Research interests: the theory and application of optimization

A role-based security information flow model in grid environment

Yihe Liu[1], Shuang Zhang[1, 2*], Yuping Qin[2]

[1]*College of computer science, Neijiang Normal University, Neijiang, 641000, China*

[2]*The engineering & technical college of Chengdu university of technology, Leshan, 614000, China*

Received 1 March 2014, www.cmnt.lv

Abstract

Security is an important component of a grid, and it directly affects the development of the grid and the practical application of the grid system software. According to the practical application problem (namely realization of the role-based management) in the role management occurring in the information application system of my school, a role-based security information flow model is proposed from the point of view of guaranteeing the information security. In this paper, the object concept in the general network environment is expended, and the organization security classification of an object and relation between the security classification and the role set are used to classify the security and define the strategy for information flow, finally a security information flow model based on the grid environment is presented. The safe classify of the object is divided by the related information of role set, the organization security classifications, and classifications etc. At the same time, the information flow role is described. A new secure information flow model based on rules and grid environment is described using these methods. It is proven from strict mathematical justification that the new model satisfies properties of the finite lattice and least upper bound operator, and it is reasonable and safe. Furthermore, it is an extension of the BLP model and the role-based information flow model as well as extension of the security information flow model in the general network environment. Therefore, it is significant to the study of grid security.

Keywords: Grid Security, BLP model, Rule, Information flow model

1 Introduction

With the rapid development of the internet technology, thousands of types of high-performance computers are distributed on the internet. How to expand and use these network resources has become the research direction of scientists in the future. This is development prospects of the grid.

The grid security is an important part of the grid, which directly affects development of the grid and practical application of the grid system software. To construct a reasonable information security model is not only the need of information security, but also the necessary means of completing an information security project. At present, there are various information security models, such as BLP model [1] and the information flow model [2]. They have different characteristics, and play a very important role in formal description of information security. Most of the existing [3-5] information flow models are developed based on the reference [2], and all are presented in the general network environment.

Current research on grid security mainly focuses on the security certificate, access control, data integrity, communication confidentiality, non-repudiation of user behaviour, as well as single sign-on [7-9] in the network environment. Grid workflow and other properties are discussed in the reference [10-11]. In the reference [12], the information flow in the network environment is studied. In this study, the information flow model proposed in the reference [12] in the network is modified. Furthermore, role and other concepts are introduced in combination with the reference [6], and information security flow in the network is discussed again from another perspective.

Content of this paper is as follows: the second section introduces related concept and known information flow models. In the third section, new concept introduced in the new model is first presented, such as role decomposition and security function; next security class, definition of expanded subjective & objective bodies, security policy, and definition of the symbol $\oplus$ are introduced finally the model description is given. In the fourth section, explanation of the model is given as well as proving of relevant properties. The fifth section summarizes the paper.

2 Related concept

2.1 GRID AND GRID SECURITY

In the grid environment, different autonomous domains or virtual organizations compose the entire grid computation environment capable of providing the service for the external, But resource nodes in each autonomous domain or virtual organization can cooperate to complete different services, if tasks submitted by the

* ***Corresponding author*** e-mail: zhangshuanghua1@126.com

grid user can not be finished in an autonomous domain or a virtual organization, then the grid server of this autonomous domain or virtual organization will request resource nodes of other autonomous domains or virtual organizations to complete it through cooperation. The physical view of grid security is shown in Figure 1 [13].

In this paper, the discussion is developed on the basis of the GSI security strategy [14] in grid Globus environment.

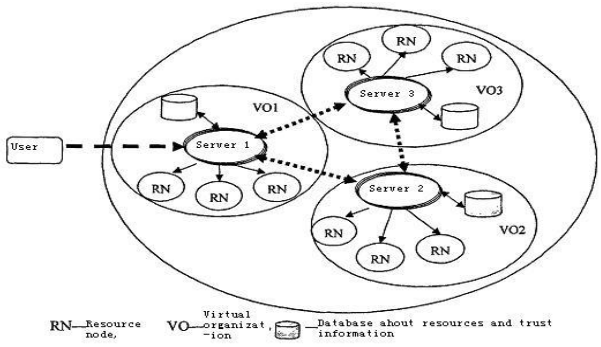


FIGURE 1 Physical view of grid security

2.2 BASIC SECURITY MODEL

2.2.1 Basic concept

Firstly, some known concepts are presented.

Subject and object: there exist a quantity of computer operation involved with the security. anything implementing operation is called the subject, such as process; they are expressed with s_1, s_2, …, s_i,…, or s, and their set is denoted as **S**. anything which is operated is called the object, denoted with o_1,o_2,…,o_j,…,or o, and their set is denoted as **O**.

Suppose there are m virtual organizations in the entire grid Globus environment, a subject represents a user or the user process in the environment. The local resource subject set corresponding to virtual organizations VO_i is denoted with S_VO_i, and the corresponding local resource object set is denoted with O_VO_i (i=1,2,…m).

Security classification: in this paper, security classifications (confidentiality grade) of the subject and object are measured in a specific way [4]. Confidentiality grade of the subject **s** is denoted with T(s), and T(o) indiactes that of the object **o**.

According to BLP model [1], the system is at security status (confidentiality), so the following needs to be satisfied: reading up and writing down are not allowed. For this reason, if T(s) is not less than T (o), then the subject **s** can read the object **o**, and if T(s) is not larger than T (o), then the subject **s** can write the object **o**.

2.2.2 Decomposition of subject and object

Decomposition of subject [12]: when the subject s needs to access simultaneously the virtual organizations VO_1, VO_2, …, VO_m, the resource proxy needs to map a Globus subject to one or more subjects belonging to the local resource, so decomposition of s in a virtual organization is expressed as $sj=(s\_VO_1,s\_VO_2,\ldots,s\_VOm)$, where the jth component indicates the subject in the jth virtual organization mapped from the Globus subject. if sj need not be mapped to the virtual organization VOj, then the corresponding component is noted with s_VOj_φ (j=1,2,..,m) .

Decomposition of Object [12]**:** when the object **o** needs to be decomposed to virtual organizations VO_1, VO_2, …, VO_m, o_j decomposed and expressed as $o_j=(o\_VO_1, o\_VO_2,\ldots, o\_VO_m)$ in the virtual organization, where the j^{th} component is the part decomposed from the object o to the virtual organization VOj, and if the object o need not be decomposed to the virtual organization VO_j, then the corresponding component may be denoted with $o_j\_VO_j\_\varphi$ (j=1,2., m).

When a subject or an object not in a virtual organization, with decomposition expression, the security classification of subject **s** /object o (T(s) /T (o)) is converted to a m-dimension vector function, in which a component can be defined like the related definition in the general network environment, such as: $T(s)=(T(s\_VO_1),T(s\_VO_2),\ldots,T(s\_VO_m))$.

2.2.3 Security classification of organization

Security classification of organization [12]: supposes the existing grid is divided into **m** independent virtual organizations VO_1, $VO_2 \ldots VO_m$ in accordance with the access request, and each virtual organization is given a different security grade, so this is called security classification of the organization, denoted with **net** (Net_name). A net (Net_name) corresponds to the organization classification of virtual organization Net_name, denoted with net (Net_name)=(net(VO_1),net(VO_2),…,net(VO_m)).

Suppose $net(VO_i)=\max_{\forall o \in O\_VO_i}\{T(o)\}$ and when $o=(o\_VO_1, o\_VO_2,\ldots, o\_VO_m)$, $net(o)=\max\{T(o\_VO_1), T(o\_VO_2),\ldots, T(o\_VO_m)\}$.

2.2.4 RBAC model

The basic idea of the RBAC model [5] is responsibility separation, and this is similar to an organization. In the RBAC model, the user is given a role, a role is awarded with permissions, and the permissions are associated with operations, the user gets the relevant permissions of the role through the role, so as to complete some operations. This paper is subject to the most basic concepts of the RBAC model.

Suppose the system has a role set $R=\{r_1,r_2,\ldots,r_n\}$;the role set of the object **o** is denoted with OR(o); While the user owned by the object o through the role is denoted OS (o).

2.2.5 Denning Information flow model

Denning's [2] information flow model is defined by FM=<N, P, SC, $\oplus$, →>, where N is an object finite set, P is a finite process set (subject set), SC is a security class set, the class-combining operator $\oplus$ is an associative and commutative binary operator. For classes of any two operands, $\oplus$ demonstrates the class to, which the result operand generated from any binary function of the two operands belongs. A flow relation → indicates information flow between pairs of security classes. For classes A and B, we write A → B, only if information in class A is permitted to flow to class B. Information is said to flow from class A to class B.

The security requirements of the model are simply stated: a flow model FM is secure if and only if execution of an operation sequence cannot produce a flow that violates the relation →. Under certain assumptions, <SC,→, $\oplus$ > forms a finite lattice:

(1) <SC,→> is a partially ordered set.(2) SC is a finite set.(3) SC has a lower bound of L such that $\forall$ A∈SC, L→A.(4) $\oplus$ is a least upper bound operator on SC.

Where the class-combining operator $\oplus$ is also a least upper bound operator, it has the following properties for all A, B, C∈SC:

(a) A→A$\oplus$B and B→A$\oplus$B.

(b) A→C and B→C$\Rightarrow$ A$\oplus$B→C.

3 A new information flow model based on grid environment

3.1 NEW MODEL DESCRIPTION

In order to give new information flow model under the grid environment, we give some new definitions in the following.

3.1.1 Role decomposition

Similar to decomposition of the object, decomposition of the role set OR (o) owned by the object o may be expressed as follows:

$$OR(o)=(OR(o)\_VO_1, OR(o)\_VO_2, \ldots, OR(o)\_VO_m).$$

Similar to decomposition of the subject,decomposition of the subject OS (o) owned by the object **o** through the role may be expressed as follows:

$$OS(o)=(OS(o)\_VO_1, OS(o)\_VO_2, \ldots, OS(o)\_VO_m)$$

3.1.2 Security function

Object Security function: it is used to describe the ultra four-dimensional vector function relevant to security classification of the virtual organization, security classification, and related roles of the object, denoted as T(net(Net_name), T(o), OR(o), OS(o)) or T(net(o),T(o), OR(o), OS(o)).

The set U indicates the complete set consisting of all ultra four-dimensional security function vectors.

3.2 DESCRIPTION OF NEW MODEL

The grid security information flow model proposed in this paper is defined as follows:

FM=<O,S,SC, $\oplus$,→> is a security information flow model based on the grid environment, in which information flows only from the virtual organization with the low security classification to the one with the high security classification, where O and S are the object set and the process (subject) set of the whole grid. They are finite sets. SC is the security class of the model, defined in Section 3.2.1.The strategy → is defined in Section 3.2.3. The operation $\oplus$ is defined in Section 3.2.4.

The relevant concepts are described and discussed as follows.

Suppose that O' is an actual object set and S' is an actual subject set (processes set) of the whole grid, they are finite sets; SC' is the set of actual security classes.

U' is the set of all actual ultra four-dimensional security function vector defined in Section 3.1.2.

SC' is the set of security classes divided according to all actual ultra four-dimensional security function value, and obviously it is a finite set, and the U'$\subseteq$ U.

3.2.1 Security class #o

Definition 1: the security class of the object o, indicated #o, #o={ o| o∈O, #o={o| o∈O,T(net(o),T(o), OR(o), OS(o)) are invariable}, as $\forall$ o'∈#o, there are T(net(o'),T(o'), OR(o'), OS(o'))＝T(net(o),T(o),OR(o), OS(o)))

Hence SC'= {#o|o∈O'}, SC' is a finite set.

We will discuss U' and U, and expand the concept of O', S', SC'.

3.2.2 Empty object and empty subject

Empty object: $\forall$ α ∈ U, but α ∉ U'. Then we will expand a object **o,** which satisfies T(net(o),T(o), OR(o), OS(o)))= α . Except name and the ultra four-dimensional security function vector, the object has no other characteristics. When a subject accesses the object, no available information is leaked. For example, an empty file has only name and we say that the object is an empty object. An empty object set is denoted with O''.

Empty subject: any object performing operation related to the above empty object is called the empty subject and its set is denoted as S''.

$\forall$ o∈O'', then SC''＝{#o|o∈O''}.

The previously described $s_i\_VO_j\_\varphi$, $o_j\_VO_i\_\varphi$ can actually be considered empty subject and empty object

respectively.

$\forall$ o $\in$ O",then SC" = {#o|o$\in$O"}.O=O' $\cup$ O", S=S' $\cup$ S", SC=SC' $\cup$ SC".

When expanding the concept of O', S', SC', we make use of them together with the general concept, don't make any distinction between them.

Now they are denoted again: O=O' $\cup$ O'', S=S' $\cup$ S'', SC=SC' $\cup$ SC''.

Where O is a object set, S is a subject set (processes set) for total grid network, SC is a security classes set, and they are finite sets.

3.2.3 Description of security policy

Definition 2: security flow policy:

For any objects o_1 and o_2, if information can flow from o_1 to o_2, then the security class of o_1 is controlled by o_2's, or they are equivalent, namely:

Axiom: $\forall$ $o_1,o_2 \in O, o_1 \rightarrow o_2$ if and only if

$net(o_1) \leq net(o_2), T(o_1) \leq T(o_2), OR(o_1) \supseteq OR(o_2)$ and $OS(o_1) \supseteq OS(o_2)$, both of them are true.

Further explanation is presented as follows:

- Meaning of $\cdot \leq, \geq, \supseteq$ operations

$\leq, \geq$ were comparison operation of "less than or equal to" and "greater than or equal to" of real numbers, $T(o_1) \leq T(o_2)$ is defined as follows: corresponding components of two **m**-dimension vectors are used for computatoin through $\leq$ operation. $T(o_1) \geq T(o_2)$ has the similar definition.

The operator $\supseteq$ represents containing operation between two sets. $OR(o_1) \supseteq OR(o_2)$ and $OS(o_1) \supseteq OS(o_2)$ are defined as follows: the sets of the corresponding components of two m-dimension vectors are used for computation through "$\supseteq$" operation.

- **The object is not defined in the decomposition of a virtual organization**

When $o_1\_VO_i\_\varphi$ is a component of the source object O_1, the definition in this paper is presented as follows:

$$T(o_1\_VO_i\_\varphi) = \min_{j\neq i, j\in\{1,2,\cdots m\}}^{m} \{T(o\_VO_j)\}$$

$$OR(o_1\_VO_i\_\varphi) = \bigcup_{j\neq i, j\in\{1,2,\cdots m\}}^{m} OR(o_j)$$

$$OS(o_1\_VO_i\_\varphi) = \bigcup_{j\neq i, j\in\{1,2,\cdots m\}}^{m} OS(o_j)$$

When a component of objective object O_2_is VOj_φis the definition in this paper is presented as follows:

$$T(o_2\_VO_j\_\varphi) = \min_{j\neq i, j\in\{1,2,\cdots m\}}^{m} \{T(o\_VO_k)\}$$

$$OR(o_2\_VO_j\_\varphi) = \bigcup_{j\neq i, j\in\{1,2,\cdots m\}}^{m} OR(o_k)$$

$$OS(o_2\_VO_j\_\varphi) = \bigcup_{j\neq i, j\in\{1,2,\cdots m\}}^{m} OS(o_k)$$

3.2.4 The definition of "$\oplus$"

Definition 3: the operator "$\oplus$" is defined as follows:

$\forall$ $o_1,o_2 \in O$,

$\#o_1 \oplus \#o_2 = \{o \mid T(net(o), T(o), OR(o), OS(o))$

$= T(\max(net(o_1), net(o_2)), \max(T(o_1), T(o_2)),$

$OR(o_1) \cap OR(o_2), OS(o_1) \cap OS(o_2), o \in O\}$.

The max $(net(o_1), net(o_2))$ for a maximum of two real numbers, $\max(T(o_1), T(o_2))$ is defined as follows:

$\max(T(o_1), T(o_2)) = \{\max(T(O_1\_VO_1), T(O_2\_VO_1)), \max(T(O_1\_VO_2), T(O_2\_VO_2)), \ldots, \max(T(O_1\_VO_m), T(O_2\_VO_m))\}$.

$OR(o_1) \cap OR(o_2)$ and $OS(o_1) \cap OS(o_2)$ have similar definition.

On the basis of the above assumption, the model FM=<O,S,SC, $\oplus$,$\rightarrow$> described in this section is a security lattice model in the grid environment.

4 Discussion of new model

We discuss the rationality of the new model from the following aspects.

4.1 THE RATIONALITY OF "$\rightarrow$" DEFINITION

According to the first hypothesis, the subject **s** and the operating objects o_1 and o_2, s= (s_VO1, s_VO2,..., s_VOm), meet the condition $s\_VO_j \in S\_VO_j$, this shows that the subject is mapped to one or more subjects belonging to the local resource through relevant strategy. It accords with the most basic requirements for grid security.

When $o_1,o_2 \in O$ and belong to the same virtual organization, $net(o_1) = net(o_2)$, then flow directions of o_1 and o_2 are determined only according to the priority levels of security classification and role natures of the two objects. According to references [3] and [6], $T(o_1) \leq T(o_2)$, $OR(o_1) \supseteq OR(o_2)$, and $OS(o_1) \supseteq OS(o_2)$ are used to define $o_1 \rightarrow o_2$, which complies with security requirements of the general network environment. At the moment, because no information flows out of the virtual organization, so it meets design requirements of the entire network.

When $o_1,o_2 \in O$ and they don't belong to the same virtual organization, according to the relevant definition, only when $net(o_1) \leq net(o_2)$, $T(o_1) \leq T(o_2)$, $OR(o_1) \supseteq OR(o_2)$ and $OS(o_1) \supseteq OS(o_2)$ are true, $o_1 \rightarrow o_2$ is true.

The first condition is satisfied only when O_1 and O_2 are from the virtual organizations with the low-level and the high-level security classification respectively. Furthermore, O_1 and O_2 should meet the second condition which is necessary for the common information flow; otherwise they cannot flow from the virtual organization with the low-level security classification to the one with the high-level security classification. So they satisfy

design requirements of the general network security. Moreover, the third or the fourth condition complies with design requirements of the entire grid security in accordance with the description in the reference [5].

In addition, when the source object or objective object is not decomposed in a virtual organization, it meets the above security requirements as per definitions of security classifications of the subject and the object.

Thus, definition of "→" in Section 3.2 is reasonable.

4.2 <SC,→> IS A PARTIALLY ORDERED SET

Theorem 1: <SC,→> is a partially ordered set.

Proof: we verify it from reflexivity, transitivity and anti-symmetry of <SC,→> .

- Reflexivity: $\forall o \in O, o \rightarrow o$.

Suppose $T(o\_VO_i) \leq T(o\_VO_i)$, $OR(o)\_VO_i \supseteq OR(o)\_VO_i$, and $OS(o)\_VO_i \supseteq OS(o)\_VO_i$ (i=1,2,…,m), $net(o) \leq net(o)$, $T(o) \leq T(o)$, $OR(o) \supseteq OR(o)$ and $OS(o) \supseteq OS(o)$ are true, so is $o \rightarrow o$.

- Transitivity: $\forall o_1,o_2,o_3 \in O, o_1 \rightarrow o_2$ and $o_2 \rightarrow o_3$, here has $o_1 \rightarrow o_3$.

$\forall o_1, o2, o3 \in O, o1 \rightarrow o_2$ and $o_2 \rightarrow o_3$.

So $net(o_1) \leq net(o_2)$, $T(o_1) \leq T(o_2)$, $OR(o_1) \supseteq OR(o_2)$,$OS(o_1) \supseteq OS(o_2)$,

and $net(o_2) \leq net(o_3)$,$T(o_2) \leq T(o_3)$,$OR(o_2) \supseteq OR(o_3)$, $OS(o_2) \supseteq OS(o_3)$. The following can be derived accordingly.

$T(o_1\_VO_i) \leq T(o_2\_VO_i)$,$OR(o_1)\_VO_i \supseteq OR(o_2)\_VO_i$, $OS(o_1)\_VO_i \supseteq OS(o_2)\_VO_i$ and $T(o_2\_VO_i) \leq T(o_3\_VO_i)$, $OR(o_2)\_VO_i \supseteq OR(o_3)\_VO_i$, $OS(o_2)\_VO_i \supseteq OS(o_3)\_VO_i$ (i=1,2,…,m).

therefore $T(o_1\_VO_i) \leq T(o_3\_VO_i)$, $OR(o_1)\_VO_i \supseteq OR(o_3)\_VO_i$, $OS(o_1)\_VO_i \supseteq OS(o_3)\_VO_i$ (i=1,2,…,m) are true.

According to the definition of decomposition, $T(o_1) \leq T(o_3)$, $OR(o_1) \supseteq OR(o_3)$, $OS(o_1) \supseteq OS(o_3)$ are true, and obviously net $(o_1) \leq net(o_3)$ is true, and so is $o_1 \rightarrow o_3$.

- Anti-symmetry: $\forall o_1,o_2 \in O, o_1 \rightarrow o_2$, and $o_2 \rightarrow o_1$, $\#o_1 = \#o_2$.

Suppose $net(o_1) \leq net(o_2)$; $T(o_1) \leq T(o_2)$, $OR(o_1) \supseteq OR(o_2)$, $OS(o_1) \supseteq OS(o_2)$ and $net(o_2) \leq net(o_1)$,$T(o_2) \leq T(o_1)$,$OR(o_2) \supseteq OR(o_1)$, $OS(o_2) \supseteq OS(o_1)$, the following can be derived:

$T(o_1\_VO_i) \leq T(o_2\_VO_i)$, $OR(o_1)\_VO_i \supseteq OR(o_2)\_VO_i$, $OS(o_1)\_VO_i \supseteq OS(o_2)\_VO_i$ 且 $T(o_2\_VO_i) \leq T(o_1\_VO_i)$, $OR(o_2)\_VO_i \supseteq OR(o_1)\_VO_i$, $OS(o_2)\_VO_i \supseteq OS(o_1)\_VO_i$ (i=1,2,…,m).

$T(o_1\_VO_i)=T(o_2\_VO_i)$, $OR(o_1)\_VO_i=OR(o_2)\_VO_i$, $OS(o_1)\_VO_i= OS(o_2)\_VO_i$ (i=1,2,…,m).

So $T(o_1)=T(o_2)$, $OR(o_1)=OR(o_2)$ and $OS(o_1)=OS(o_2)$ are true. obviously $net(o_1)=net(o_2)$ is true.

It can be seen from the above, derivation that <SC,→> is a partially ordered set.

4.3 SC HAS A LOWER BOUND $\#O_{LOWEST}$

SC has a lower bound, denoted as $\#o_{lowest}$.

$\#o_{lowest}=\{o|T(net(o),T(o),OR(o),OS(o))=T(\min(net(o), \min(T(o)), \bigcup_{o' \in O} OR(o'), \bigcup_{o' \in O} OS(o')),o \in O\}$.

According to the above definition $\#o_{lowest}$ is always present and $\#o_{lowest} \rightarrow \#o$.

4.4 SOME PROPERTIES OF $\oplus$

Property 1: $\forall o_1,o_2,o_3 \in O, \#o_1 \rightarrow \#o_1 \oplus \#o_2, \#o_2 \rightarrow \#o_1 \oplus \#o_2$

Proof: (1) because of the definition of $net(o_1)$ and $net(o_2)$

So $net(o_1) \leq \max(net(o_1),net(o_2))$, $net(o_2) \leq \max(net(o_1),net(o_2))$ is true.

(2) because of the definition of $\max(T(o_1),T(o_2))$,

So $T(o_1) \leq \max(T(o_1),T(o_2))$, $T(o_2) \leq \max(T(o_1),T(o_2))$ are true.

(3) when i=1,2,…,m

$OR(o_1)\_VO_i \supseteq OR(o_1)\_VO_i \cap OR(o_2)\_VO_i$,$OS(o_1)\_VO_i \supseteq OS(o_1)\_VO_i \cap OS(o_2)\_VO_i$

so $OR(o_1) \supseteq OR(o_1) \cap OR(o_2)$ and $OS(o_1) \supseteq OS(o_1) \cap OS(o_2)$ are true.

Similarly, $OR(o_2) \supseteq OR(o_2) \cap OR(o_2)$ and $OS(o_2) \supseteq OS(o_2) \cap OS(o_2)$ are true.

On the basis of the definition of $\oplus$ and the above analysis, the following can be derived:

$\#o_1 \rightarrow \#o_1 \oplus \#o_2$, $\#o_2 \rightarrow \#o_1 \oplus \#o_2$.

Property 2: $\forall o_1,o_2,o_3 \in O, \#o_1 \rightarrow \#o_3, \#o_2 \rightarrow \#o_3$, **so** $\#o_1 \oplus \#o_2 \rightarrow \#o_3$

Proof: (1) Suppose $net(o_1) \leq net(o_3)$, $net(o_2) \leq net(o_3)$, Max $(net(o_1), net(o_2)) \leq net(o_3)$ is true.

(2) Suppose $T(o_1) \leq T(o_3)$, $T(o_2) \leq T(o_3)$,

$\max(T(o_1\_VO_i),T(o_2\_VO_i)) \leq T(o_3\_VO_i)$ (i=1,2,…,m) is true.

so is $\max(T(o_1), T(o_2)) \leq T(o_3)$.

(3) $\#o_1 \rightarrow \#o_3$

$OR(o_1)\_VO_i \supseteq OR(o_3)\_VO_i$, $OS(o_1)\_VO_i \supseteq OS(o_3)\_VO_i$ (i=1,2,…,m)

$\#o_2 \rightarrow \#o_3$

$OR(o_2)\_VO_i \supseteq OR(o_3)\_VO_i$, $OS(o_2)\_VO_i \supseteq OS(o_3)\_VO_i$ (i=1,2,…,m),

so the following can be derived:

$OR(o_1)\_VO_i \cap OR(o_2)\_VO_i \supseteq OR(o_3)\_VO_i$, $OS(o_1)\_VO_i \cap OS(o_2)\_VO_i \supseteq OS(o_3)\_VO_i$ (i=1,2,…,m).$OR(o_1) \cap OR(o_2) \supseteq OR(o_3)$, $OS(o_1) \cap OS(o_2) \supseteq OS(o_3)$

According to the definition of $\oplus$ and the above analysis, $\#o_1 \oplus \#o_2 \rightarrow \#o_3$ is true.

5 Conclusions

In this study, on the basis of elements of grid security, the concept of role decomposition, the ultra four-dimensional security function of the object are defined by means of the characteristics of the security formation flow model in general network environment.

A new information flow model based on the grid environment is described by defining the information flow strategy by means of the ultra four-dimensional safety function value of the object.

According to the analyses, it is proved that the model is safe and reasonable. The model can reflect the information flow in the grid environment well, and this is helpful for description of security information flow in the grid environment.

Because the grid environment is too complicated and the thesis length is limited, practical application of the information flow model in this study will be detailed in a separate paper.

Acknowledgment

This work is supported by Sichuan province academic and technical leader training funded projects 12XSJS002.Sichuan provincial department of science and technology project 2013ZR0089. Sichuan province department of education natural science point item 12ZB192, 13ZA0003, 14ZB0360, 14ZB0363.

References

[1] Bell D E, Lapadula L J 1973 Secure compter system[R]:mathematical foundation MTR-2527, Mitrecorp, Bedford, MA (NTIS AD771543)
[2] Denning D E 1976 A lattice model of secure information flow *Communications of the ACM* **19**(5) 236-43
[3] Sandhu Ravi S 1993 Lattic-Based Access Control Model *IEEE computer* **26**(11) 9-19
[4] Liu Yihe, Shen Chang Xiang 2005 An Information Security Function and Application Model *Journal цf Computer Aided Design & Computer Graphics* **17**(12) 2734-38 *(In Chinese)*
[5] Sandhu R S, Samarati P 1994 Access control: principles and practice *IEEE communications* **32**(9) 40-8
[6] Liu Yihe, Liu Jiayong 2004 An Information Flow Model Based on Roles and Applications *Journal of Sichuan University (Engineering Science Edition)* **36**(5) 94-7 *(In Chinese)*
[7] May Phyo Oo, Thinn Thu Naing 2007 Access Control System for Grid Security Infrastructure *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology – Workshops* 299-302
[8] Han Bing 2006 Study on data management and the security problem Under the grid environment. Chinese Scientific and Technical University master's degree paper, in Chinese, 2006
[9] Xiaoqin Huangetc 2005 *An Identity-Based Model for Grid Security Infrastructure* ISSADS2005, LNCS3563 258-66 Springer-Verlag Berlin Heidelberg
[10] Bivens H 2001 *Grid work flow* Albuquerque:Sandia National Laboratory
[11] Zhou Jian-tao 2008 A Review on the Grid Workflow and Its Key Technologies *Journal of jnner Mongolia University* **39**(5) 581-9 *(In Chinese)*
[12] Liu Yi-he Security Information Flow Model Based on Grid Environment *Computer Science* **38**(6) 157-60 *(In Chinese)*
[13] Wфтп Fang 2009 *Study under the grid environment's trust mechanism* Nanjing Posts and telecommunications University master's degree paper *(in Chinese)*
[14] *The Globus Project [EB ／OL]* http://www.globus.org/

Authors

Yi He Liu, born in April, 1964, Neijiang, China

Current position, grades: Professor in Neijiang Normal University
University studies: Doctor of Cryptography in Sichuan University
Scientific interest: Intra-body communication, Cryptography
Publications: 2 Patents, 46 Papers
Experience: He received the Ph.D. degree in Sichuan University, Chengdu, Sichuan, China, in 2005.Since 2009, he has been involved in research in the areas of intra-body communication. He is currently a Professor, Neijiang Normal University.

Shuang Zhang, born in May, 1983, Leshan, China

Current position, grades: Lecturer in The Engineering & Technical college of Chengdu University of Technology
University studies: Doctor of Electrical & Electronic Engineering in University of Macau.
Scientific interest: Intra-body communication, Cryptography
Publications: 2 Patents, 25Papers
Experience: He received the Master degree in control Engineering from the Graduate University of Chinese academy of sciences, Beijing, China in 2011.Since 2011, he has performed research in the areas of Digital Image Processing, Digital signal processing, Intra-body communication. He is now a lecturer, department of electronic information and computer technologies, The Engineering & Technical college of Chengdu University of Technology. At the same time, He is now a Ph.D. student in the Department of Electrical and Electronics Engineering, Faculty of Science and Technology, University of Macau.

Yu Ping Qin, born in March, 1984, Leshan, China

Current position, grades: Lecturer in The Engineering & Technical college of Chengdu University of Technology
University studies: Master of basic mathematics in Sichuan Normal University.
Scientific interest: Intra-body communication, Cryptography
Publications: 2 Patents, 25Papers
Experience: She received the Master degree in control Engineering from the Sichuan Normal University, Chengdu, China in 2011.Since 2011, his has performed research in the areas of Partial differential equation, Intra-body communication. His is now a lecturer, The Engineering & Technical college of Chengdu University of Technology.

Fashion colour forecasting based on BP neural network

Xiaohong Kong[1*], Lixia Chang[2], Junpeng Xu[1]

[1]*School of Mechanical and Electrical Engineering, Henna Institute of Science and Technology, China, 453003, Henan, China*

[2]*School of Fashion, Henna Institute of Science and Technology, China, 453003*

Received 12 July 2014, www.cmnt.lv

Abstract

Fashion colour forecasting is a hot issue in fashion industry and also a hard problem because of much uncertain information. Utilizing strong mapping capability of BP Neural Network (BPNN) for nonlinear function, this paper investigated the forecasting model of fashion colour. Based on colour data of recent several years, the forecasting model for future colour trend is discussed and built. The historical data are input to train the Neural Network weights and the different parameters of BPNN were investigated to find how to affect the forecasting performance. The results demonstrate the algorithm is very efficient in colour forecasting and can approximate nonlinear relationship of fashion colour very close.

Keywords: Fashion colour, BP neural network, Trend forecasting, Network parameters

1 Introduction

Fashion colour is the colour or colours that the most people prefer in a given period of time. It is very important to forecast the future colour trend that will be popular in 24 months later during the fashion selling seasons and offer the trend palettes for clothing enterprises [1-2]. The obtained palettes will offer the industries useful information and guidance in marketing and manufacturing [3]. And it has long been considered as the most powerful driving forces. Therefore, accurate forecast of fashion colour is not only a marketing strategy of garment industry, but also a hot issue in international fashion market. If the market products do not meet pop elements, it would be unsalable or have to be withdrawn from the market [3-4].

Currently, many authoritative associations on colours release fashion colour palettes every year such as INTER-COLOUR, JAFCA CFCA, PANTONE®, and so on, but little detailed information is known about the forecasting methodologies dealing with this complex and uncertain process [5]. What's more, the lag and confidence of palettes during propagating restrict its commercial value, which stimulates the development of research on fashion colour forecasting. So far, several theories and methods have been applied into this field by using statistical analysis [6-7], grey theory [5, 8], artificial neural network [9-12] and hybrid models [13-14]. Although some achievements obtained, fashion colour prediction is still in exploring stage on the whole and few convincing forecasting systems had been established. Deficiencies and controversies are still existed and the validity of prediction has yet been a pressing problem of the day in the textile and clothing industries [15].

Fashion colour forecasting involves various factors such as sales psychology, geographical, social psychology, political situation, climate, season, skin colour, education level [2-3]. These factors are all interrelated to each other and the fashion colour palettes usually have non-linear features, so forecast of fashion colour becomes a very difficult task. Therefore, an effective forecasting system is required to be established. Utilizing strong function nonlinear mapping capability of Back Propagation neural network (called BPNN)[16], this paper carried out simulation and forecast of fashion colour and built the forecasting model. It is well-known that a single hidden layer BP neural network can approximate arbitrary nonlinear mapping relationship [9, 15-16]. The successive 6 years' fashion colour palettes, released by INTER-COLOUR, from 2007 to 2012 are taken as the objects in this study. The colours are pre-treated using PANTONE colour system and used as input parameters. In some degree, colour data of recent years reflect the trend and greatly reduce the uncertain character. Finally, the different parameters of BPNN were investigated to find how to affect the forecast performance.

2 Algorithm model

2.1 BP NEURAL NETWORK (BPNN)

The artificial Neural Network algorithm mimics biological neural networks, and it is composed of many neurons and adjustable weights connection [16]. Artificial neural network system is made with a large-scale parallel processing, distributed information storage, good self-organizing and self-learning ability and other characteristics. It is widely used in the information processing, pattern

* ***Corresponding author*** e-mail: nancykong@hist.edu.cn

recognition, intelligent control and system modelling and other areas to adapt systems to the environment, find the law of things and recognize or control object [9-12, 17-19]. In particular, the error back propagation algorithm can approximate nonlinear function with strong mapping ability and the colour forecasting is the case. In this network, data stream propagate forward and error signal back-propagation. The network can be represented by a directed acyclic graph $G=(N,W)$, where N is a set of neurons and W is a set of weights, corresponding to the data dependence among neurons. The weights are associated with the amount of data units to be transferred from one neuron to another neuron.

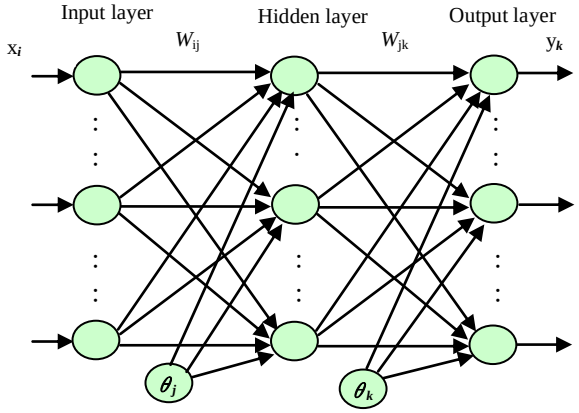


FIGURE 1 The architecture of a BPNN model

BP neural network is suitable to solve the problems of fuzzy and uncertain system and the most commonly used network is a three-layer feed-forward model shown in Figure 1 [16]. It consists of one input layer, one hidden layer and one output layer. The input layer has n inputs and the output layer m outputs, denoted as $x_i(i=1,\cdots,n)$, and $y_k(k=1,\cdots,m)$ respectively. Suppose the hidden layer has h neurons. The neurons of adjacent layers are fully connected with weights W_{ij} and W_{jk} and the data only flow from the input layer to output layer. The neurons of hidden layer and the output have different transfer function named as f_h and f_o. When the $x_i(i=1,\cdots,n)$ is input to network, $\sum_{i=1}^{n} w_{ij}x_i(j=1,\cdots,h)$ are applied to h neurons together with bias $\boldsymbol{\theta}_j$ of hidden layer. The transfer function determines the output of every neuron. The output layer has the same process and finally $y_k(k=1,\cdots,m)$ comes out. θ_k represents the bias of the output layer.

BP algorithm is a supervised learning algorithm. In this case, the weights are not known ahead and obtained from sample data. The main idea is: the BP model must be trained by input samples; During the training process, error back-propagation algorithm is used to adjust weights and the biases repeatedly, and it is ensured that the output vector and the desired vector is as close as possible; When the error of output layer is less than a specified error, the training is finished and the network weights and biases are saved for further forecasting. The specific steps are as follows [16]:

(1) Initialize the weights (W_{ij}, W_{jk}) and bias (θ_j, θ_k).

(2) Calculate the output of the hidden layer neuron according to Equation (1).

$$a_j = f_h\left[\sum_{i=1}^{n} w_{ij}x_i - \theta_j\right]. \tag{1}$$

(3) Calculate the output of output neuron according to Equation (2).

$$y_k = f_o\left[\sum_{j=1}^{h} w_{jk}b_j - \theta_k\right], \tag{2}$$

if the bias $\boldsymbol{\theta}_k$ is considered as $-1\times w_{0j}(w_{0j}=\theta_k)$, Equation (2) can be rewritten as $y_k = f_o\left[\sum_{j=0}^{h} w_{jk}b_j\right]$.

(4) Calculate the error between the desired output and network output according to Equation (3), t_k is the target output.

$$E=\sum_{k=1}^{m}(t_k-y_k)^2. \tag{3}$$

(5) Error back propagation and the weights are updated according to Equation (4).

$$\begin{aligned}\Delta W(t+1)&=\eta\delta[W(t),X(t),T(t)]X(t)\\ W(t+1)&=\Delta W(t+1)+W(t)\end{aligned}, \tag{4}$$

where W denotes the connection weight matrix, X denotes the input vector and T the target vector, $\boldsymbol{\delta}$ represents the back propagation error based on different learning rule and η denotes the learning rate, $\eta>0$.

(6) Repeat step (2) –step (5) until the error satisfy the performance.

2.2 BPNN MODEL FOR THE FASHION COLOUR PREDICTING

In this study, a BP neural network with three-layer is adopted. The neurons of the hidden layer use sigmoid transfer function and the output layer linear function [20]. The nonlinear transfer functions can represent nonlinear and linear relationships between input and output vectors of the network. The linear neurons of the output layer allow the network to produce values outside the range [–1,+1]. BPNN is a negative gradient descent algorithm. There are two ways to realize the gradient descent algorithm, incremental mode and batch mode. In the incremental model, the gradient is calculated and the weights are updated after each input is applied to the network. In the batch mode, only when the whole sample

has been trained, the weights and bias are updated. The negative error gradients of training samples determine the change of the weights and bias.

BPNN is powerful to distinguish the complex mode with its strong learning ability and generalization ability. The algorithm transform a set of the input/output samples into a nonlinear optimization problem by iterative learning method, but its convergence speed is slow and easy to fall into local minima. In order to solve the network instability of learning defects, many improved algorithms are raised. In this paper, several considered versions are shown in Equation (5).

$$\Delta W(t+1) = \eta\delta[W(t), X(t), T(t)]X(t) + \alpha\Delta W(t), \quad (5)$$

where α is the momentum factor, $\alpha \in (0,1)$ and $\alpha\Delta W(t)$ reflects the experience accumulated previously.

The learning rate is a multiple of the negative gradient to determine the changes of the weights and biases. An adaptive algorithm will be adopted and learning rate η will be adjusted during the training period. The weights are updated after all of the inputs are applied to the network. A flow chart of the BP is shown in Figure 2.

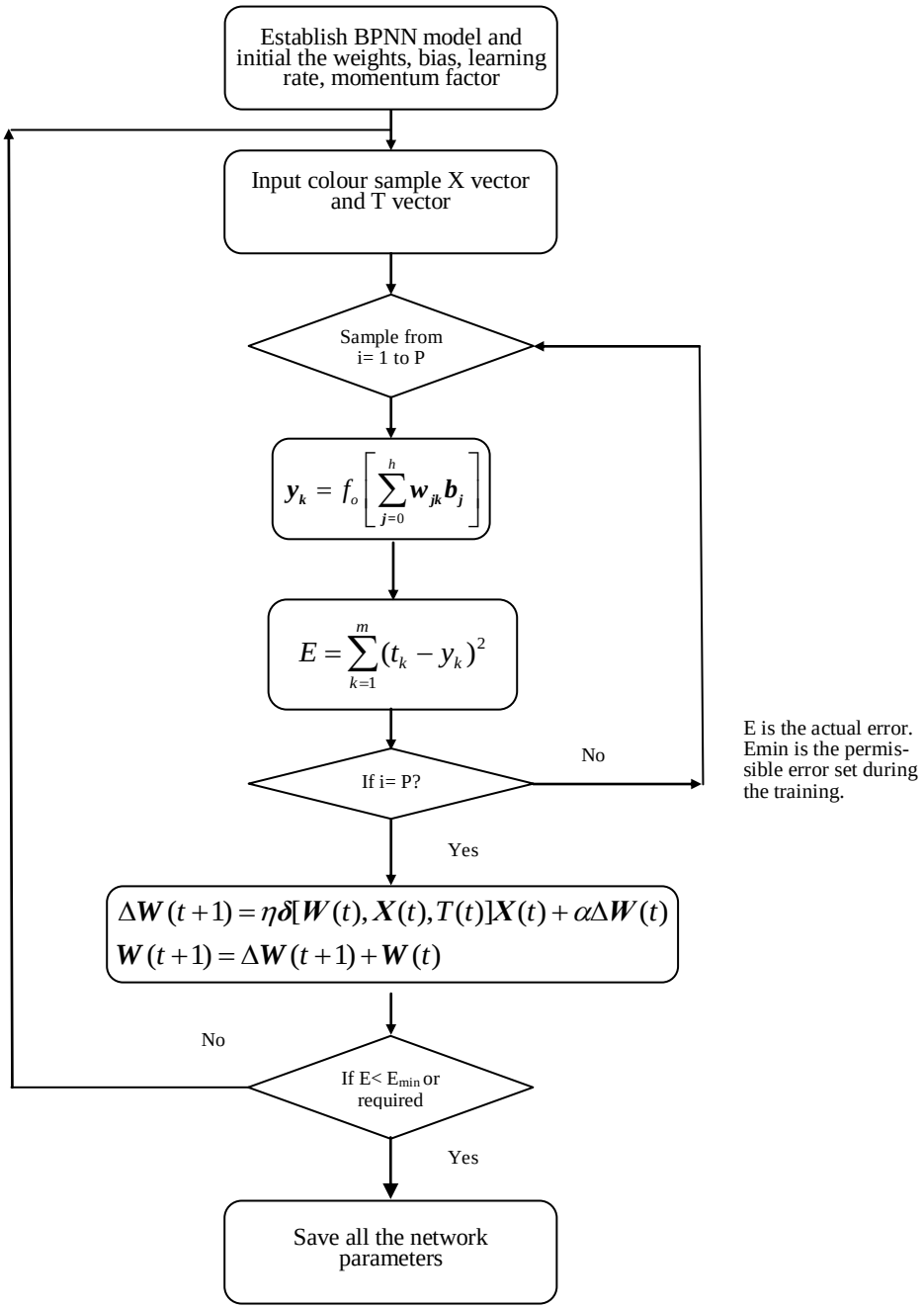


FIGURE 2 The algorithm flow diagram

3 Experiments based on BPNN algorithm

3.1 TEST BENCHMARK

Fashion colour conveys consumers a tendency of tones that will be widely accepted while not specific one or several colours. Generally, more than 40 colours can appear in every palette, and there is no exact corresponding colours in every year, therefore, the colours must be classified and quantified according to the colour theory on hue, lightness and chroma, respectively. Fashion Colour palettes for women's Spring/summer, released by INTER-COLOUR are taken as the test benchmark in this study. These successive fashion colour palettes from 2007 to 2012 were collected. In order to exploit the historical data in BP algorithm, the colours are quantified to numeric data by applying Pantone colour system by PANTONE INC, which is influential in colour specification. The detailed quantification theory and method can be seen in our previous research results [6, 11].

The study only discussed the hue feature. The colours are firstly reflected as their corresponding domains in the Pantone colour space, shown in Table 1[6]. And the colours are transformed further according to Equation (6) and here P_i represents the proportion in the palette.

$$P_i = \frac{n_i}{N}, \tag{6}$$

where n_i depicts the numbers of the colours reflected into the corresponding domains. i denotes the different colour domains; and N indicates the whole numbers of colours in every palette. The quantified results were listed in Table 2.

TABLE 1 Colour domain in the Pantone colour space

Domain	Colour	Domain	Colour	Domain	Colour	Domain	Colour	Domain	Colour
[6-10]	Yellow	[11-14]	Yellow/red	[15-16]	Red	[17-24]	Red/Purple	[25-33]	Purple
[34-38]	Blue /Purple	[39-47]	Blue	[48-52]	Blue /Green	[53-59]	Green	[60-64] [0-5]	Yellow /Green

TABLE 2 Colour ratio in the palettes from 2007 to 2012

Year	Yellow	Yellow/Red	Red	Red/Purple	Purple	Purple/Blue	Blue	Blue/Green	Green	Green/ Yellow
2007	0.242	0.069	0.104	0.034	0.103	0.034	0.242	0.034	0.034	0.104
2008	0.279	0.047	0.07	0.023	0.023	0.093	0.233	0.023	0.047	0.162
2009	0.130	0.218	0.043	0.065	0.022	0.043	0.217	0.001	0.108	0.153
2010	0.244	0.245	0.019	0.113	0.001	0.001	0.188	0.038	0.057	0.094
2011	0.190	0.119	0.072	0.095	0.001	0.001	0.214	0.095	0.025	0.190
2012	0.166	0.194	0.001	0.083	0.028	0.001	0.222	0.056	0.027	0.222

3.2 MODEL PARAMETERS SELECTION

The numbers of the hide layers, the number of processing units and network learning coefficient and other parameters have great flexibility and must be set depending on the circumstances. In many applications, these parameters play an important role. There is no theoretical guidance for selecting the hide layer of the network and the number of units. If BP neural network parameters are not fit, the network cannot guarantee convergence to the global optimal point.

The algorithm is implemented and simulated with neural network toolbox in Matlab 7 [20]. In order to investigate network performance, different parameters were considered as the following:

1) the weights (W_{ij}, W_{jk}) were randomly initialized between (-0.5, +0.5).
2) the learning rate η is adaptive with the training process.
3) the momentum factor α uses the default value.
4) Mean squared error performance function is chosen as target function and target error is set as 0.001.
5) maximum iterations =1000
6) The size of hidden layer neurons, denoted as *Nsize*. *Nsize* is determined by Equation (7).

$$Nsize = \sqrt{n+m} + (1 \sim 8). \tag{7}$$

Here, n and m represent the number of input layer neurons and output layer neurons.

Sigmoid and linear functions were selected as the transfer function of hidden layer and output layer. In Matlab environment, there are several variations of gradient descent algorithms Different training function and learning function were selected from the following functions [20]:

traingd – Gradient descent backpropagation.
Traingdm – Gradient descent with momentum backpropagation.
Traingda – Gradient descent with adaptive lr backpropagation.
Traingdx – Gradient descent w/momentum & adaptive lr back propagation.
trainlm – Levenberg-Marquardt back propagation.
Learngd – Gradient descent weight/bias learning function.
Learngdm – Gradient descent w/momentum weight/bias learning function.

3.3 TRAINING AND SIMULATING

When the BPNN was trained with the colour samples, the weights were modified continuously. The training stops if the number of iterations exceeds epochs or if the error $E < E\min$.

3.3.1 Single factor and single colour predicting

When the BPNN is trained, the three successive years' fashion colour data is input and the fourth year's data is the output. That is to say, the recent three years of data roll forecast the fourth year trends. In this mode, only one colour of the past three years influences the same colour of the fourth year. For example, the data of 2007, 2008, and 2009 is the input data and the data of 2010 is the output. And so on, the training samples of the yellow are listed in Table 3.

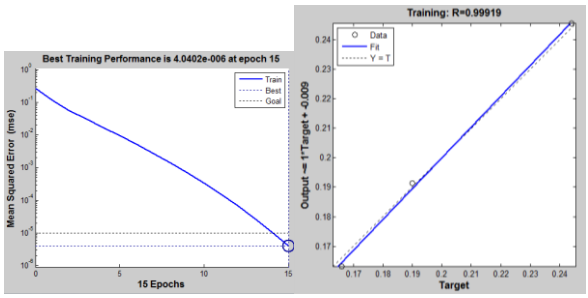
After training the network, the original input vectors were applied the trained network and the simulating results are listed as Table 4.

TABLE 3 Training samples of the hue yellow

Colour	Inputs			Output
Yellow	0.242	0.279	0.130	0.244
	0.279	0.130	0.244	0.190
	0.130	0.244	0.190	0.167

TABLE 4 Single factor forecast of the hue yellow

Value \ Year	2010	2011	2012
Actual value	0.2440	0.190	0.1660
Forecast value	0.2454	0.1914	0.1632
Error	-0.0014	-0.0014	0.0028



(a) (b)

FIGURE 3 (a) Performance plot; (b) Regression plot

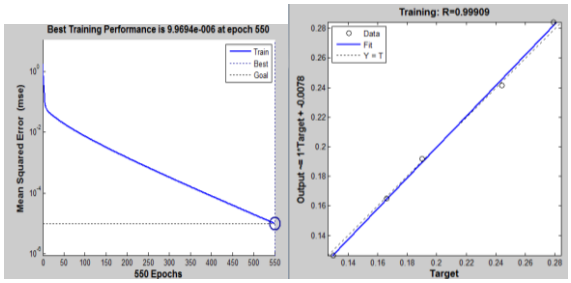
There are two plots in Figure 3 that describe the performance plot and the regression plot respectively. The performance plot shows the mean squared error of the network began with a large value and gradually decreased to a smaller value. It also shows the learning process. The regression plot illustrates regression analysis between the network output and the expecting targets. A dashed line indicates the best linear fit and the solid line shows the perfect fit (output equal to targets). In this example, the best linear fit line is very close to the perfect fit line and the algorithm is very good.

3.3.2 Multiple factors adjacent predicting

Colour forecast is a different task and is affected by many factors. It is not enough to use one colour data of the past three years. In this mode, the next year colour is determined by the ten colours of the former year. In this mode, initial η =0.1, *Nsize* =8, training function= 'traingd', learning function='learngdm' in first experiment. The simulation results were listed in Table 5 and Figure 4.

TABLE 5 Multiple factor forecast of the hue yellow

Value \ Year	2008	2009	2010	2011	2012
Actual value	0.2790	0.1300	0.2440	0.1900	0.1660
Forecast value	0.2841	0.1265	0.2416	0.1921	0.1649
Error	-0.0051	0.0035	0.0024	-0.0021	0.0011



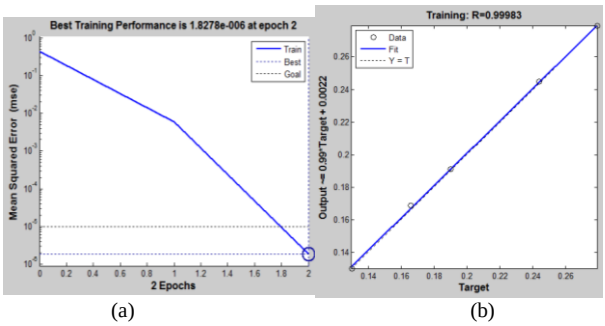
(a) (b)

FIGURE 4 (a) Performance plot; (b) Regression plot

In the second experiment, the training function is modified as 'trainlm' and the other parameters remain unchanged. The experimental results are shown in Table 6 and Figure 5.

TABLE 6 Multiple factors forecast of the hue yellow

Value \ Year	2008	2009	2010	2011	2012
Actual value	0.2790	0.1300	0.2440	0.1900	0.1660
Forecast value	0.2841	0.1265	0.2416	0.1921	0.1649
Error	-0.0051	0.0035	0.0024	-0.0021	0.0011



(a) (b)

FIGURE 5 (a) Performance plot (b) Regression plot

From the above results, the multiple factors forecast was not good as single factor forecast because the former only use one year data and the learning information was less.

3.3.3 Multiple factors and single colour predicting

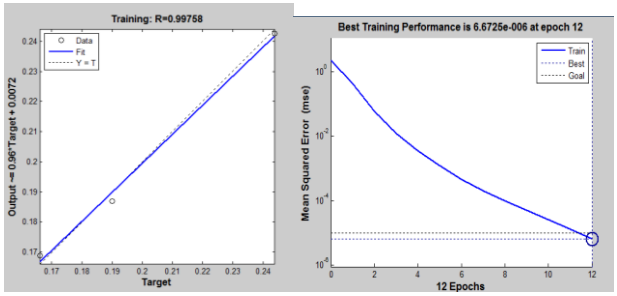
In order to improve the forecast performance, the above two situations were combined. So in this mode, ten colours of the past three years were considered to predict the fourth colour trend.

(1) Different learning rate

In this case, *Nsize* =8, train function='traingdx', learn function='learngdm'.

TABLE 7 Multiple factors forecast of the hue yellow (η =0.05)

Value \ Year	2010	2011	2012
Actual value	0.2440	0.190	0.1660
Forecast value	0.2427	0.1868	0.1688
Error	0.0013	0.0032	-0.0028

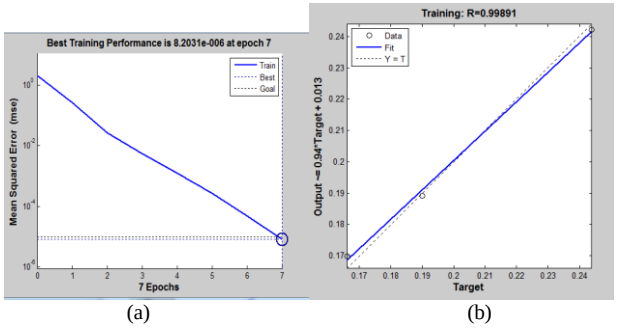


(a) (b)

FIGURE 6 (a) Performance plot; (b) Regression plot

TABLE 8 Multiple factors forecast of the hue yellow (η =0.1)

Value \ Year	2010	2011	2012
Actual value	0.2440	0.1900	0.1660
Forecast value	0.2435	0.1886	0.1623
Error	0.0005	0.0014	0.0037



(a) (b)

FIGURE 7 (a) Performance plot; (b) Regression plot

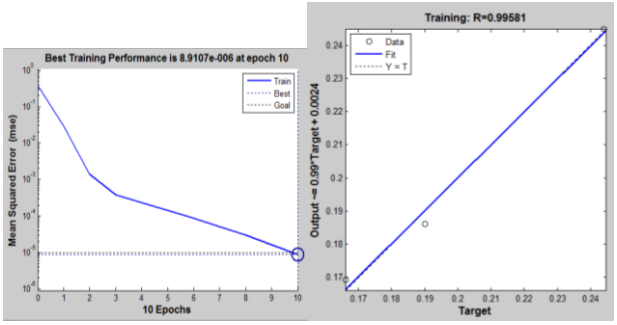
The experimental results of different learning rates are shown in Table 7, Table 8 and Figure 6, Figure 7. Bigger learning rates may make the network weight change too much each time, or even lead to weights jump without convergence during the training process; on the other hand, smaller learning rates lead to learning time too long and can't guarantee convergence to a minimum. In this example, η =0.1 has better results.

TABLE 9 Multiple factors forecast of the hue yellow (train function='traingdx', learning function='learngd')

Value \ Year	2010	2011	2012
Actual value	0.2440	0.1900	0.1660
Forecast value	0.2448	0.1861	0.1693
Error	-0.0008	0.0039	-0.0033

(2) Different training function and learning function

In this experiment, η =0.1, *Nsize* =8.

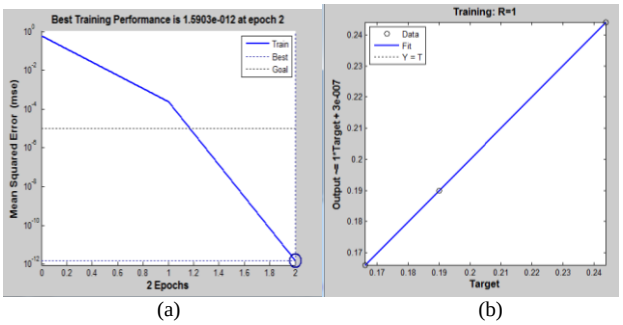


(a) (b)

FIGURE 8 (a) Performance plot; (b) Regression plot

TABLE 10 Multiple factors forecast of the hue yellow (training function='trainlm', learning function='learngdm')

Value \ Year	2010	2011	2012
Actual value	0.2440	0.1900	0.1660
Forecast value	0.2440	0.1900	0.1660
Error	0	0	0



(a) (b)

FIGURE 9 (a) Performance plot; (b) Regression plot

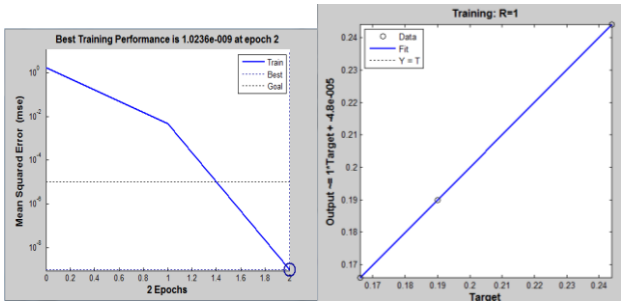
Form the results in Table 9, Table 10 and Figure 8, Figure 9, samples were perfectly fit and the error was zero. TRAINLM training function uses Levenberg – Marquardt algorithm and has the fastest convergence speed, but requires more storage space. Learngdm is a Gradient descent function with momentum back propagation and considers the learning experience.

(3) Different neurons in hidden layer

η =0.1, training function='trainlm', learning function= 'learngdm'. *Nsize* has different value.

TABLE 11 Multiple factors forecast of the hue yellow (*Nsize* =7)

Value \ Year	2010	2011	2012
Actual value	0.2440	0.1900	0.1660
Forecast value	0.2440	0.1900	0.1660
Error	0	0	0

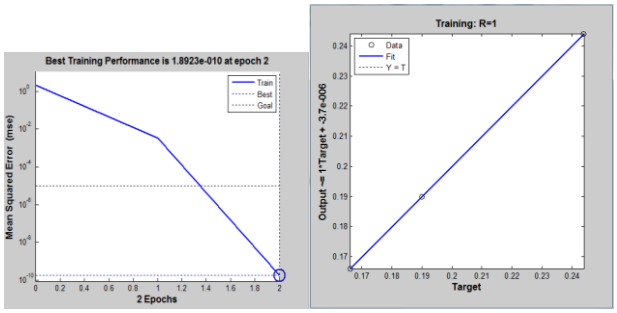


(a) (b)

FIGURE 10 (a) Performance plot; (b) Regression plot (*Nsize* =7)

TABLE 12 Multiple factors forecast of the hue yellow (*Nsize* =6)

Value \ Year	2010	2011	2012
Actual value	0.2440	0.1900	0.1660
Forecast value	0.2440	0.1900	0.1660
Error	0	0	0

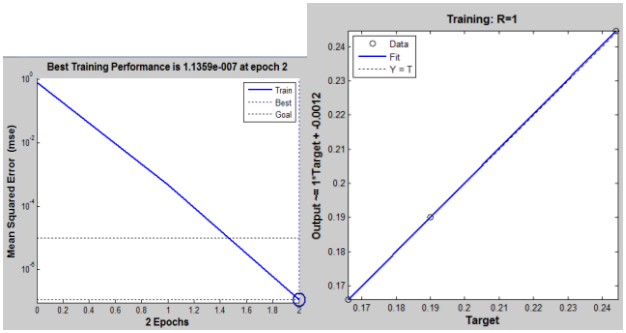


(a) (b)

FIGURE 11 (a) Performance plot; (b) Regression plot (*Nsize* =6)

TABLE 13 Multiple factors forecast of the hue yellow (*Nsize* =2)

Value \ Year	2010	2011	2012
Actual value	0.2440	0.1900	0.1660
Forecast value	0.2440	0.1900	0.1660
Error	0	0	0



(a) (b)

FIGURE 12 (a) Performance plot; (b) Regression plot (*Nsize* =2)

The above results under different hidden layers are given in Table 11, Table 12, Table 13 and Figure 10, Figure 11, Figure 12. The results told that the number of hidden layer neurons has less effect to forecast performance. The main reason is that less size of input and output of samples in this example and less neuron can give better results.

4 Conclusions

The above results have demonstrated that the BPNN algorithm can forecast the fashion colour very close. When selecting a different training and learning functions, and other parameters, it can adjust the forecasting performance. In this example, the appropriate samples and network parameters, can achieve zero error prediction and it confirmed the superiority of the neural network in fashion colour forecasting. In the same time, some parameters have little influence on forecasting performance. Especially, multi factors model of the colour data of past three years has considered much information and done better work.

Acknowledgments

The work is supported by Scientific and Technological research projects of Henan Province (102102210190) and Foundation of Henan Province Educational Committee (2011A120005).

References

[1] Diane T, Cassidy T 2005 *Colour Forecasting* Oxford: Blackwell Science 6-23
[2] Diane T C 2007 Personal Colour Analysis, Consumer Colour Preferences and Colour Forecasting for the Fashion and Textile Industries *J Colour Design & Creativity* **1**(1) 1-14
[3] Nenni M E, Giustiniano L, Pirolo L 2013 Demand forecasting in the fashion industry: a review *J Engineering Business Management* ahead of print December 2013 DOI: 10.5772/56840
[4] Na Liu, Shuyun Ren, Tsan-Ming Choi, Chi-Leung Hui, Sau-Fun Ng 2013 Sales Forecasting for Fashion Retailing Service Industry: A Review *J Mathematical Problems in Engineering* ahead of print DOI:10.1155/2013/738675
[5] Lin J J, Sun P T, Chen J J, et al. 2010 Applying Gray Model to Predicting Trend of Textile Fashion Colours *J the Textile Institute* **101**(4) 360-8
[6] Chang L X, Gao W D, Lu Y Z, et al. 2010 Analysis on Spring/summer Fashion Colours *J Textile* **31**(3) 98-103
[7] Stanstield J, Whitfield T W 2005 Can Future Colour Trends be Predicted on the Basis of Past Colour Trends?: An Empirical Investigation *J Colour Research & Application* **30**(3) 235-42
[8] Chang L X, Gao W D, Pan R R 2011 Applying grey model for international fashion colour trend forecasting *J Grey System* **14**(4) 159-64
[9] Kenji Suzuki 2011 *Artificial Neural Networks - Industrial and Control Engineering Applications* InTech
[10] Yu Y, Hui C L, Choi T M 2012 An Empirical Study of Intelligent Expert Systems on Forecasting of Fashion Colour Trend *J Expert Systems with Applications* **39** 4383-9
[11] Chang L X, Pan R R, Gao W D 2013 Fashion colour forecasting by applying an improved BP neural network *J Donghua University* **30**(1) 58-62
[12] Bhambure S D, Dhavale A J, Kadole P V, et al. 2013 Artificial Neural Network & its Applications in Textiles *Journal of the textile association* 31-7

[13] Wu Y Z, Zhai Y C, Sun L 2011 Fashion Colours Forecast Based on Gray Back-propagation Neural Network Model *J Donghua University: Natural science* **37**(4) 199-204
[14] Hui C, Ng S, Yu Y 2012 Colour Trend Forecasting of Fashionable Products with Very Few Historical Data *J IEEE Transactions on Systems, Man and Cybernetics, Part C: Application and Review* **42**(6) 1003-10
[15] Adya M, Collopy F 1998 How effective are neural networks at forecasting and prediction? A review and evaluation *J Forecasting* **17** 481–95
[16] Rojas R 1996 *Neutral Networks: A Systematic Introduction* Springer
[17] Xiaoshu S, Hong Z, Xuemin H 2012 Fabric Defect Detection Based on Regional Growing PCNN *J Multimedia* **7**(5) 372-9
[18] Habib M, Rokonuzzaman M 2013 An Empirical Approach to Optimize Design of Back propagation Neural Network Classifier for Textile Defect Inspection *J Mathematics & Computer Science* **3**(4) 617-34
[19] Wu Y Z, Sun L, Le J J 2010 G-LMBPNN: A New Fashion Colours Prediction Model In: *2010 International Conference on Computational Aspects of Social Networks*, Taiyuan, China, 501-4
[20] Häusler S 2010 *Neural Network Toolbox*

Authors

Xiaohong Kong, born on December 14, 1972, China

Current position, grades: researcher at Henna Institute of Science and Technology, China.
University studies: Ph.D. degree in Information Technology and Engineering from Jiangnan University, China in 2007.
Scientific interests: Automatic Control and Intelligent Algorithm.

Lixia Chang, born on January 22, 1972, China

Current position, grades: researcher at Henna Institute of Science and Technology, China.
University studies: Ph.D. degree in Textile engineering from Jiangnan University, China in 2014.
Scientific interests: apparel technology and culture.

Junpeng Xu, born on March 19, 1979, China

Current position, grades: Teacher at Henan Institute of Science and Technology, China.
University studies: master's degree in Fundamental Mathematics from Henan University, China in 2011.
Scientific interests: information technology and Automatic Control.

Gender impact on the identification based on EEG

Jinghai Yin, Zhendong Mu*

Institute of Information Technology, Jiangxi University of Technology, Nanchang 330098, China

Received 1 July 2014, www.cmnt.lv

Abstract

To study the genders impact on identification, this paper analysis the electroencephalograph (EEG) of eight male subjects and seven female subjects. In order to reduce the noise signal interference, the high pass and low pass were used to cut extra frequencies, and in order to prominent the feature signal, the power spectrum method was used to convent the time domain signal to frequency domain, and then fisher distance was used to extraction the feature. All EEG signal was acquired by neurescan, and the EEG signal was evoked by VEP method used subjects photo. The experiment was divided into three models: all subjects were the same sex, added some opposite sex, added some stranger. The analysis results show, to model 1, the correct recognition rate for male subjects, average is 88.50, and this for female, average is 92.51%; the false recognition rate for male subjects, average is 30.84%, and this for female, average is 27.67%,this result indicates VEP can be used as identification tool, the results of model 2 and model 3 show weather opposite sex or stranger should affect the correct recognition rate, but to male subject, the opposite sex effect is greater than stranger, to female, the result were reversed. The results also show noise photo affected female lower than male.

Keywords: Genders Impact, EEG, gender, Fisher Distance

1 Introduction

EEG brain of external things to make a biological reaction process to produce electricity, and when received outside special stimulation, the brain will produce features with this stimulation, the use of this nature, by brain researchers electrical signals to the brain as a bridge launched a variety of studies, such as the use of EEG studies BCI system control peripheral devices, the use of EEG studies of various diseases, the use of EEG to study various psychological phenomena, etc.. With the development of a variety of camouflage technology, the traditional identification technology increasingly difficult to do today's information security applications, looking for new biometric technology becomes increasingly important.

EEG advantages due to its difficult to counterfeit, and therefore the use of EEG as an identification tool for many a research investigator in the study of electrical signals in the brain, evoked potentials, especially the visual evoked potential cut is easy to implement because of the stability characteristics which is widely used in the identification same study, the researchers also used a lot of evoked potentials identification study, more successful in these studies. Paranjape et al Acquisition subjects wide open, EP signal is generated when the eyes closed, use AR model parameters as recognizable features [1]. In a recent research results, Riera et al chose the two electrodes, collecting subjects wide open, EP eyes closed when the signal is analyzed using autoregressive, Fourier transform, mutual information, etc. multi- feature fusion achieved 87.5% to 98.1% recognition rate[2]. When there is a specific visual stimulus will produce visual evoked potential (Visual Evoked Potential, VEP), which is a good entry point, Singhal et al use VEP peak there are individual differences in characteristics, setting relevant peak potential matching algorithm identifying subjects also achieved good results[3]. In study Biel, etc., then the design of the subjects viewed a picture, the choice of the three electrodes, using VEP as a signal source analysis, implementation verification using multivariate statistical methods. Palaniappan use the definition of an automatic identification method using Davies-Bouldin index information to select the largest electrode, the black and white pictures of common objects as visual stimuli, using a neural network classification [4, 5]. Touyama et al study is unfolding from the perspective of ERP[6], they take advantage of the characteristics of the P300, P300 evoked that will only appear when the target stimulus, the experiment they let 9 photos randomly, participants selected targets stimulation, selected different subjects of different target stimulus, they choose a password that is [7]. Wardzinski et al while the use of linear and nonlinear methods for EEG signal analysis, the model parameters as features, and based on Mahalanobis distance classifier, the recognition rate of 88% [8, 9]. Poulos et al uses of wave EEG signals were identified, using the AR model for feature extraction, the final recognition rate between 72 -84%[10,11,12,13].

Based on existing research, we use visual evoked potential as a tool for identity recognition research, the first of the collected EEG filtering, select a specific wavelength as the signal source, and then on the EEG power spectrum estimation, calculation different band EEG power spectrum, and finally the use of distance between the characteristics of different subjects for feature selection, feature matching validated by the final

* ***Corresponding author*** e-mail: 418623577@qq.com

results achieved recognition rate 80%-96%[14,15], the results show that the use of visual evoked potentials can be very good for identity recognition research.

2 Visual evoked potentials and experimental models

Visual evoked potential (Visual Evoked Potential, EVP) is occipital cortex electrical response to visual stimuli occurred District, on behalf of the retina receive stimulation through visual pathway conduction to the occipital cortex caused by potential changes. Visual evoked potential (VEP) in respect of its stimulation divided into non graphical stimulus (flash stimulation) and graphic stimuli into two categories, the current use of the latter in particular Othello plate grid, grating pattern image. Their black and white compositions at a rate conversion constitute a valid alternative stimulation induced electrical activity called pattern reversal visual evoked potentials (PRVEP). With visual stimulation in the occipital scalp recorded VEP main representative of the central field of vision visual impulses 6 ° ~ 12 °, and by the lateral geniculate body projected onto the occipital pole electrical activity from the state after the split rear and pillow . Today, visual evoked potentials, is the most successful application of visual evoked potentials for detection of retinal ganglion cells and optic nerve diseases such as central as the passage of optic neuritis, multiple sclerosis, cancer and other oppressed optic neuropathy disease ; graphics VEP can be used for an objective determination of visual acuity. In the visual evoked potential studies often use ERP component analysis, the main object of study for the N75, P100, N145 and other ingredients.

As used herein, the basis of the different subjects of the experiment is to familiarize yourself with the different responses favorite pictures and other pictures reactions, resulting in different brain wave signal is mainly based on, for example, a photo of yourself and a photo of a different reaction to the others collected in accordance with the visual EEG evoked potentials data . Experiments in Jiangxi Institute of Technology Institute of Information Technology BCI laboratory, Jiangxi University of Science subjects in school, subjects were placed in a quiet shielded room, sitting on soft chairs without armrests experiment during the experiment, subjects in accordance with the test requirements, looking at the computer screen in front of the related operations, 15 experimenters were divided into three groups of five people, the experimenter experimental models include the following three modes : experimental background for the next gay familiar experiments have become familiar with the background to participate in sex case experiments and experimental background in the case of a stranger to participate in the experiment. Under each group mode, the experimenter was asked to do attention be collected under their own photos EEG circumstances and concerns of others cases, each experiment, the picture display 750ms, then the interval 1s.

EEG acquisition is the use of lead Neuroscan amplifier 40 through scan4.3 software for acquiring, using the right of way of the mastoid reference electrode as the reference electrode, the use of 1000Hz sampling rate, the use of 200Hz low-pass band collection, 0.05Hz 50Hz high- pass and notch.

3 Method

3.1 FILTER

Intermediate signals received useful signal selection process is called filtering, "received signal" corresponds to the observed random process, the "useful signal" corresponds to the random process is estimated. E.g., aircraft tracking radar, the measured position data of the aircraft, the other containing the measurement errors and random noise, how to use them to estimate as accurately as possible the position of the aircraft at each moment, velocity, acceleration, and to predict future aircraft position, is a filtering and prediction problems. Such problems in electronics, aerospace science, control engineering and other scientific and technical sectors are abounding.

Filtering is carried out in accordance with, or just on some sampling points can be divided into a continuous-time filter and a discrete-time filtering on the entire time. The former set time parameters T will be desirable half real axle $[0,\infty)$ or the real axis; latter T desirable non-negative integers {0, 1, 2, 3…} or integers {…,-2,-1, 0, 1, 2…}.

Set $X=\{X,t\in T,t\in T\}$ is limited, that is to say:

$$C(H_j\mid r)=\sum_{i=1}^{M-1}C_{ji}P(H_i\mid r), \tag{1}$$

where X is a process to be estimated, it can not be directly observed; Y is the observed process, which contains some information of X.

Use $y^t=\{y_s:s\in T,s\le t\}$ to represents the observed data with the time t until the whole. If you can find a function f(x) about a variable in y making it reach the minimum mean square error of $E\mid X_t-f(y)^t\mid$, $X_t-f(y)^t$ is said to the optimal filtering $X_{t.}$

In order to facilitate the application and narrative, sometimes also above definition to classify more detail. Let τ t is a real number or integer determined, and the process is considered to be estimated as

$$(X_{\tau+t},t\in T),X_{\tau+t}=\tilde{E}(X_{\tau+t}\mid y). \tag{2}$$

Or

$$\tilde{E}(X_{\tau+t}\mid y^t),\tilde{X}_{t+\tau|t}=X_{t+\tau}-\tilde{X}_{t+\tau|t},D_{t+\tau|t}. \tag{3}$$
$$=E(X_{t+\tau}\mid y^t)$$

According to τ = 0, τ> 0, τ <0, are called optimal filtering, (τ-step) prediction or extrapolation, (τ step)

smoothing or interpolation. $\tilde{X}_{t+\tau|t}$, $D_{t+\tau|t}$ corresponding to each error and the mean square error, and collectively these problems for the filter problem.

According to brain wave frequency characteristics, this paper 1Hz high pass and 55Hz low-pass method of combining the signal source is filtered, collected EEG first by 1Hz high-pass filter to reduce direct impact, and then conduct a 55Hz low-pass filter, so You can select 1-55hz useful EEG.

3.2 POWER SPECTRUM CALCULATION

The original EEG is a signal domain, the features are hidden between the various noise data, through the original EEG feature extraction speed and accuracy will be affected, in order to improve the speed of analysis and analysis accuracy, can be converted to the original EEG power spectrum, the frequency analysis.

A simple way to estimate the power spectrum of the stochastic process is a direct request sampling DFT, and then takes the results of the square of the amplitude. Such a method is called period gram

A length L of the signal $x_L[n]$ is estimated period gram PSD as follow

$$\tilde{P}_{xx}(f)=\frac{|X_L(f)|^2}{f_s L}, \tag{4}$$

where $X_L(f)$ is defined using the FFT matlab inside without normalization coefficients, so dividing L.

$$X_L(f)=\sum_{n=0}^{L-1}x_L[n]e^{-2\pi jfn/fs}. \tag{5}$$

The actual calculation of $X_L(f)$ may be performed using FFT, and only on a limited frequency. Applications in practice are most period gram PSD estimation to calculate the N-point:

$$\tilde{P}_{xx}(f)=\frac{|X_L(f)|^2}{f_s L}, f_k=\frac{kf_s}{N}, k=1,2,3...,N-1$$

$$X_L(f_k)=\sum_{n=0}^{L-1}x_L[n]e^{-2\pi jfn/N}. \tag{6}$$

3.3 FEATURE EXTRACTION METHODS

EEG feature extraction in order to extract useful brain signals, thereby to classify the feature extraction method to identify subjects used herein are described below:

Step1. Common average: In this paper, we use Hjort derivation to reduce interference from the neighbouring electrode, The Hjort derivation C_i^H is calculated as

$$C_i^H=c_i-\frac{1}{4}\sum_{j\in S_i}sc_j, \tag{7}$$

where ci is the reading of the centre electrode sc_j, with i=1…30 and j is the set of indices corresponding to the eight electrodes surrounding electrode c_i.

Step2. Filter: EEG signal acquisition band from 0.05Hz to 200Hz, in order to extract features, we filter EEG signal band from 0.05Hz to 50Hz.

Step3. AR conversion: time-domain EEG data disorganized EEG in order to better highlight the characteristics of EEG signal, we use AR model to convert the time domain signals into frequency domain, and extract the feature from the frequency domain signals.

Step4. Calculate add of Fisher' distance: Calculated the Fisher distance between two classes. Fisher distance $F_{i,j}$ is calculated as

$$F_{i,j}=\frac{(\mu_i-\mu_j)^2}{\sigma_i^2+\sigma_j^2}, \tag{8}$$

where $F_{i,j}$ is Fisher' distance between the subject I and the subject j. μ and σ are the mean and the standard deviation of the feature they correspond to.

The add of Fisher' distance F_i is calculated as

$$F_i=\sum_{j\neq i1}F_{i,j}. \tag{9}$$

Step5. Sort the Fisher's distance adds: The Fisher distance was often used to denote differences between classes in classification research. The bigger the fisher distance was the more notable the difference was. So we sort the Fisher's distance on descending order.

Step6. Feature extraction: Through analysis the five subjects Fisher's distance, we select three thousand as feature range, all feature are selected from this range.

Step7. Get the classifier: Use BP neural network to calculate the feature then can get these five subjects classifier. To the test sample of these subjects, if the result is in credibility, that is to say, this sample is right

In this paper, the correct recognition rate, error recognition rate, false recognition rate and credibility used to analysis the results, the following were defined them:

1. Correct recognition rate: the correct recognition rate is defined as the rate of correct recognizes the iden-

tity of subjects through use subjects' classification to test their own samples.

2. Error recognition rate: the error recognition rate is defined as the rate of incorrect recognizes the identity of subjects through use subjects' classification to test their own samples.

3. False recognition rate: the false recognition rate is defined as the rate of recognize the identity of subjects is himself through use subjects' classification to test others samples.

4. Credibility. The credibility is defined as acceptable error range (in this paper, the credibility is [-0.02 0.02]).

4 Results and discussion

In this paper, experiments using three modes: all same group subject is same gender and familiar, some subject in group is different gender, some subjects in group is stranger. Participation has sex with strangers during the experiment, namely to study the impact of the opposite sex and strangers on the classification accuracy, each of which is divided into a group of five people to experiment, three groups were collected from 15 individuals of brain waves, which 8 males, 7 females, each subject collected EEG data 183 trial, excluding invalid trial, each subject was eventually carried out using 156 trial identification calculations, this paper is divided into these 156 trial characteristics extraction collection and validation collection, which features a collection of 50 extraction trial, the validation set 106 trial.

After the same background, the same group of subjects were EEG analysis, the feature set of the first to use the method described earlier for data conversion, data conversion After feature extraction, feature subjects extracted using the same data conversion method for EEG data conversion, and then select the appropriate feature set, classification calculated by feature matching, feature extraction for a set T, and a test set tc, electrical test matches z formula can be expressed in the brain as follows:

$$z = (\sum_{i=1}^{n} | x_i - y_i |) / n, x_i \in T, y_i \in TC, \quad (10)$$

where n is the number of features.
Classification accuracy of the eight male subjects show as following.

As described in Figure 1, the x axis indicates that the three experimental model, 1 is all same group subject is same gender and familiar, 2 is some subject in group is different gender, 3 is some subjects in group is stranger. The Y axis indicates that the value of correct recognition rates, and corresponding to the histogram of each model is eight male subjects.

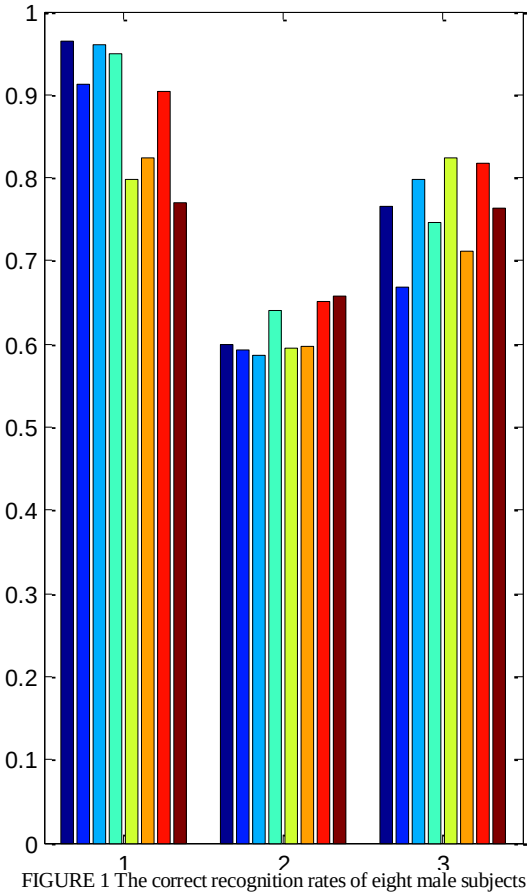


FIGURE 1 The correct recognition rates of eight male subjects

As figure 1 shows, when input the test samples, the correct recognition rate of model 1 are 96.33%, 91.18%, 96.03%, 94.96%, 79.71%, 82.41%, 90.04%, 77.05%. The correct recognition rate of model 2 are 59.89%, 59.20%, 58.70%, 63.96%, 59.52%, 59.79%, 65.21%, 65.59%. The correct recognition rate of model 3 are 76.53%, 66.85%, 79.85%, 74.71%, 82.30%, 71.22%, 81.64%, 76.37%. As correct recognition rate result shows, if the same group subject all is male, the lowest result is 77.05%, and the highest result is 96.33%, that is to say, use EEG signal as tool of identification can recognize them, if test in model 2, the lowest result is 58.70%, the highest result is 65.59%, compare to model 1, the correct recognition rate is decreased, if test in model 3, the lower result is 66.85%, the highest result is 82.30%, compare to model 1, the correct recognition rate is decreased too.

Figure 2 show the eight male subjects false recognition rates.

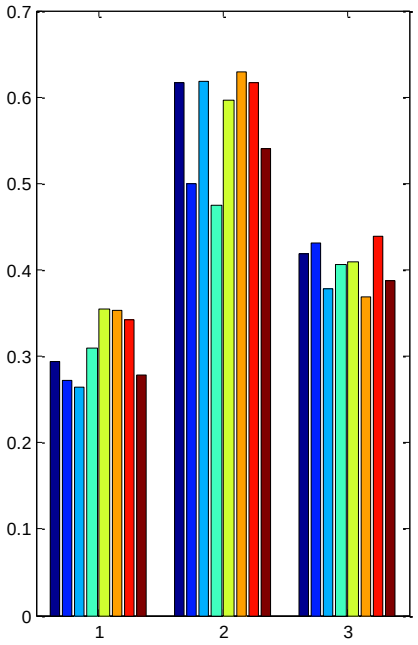


FIGURE 2 The false recognition rates of eight male subjects

As described in Figure 2, the x axis indicates same as Figure 1; y axis indicates that the value of false recognition rates of eight male subjects.

As figure 2 shows, when input the test samples, the false recognition rate of model 1 are 29.31%, 27.20%, 26.45%, 30.98%, 35.48%, 35.36%,34.17%, 27.75%. The false recognition rate of model 2 are 61.66%, 49.90%, 61.89%, 41.49%, 59.65%, 62.87%, 61.73%, 54.06%. The false recognition rate of model 3 are 41.79%, 43.13%, 37.86%, 40.56%, 40.95%, 36.91%, 43.86%, 38.67%.

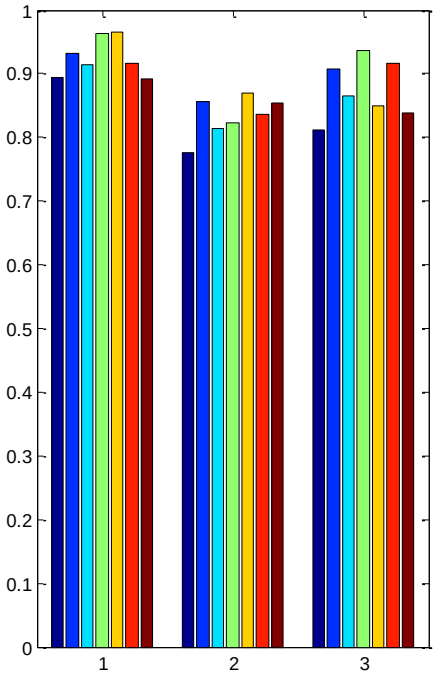


FIGURE 3 The correct recognition rates of seven female subjects

The x axis indicates same as Figure 1; y axis indicates that the value of correct recognition rates of seven female subjects.

As figure 3 shows, when input the test samples, the correct recognition rate of model 1 are 89.46%, 93.20%, 91.32%, 96.24%, 96.46%, 91.69%, 89.21%. The correct recognition rate of model 2 are 77.66%, 85.70%, 81.31%, 82.31%, 86.68%, 83.63%, 85.38%,. The correct recognition rate of model 3 are 81.11%, 90.78%, 86.43%, 93.69%, 84.96%, 91.74%, 83.92%,. As correct recognition rate result shows, if the same group subject all is male, the lowest result is 89.21%, and the highest result is 96.46%, that is to say, use EEG signal as tool of identification can recognize them, if test in model 2, the lowest result is 77.66%, the highest result is 86.88%, compare to model 1, the correct recognition rate is decreased, if test in model 3, the lower result is 81.11%, the highest result is 93.69%, compare to model 1, the correct recognition rate is decreased too.

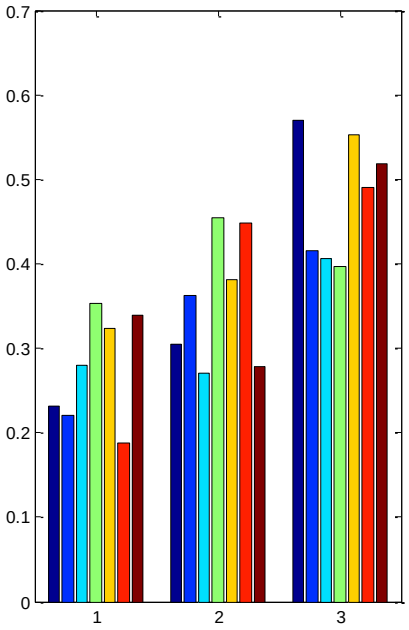


FIGURE 4 The false recognition rates of seven female subjects

As described in Figure 4, the x axis indicates same as Figure 1; y axis indicates that the value of false recognition rates of seven female subjects.

As figure 4 shows, when input the test samples, the false recognition rate of model 1 are 23.22%, 22.04%, 27.99%, 35.32%, 32.42%, 18.73%, 33.95%,. The false recognition rate of model 2 are 30.48%, 36.23%, 27.08%, 45.44%, 38.15%, 44.87%, 27.76%. The false recognition rate of model 3 are 56.91%, 41.50%, 40.65%,39.63%, 55.31%, 49.00%, 51.80%,.

Experimental results show that the time when the subject of the same sex, male subjects being able to confirm the correct recognition rate their average of 88.50%, and the average of female correct recognition is

92.51%, the test sample is added when the others after EEG sample misidentification rate male subjects with an average of 30.84%, the error recognition rate of female subjects with an average of 27.67%, the experimental data show that in their own photo as a reference, the female subjects than recognition rate is generally higher than male subjects male subjects, while female subjects to others EEG signals recognized as its own recognition rate is also lower than male subjects, VEP feature shows the subjects picture familiarity observed, the average correct recognition rate and error recognition rate results showed photos of female subjects for their degree of concern than male subjects, the correct identification rate data also show the use of visual evoked potentials can be used as identity identification tool.

In order to study the change of identification rate, this paper introduced opposite sex and strangers photo as noise. in the same experimental model and methods of analysis, compared to model 1, table 1 show the correct recognition rate of eight male change of add opposite sex and strangers experimental model. the classification accuracy decreased average: 0.2702, wherein the maximum change is 0.3733, min changes is 0.1146, when the stranger cases, the average correct classification rate dropped 0.1232, the biggest change is 0.2433, the smallest change is 0.0259.

TABLE 1. Male affect recognition rate

number	correct recognition rate of mode 1	model2 change	model 3 change
1	0.9633	0.3644	0.1980
2	0.9118	0.3198	0.2433
3	0.9603	0.3733	0.1618
4	0.9496	0.3100	0.2020
5	0.7971	0.2019	-0.0259
6	0.8241	0.2262	0.1119
7	0.9034	0.2513	0.0870
8	0.7705	0.1146	0.0068

The result of Table 1 show that male subject easer affected by opposite sex than stranger.

The same experiment for female, we can get the result table 2.

TABLE 2 Female affect recognition rate

number	correct recognition rate of mode 1	model2 change	model 3 change
1	0.8946	0.1180	0.0835
2	0.9320	0.0750	0.0242
3	0.9132	0.1001	0.0489
4	0.9624	0.1393	0.0255
5	0.9646	0.0958	0.1150
6	0.9169	0.0806	-0.0005
7	0.8921	0.0383	0.0529

As table 2 shows, to the seven female subjects, if add opposite sex subject, the correct recognition rates would decrease, the average rate dropped 0.0924, the biggest change is 0.1393 and the smallest change is 0.0383.
The result of Table 2 show that the opposite sex and strange also affected the female subjects focus.

Compare table 1 and table 2 results, gender impact on the experiment as show in Figure 5, to opposite sex, this effect is more than female, and to strange, this effect also male more than female, this VEP result show that female is focus on their own photo more than male, so the correct recognition rate effect to noise photo is smaller than male. Whether male or female, the correct recognition rate show use this way to identification is effective.

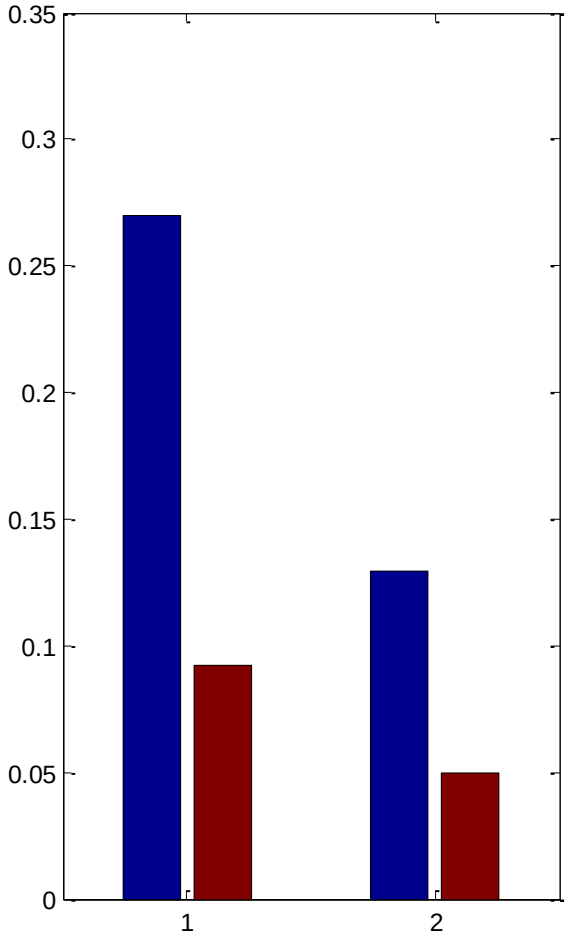


FIGURE 5 Gender impact on correct recognition rate

The correct recognition rate is affected by noise photo; weather false recognition rate would effect by noise too?

Table 3 and table 4 is male and female result of this effect.

TABLE 3 Male affect false recognition rate

number	false recognition rate of mode 1	model2 change	model 3 change
1	0.2931	-0.3235	-0.1248
2	0.2720	-0.2270	-0.1593
3	0.2645	-0.3544	-0.1141
4	0.3098	-0.1651	-0.0958
5	0.3548	-0.2417	-0.0547
6	0.3536	-0.2751	-0.0155
7	0.3417	-0.2756	-0.0969
8	0.2775	-0.2631	-0.1092

TABLE 4 Female affect false recognition rate

number	false recognition rate of mode 1	model2 change	model 3 change
1	0.2322	-0.0727	-0.3369
2	0.2204	-0.1420	-0.0194
3	0.2799	-0.0091	-0.1266
4	0.3532	-0.1011	-0.0430
5	0.3242	-0.0573	-0.2289
6	0.1873	-0.2613	-0.3027
7	0.3395	-0.0619	-0.1785

As show in Table 1 and Table 2, under model 2, that is to say if a opposite sex subject be added, the false recognition rate, to male subject will ascend, the average number is raise 0.2657, and under model, the false recognition rate is also ascend, this number average raise 0.0963, this result show, to male subject, the opposite sex is more attractive than stranger. The same analyses method to female, under this two models, the false recognition rate are all ascend, to model 1,this number raise 0.0805, and to model 2,this number raise 0.2016, that is to say, to female subjects, the stranger is more attractive than opposite sex, this result can describe in Figure 6.

To the study of VEP, if uses the method of overlap some trial data, the gender impact is not big, but to the study of identification, weather this impact is no big too? Through the eight male subjects and seven female subjects research results show that the noise of opposite sex and strangers subjects will be affected, but the impact of female subjects than male subjects suffered by little affected.

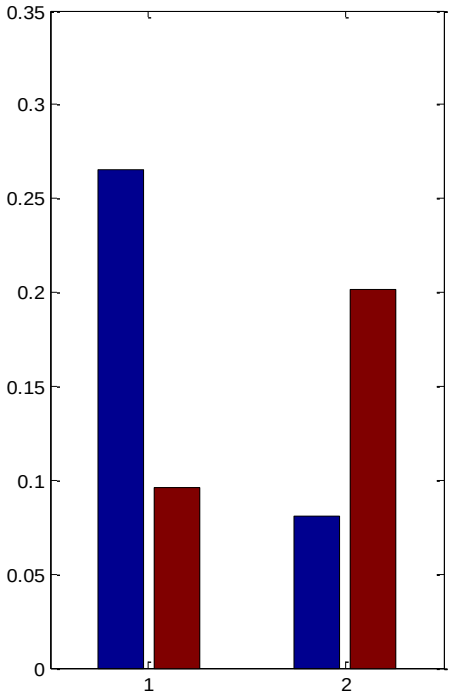


FIGURE6. Gender impact on false recognition rate

Acknowledgments

This work was supported in part by a grant from natural science foundation of Jiangxi Province [No 20122BAB201049] and IT projects of Jiangxi Office of Education [No. GJJ14765].

References

[1] Wang Bei, et al. 2014 Automatic reference selection for quantitative EEG interpretation: Identification of diffuse/localised activity and the active earlobe reference, iterative detection of the distribution of EEG rhythms *Medical engineering & physics* **36**(1) 88-95

[2] Zuhair Mohd, et al. 2014 Automatic Identification of an Epileptic Spike Pattern in an EEG Signals Using ANN *Proceedings of the Third International Conference on Soft Computing for Problem Solving* Springer India

[3] Kidokoro Hiroyuki, et al. 2013 Neonatal seizure identification on reduced channel EEG *Archives of Disease in Childhood-Fetal and Neonatal Edition* **98**(4) F359-F361

[4] Xu Rui, et al. 2013 Feature Extraction of Individual Differences for Identification Recognition Based on Resting EEG *Cross-Cultural Design. Methods, Practice, and Case Studies* Springer Berlin Heidelberg 503-9

[5] Pan Yaozhang, et al. 2013 Common frequency pattern for music preference identification using frontal EEG *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on* IEEE

[6] Wang Bei, et al. 2013 Automatic identification of diffuse and localized activity for topographical distribution of EEG rhythm based on suitable reference selection with pre-judgments *Complex Medical Engineering (CME), 2013 ICME International Conference on IEEE*

[7] Rocca Daria La Patrizio Campisi, Jordi Sole-Casals 2013 EEG based user recognition using BUMP modelling *Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the IEEE*

[8] Wang Ting, et al. 2014 EEG Eye State Identification Using Incremental Attribute Learning with Time-Series Classification *Mathematical Problems in Engineering*
[9] Toma Junya, Tadanori Fukami, Takamasa Shimada 2013 Character identification by maximizing the difference between target and non-target responses in EEG without sophisticated classifiers *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*
[10] Lu Peng, et al. 2013 Single-Trial Identification of Motor Imagery EEG based on HHT and SVM. *Proceedings of 2013 Chinese Intelligent Automation Conference. Springer Berlin Heidelberg*
[11] Qiao Xiao Yan, Chun Hui Wang 2013 The Identification of EEG Feature Evoked by Imaginary Movement. *Applied Mechanics and Materials* **427** 2059-63
[12] Sanchez S M, et al. 2013 Pediatric ICU EEG monitoring: current resources and practice in the United States and Canada *Journal of Clinical Neurophysiology* **30**(2) 156-60
[13] Zhang Xue Min, Zeng Gang Xiong, Zhen Dong Mu 2013 Authentication Based on Eyes-Closed EEG *Advanced Materials Research* **811** 426-9.
[14] Zeng Changqing, Zhendong Mu 2013 A Mathematical Model of Authentication under Normal State EEG *International Journal of Applied Mathematics and Statistics* **49**(19) 366-74
[15] Mu Zhen-dong, De-rong Jiang, Jing-hai Yin 2012 EEG identification computing based on photo stimulation *International Journal of Digital Content Technology and its Applications* **6**(13) 446-53

Authors

Jinghai Yin, born on January 28, 1977, China

Current position, grades: researcher at Jiangxi University of Technology, China.
University studies: M.E degree in computer science and technology from Nanchang University, China in 2005.
Scientific interests: information security, algorithm design and EEG analysis.

Zhendong Mu, born on November 1, 1975, China

Current position, grades: researcher at Hunan University of Commerce, China.
University studies: M.E degree in computer science and technology from Nanchang University, China in 2004.
Scientific interests: information security, algorithm design and EEG analysis.

Sport service evaluation of urban community based on fuzzy comprehensive evaluation

Zhengmei Lin[1*], Jiwei Yao[2]

[1] *Department of Physical Education and Sport Science, Fuqing Branch of Fujian Normal University, Fuqing 350300, Fujian, China*

[2] *Physical Education Institute, Hunan University of Science and Technology, Xiangtan 411201, Hunan, China*

Received 1 June 2014, www.cmnt.lv

Abstract

Fuzzy comprehensive evaluation refers to a comprehensive assessment method based on fuzzy mathematics through quantifying the factors difficult to quantify with obscure boundaries and using the principle of fuzzy relation synthetic. Fuzzy comprehensive evaluation can comprehensively evaluate the object system involving fuzzy factors and is widely applied in the fields of economy and society. This paper takes new public service theory as the research perspective; designs evaluation index system which can comprehensively reflect sport development level of urban community applies fuzzy comprehensive evaluation to evaluate sport development level of urban community in Hunan province and provides the evaluation process. The evaluation results show leadership organization and team construction of the urban community fail to develop to certain level and that the health index of community residents is slightly low. This paper presents suggestions to improve this situation.

Keywords: Fuzzy Mathematics, Fuzzy Comprehensive Evaluation, Evaluation Index System, Urban Community, Sport Service

1 Introduction

Fuzzy comprehensive evaluation (FCE) was first put forward by an American L.A. Sade in 1965 [1]. It is a method to describe fuzzy mathematics written in Fuzzy Sets. This method declared the birth of fuzzy mathematics. Since then, fuzzy phenomena have entered the field of human scientific research [2]. FCE is an important aspect of fuzzy mathematics applied in natural science field and social science field. Evaluation objects of comprehensive evaluation problem are decided by factors in multiple aspects, so it is required to evaluate every factor [3]. With regard to evaluation of sport development level of urban community, the influencing factors are very complex and fuzzy. Thus, fuzzy means is used to deal with fuzzy problems, which will make the evaluation results truer and more rational [4].

Community sport as the main base for national fitness and an important constituent part of community building is highly valued by the government [5]. However, it is found in practice real sport service system of urban community has not been established currently [6]. Various mass physical fitness organizations, grass-root fitness clubs, fitness stations and grass-root sport associations as well as the network formed on this basis are not perfect [7]. The development is unstable and disorderly [8]. How to make the input in community sport more scientific and rational, to make community sport resources utilized more fully and effectively, to make organizational measure of community sport more beneficial and to make community sport produce greater social benefits requires scientific evaluation of community sport resources [9].

One thing often needs multiple indexes to depict its nature and characteristics. Besides, the people's evaluations for one thing are not simply good or bad, but they adopt fuzzy language to evaluate it with different degrees [10]. Since the relation of evaluation levels are fuzzy without absolutely explicit dividing line, so the evaluation is fuzzy [11]. Obviously, for this fuzzy evaluation problem, classical evaluation method is not rational. But the application of fuzzy mathematics for comprehensive judgment will gain better practical effects. So, this paper utilizes FCE method to evaluate community sport development level of a city on the basis of analysing the factors influencing community sport development level.

2 Public service level index system of community sport

With the advancement in networking and multimedia technologies enables the distribution and sharing of multimedia content widely [12]. In the meantime, piracy becomes increasingly rampant as the customers can easily duplicate and redistribute the received multimedia content to a large audience [13].

The hierarchy, regional nature, overall comprehensiveness and trend of public service of community sport decide public service level index system of community sport has the following characteristics [14]:

(1) Universality of evaluation content. There are many factors influencing public services of community sport, including two aspects: "hardware" and "software". Since the contents involved are many, index system design,

[*] ***Corresponding author*** e-mail: zhengmei1111@yeah.net

standard formulation and method selection should follow the principle of combining subjectivity and objectivity, nature determination and quantification as well as self-evaluation and expert evaluation.

TABLE 1 Opinions of community residents on national fitness paths and equipment

	Able to exercise	Interesting	Good shape	Useless	Danger
n	115	27	28	20	10
%	57.5	13.5	14.0	10.0	5.0

(2) Regional correlation. Public service level of community sport is also restricted by social economic development level of the community. Construction environment of index system should start from practical situations and realistic demand, ensure combination of scientificalness and practical applicability and meanwhile consider the imbalance of regional development. So, different things should not be treated as the same simply.

TABLE 2 Construction situations of fitness centers of 6 cities

	City A	City B	City C	City D	City E	City F
Fitness centers	10	8	7	10	17	4
Fitness sites	55	40	61	54	66	18
Communities	51	65	45	66	96	6
Proportion (fitness centers)	1.7	1.0	1.8	1.3	1.2	5.0

(3) Objective consistency. Consistency contains the maiming at two levels. At the first level, the objective to establish evaluation index system should be consistent with the demands of community members. The establishment of system index is to better serve community residents for physical fitness. So, the system should be able to reflect the common desire of the community residents and be oriented to their demands. At the second level, evaluation objective and construction objective should be consistent.

Overall structure of public service indexes of community sport refers to the core problem of index system design. During construction of this system, all indexes of public services of community sport should form an organic whole and coordinate mutually. There are many factors influencing public service of community sport. These factors involve various aspects. The influences of some indexes are very small and can be neglected. There are many evaluation indexes for sport development level of urban community, including qualitative index, quantitative index, macroscopic index and microcosmic index. In comprehensive evaluation, corresponding indexes are often selected from these for permutation and combination.

3 FCE of sport development level of urban community

FCE can comprehensively evaluate the object system involving fuzzy factors and is widely applied in economic and social fields [15]. FCE has strong subjectivity in evaluation and cannot solve the problem of repeated evaluation information caused by evaluation indexes. In addition, there is no systematic method to confirm membership function [16]. Especially when faced with a complex system, since there are many factors in need of consideration; it is very difficult to confirm weight allocation of each factor when FCE is applied. Meanwhile, FCE applies normalization handling method. After weight allocation is confirmed, it is still necessary to comprehensively evaluate Sade operator for the index. Thus, the results cannot meet the due value. This paper based on the above considerations optimizes FCE and adopts multi-level FCE for multi-level processing of weight allocation. They are classified into several levels according to factors or indexes. Firstly, low-level factors are comprehensively evaluated; the evaluation results are used for higher-level comprehensive evaluation. Single factor evaluation of each level is multi-factor comprehensive hiragana at lower-level. So, the evaluation is carried out from low level to high level one by one. Moreover, to consider from different perspectives, we can first classify the personnel participating in judging. In line with the steps of FCE, fuzzy statistics matrix of each type of judging personnel on the evaluation object is provided to calculate the judging results of each type of judging personnel on the evaluation object. The influences of the judges with different perspectives are considered through "quadric weighting". The steps for construction of multilevel FCE model are shown in Fig.1:

(1) To confirm multilevel factor set according to original data of the evaluation object. The factor set includes all indexes used to evaluate the object. The factor set is expressed through collective concept. The matrix form of the collection is as follows:

$$U=(U_1,U_2,\cdots,U_n)=\begin{pmatrix}(u_{11},u_{12},\cdots,u_{1m})\\ \vdots \\ (u_{n1},u_{n2},\cdots,u_{nm})\end{pmatrix}=\begin{pmatrix}\begin{pmatrix}(u_{111},u_{112},\cdots,u_{11i})\\ \vdots \\ (u_{1m1},u_{1m2},\cdots,u_{1mi})\end{pmatrix}\\ (\vdots)\\ \begin{pmatrix}(u_{n11},u_{n12},\cdots,u_{n1i})\\ \vdots \\ (u_{nm1},u_{nm2},\cdots,u_{nmi})\end{pmatrix}\end{pmatrix}, \quad (1)$$

where, U is factor set; n is the number of the first-level factors; each first-level indexes can be divided into $U_i=(u_{i1},u_{i2},\cdots,u_{im})$ second-level sub-indexes; m is the number of second-level indexes. By parity of reasoning, all lower-level I can be reasoned out.

(2) To catty out weight assignment for each index in the above index sets and the weight sets gained are as follows:

$$W=(W_1,W_2,\cdots,W_n)=\begin{pmatrix}(w_{11},w_{12},\cdots,w_{1m})\\ (w_{21},w_{22},\cdots,w_{2m})\\ \vdots \\ (w_{n1},w_{n2},\cdots,w_{nm})\end{pmatrix}=\begin{pmatrix}\begin{pmatrix}(w_{111},w_{112},\cdots,w_{11i})\\ \vdots \\ (w_{1m1},w_{1m2},\cdots,w_{1mi})\end{pmatrix}\\ (\vdots)\\ \begin{pmatrix}(w_{n11},w_{n12},\cdots,w_{n1i})\\ \vdots \\ (w_{nm1},w_{nm2},\cdots,w_{nmi})\end{pmatrix}\end{pmatrix}, \quad (2)$$

where, $W=(W_1,W_2,\cdots,W_n)$ is the weight of the first-level index $U=(U_1,U_2,\cdots,U_n)$; $W=(w_{i1},w_{i2},\cdots,w_{im})$ is the weight of second-level index $U_i=(u_{i1},u_{i2},\cdots,u_{im})$ relative to the first-level index $U=(U_1,U_2,\cdots,U_n)$.

(3) To conduct level evaluation for all indexes. Usually, $V=(v_1,v_2,\cdots,v_k)$ is expressed as the evaluation level of each index, where k is the number of evaluation level. In FCE, each level evaluation is classified into five levels {very good, good, general, bad, very bad}, and it is expressed in figure is $V=\{5,4,3,2,1\}$.

(4) After the above work is finished, confirm fuzzy mapping relationship between evaluation index and evaluation set in the index sets. The fuzzy mapping matrix is as follows:

$$W_j^{carbon}=(\alpha_1+\alpha_2+\alpha_3)\sigma\gamma_j q_j^s=M\gamma_j q_j^s \quad (3)$$

where, n is the number of the first-level indexes; R_i is fuzzy mapping if i indexes in the first-level indexes ($i=1,2,3,\cdots,n$); k_i is the number of the last-level indexes corresponding to every first-level indexes; r_{ij} means membership degree of last indexes to evaluation set V.

(5) Finally, implement matrix operations for matrix fuzzy relation and weight set $B=RW$. Then, fuzzy evaluation results are gained.

FCE is a very effective multi-factor decision-making method to comprehensively evaluate the things influenced by multiple factors. Evaluation of public services of community sport often involves multiple factors or indexes. At this moment, it is required to evaluate the thing according to multiple factors, instead of single factor. This paper establishes FCE model for public services of community sport through referring to pertinent literatures.

4 Application example

This paper takes urban community in Hunan province for example and conducts empirical analysis of urban community in Hunan province. In accordance with characteristics of urban community in Hunan province, software and hardware conditions of existing sport, an evaluation system including three second-level indexes (infrastructure and service, resident participation degree and sport benefit) and 11 third-level indexes under the basic principle of SERVQUAL. The indexes are shown in Table 3.

TABLE 3 Evaluation index system

Level 1 index	Level 2 index	Level 3 index
Sport service level of urban community	Infrastructure and service	Management rules
		Sport service team
		Sport field
		Sport funds
	Resident participation degree	Form of sport activities
		Resident participation time
		Sport consumption expenditure
		Large fitness activity
		Participation in superior sport activity team
	Sport benefit	Physical fitness measurement index
		Results of athletic contest

Confirm the weight of each index, evaluation level and membership degree through analytic hierarchy process, as shown in Table 4.

TABLE 4 Evaluation index weight of urban community sport

Second-level		Third-level		Evaluation index and membership degree				
Index	*Weight*	*Index*	*Weight*	*V1*	*V2*	*V3*	*V4*	*V5*
U1	0.35	U11	0.25	0.1	0.25	0.3	0.2	0.15
		U12	0.3	0.2	0.2	0.1	0.3	0.2
		U13	0.25	0.3	0.2	0.4	0.1	0
		U14	0.2	0	0.3	0.2	0	0.5
U2	0.35	U21	0.3	0.4	0	0.1	0.3	0.2
		U22	0.2	0.2	0.4	0.1	0.1	0.2
		U23	0.25	0	0.4	0.2	0.3	0.1
		U24	0.25	0.3	0.2	0.2	0.2	0.1
U3	0.3	U31	0.3	0.3	0.1	0.1	0.3	0.2
		U32	0.4	0.2	0.2	0.4	0.1	0.1
		U33	0.3	0.1	0.1	0.3	0.3	0.2

The result matrix of each step is gained according to the process of FCE as follows:

1) Index weight at each level

$$W=(W_1,W_2,\cdots,W_n)=(0.35,0.35,0.3)=\begin{pmatrix}(w_{11},w_{12},w_{13},w_{14})\\(w_{21},w_{22},w_{23},w_{24})\\(w_{31},w_{32},w_{33})\end{pmatrix}=\begin{pmatrix}(0.25,0.3,0.25,0.2)\\(0.3,0.2,0.25,0.25)\\(0.3,0.4,0.3)\end{pmatrix}$$

2) Membership degree and evaluation matrix of the third-level index

$$R_1=\begin{bmatrix}0.1&0.25&0.3&0.2&0.15\\0.2&0.2&0.1&0.3&0.2\\0.3&0.2&0.4&0.1&0\\0&0.3&0.2&0&0.5\end{bmatrix}$$

$$R_2=\begin{bmatrix}0.4&0&0.1&0.3&0.2\\0.2&0.4&0.1&0.1&0.2\\0&0.4&0.2&0.3&0.1\\0.3&0.2&0.2&0.2&0.1\end{bmatrix},$$

$$R_3 = \begin{bmatrix} 0.3 & 0.1 & 0.1 & 0.3 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.3 & 0.3 & 0.2 \end{bmatrix}$$

3) Comprehensive evaluation

$$R = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} W_1 R_1 \\ W_2 R_2 \\ W_2 R_2 \end{bmatrix} = \begin{bmatrix} 0.16 & 0.23 & 0.25 & 0.17 & 0.20 \\ 0.24 & 0.23 & 0.15 & 0.24 & 0.15 \\ 0.20 & 0.14 & 0.28 & 0.22 & 0.16 \end{bmatrix}$$

$$B = WR = (0.20\ 0.20\ 0.23\ 0.21\ 0.17)$$

Carry out five-level FCE of the above results and gain the valuation result of community sport development level as follows:

$$BV = (0.20\ 0.20\ 0.23\ 0.21\ 0.17)\begin{pmatrix} 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} = 3.05$$

The sport development of this urban community is 3.05, above general and below good. This analysis indicates sport development of urban community in Hunan province has large development space. This is mainly because the infrastructure and service level remain improving; resident participation awareness remains improving; resident health is at a low level. Thus, the community sport development needs urgent enhancement. The sport input and education should be highly valued by leadership organization of the community.

5 Conclusions and suggestions

5.1 CONCLUSIONS

This paper designs evaluation index system which can comprehensively reflect community sport service level, applies FCE to evaluate sport development level of an urban community and provides the evaluation process.

Evaluation system for sport service level of urban community contains 3 second-level indexes and 11 third-level indexes.

This paper adopts FCE to construct FCE mathematical model for public services of community sport and implements empirical research by taking urban community in Hunan province as example. The results show community sport development level of Hunan province is above average and remains further improving.

5.2 SUGGESTIONS

To establish and perfect public service performance evaluation system for community sport as soon as possible according to the requirement for community sport service construction put forward by modern social and economic development by taking evaluation as the means and improvement of public services of community sport as the objective;

In order to enhance service-oriented government construction. The government should position the role accurately in public services of community sport, specify the responsibility and improve service level.

To regard service as the common goal, standardize and perfect public service system of community sport, construct community sport public service convenient for people, favourable to people and friendly to people based on all residents and improve health level and life quality of residents through diversified public sport services.

Acknowledgments

This work was supported by A Project of National Social Science Fund [13CTY009], A Project of Hunan Provincial Education Department (12C0134) and A Project of Hunan Provincial Sport Bureau [KT12-030].

References

[1] Ali Bahrami, Cihan H. Dagli 1993 From fuzzy input requirements to crisp design *The International Journal of Advanced Manufacturing Technology* **8**(1) 52-60
[2] Dimitriyadis Irini, Kahraman Cengiz 2012 A Primer in Understanding Retirement Planning with Fuzzy Mathematics. *Journal of Multiple-Valued Logic and Soft Computing* **18**(3 - 4) 267-289
[3] Harrison J P, Hudson J A 2010 Incorporating Parameter Variability in Rock Mechanics Analyses: Fuzzy Mathematics Applied to Underground Rock Spalling *Rock Mechanics and Rock Engineering* **43**(2) 219-224
[4] Jiang Weiguo, Deng Lei, Chen Luyao 2009 Risk assessment and validation of flood disaster based on fuzzy mathematics. *Progress in Natural Science* **19**(10) 1419-25
[5] Liang Junxiong 2003 Study on urban community sport construction. *China Sport Science and Technology* **39**(9) 23-25
[6] Qian Wenjun 2010 Construction of community sport construction index system *Journal of Nanyang Normal University* **9**(3) 72-7
[7] Salim Labiod, Thierry Marie Guerra 2009 Anytime measures for top- k algorithms on exact and fuzzy data sets *The VLDB Journal* **18**(2) 407-27
[8] Xia Chongde, Chen Po, Yin Ying 2007 Study on comprehensive evaluation system for sustainable development of competitive sports. *Journal of Beijing Sport University* **30**(11) 1564-7
[9] Xiao Linpeng 2007 Concept and theoretical analysis of public sport service *Journal of Tianjin Institute of Physical Education* **22**(2) 97-101
[10] Xinping Wu 2012 The Contract and Project Quality Management System Based on Fuzzy Mathematics Methods *JDCTA*, **6**(14) 77 - 85
[11] Yin Daiyin, Zhang Xiaoran, Wang Xueyan 2013 Application of Fuzzy Analysis Mathematics Method on Optimizing Fracturing Well in Polymer Flooding Period *International Journal of Applied Mathematics and Statistics* **42**(12) 463-73.
[12] Yi-Fei Chen, Xiao-Lin Qin, Liang Liu, Bo-Han Li 2012 Fuzzy Distance-Based Range Queries over Uncertain Moving Objects *Journal of Computer Science and Technology* **27**(2) 376-96

[13] Kim Y M, Choi J C, Kim J H, Kim C 2002 Development of a System for Progressive Working of an Electric Product by Using Fuzzy Set Theory. *The International Journal of Advanced Manufacturing Technology* **20**(10) 765-79
[14] Yi Yin 2013 A Fuzzy Comprehensive Evaluation Model for Channel Selection in Telecom Industry *International Journal of Applied Mathematics and Statistics* **39**(9) 196-204
[15] Zhang Hongya 2001 On title, feature and function of community sport *Sports and Science* **22**(2) 25-30
[16] Zsolt Csaba Johanyák 2010 Student Evaluation Based On Fuzzy Rule Interpolation *International Journal of Artificial Intelligence* **5**(A10) 37-55

Authors

Zhengmei Lin, born on July 1, 1973, China

University studies: majored in Sports Education, won the Bachelor Degree of Science in Education from Fujian Normal University in June 1997, and acquired Master of Science in Education from Fujian Normal University in December 2006.
Scientific interests: the science of physical culture and sports
Experience: In July 1997, Zhengmei Lin served as a teacher in Department of Physical Education and Sport Science, Fuqing Branch of Fujian Normal University. In september 2012, he made paper Application of Factor Analysis and Pareto Analysis for the Value of College Students Regarding Playing Basketball-a Periodical which named International Journal of Digital Content Technology and its Applications. His previous research area is College sports education.

Jiwei Yao, born on September 1, 1976, China

University studies: majored in Sports Education, won the Master Degree of Education from Guangxi Normal University in China in June 2004, and acquired Doctor of Science in Education from Fujian Normal University in June 2013.
Scientific interests: the science of physical culture and sports
Experience: In July 2004, Jiwei Yao served as a teacher in the Physical Education Department of Hunan University of Science and Technology. In September 2012, he made paper A VEC-Model-Based Sports Public Service Analysis-a Periodical which named International Journal of Digital Content Technology and its Applications. His previous research area is College sports education.

An evaluation approach based on word-of-mouth for trust models in recommendation systems

Feng Jiang[1,2], Min Gao[3,4,*], Hui Xia[3], Qingyu Xiong[3,4], Junhao Wen[3,4]

[1] *College of Civil Engineering, Chongqing University, Chongqing, 400044, China*

[2] *College of Construction Engineering, Chongqing Technology and Business Institute, Chongqing, China*

[3] *School of Software Engineering, Chongqing University, Chongqing 400044, China*

[4] *Key Laboratory of Dependable Service Computing in Cyber Physical Society, Ministry of Education, Chongqing 400044, China*

Received 1 June 2014, www.cmnt.lv

Abstract

Recommendation systems have been recognized as an effective approach to heavy information load. In recommendation systems, trust/reputation has attracted increasing attention because it helps to improve the precision of recommendation and the robustness of systems to shilling attacks. Recommendation system oriented trust models, mostly rating-based, used to build the reputation and trustiness among users. They are often evaluated in terms of how accurately they help to predict user ratings and how robustly they resist shilling attacks. However, those evaluation techniques disregard the trust values themselves: how accurately they calculate the trust values themselves is not measured. To solve the problem, in this work, we propose an approach to measure the trust values based on electronic word-of-mouth (eWoM) theory. The eWoM believes a user is reliability if he is of good public praise. In our approach, firstly, according to eWoM, the reliability value of a user can be judged by other users' votes - whether the user's ratings or feedbacks are positive or negative. Secondly, the trust values of users can be calculated by a trust model. Finally, we compare trust values and reliability values. As a case study, we propose a simple rating-based trust model and then evaluate the trust model based on the proposed evaluation approach and Amazon dataset.

Keywords: Recommendation Systems, Evaluation Approach, Trust Model

1 Introduction

Trust has been a spot in recommendation systems, because combining trust and recommendation systems can improve the accuracy and stability of recommendations and improve the user's experience [1]. Trust models have been used to recommend movies, songs, products in e-commercial websites, and friends in social networks.

Trust models, usually rating-based in recommendation systems, are often evaluated according to how accurately they help to predict these ratings [1, 2], or by measuring the stability of rating prediction and hit ratio [3, 8]. However, the set of evaluation techniques disregard the trust values themselves: how accurately they calculate the trust values themselves is not measured. To solve the problem, we proposed an evaluation approach to measure trust models according to eWoM theory [4, 5]. The approach can evaluate the accuracy of a trust model by users' assessment. Applied this evaluation approach, the well-performed trust model can be applied to other recommendation systems without user's assessment. For example, there is a trust model m1, according to the evaluation approach and datasets with user assessments (e.g. the datasets driving from Amazon.com), we can evaluate m1. If m1 is good enough, then it can be used to other recommendation systems without user assessments information, such as Movielens and Netflix.

In this work, we first provide an overview of trust model, evaluation approaches and their limitations (Section 2) to explain why a novel evaluation approach may be important in recommendation systems. We proposed a eWoM-based approach to evaluate trust models (Section 3): by proposing a eWoM-based algorithm for reliability values of users, and then comparing those reliability values and trust model based trust values. We finally to illuminate the evaluation approach using a case study (Section 4), comparing the trust values calculated by a trust model to the public reliability values on datasets deriving from Amazon. We draw conclusion in Section 5.

2 Related works and associated problems

With the advancement in networking and multimedia technologies enables the distribution and sharing of multimedia content widely. In the meantime, piracy becomes increasingly rampant as the customers can easily duplicate and redistribute the received multimedia content to a large audience.

Trust/reputation is what is generally said or believed about a person's or thing's character or standing [6, 7].

[*] ***Corresponding author*** e-mail: skycat1980@hotmail.com

Giorgos and Zacharia [8] proposed Sporas model based on eBay and consider the trustworthiness of the graded counterpart to mitigate influences of malicious behaviours; and then they propose a novel collaborative reputation mechanism for electronic marketplaces [9]. Guo, Jiang, and Cai [10] modified Sporas into E-Sporas by adding the factors of volume and number of trades to make Sporas better in C2C. Wang, Zhang, and Chen [11] proposed a cloud-based reputation reporting mechanism to solve the fake and joint cheat recommendations of malicious customers. Li, Liang, and Zhang [12] gave a systematic study on trustworthiness and reputation management model in e-commerce. Wu, Li, and Kuo [13] proposed several metrics for online auctions to evaluate the trust values of participants. Abdel-Hafez, Xu, and Tjondronegoro [14] provided a product reputation model for recommendation based on opinion mining techniques.

In recommendation systems, Golbeck et al. looked specifically at trust between people who have no directly links to another and proposed TidalTrust [15] mechanism to infer trust in continuous trust networks. Paolo Avesani et al proposed a time-efficient trust metric named Moletrust [16] and applied it in Moleskiing application to compute the trustworthiness of users. Levien et al. [17] investigated the role of trust metrics in attack-resistant public key certification. Kuter et al. described SUNNY [2], a new trust inference algorithm using a probabilistic sampling technique to estimate confidence. Cai-Nicolas Ziegler et al. provide a novel trust metric Appleseed [18] for local group trust of semantic web issues. Massa and Avesani [19, 20] proposed approaches to estimate all trust weights of users by propagate trust over a trust network. Verbiest et al. [21] proposed a new approach to calculate trust and distrust by introduce path length incorporating aggregation strategies. Victor et al. proposed bilattice-based aggregation approaches for gradual trust and distrust [22] and then incorporated the gradual trust and distrust in recommender systems [23].

Large numbers of literatures concern about trust and reputation recently. However, they are mainly proposed for constructing better recommendations, such as decreasing the accuracy (e.g. MAE, RMSE) of rating prediction, improving precision and recall of recommendations [9, 10, 14], and improving robustness [8, 11, 13] (especially stability ability under shilling attacks) and users' satisfactions [8-14]. However, the set of evaluation techniques disregard the trust values themselves: how accurately they calculate the trust values themselves is not measured. Thus, we propose a new evaluation approach to solve the problem.

3 An EWoM-based evaluation approach

Traditionally, WoM is to pass information from person to person by oral communication. When WoM is mediated through electronic means, the eWoM refers to any statement consumers share via the Internet about a product, service, brand, or company [4]. The eWoM proves to have more effectiveness in leading purchasing decisions, as the wide scale of online broadcasting and weak correlations between online customers and marketers [5]. The enormous impact of eWoM on consumer behaviour and product success thus attracts many researchers.

The steps for the evaluation approach are show as Figure 1. We can get reliability values from eWoM-based algorithm (subsections 3.1 and 3.2) and then compare those reliability values to trust values based on trust models (subsection 3.3).

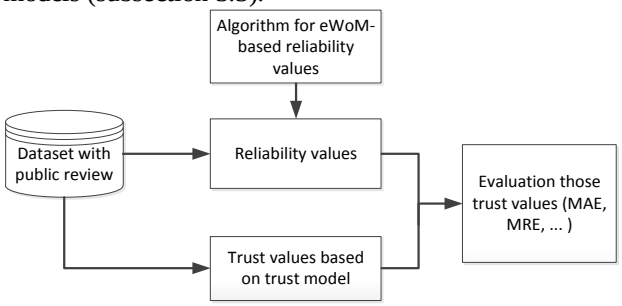


FIGURE 1 The steps for the evaluation approach

3.1 EWOM-BASED RELIABILITY VALUES

As known to us all that, people can rate their consumed product after each transaction. However, the reliability [7] of those consumers and their ratings is still a problem. Inspired by the effectiveness of eWoM, an empirically convincing evaluation approach under mass attitude is proposed to assess reputation of customer with their ratings.

Firstly, we collect support ϕ_j of the whole ratings from a customer u_j; at the same time, compute the weight φ_j of u_j's ratings over the whole system. Then the reliability value of u_j's from public views can be simply denoted as Equation (1).

$$Reli_j = \alpha f_j + \beta j_j, \tag{1}$$

where the weights of ϕ_j and φ_j are α and β, $\alpha + \beta = 1$; $\phi_j = \sum_{0 \le i \le k} h_{i,j} / v_{i,j}$, where $h_{i,j}$ is the helpful votes from u_j to item i, $v_{i,j}$ is the corresponding total votes; $\varphi_j = k_j / R$, $k_j = |r(u_j)|$, k_j is the number of u_j's all ratings; R is the number of all ratings in the dataset.

The weights α and β can be set to suitable values by experiments. We will illuminate how to find the suitable values in Section 3.2.

3.2 SET THE WEIGHTS USING AMAZON DATASET

The dataset is Amazon product co-purchasing network metadata* collected in summer 2006. The dataset contains 548,552 different items (Books, music CDs, DVDs, and VHS video tapes). Each item contains reviews from customers, including ratings, the number of helpful votes for ratings, and total number of votes.

* http://snap.stanford.edu/data/amazon-meta.html

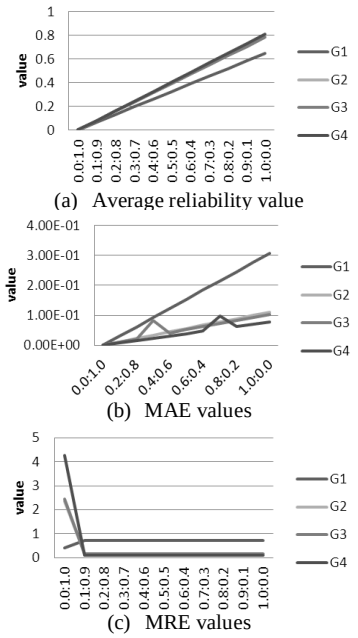
To get suitable values of the weights, we divided the datasets to four subsets G1, G2, G3, and G4 according to the number of all ratings of a user k (k>0, k>5, k>10, k>20). We calculated the average reliability value avg_R_{ci}, mean absolute error (MAE) MAE_{ci}, and mean relative error (MRE) MRE_{ci} (see Equations 2, 3, and 4), using ten combinations of α and β, C = (0.0, 1.0; 0.1, 0.9; 0.2, 0.8; 0.3, 0.7; 0.4, 0.6; 0.5, 0.5; 0.6, 0.4; 0.7, 0.3; 0.8, 0.2; 0.9, 0.1; 1.0, 0.0), marked by c_1, c_2, ..., and c_n.

$$avg\_Reli_{c_i} = \sum_{i=1}^{n} \left| Reli_{u_i}(c_i) \right| \tag{2}$$

$$MAE_{c_i} = \frac{\sum_{i=1}^{n} \left| Reli_{u_i}(c_i) - R_{u_i} / all\_avg\_I * avg\_Reli_{c_i} \right|}{n} \tag{3}$$

$$MRE_{c_i} = \frac{\sum_{i=1}^{n} \left(\left| Reli_{u_i}(c_i) - R_{u_i} / all\_avg\_I * avg\_Reli(c_i) \right| / Reli_{u_i} \right)}{n}, \tag{4}$$

where n means the total number of users in group; $Reli_{ui}(c_i)$ can be calculated by Equation (1); Ru_i is a set of u_j's all ratings; all_ave_I is the average of all items average rating scores.



(a) Average reliability value

(b) MAE values

(c) MRE values

FIGURE 2 Four groups of average E, absolute error, and relative error

The avg_Reli_{ci} results are shown as Figure 2 (a). It is simply that 4 curves grow linear with the value of α increase, though β decreases. Actually, k / all_num is normally too small to have nearly no impacts on the results.

From sampling ratings from Amazon dataset, the value of all_avg_I is 4.18, thus we could get the absolute and relative error values according to Equations (3) and (4) respectively, and then have the corresponding curves as Figure 2 (b) and (c) denotes.

In Figure 2 (b), all groups' absolute errors are very small even though the proportion of α and β are different, which seem to remind us that proportion of α and β takes no effect. In Figure 2 (c), we can find a phenomenon that there are 4 outlier points where the value of (α, β) is (0.0, 1.0). The phenomenon gives us a further authentication that the first part of the equation (1) plays an important role in computing the value of $Reli_{ui}$ in another direction. Another obvious phenomenon in (c) is that all curves rise and fall slightly around certain numbers, which look like 4 straight lines no matter what happens to the value of (α, β) later. From the above two figures, we could make sure that any proportion of α and β can work except α equals 0.0. Thus, in the next experiments, we keep (α, β) is (0.8, 0.2).

3.3 EVALUATION APPROACH

To evaluate a trust model, we calculate *wMAE* (weighted MAE) and *wMRE* (weighted MRE) for the trust model based trust values and eWoM-based reliability values (See Equations (5) and (6)).

$$wMAE = \sum_{j=1}^{n} (T_j * \overline{Reli} / \overline{T} - Reli_j) / n \ , \tag{5}$$

$$wMRE = \sum_{i=1}^{n} ((T_j * \overline{Reli} / \overline{T} - Reli_j) / Reli_j) / n \cdot \tag{6}$$

In the Equations,

$$\overline{T} = \sum_{j=1}^{n} T_j / n \ , \tag{7}$$

$$\overline{Reli} = \sum_{j=1}^{n} Reli_j / n \ \cdot \tag{8}$$

T_j is the trust value for user j. $\overline{T}$ is the average value of T_j. $Reli_j$ is the reliability value for user j.

4 Promoting the proposed evaluation approach

To illuminate the steps for applying the evaluation approach, we will use a simple trust model first as an example, then evaluate the model using the proposed approach.

4.1 A SIMPLE TRUST MODEL – AS A CASE STUDY

Here we have a simple but effective trust model for customer according to a basic truth, that is, each rating has a potential bias since each rating is affected by users' sentimental factor. Thus we should use correct function (5) to objectify the biased rating r_{ij}.

$$r_{ij}^{'} = r_{ij} * \frac{\bar{r}}{\bar{r}_i}, \tag{9}$$

where $\bar{r}$ is the average rating of the whole system, while $\bar{r}_i$ is the average rating of item i. The objective part of r_{ij} is $r_{ij}^{'}$.

After the correction, we could modify the existed trust value of u_j iteratively (See Equation 10).

$$T_j = \frac{T_j^{'} * k + r_{ij}^{'}}{k+1}, \tag{10}$$

where $T_j^{'}$ is the previous trust value before u_j rated item *i*.

4.2 THE COMPARISON AND ANALYSIS

We derived two random selected groups from Amazon dataset to validate the tendency: the consistency - how the average trust values and average reliability values looked like. Experimental results are shown in Figure 3.

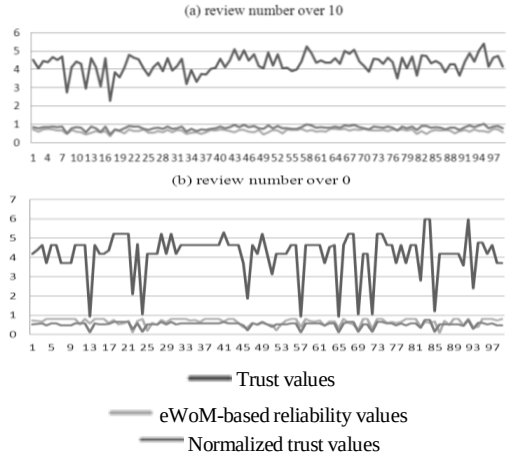
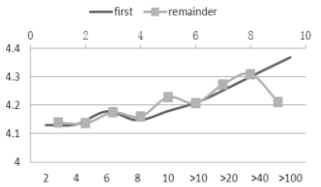
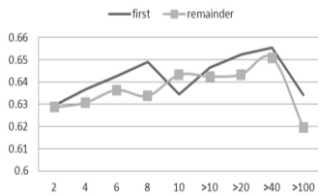


FIGURE 3 Reputation trends in review number (a) over 10, (b) over 0, with initial trust values, the normalized trust values from equation and reliability values from equation (1) with α=0.8 and β=0.2

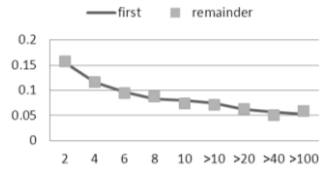
We then derived from selected sample data with 40000 items. We divided the data to two groups according to the number of users' total reviews *k*. One group was the *first* 20000 sample items, and the other one was the *remaining* 20000 sample items. We firstly grouped 9 classes to perform experiments where k=2, k=4, k=6, k=8, k=10, k>10, k>20, k>40, and k>100 respectively, and then calculated the average trust values (*T*), average reliability values (*ave_Reli*), *wMAE*, and *wMRE* according to Equation (5) - (8). The results are shown in Figure 4.



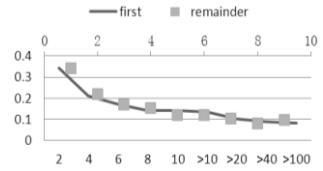
(a) Average trust values



(b) Average reliability values



(c) Weighted *MAE*



(d) Weighted *MRE*

FIGURE 4 The comparisons of two sets between (a) Average trust values (*T*), (b) average reliability values (*ave_Reli*), (c) *wMAE*, (d) *wMRE*

From Figure 4 (a) and 4 (b), each two lines have small fluctuations respectively, but still can fit to others to some extent, which confirms that both of the two samples can reflect the overall dataset. From Figure 4 (c) and 4 (d), we can see that both two sets' curves of *wMAE* and *wMRE* firstly decrease quickly before k = 8 or so, then slowly and have signs of tending towards stability after k = 8.

The largest *wMAE* in the first group is 0.156 at the first point, and falls down to 0.1 at the fifth point, and preserves stable around 0.08 after 8. It is almost the same trend to the second group's curve.

Therefore, we can conclude that, most *wMAE* is smaller than 0.1, and larger value of k means lower absolutely error. Similarly, we can analyse that the range of relative error is 0.0 - 0.2, and larger k means lower relative error.

5 Conclusions and future work

In this paper, we have analysed up-to-date trust/reputation models especially in recommendation systems. We have introduced evaluation approaches of the models and their problem: disregard the accuracy of trust values themselves. To solve the problem, we have proposed an evaluation approach based on WoM theory and given a case study and experiments on Amazon dataset to illuminate how to apply and analyse the proposed evaluation approach.

We have just used a trust model to show how to apply the proposed approach in the paper. We will try to apply the evaluation approach to more trust models and further analyse and improve the approach.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (71102065), the China Postdoctoral Science Foundation (2012M521680), the Fundamental Research Funds for the Central Universities (CDJZR12090001), the Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ121607 and KJ131603), and the Frontier and Application Foundation Research Program of CQ CSTC (cstc2014jcyjA40037).

References

[1] Golbeck J 2009 Tutorial on using social trust for recommender systems. *Proc. Conf. on Recommender systems* ACM New York USA 425-6

[2] Kuter U, Golbeck J 2007 Sunny: A new algorithm for trustence in social networks using probabilistic confidence models. *Proc. Conf. on Artificial Intelligence* AAAI Vancouver British Columbia Canada 1-6

[3] Gao M, Wu Z, Jiang F 2011 Userrank for item-based collaborative filtering recommendation *Information Processing Letters* **111**(9) 440-6

[4] Kietzmann J H, Canhoto A 2013 Bittersweet! Understanding and Managing Electronic Word of Mouth *Journal of Public Affairs* **13**(2) 146–59.

[5] He C 2009 *Research of Source Credibility of Online Review Based on Relationship Stages, Master Thesis* Dalian University of Technology

[6] McLeod Carolyn 1999 Trust *The Stanford Encyclopedia of Philosophy* 2011 Edition Edward N Zalta (ed.) URL=http://plato.stanford.edu/archives/spr2011/entries/trust

[7] Jøsang A, Ismail R, Boyd C 2007 A Survey of Trust and Reputation Systems for Online Service Provision *Decision Support Systems* **43**(2) 618-44

[8] Zacharia G 1999 *Collaborative Reputation Mechanisms for Online Communities* Massachusetts Institute of Technology USA

[9] Zacharia G, Moukas A, Maes P 2000 Collaborative Reputation Mechanisms in Electronic Market places *Decision Support Systems* **29**(4) 371-88.

[10] Guo H, Jiang J, Cai H 2009 Modelling for Reputation Computing in C2C Communities *Chinese Journal of Management* **6**(8) 1056-60

[11] Wang P, Zhang S, Chen X 2011 A Novel Reputation Reporting Mechanism Based on Cloud Model and Grey System Theory *International Journal of Advancements in Computing Technology* **3**(10) 75-84

[12] Li D, Liang Y, Zhang W 2010 Survey on trust management for e-commerce systems, *Application Research of Computers* **27**(4) 1208-11

[13] Wu F, Li H H, and Kuo Y H 2011 Reputation evaluation for choosing a trustworthy counterparty in C2C e-commerce. *Electronic Commerce Research and Applications* **10**(4) 428-36

[14] Abdel-Hafez A, Xu Y, Tjondronegoro D 2012 Product Reputation Model: An Opinion Mining Based Approach SDAD *Proc. Conf. on Sentiment Discovery from Affective Data* Bristol UK 16-27

[15] Golbeck J 2005 *Computing and Applying Trust in Web-based Social Networks* PhD thesis, University of Maryland, College Park, MD, USA

[16] Avesani P, Massa P, Tiella R 2005 Moleskiing.it: a trust-aware recommender system for ski mountaineering *International Journal for Infonomics* **20**(1) 1-19

[17] Levien R, Aiken A 1998 Attack-resistant trust metrics for public key certification *The 7th USENIX Security Symposium* San Antonio, Texas 229-42

[18] Ziegler C-N, Lausen G 2004 Spreading activation models for trust propagation *Proc. IEEE Conf. on e-Technology, e-Commerce, and e-Service* Taipei, Taiwan 83-97

[19] Massa P, Avesani P 2007 Trust-aware recommender systems *Proc. Conf. on Recommender Systems* Minnesota, USA 17-24

[20] Massa P, Avesani P 2009 Trust metrics in recommender systems *Computing with Social Trust* Springer: London

[21] Verbiest N, Cornelis C, Victor P, Herrera-Viedma E 2012 Trust and distrust aggregation enhanced with path length incorporation *Fuzzy Sets and Systems* **202** 161-74

[22] Victor P, Cornelis C, Cock M De, Herrera-Viedma E 2011 Practical aggregation operators for gradual trust and distrust *Fuzzy Sets and Systems* **184**(1) 126-47

[23] Victor P, Cornelis C, Cock M De, Pinheiro da Silva P 2009 Gradual trust and distrust in recommender systems *Fuzzy Sets and Systems* **160**(10) 1367-82

Authors

Feng Jiang, born on February 13, 1979, China

Current position, grades: researcher at Chongqing Technology and Business Institute, China.
University studies: MS degree in Civil Engineering from Chongqing University, China in 2006. Currently, he is a PhD candidate in the College of Civil Engineering, Chongqing University.
Scientific interests: information processing and Construction Engineering.
Publications: more than 10 refereed journal and conference papers in these areas.

Min Gao, born on April 08, 1980, China

Current position, grades: professor at the School of Software Engineering, Chongqing University.
University studies: MS and PhD degrees in computer science from Chongqing University in 2005 and 2010 respectively.
Scientific interests: recommendation system, service computing, and data mining.
Publications: more than 30 refereed journal and conference papers in these areas.
Experience: visiting researcher at the School of Business, University of Reading. Grants from the National Natural Science Foundation of China, the China Postdoctoral Science Foundation, and the China Fundamental Research Funds for the Central Universities.

Hui Xia, born on September 24, 1989, China

University studies: PhD candidate in the College of Computer Science, Chongqing University.
Scientific interests: recommendation system and computer networking.
Experience: participated in 5 industrial projects and developed many commercial systems and software tools.

Qingyu Xiong, born on October 03, 1965, China

Current position, grades: dean of the School of Software Engineering, Chongqing University.
University studies: BS and MS degrees from the School of Automation, Chongqing University in 1986 and 1991, respectively, and the PhD degree from Kyushu University of Japan in 2002.
Scientific interests: neural networks and their applications.
Publications: more than 100 journal and conference papers in these areas.
Experience: more than 20 research and applied grants.

Junhao Wen, bor on May 01, 1969, China

Current position, grades: deans of the School of Software Engineering, Chongqing University. Professor in the School of Software Engineering.
University studies: MS and PhD degrees in computer science from Chongqing University in 1999 and 2008 respectively.
Scientific interests: computing and recommendation systems.
Experience: more than 80 journal and conference papers in these areas.

Application of principal component analysis in fire risk prediction of stadium

Huajie Chen[1], Mingchang Liu[2*]

[1] *Physical Education Department, Jinggangshan University, Jian 343009, Jiangxi, China*

[2] *Department of Physical Education, Huazhong Agricultural University, Wuhan 430070, Hubei, China*

Received 1 June 2014, www.cmnt.lv

Abstract

The security of stadium is one of the problems increasingly concerned by people. Among various security problems, the fire risk of stadium is the most important. By using principal component analysis (PCA) and combining the influencing factors of risk prediction of stadium, several influencing factors of the fire risk of stadium are evaluated and analysed, then the uppermost principal component factors are selected, the inconspicuous components are eliminated, the influence of relevant factors on the fire risk prediction of stadium is analysed, and the correlation of various indicators with fire risks of stadium is understood to find the potential best security management prediction system, for the purpose of taking pertinent prediction measures of fire accident risks to overcome the unfavourable factors in the security management, and ensure the safe and normal operation of stadium as well as the personal and property security of sporters.

Keywords: Principal Component Analysis, Stadium, Fire Risk, Prediction

1 Introduction

The principal component analysis (PCA) is a kind of statistical analysis method that converts multiple original variables into several comprehensive indicators [1]. From the mathematical perspective, it is a dimension reduction technology [2]. Its general idea is: a research object is often a multi-element complex system. Too many variables will certainly increase the difficulty and complexity of problem analysis [3]. It is a way to simplify the problem by using the correlation between original variables; replacing more original variables with less new variables and making the fewer variables preserve the information reflected by more original variables as more as possible.

The security management of large stadium is an important issue [4]. As a large public place, the large stadium is an indispensable part of people's spiritual entertainment, characterized by highly populated people and big liquidity, so it often causes significant casualty and property loss in case of any fire, thus bringing great negative impacts [5]. Therefore, the fire accident in the safety management is the most serious. In recent years, the stadiums in China have suffered more than 200 fires, with a death of more than 1000 people, which fully reveals the serious problem in the security management of stadium [6]. The academic circles and manager pay more and more attention to seek the way of reducing fire risks by analysing the influencing factors of the fire risk prediction of stadium. To better ensure the security of stadium, it is required to strengthen the research on the reasonability and scientific of the fire risk prediction of stadium [7]. In order to analyze the influencing factors of the fire risk prediction of stadium, it is required to conduct research and analysis from different theories, thus introducing the application of PCA in the fire risk prediction of stadium [8]. PCA converts many indicator factors into several important comprehensive indicators through the thought of dimension reduction. In practical problems, the primary and secondary cannot be mastered because of the complicated and changing conditions as well as more influencing factors, so it is hoped that there are less variable indicators in the quantitative analysis and more information can be got from these indicators [9]. This can be met by analytic hierarchy process (AHP) as it can explains the most original variables with less indicator variables, converts the indicators with high correlation into independent indicator variables, without any correlation and also can well explain most variables [10].

PCA is one of the important methods of analysing problems in the academic circles at present, which can select according to the correlation degree of influencing factors to reveal the intrinsic connection and difference of objective things. By studying the relevant factors that influence the fire risk of stadium, the application of PCA in the fire risk prediction of stadium can be explored, which is of great realistic significance in the fire risk management and security system of stadium.

[*] ***Corresponding author*** e-mail: liumchang@yeah.net

2 Overview of PCA

2.1 BASIC IDEA OF PCA

The basic idea of PCA is to recombine, divide and convert more original P indicators X1, X2, …, XP with correlation into a set of less representative indicators, without any correlation. Finally, the fewer indicators without any correlation are used to replace more original indicators.

Suppose that F1 means the principal component indicator formed by the first linear combination of original variables, that is, $F_1 = a_{11}X_1 + a_{21}X_2 + ... + a_{p1}X_p$, it is known by mathematical knowledge that the information got from each principal component can be measured by its variance. The larger the variance Var(F1) is, the more the information contained in F1 will be. Generally, it is hoped that the principal component F1 contains the largest information, so F1 selected from the linear combination should be the one with the largest variance in the linear combinations of X1, X2, …, XP, so F1 is the first principal component. If the first principal component fails to represent the information of original p indicators, the second principal component indicator F2 can be selected. To effectively reflect the original information, the existing information of F1 will not occur in F2, that is, F2 and F1 should be independent and uncorrelated. In the mathematical language, Cov(F1, F2)=0, so F2 is one with the largest variance in the linear combinations of X1, X2, …, XP without any correlation with F1, and F2 is called the second principal component. F1, F2, ……, Fm got by this analogy are the first, second,..., m principal component of original variable indicators X1, X2……XP.

$$\begin{cases} F_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p \\ F_2 = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p \\ \quad ..... \\ F_m = a_{m1}X_1 + a_{m2}X_2 + ... + a_{mp}X_p \end{cases} \quad (1)$$

It is known from the above analysis that:

(1) Fi and Fj are uncorrelated, that is, Cov(Fi, Fj) = 0 and Var(Fi)=ai'Σai, in which Σ is the covariance matrix of X.

(2) F1 is of the largest variance in the linear combinations of X1, X2, …, Xp (the coefficient meets the above requirements), that is, Fm is of the largest variance in the linear combinations of X1, X2, …, Xp, without any correlation with F1, F2, ……, Fm-1.

F1, F2, …, Fm (m≤p) are the new variable indicators, that is, the first, second, ..., m principal component of original variable indicators.

2.2 MAIN FEATURE OF PCA

When PCA is used to analyse the research problems, there are two main characteristics:

(1) Obtain the equation of each principal component Fi (i = 1, 2,..., m) relative to the original variable Xj (j = 1, 2,..., p), namely, coefficient a_{ij} (i = 1, 2,..., m; j = 1, 2,..., p). By using the mathematical knowledge, it is proved that the variance of principal component is the characteristic root of original variable covariance matrix, so the front m larger characteristic roots represent the front m larger principal component variances; the corresponding eigenvectors of front m characteristic roots λi of original variable covariance matrix (to ensure the largest variance of principal component in turn) is the coefficient of corresponding principal component Fi expression ai. To limit, the coefficient of ai adopts the corresponding unit eigenvector of λi.

(2) Calculate the principal component load, which reflects the correlation degree of principal component Fi and original variable Xj:

$$P(Z_k, x_i) = \sqrt{\lambda_k}\, a_{ki} (i, = 1, 2, \cdots, p; k = 1, 2, \cdots, m).$$

2.3 CALULATION STEPS OF PCA

2.3.1 Obtain the covariance matrix

The original variable covariance matrix of sample is Σ=(sij)p×p, where

$$s_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j) \quad I, j = 1, 2, \ldots, p\,. \quad (2)$$

2.3.2 Obtain the characteristic value λ_i of covariance matrix and corresponding eigenvector a_i under orthogonally conditions

The front m larger characteristic values of matrix, λ1≥λ2≥…λm>0 are the corresponding variances of front m principal components, the corresponding unit eigenvector a_i of λ_i is the coefficient of principal component Fi relative to the original variable, so the principal component Fi of i original variable is:

$$F_i = X \quad (3)$$

The variance contribution of principal component generally represents the reflection information, the greater the contribution is, the more the information will be. α_i is:

$$\alpha_i = \lambda_i / \sum_{i=1}^{m}\lambda_i \quad (4)$$

2.3.3 Choose the principal component

The most principal component among various indicators, that is, m in F1, F2,……,Fm should be confirmed by calculating the accumulated variance distribution.

$$G(m) = \sum_{i=1}^{m}\lambda_i / \sum_{k=1}^{p}\lambda_k \quad . \quad (5)$$

Generally, when the accumulated variance distribution is more than 85%, the information of original variable can be better reflected and m is the number of principal components selected finally.

2.2.4 Confirm the principal component load

The principal component load presents the correlation between principal component Fi and original variable Xj and the load lij (i=1, 2, ..., m; j=1, 2, ..., p) of original variable Xj (j=1, 2, ..., p) on the principal component Fi ((i=1, 2, ..., m) is as follows:

$$l(Z_i, X_j) = \sqrt{\lambda_i} a_{ij} (i = 1, 2, \cdots, m; j = 1, 2, \cdots, p) . \quad (6)$$

Then, SPSS software can be used for calculation and analysis and the "component matrix" obtained is the principal component load matrix.

2.2.5 Calculate the score of principal component

Based on the calculation of principal component, the score of principal component is calculated:

$$F_i = a_{1i} X_1 + a_{2i} X_2 + ... + a_{pi} X_p \quad i = 1, 2, \ldots, m \quad (7)$$

However, in the actual problem, each indicator has different dimensional effects, so it is required to eliminate the influence of dimension on it before calculating the score. Generally, the dimensional influence of data is eliminated through the standardization treatment of original data, that is, the data conversion is transformed as follows:

$$x_{ij}^* = \frac{x_{ij} - \overline{x}_j}{s_j} \qquad i = 1, 2, ..., n; \ j = 1, 2, ..., p \quad (8)$$

According to the mathematical expression and relevant mathematic principle, on one hand, after standardized transformation, the variable covariance matrix can express its relevant coefficient matrix; on the other hand, the standardized covariance is the relevant coefficient of original variable, so it is concluded that after the standardized treatment of original variable, the relevant coefficient matrix is of no change.

All in all, the corresponding characteristic value λi of the relevant coefficient matrix of original indicator is the variance contribution of principal component and the variance distribution is $\alpha_i = \lambda_i / \sum_{i=1}^{p} \lambda_i$. The larger α_i is, the stronger the ability of corresponding principal component to reflect the comprehensive information. The principal component can be selected according to λi. The combination coefficient (load of original variable on the principal component) a_i of each principal component is the corresponding unit eigenvector of characteristic value λi.

3 Comprehensive application of PCA in fire risk warning of stadium

In the research on fire accident of stadium, the evaluation and assessment of influencing factors of fire risk warning are an essential part in our daily warning. With the social development, the evaluation system and method for fire risk warning become diversified, however, most of which are limited to the investigation and analysis phase and rarely conducts further analysis and verification from the theoretical perspective. According to the above situation, the fire risk system of PCA is established and the subjective analysis is integrated with the scientific component of mathematical model to realize the change from qualitative analysis accurate quantitative analysis.

3.1 INDICATOR SELECTION OF STADIUM

The influencing factors of fire risk of stadium from 1990-2013 are calculated by using PCA and the main indicators are selected. According to the actual situation of stadium, 9 main indicators are selected finally as follows:

TABLE 1 Indicator of fire rusk prediction model of stadium

Indicator	Meaning
X1	Electric fire
X2	Poor internal fire security inspection
X3	Fail to build the stadium according to the fire design
X4	Incomplete patrol
X5	Poor automatic fire warning system
X6	Poor fire acceptance of stadium
X7	Poorly implemented fire security and regulations of stadium
X8	Insufficient fire extinguisher or failure of fire extinguisher
X9	Fail to take actions immediately

3.2 SPECIFIC STEPS OF PCA IN FIRE RISK

According to the above elaboration, to eliminate the dimensional influence, the covariance matrix is calculated after standardization of variable, that is, the relevant coefficient matrix of original variable is calculated directly, so the calculation steps of PCA in fire risk of stadium are as follows:

(1) Calculate the relevant coefficient matrix;
(2) Obtain the characteristic value λ_i and corresponding orthogonal unit eigenvector a_i of relevant coefficient matrix;
(3) Calculate the weight of each component;
(4) Choose the principal component.

3.3 APPLICATION EXAMPLE OF PCA IN FIRE RISK

In the research on fire risk prediction of stadium, 9 main indicators among various risk influencing factors are selected to discuss the application of PCA.

According to the specific steps of PCA, the relevant coefficient matrix is obtained after standardizing the

original indicator data and then SPSS software is used to calculate the relevant coefficient matrix and characteristic value of 9 evaluation indicators to confirm the main factor number of evaluation. The number of principal component is confirmed according to the accumulated variance contribution of characteristic value.

After eliminating the dimensional influence of original data of fire risk influencing factors of stadium, the standardized data is obtained (see table 2). The relevant matrix coefficient is shown in table 3.

The characteristic root, contribution and accumulated contribution of relevant coefficient matrix are obtained through further calculation.

It is known from the characteristic equation $|R-\lambda I_9|=0$ that 9 indicators are non-negative characteristic roots, that is, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_9 \geq 0$, where λ1=7.7719, λ2=0.756, λ3=0.3796, λ4=0.077 and λ5=0.0155. The results are shown in table 4.

According to the characteristic root of matrix λi (i=1, 2, 3, ····, 9), the eigenvector is further obtained as shown in table 5.

9 new components are obtained by using PCA. It can be seen from table 3 that the characteristic root λ4, λ7, λ8, λ9 =0 and the corresponding contribution and accumulated contribution are also 0, indicating that some indicators among original indicators are of great correlation and 5 variable indicators can completely represent the information reflected by 9 original indicators. so 5 indicators are selected as the principal components and other 4 are eliminated. Next, by calculating the weight of original 9 indicators in reflecting the influencing factors of fire risk of stadium, the indicator with the largest weight is obtained and the principal component is confirmed.

$$\xi_1 = 0.2978\times 7.7719 + 0.1622\times 0.756 + 0.104\times 0.3796 + 0.5274\times 0.077 - 0.7079\times 0.0155 + 0.056\times 0 + 0.03\times 3 - 0.2352\times 0 - 0.0708\times 0 = 2.4987.$$

By this analogy, the weight of X2,X3,X4,…,X9 in the comprehensive information can be obtained, that is, ξ2=2.1468, ξ3=2.2602, ξ4=2.3766, ξ5=2.7844, ξ6=2.604, ξ7=2.8274, ξ8=2.5632, ξ9=2.1178.

The weight in a descending order is X7, X6, X5, X8, X1, X4, X3, X2, X9. Therefore, 4 corresponding indicators with the minimum weight to be eliminated are X4, X3, X2, X9.

Through the above calculation and analysis, the main influencing indicators that influences the fire risk prediction of stadium are screened, which are poorly implemented fire security rules and regulations of stadium X7, poor fire acceptance of stadium X6, poor automatic fire warning system X5, insufficient fire extinguisher or failure of fire extinguisher X8 and electric fire X1. These indicators considers the critical influencing indicator factors of fire risk prediction of stadium and are also greatly concerned by the relevant staff of enterprise upon management because they play a decisive influence role in the fire risk prediction of stadium, thus providing important and reliable direction for the security management decision of stadium. If the indicator that influences the risk prediction of stadium is uncertain and complex, the PCA can be used to screen the main indicators and eliminate the relevant indicators in the comprehensive evaluation. This method is of good value in the fire risk prediction of stadium and great theoretical and realistic significance in the security management of stadium.

TABLE 4 Characteristic root, contribution and accumulated contribution

H（I）	GXL（%）	LJGXL（%）
7.7719	86.3541	86.3541
.756	8.3997	94.7538
.3796	4.2177	98.9715
.077	.8559	99.8274
.0115	.1726	100
0	0	100
0	0	100
0	0	100
0	0	100

M=9, LJGXL=.8635, D=9

H(I) is the characteristic root, GXL(%)is the contribution and LJGXL(%) is the accumulated contribution.

TABLE 2 The standardized data

	X1	X2	X3	X4	X5	X6	X7	X8	X9
2008	60.36	17.59	25.64	31.17	12.53	70.06	32.87	22.47	33.91
2009	68.48	21.96	20.38	37.02	18.55	65.76	25.93	31.52	39.02
2010	69.38	30.77	22.19	29.08	19.00	79.24	38.77	42.08	38.32
2011	70.27	33.58	26.78	49.05	21.01	68.65	35.95	47.21	38.31
2012	74.10	29.56	28.85	37.08	13.35	63.32	39.08	58.64	35.96
2013	81.60	37.96	23.64	32.79	12.98	59.87	40.65	51.32	44.05

TABLE 3 The relevant matrix coefficient

Indicator	X1	X2	X3	X4	X5	X6	X7	X8	X9
X1	1.000	0.509	0.509	0.206	0.207	0.206	0.122	0.483	0.483
X2	0.509	1.000	1.000	1.000	0.604	0.345	0.543	0.429	0.728
X3	0.509	1.000	1.000	1.000	0.604	0.345	0.543	0.429	0.728
X4	0.206	1.000	1.000	1.000	0.604	0.345	0.543	0.429	0.728
X5	0.207	0.604	0.604	0.604	1.000	0.764	0.288	0.476	0.237
X6	0.206	0.345	0.345	0.345	0.764	1.000	-0.2	-0.30	-0.57
X7	0.122	0.543	0.543	0.543	0.288	-0.2	1.000	0.367	0.456
X8	0.483	0.429	0.429	0.429	0.476	-0.30	0.367	1.000	0.456
X9	0.483	0.728	0.728	0.728	0.237	-0.57	0.456	0.456	1.000

TABLE 4 Characteristic root, contribution and accumulated contribution

H（I）	GXL（%）	LJGXL（%）
7.7719	86.3541	86.3541
.756	8.3997	94.7538
.3796	4.2177	98.9715
.077	.8559	99.8274
.0115	.1726	100
0	0	100
0	0	100
0	0	100
0	0	100

M=9, LJGXL=.8635, D=9, H(I) is the characteristic root, GXL(%)is the contribution and LJGXL(%) is the accumulated contribution.

TABLE 5 Eigenvector

	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
X1	0.2978	0.3022	0.3127	0.3226	0.3428	0.3428	0.3517	0.3301	0.2831
X2	0.1522	-0.332	0.0097	-0.373	0.21	0.0693	-0.032	-0.429	0.728
X3	0.104	0.2044	-0.477	-0.504	0.0313	-0.514	0.3153	-0.43	-0.567
X4	0.5274	-0.338	0.0182	0.2506	-0.514	-0.601	-0.04	0.476	0.7285
X5	-0.707	-0.161	0.1755	-0.013	-0.036	-0.225	0.1343	0.3679	0.2376
X6	0.056	0.036	-0.683	0.4072	0.4206	-0.610	0.0973	-0.30	-0.57
X7	0.033	-0.750	0.3104	-0.610	0.353	-0.062	0.0595	0.1095	-0.134
X8	-0.235	0.4684	-0.377	-0.054	-0.231	-0.054	-0.088	-0.074	0.2991
X9	-0.070	-0.163	0.0547	-0.227	-0.08	-0.57	-0.857	-0.404	0.3621

4 Conclusions

PCA, with simple principles, is simple and convenient in the practical application. Although the computation task is large in case of a large amount of data, with the popularization and application of computer, supported by computer program and relevant software, the results can be obtained accurately and rapidly. PCA is a widely used among various analysis methods, which chooses the most principal component from the multiple indicators (variables) and multiple observation data of sample and quantitatively confirms the relevant influence between sample and indicators. In studying the fire risk prediction of stadium, it is allowed to analyze the influencing factors by using the mathematics and computer tool as well as PCA and choose the primary and secondary component to make relevant management personnel directly know the correlation between indicators in the fire risk of stadium and find the potential best security management prediction system, for the purpose of taking pertinent prediction measures of fire accident risks to overcome the unfavourable factors in the security management and ensure the safe and normal operation of stadium as well as the personal and property security of sporters. In addition, in the analysis and applica-

tion, the actual situation is combined. PCA in this paper provides some new ideas for other analysis method, which increases the theoretical research basis of PCA and provides some references to the further research, with theoretical and realistic significance.

Acknowledgments

Mingchang Liu is corresponding author. This work was supported by 2013 Hubei Provincial Education Science Twelfth Five Year Plan Project [2013B032].

References

[24] Xiao Zenan 2008 Smoke Control System Performance Design of Beijing Olympic Stadium *Journal of Agricultural Science and Technology* **26**(15) 45-9
[25] Chen Yonggao, Sun Wenjian 2007 Research on Competitiveness Evaluation of Construction Enterprise Based on PCA *Construction Technology* **36**(12) 33-4
[26] Yu Xia 2010 Wang Dongqiang. Comprehensive Evaluation on Fire Performance of stadium *Journal of the Chinese People's Armed Police Force Academy* **26**(6) 34-9
[27] Mohammad Saadatseresht, Ali Mansourian, Mohammad Taleai. 2009 Evacuation planning using multiobjective evolutionary optimization approach *European Journal of Operational Research* **198**(1) 305-14
[28] Lämmel G, et al. 2009 The representation and implementation of time-dependent inundation in largescale microscopic evacuation simulations *Transportation Research Part C* **18**(1) 84-98
[29] Zheng X P, Zhong T K, Liu M T 2009 Modeling crowd evacuation of a building based on seven methodological approaches *Building and Environment* **44**(3) 437-45
[30] Deb K, Pratap A, Agarwal S, Meyarivan T 2002 A fast and elitist multi-objective genetic algorithm: NSGA-II *IEEE Transaction on Evolutionary Computation* **6**(2) 181-97
[31] Salim Labiod, Thierry Marie Guerra 2009 Anytime measures for top- k algorithms on exact and fuzzy data sets *The VLDB Journal*, **18**(2) 407-27
[32] Yi-Fei Chen, Xiao-Lin Qin, Liang Liu, Bo-Han Li 2012 Fuzzy Distance-Based Range Queries over Uncertain Moving Objects *Journal of Computer Science and Technology* **27**(2) 376-96
[33] Kim Y M, Choi J C, Kim J H, Kim C 2002 Development of a System for Progressive Working of an Electric Product by Using Fuzzy Set Theory. *The International Journal of Advanced Manufacturing Technology* **20**(10) 765-79

Authors

Huajie Chen, born on January 1, 1977, China

Current position, grades:
University studies: majored in Sports Education, won the Bachelor Degree of Education from Wuhan institute of sport in China in June 2001, and acquired Master of Science in Education from Shanghai University of sport in 2010.
Scientific interests: science of physical culture and sports is his research field, previous research area is Sports Science and school Physical education
Publications: In April 2012, he made paper "Research on Jiangxi province coach continued education and post training present situation" - a Periodical which named Popular literature and art.
Experience: In July 2001 served as a teacher in the Physical Education Department of Jinggangshan University located at Jian sity in China.

Mingchang Liu, born on December 1, 1977, China

Current position, grades: P.E teacher in the Department of Physical Education, Huazhong Agricultural University in China.
University studies: Majored in Table Tennis Teaching and Training at the College of Physical Education, studied in Wuhan Sports University from 1997 to 2001, and graduated in 2001, master's degree in education at the School of Physical Education & Sports Science, Nanjing Normal University in 2007. Master's Thesis entitled "Study on the Curriculum Arrangement and Teaching Mode of University Physical Education in Wuhan" was awarded national outstanding Master's thesis.
Scientific interests: School sports and community sports, and how school sports better assisted and served community sports.

Application of factor analysis in risk evaluation of basketball arena project construction

Ronghui Hu*

Physical Education Institute, Hunan University of Technology, Zhuzhou 412007, Hunan, China

Received 1 June 2014, www.cmnt.lv

Abstract

In basketball arena project construction, the project safety is an important topic. Aimed at the current situation of basketball arena project construction, the risk factors in the project construction are evaluated from the aspect of factor analysis method and the data in the process are processed according to relevant theory of factor analysis method and by virtue of SAPP software, so as to verify the scientific reasonableness of the method. In addition, in the process of evaluation, the subjectivity is linked with the objectivity in combination with the reality to provide main basis for the safety risk problems in the basketball arena project construction, which is beneficial for the safety management in the project construction. The application of factor analysis method in the basketball arena project construction can provide some new thoughts for other analysis methods, so as to enhance the theoretical research basis for the factor analysis method.

Keywords: Factor analysis method, Basketball arena, Project construction risk, Evaluation

1 Introduction

The concept of factor analysis originated from the intelligent test statistical analysis experiment of Karl Pearson and Charles Spearmen etc. [1]. In recent decades, with the rapid development of computer network, factor analysis method has been widely applied in various fields, so as to further enrich the theory and method of factor analysis [2]. Factor analysis method is a model analysis method to find out the relevant public factors, which is to construct several public factors with clear significance on the basis of main components, and to analyze and explain the original variables based on these public factors, so as to get the relation with the original variables [3].

With the continuous deepening of the reform and opening up, more and more basketball arenas are required, and the project construction of basketball arena has also welcomed infinite vitality, developing at an unprecedented scale and speed [4]. However, in recent years, the project construction has become "the disaster area of corruption", with a lot of risk safety crisis, mainly reflected in the following two aspects: first in the risk management of project construction, the construction unit selects staff with low quality and skill level, the machinery and equipment are operated unreasonably and there is an improper management for the material purchase and quality; second, the supervisors do not strengthen supervision to the project, with cheating on workmanship and materials [5-6]. Therefore, the researching and analysis on the risk problems in the basketball arena project construction and effectively finding out a risk control approach have been increasingly emphasized by the academic circle and political circle [7]. In order to construct the project construction better, it is required to strength the research on the risk evaluation problem of basketball arena project construction [8]. In order to comprehensively evaluate the basketball arena project construction, we research and discuss from different theories, so as to introduce the application of factor analysis method in the risk evaluation of basketball arena project construction [9].

Factor analysis method is a multi-variable statistical analysis method of including a lot of variables with complicated relations into several comprehensive factors by starting from the research on the dependence of internal correlation [10]. Its key is to detect and classify the correlation between variables and finally use a public factor to express the original variable, and it is widely used for limited unforeseeable invisible variables to explain the correlation between the original variables [11]. Factor analysis method is one of the important methods in the academic circle to analyse a lot of problems. In this research, the factor analysis method is used to comprehensively evaluate the risk problems of basketball arena project construction, so as to further discuss the application of factor analysis method.

2 Overview of factor analysis method

The main purpose of factor analysis method is to use a few factors to describe more indicators and analyze the correlation among these factors, and include several variables with close relation into the same category, each category of variable is one of the factors (because it is not the specific variable, and cannot be observed), and use a few factor to reflect most of the information of the things evaluated.

* ***Corresponding author*** e-mail: huronghui@yeah.net

Using this research method, we can analyze and solve a lot of problems better conveniently.

2.1 BASIC MODEL OF FACTOR ANALYSIS METHOD

The basic model of factor analysis method is described as follows:

(1) $X=(X_1, X_2, \ldots, X_P)$, in which X is an observable random vector, with mean factor $E(X)=0$, covariance matrix $Cov(X)=\sum$, and the covariance matrix $\sum$ is equal to relevant matrix R (the variable should be standardized to achieve equality of the two).

(2) $F=(F_1, F_2, \ldots F_m)$, in which $F(m<p)$ is an unobservable vector, with mean factor $E(F)=0$, covariance matrix $Cov(F)=I$, and at this moment, each component is mutually independent.

(3) $e=(e_1, e_2, \ldots, e_p)$, in which e and F are mutually independent, and $E(e)=0$, e is the covariance matrix $\sum$, which is a diagonal matrix, and at this moment, each component e is also mutually independent, so the matrix can be expressed as:

$$
\begin{aligned}
X_1 &= A_{11}F_1 + A_{12}F_2 + \ldots + A_{1m}F_m + e_1 \\
X_2 &= A_{21}F_1 + A_{22}F_2 + \ldots + A_{2m}F_m + e_2 \\
&\ldots \\
X_P &= A_{p1}F_1 + A_{p2}F_2 + \ldots + A_{pm}F_m + e_p
\end{aligned}
$$

It is called factor analysis model, this model is aimed at the variable for calculation, and each factor is mutually orthogonal, so it is also called R-type orthogonal factor model.

Its model matrix form is $x=A_F+e$

Where, X, A, F, e are the known correlatives.

Here,

(1) m is unequal to p;

(2) $Cov(F, 0)=0$, i.e. F is unrelated to e;

(3) $D(F)=I_m$, i.e. $F_1, F_2, \ldots, F_m$ are unrelated and their variance is 1;

The quadratic sum of each element of column $j(j=1, 2, \ldots, m)$ of the factor load matrix A is recorded as g_{j2}, called the variance of the public factor F_j to X. g_{j2} is the sum of variance provided by the j_{th} public factor F_j to each component $X_i(i=1, 2, \ldots, p)$ of X, which is an important factor to weigh the public factor. The more g_{j2} is, the more contribution of the public factor F_j to X will be, or the larger influence on X will be. If we calculate all $g_{j2}(j=1, 2, \ldots, m)$ of the factor load matrix A, and sort them in order of size, finally we can get the most influential public factor based on this.

2.2 BASIC CALCULATION STEP OF FACTOR ANALYSIS METHOD

During research and analysis on the factor analysis method, the main problem faced is to reasonably construct the factor variable and name and interpret the factor variable, and the general steps are:

(1) Input the raw data, detects their reasonable correlation, calculate the sample covariance and carry out standardized processing;

(2) Calculate the sample correlation coefficient matrix R;

(3) Calculate the characteristic root $\lambda_i(\lambda_1, \lambda_2, \lambda_3 \cdots \succ 0)$ of correlation coefficient matrix, and the characteristic vector.

(4) Determine the number of public factor, and calculate its communality variance;

(5) Rotate the load matrix, so as to better explain the public factor;

(6) Have a special explanation on the public factor;

(5) Calculate the final score with the variance contribution rate weight of each factor.

3 Application of factor analysis method in risk evaluation of basketball arena project construction

In the research on the fire accident in gymnasium, evaluation and assessment on the warn of fire disaster risk is an inseparable link in our daily warning. With the development of society, the assessment systems and methods for the fire risk warning have been diversified increasingly, but most of them are limited to the investtigation and analysis stage, and there is seldom further analysis and affirmation from theory. According to the above situation, we establish the fire risk system of main component analysis method and add the scientific component of mathematical model to the subjective analysis, so as to realize the transformation from qualitative analysis to the accurate quantitative analysis.

3.1 SELECTION OF RISK EVALUATION INDICATORS IN BASKETBALL ARENA PROJECT CONSTRUCTION

The factor analysis method is applied to consider the comprehensive factors in the project construction in recent 10 years, from the initial construction to the later completion of the basketball arena, including many risks, 10 specific risk manifestation indicators are selected as the raw data, see Table 1;

TABLE 1 Risk manifestation indicators

Raw data indicator	
X_1	Personnel physical quality risk
X_2	Personnel professional skill risk
X_3	Personnel knowledge and common sense risk
X_4	Mechanical technology risk
X_5	Project design risk
X_6	Material technology risk
X_7	Project supervision risk
X_8	Artificial environment risk
X_9	Natural condition risk
X_{10}	Geographical condition risk

3.2 PROCESS ANALYSIS IN RISK EVALUATION OF BASKETBALL ARENA PROJECT CONSTRUCTION

With the popularization of computer and its use in various aspects, in the process SAPP data analysis software can be used for the factor analysis method.

(1) Analyse whether the selection of evaluation indicator in basketball arena project construction is appropriate for the factor analysis. Bartlett's test of sphericity and KMO detection are used for detection and SPSS software is used for the calculation to obtain the raw data correlation examination, as shown in Table 2.

TABLE 2 KMO detection result

Bartlett and KMO detection		
Sufficient sampling through KMO weighing		0.708
Bartlett's test of sphericity	About Chi-square	253.875
	Degree of freedom	120
	Examination value of difference significance	0.000

KMO detection value ranges from 0 to 1, the more KMO value indicates that there are more communities among variables and it is more appropriate for factor analysis. Generally, when the detection value is more than 0.5, factor analysis is appropriate, whole not appropriate when less than 0.5. The result in Table 1 sows that the KMO value is 0.708, Bartlett examination value is 0.00, less than 1%, indicating that the evaluation indicator is appropriate for the factor analysis, but there is a correlation among the variables.

(2) Carry out standardization processing to the primitive matrix, the standardization expression formula is $Zi = \frac{Xj - \bar{X}j}{Sj}$ in which X_j is the j_{th} risk evaluation indicator of the i_{th} basketball arena, S_j is the mean value and sample covariance of X_j, so as to calculate the relevant coefficient matrix of the standardization data and get the characteristic value.

TABLE 3 Factor variance contribution rate

Explanation on all variables						
Component	*Initial characteristic value*			*Factor rotation load*		
	Component characteristic value	Percentage of factor variance in total variance /%	Accumulative percentage of factor variance in total variance/ %	Component characteristic value	Percentage of factor variance in total variance /%	Accumulative percentage of factor variance in total variance/ %
1	4.467	27.921	27.921	3.92	24.502	24.502
2	2.078	12.99	40.911	1.968	12.302	36.804
3	1.399	8.741	49.652	1.757	10.98	47.784
4	1.240	7.752	57.404	1.312	8.190	55.983
5	1.066	6.664	64.067	1.293	8.084	64.067
6	0.935	5.845	69.913			
7	0.867	5.422	75.334			
8	0.722	4.511	79.846			
9	0.669	4.182	84.028			
10	0.624	3.9	87.928			

TABLE 4 Rotated factor load matrix

	Component				
	1	*2*	*3*	*4*	*5*
X_1	0.709	0.14	0.039	0.092	-0.073
X_2	0.697	0.11	-0.308	0.166	-0.091
X_3	0.826	0.149	0.04	0.054	0.071
X_4	0.102	-0.454	-0.147	0.089	0.213
X_5	0.395	0.602	-0.278	-0.31	0.031
X_6	-0.2	0.694	0.09	-0.018	0.099
X_7	0.048	0.168	-0.782	-0.148	0.196
X_8	0.125	-0.191	0.14	0.974	0.083
X_9	0.183	0.151	-0.043	-0.849	0.053
X10	0.026	0.6	0.118	0.068	0.706

Generally, according to the calculation of variance contribution rate and the accumulative variance contribution rate, the total variance contribution rate should be above 80%.

(3) Rotate and determine the factors. Imagine F_1, F_2, …, F_m are m factors, the first n factors with accumulative variance contribution rate above 80% are obtained, and these n factors can represent all information volume. The actual significance of these n factors cannot be determined clearly, so they are often rotated to obtain the significant actual meaning.

TABLE 6 Factor scoring coefficient matrix

	Component				
	1	*2*	*3*	*4*	*5*
X1	-0.098	0.495	-0.073	0.092	-0.081
X2	0.187	-0.011	-0.178	0.2	-0.12
X3	-0.002	0.104	-0.128	0.69	0.058
X4	0.205	-0.137	-0.049	0.11	0.121
X5	0.189	-0.095	-0.071	-0.192	-0.025
X6	-0.025	0.002	0.473	-0.143	0.082
X7	-0.081	0.086	-0.053	-0.095	0.627
X8	0.038	-0.158	0.066	0.202	0.539
X9	-0.014	0.099	-0.224	-0.21	0.031
X10	-0.063	0.301	0.279	-0.157	0.064

The factor matrix is rotated with orthogonal rotation method, the results obtained show that the first factor has a large load in X_1, X_2 and X_3, i.e. the influential factors represent these aspects; the second one has a relatively large load in X_4, X_5 and X_6; the third one has a large load in X_7, the fourth one has a large load in X_8 and X_9, the fourth factor has a large load in X_8 and X_9 and the fifth one has a relatively large load in X_{10}. The first factor represents the staff, the second one the influence of project construction machinery, facilities and materials, the third one the influence of construction site, the fourth factor the influence of environmental conditions, and the fifth factor the influence of geology selected for the project construction, so it is concluded that the variables represented by the factors are as shown in Table 5 below, consistent with the actual condition.

TABLE 5 Factor structure

Factor	Raw data	Variable
F_1	X_1	Staff physical condition: disease, emotion, attitude
	X_2	Staff skills: Proficiency, mastering degree of professional knowledge
	X_3	Staff knowledge: special and hazardous work, equipment safety, electrical power, open fire operation site, construction technology, protective goods, mastering and defect of protection technology.
F_2	X_4	Machinery: obsolete or problematic machinery,
	X_5	overloading operation, mixed operation of
	X_6	personnel and machinery. Facilities: setup, overloading operation Materials: transmission method, ordering degree of stacking, use of nonconforming materials, storage of special materials, use of members, accessories and factor prefabricated members.
F_3	X_7	Site: size of working face, whether work platform or not, arrangement of construction site, whether protective facilities or not, overhead operation or deep foundation pit operation, clearing of construction site.
F_4	X_8	Condition: special environment (oxygen deficit,
	X_9	toxic gas, smoke, noise, vibration), on-site lighting and obstacle influencing vision etc. Climate: severe natural environment
F_5	X_{10}	Geology: stratum, water quality, construction

(4) Calculate comprehensive score: carry out linear combination with raw indicator data, so as to get the weight of covariance contribution rate of each factor, the comprehensive evaluation of the linear combination of factor indicator is expressed as:

$F = \omega_1 F_1 + \omega_2 F_2 + \omega_3 F_3 + \cdots + \omega_m F_m$, in which ω_m represents the covariance contribution weight of the rotated factor.

The score of each factor can be calculated according to the formula $F = \omega_1 F_1 + \omega_2 F_2 + \omega_3 F_3 + \cdots + \omega_m F_m$.

$$\begin{cases} F_1 = -0.098X_1 + 0.187X_2 - 0.002X_3 + \cdots - 0.063X_{10} \\ F_2 = 0.495X_1 - 0.011X_2 + 0.104X_3 + \cdots + 0.301X_{10} \\ F_3 = -0.073X_1 - 0.178X_2 - 0.128X_3 + \cdots + 0.279X_{10} \\ F_4 = 0.092X_1 + 0.200X_2 - 0.690X_3 + \cdots - 0.157X_{10} \\ F_5 = -0.081X_1 - 0.120X_2 - 0.058X_3 + \cdots + 0.064X_{10} \end{cases}$$

Then SPSS is used to calculate the function scoring result based on the above matrix formula, and the scores of the 5 factors are converted into new variables, then the covariance contribution rate weight of each factor is calculated, and finally it is concluded that the comprehensive situation of the risk evaluation of project ball arena project construction is:

$$F = 0.179F_l + 0.230F_2 + 0.187F_3 + 0.078F_4 + O.067F_5 .$$

3.3 RESULT ANALYSIS OF RISK EVALUATION OF BASKETBALL ARENA PROJECT CONSTRUCTION

The comprehensive score of each factor can be obtained according to the covariance contribution rate of the rotated factor, so as to analyze the influence of the factors on the risk evaluation of basketball area project construction according to the score. The result shows that the second factor has the highest score, with coefficient 0.230, followed by the third factor 0.180, and then successively the first one 0.179, the fourth one 0.078 and the fifth one 0.067. As the second factor represents the machinery (obsolete or problematic machinery, overloading operation, mixed operation of personnel and machinery), facilities (setup, overloading operation) and materials (transmission method, ordering degree of stacking, use of nonconforming materials, storage of special materials, use of members, accessories and factory prefabricated units), which are the skeleton of basketball arena construction in actual application and play a vital important role for the project safety; the third factor site (size of working face, weather work platform or not, arrangement of construction site, whether protective facilities or not, overhead operation or deep foundation pit operation and clearing of construction site) also directly concerns the safety in project construction, while the objective factors such as geology and environmental condition have a small risk influence on the project construction.

Therefore, in the basketball area project construction, first, strictly control the quality and operation of machinery, equipment and raw materials, when supervising the project, relevant departments must detect the project strictly according to the national regulations and systems; second, the project construction personnel should have professional technical level and basic occupational moral

quality, prevent cheating in craftsmanship and materials and illegal operation. Finally, in the basketball area project construction, it is also required to pay attention to the selection of geological condition, and select the high-quality geographical environment as far as possible.

4 Conclusions

The factor analysis method has a simple principle, with relatively complicated calculation process. With the wide application of computer, using SAPP to process and analyze the step calculation process greatly reduces the work amount of data calculation and improves the accuracy, so factor analysis method had been gradually applied in various fields. It is possible to objectively analyze the safety of basketball arena project construction, discover the potential risks timely and overcome the adverse factors unfavorable for the project construction by using factor analysis method to comprehensively evaluate the basketball arena project construction risks, carrying standardization processing to the raw data and rotating the factor to get the comprehensive score of each factor, so as to promote the normal, healthy and continuous development of the basketball arena project construction. Besides, this evaluation and analysis method is to used data for statistical processing via computer, the influences of subjective factors are overcome in the process of evaluation and the result obtained is more objective, conforming to the actual project construction condition more. The application of factor analysis method researched in this paper in the risk evaluation of basketball arena project construction provides some new thoughts for other analysis methods in a certain extent, increasing the theoretical research basis for the factor analysis method and providing some reference values for the further research.

References

[1] Jianmin Wu, Peter M Bentler 2012 Application of H-likelihood to factor analysis models with binary response data *Journal of Multivariate Analysis* **106** 72-9.

[2] Xiaowei Zhang, Ying Zhang, Dasong Deng 2012 The Efficiency Evaluation of China's Urban Pension System: Using Factor Analysis and the Analytic Hierarchy Process *JDCTA* **6**(4) 33-41

[3] Unkel S, Trendafilov NT 2010 A majorization algorithm for simultaneous parameter estimation in robust exploratory factor analysis. *Computational Statistics & Data Analysis* **54**(12) 3348-3358

[4] Donald Robertson, James Symons 2007 Maximum likelihood factor analysis with rank-deficient sample covariance matrices *Journal of Multivariate Analysis*,**98**(4) 813-28

[5] Hiroki Okubo, Mont Hubbard 2010 Identification of basketball parameters for a simulation model *Procedia Engineering* **2**(2) 3281-86

[6] Hamid Jazayeriy, Masrah Azmi-Murad, Nasir Sulaiman, Nur Izura Udizir 2011 Pareto-optimal Algorithm in Bilateral Automated Negotiation *JDCTA* **5**(3) 1-11

[7] Chao Lu 2012 Analyzing the Pareto Principle Based on the Artificial Stock Market *JCIT* **7**(16) 78-86

[8] Bhat N, Barrans S M, Kumar A S 2010 Performance analysis of Pareto optimal bearings subject to surface error variations. *Tribology International* **43**(11) pp. 2240-49

[9] Ma Dan 2009 Using Factor Analysis Method to Evaluate Students' Scores *Journal of Qiqihar University* **25**(3) 79-82

[10] He Shiyou Xu Wenqin 2003 Application of Factor Analysis Method in Comprehensive Evaluation of Economic Benefit of Industrial Enterprises. *Mathematical Statistics and Management*, **22**(1) 123-5

[11] Li Fenhong, Wang Xiao 2008 Factor Analysis and Its Application in Comprehensive Evaluation of Students *Journal of Shangluo College* **22**(5) 56-9

Authors

Ronghui Hu, born on October 1, 1979, China

Current position, grades: teacher at Physical Education Institute, Hunan University of Technology, China.
University studies: Bachelor Degree of Science in Education from Wuhan Sports University in 2002 and Master degree of Science in Education from Wuhan Sports University, China in 2010.
Scientific interests: physical education and physical exercise.

Short-term traffic flow forecast of highway network based on chaos time series method

Hao Zhang*

Department industrial engineering, Tsinghua University, Beijing, China, 100084

Received 1 August 2014, www.cmnt.lv

Abstract

With the construction of the highway network and the growing traffic flows, demand for real-time control and guiding service have become increasingly prominent. As for short-term forecast of highway network, it is not only the basis and foundation of the real-time control and guiding service for traffic flows, and the precise forecast result will have a magnificent impact on improving the traffic capacity and service levels. This paper builds a short-term traffic flows forecast model for highway network based on chaotic time series analysis and prediction theory. The forecast of the traffic flows in given areas can be calculated. Results show that this model is feasible and has a high accuracy.

Keywords: highway network, traffic flow, chaos theory, time series, prediction

1 Introduction

With the increasing completed highway network and traffic flow, the demand for real time control and guidance services of highway traffic flow becomes increasingly prominent. Short-term traffic flow forecast of highway network not only is the foundation and basis of real time control and guidance services of highway traffic flow, and the accuracy of prediction results is important to improve traffic capacity and level of service of highway network.

According to the traffic flow prediction for different purposes, the focus and requirements of the prediction are also different. On the one hand, for long-term traffic flow forecasting, the demand for prediction accuracy is not high. On the other hand, in order to make network traffic control scheme, especially to make traffic control and guidance service, the prediction period is greatly reduced. Known for short-term traffic flow forecast, it requires high precision. At present, short-term traffic flow forecast mainly serves for real-time control and guidance services for city traffic so it requires high real-time control. The maximum period of traffic control is 2.5-3min generally and traffic guidance is 5min. Thus, the period of short-term traffic flow forecast of city traffic is generally no more than 15min. However, with the development of highway network in China, traffic congestion is increasingly frequent and serious in the local highway. Therefore, the demand for real time control and guidance services of highway traffic flow is increasingly urgent. Due to the difference between the highway traffic flow characteristics and city road network, it is necessary to study short-term traffic flow forecast of regional highway network.

2 Choosing short-term traffic flow forecast method of highway network

Traffic flow forecast is to speculate the future traffic flow according to the present traffic flow data, namely to make real-time prediction of traffic flow for the next time $t+\Delta T$ and subsequent several times at the moment t. Usually, the traffic flow within 15min is called the short-term traffic flow, so to predict the traffic flow between t and $t+\Delta t$ which does not exceed 15min is short-term traffic flow forecast.

There are many prediction methods for short-term traffic flow forecast [1-3,5,7], but generally can be divided into two categories: one is based on traditional statistics methods, such as regression prediction, time series prediction, Kalman Filtering model and so on; the other category is the forecast method based on neural network, fuzzy mathematics and nonlinear theory, which does not pursue mathematical derivation strictly and clear physical meaning of the object but pay more attention to the fitting effect of the real traffic flow phenomena. Among the above prediction methods, historical mean method and regression method use least square estimation to compute parameters, which is a simple calculation but the uncertainty and nonlinear of the traffic flow are difficult to reflect and the influence of random factors are difficult to overcome; neural network forecasting method, including BP network, fuzzy neural network and the high order neural network trains parameters complicatedly and needs long computing time; Kalman Filtering method and nonlinear time series method are characterized as high real-time and accuracy. Due to the complex feature of highway traffic flow whose change process is nonlinear and uncertain, and it is difficult to find a mathematical model to reflect the traffic flow characteristics in practice, this paper adopts second category of forecasting methods,

* ***Corresponding author's*** e-mail: statistics.zhang@gmail.com

namely short-term traffic flow forecast of highway network based on chaos theory.

To predict short-term traffic flow of highway network based on chaos theory, it is primarily to distinguish whether the road network traffic flow system is a chaotic system. If it is a chaotic system, there can be short-term prediction, but not for long-term forecast; secondly, to determine the embedding dimension and time delay parameters and to find out the last known points in phase space, we need to use phase space reconstruction technique to make phase space reconstruction for traffic flow time series data; then we find out several adjacent points in the phase space and determine the predictive value of fitting function and separation in a known point as the center.

At present, there are many achievements of chaos theory to predict short-term traffic flow in domestic and foreign application [4,6,13,18], but most of these studies are used for traffic flow forecasting on a section in the city road network, while there needs further research for multiple sectional short-term traffic flow forecasting methods and models for highway network. The following will discuss the use of multidimensional chaotic time sequence to construct multiple sectional short-term traffic flow forecast model of highway network.

3 Short-term traffic flow forecast model of highway network based on chaos theory

3.1 BUILDING MULTIDIMENSIONAL CHAOTIC TIME SERIES PREDICTION MODEL

First, construction of multidimensional time series matrix.

Assumption (1): choosing traffic flow statistical data of M sections, corresponding M time series is obtained. X_i, i = l, 2, …, M.

Assumption (2): the length of each traffic flow time series sample is N. Thus, a multidimensional time series matrix as follows:

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,N} \\ x_{2,1} & x_{2,1} & \cdots & x_{2,N} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N,1} & x_{N,1} & \cdots & x_{N,N} \end{bmatrix}. \tag{1}$$

Second, identification of chaotic time series. To make chaotic time series analysis of the traffic flow time series X_i respectively above, and select time delay σ_i and embedding dimension d_i to make phase space reconstruction, calculating Lyapunov index of each time series. If the index is greater than zero, it indicates that the time series is a chaotic time series, where the related theory and the method of chaotic time series can be used for analysis and modeling.

Third, reconstruction of multidimensional phase space. Based on the chaotic time series, according to multidimensional phase space reconstruction method proposed by Liangyue Caoetal (1998), the reconstruction of phase space is obtained:

$$\begin{aligned} V_n = (&x_{1,n}, x_{1,n-r_1}, ..., x_{1,n-(d_1-1)r_1}, \\ &x_{2,n}, x_{2,n-r_2}, ..., x_{1,n-(d_1-1)r_2}, \\ &..., \\ &x_{M,n}, x_{M,n-r_m}, ..., x_{M,n-(d_1-1)r_m}), \end{aligned} \tag{2}$$

$$n = \max_{1 \le i \le M}(d_i - 1)\tau_i + 1, \ldots, N, \tag{3}$$

Forth, construction of prediction model. According to the embedding theorem, there is a smooth mapping function in a D dimensional space like $F\left(d = \sum_{i=1}^{M} d_i\right)$: $R^{\mathrm{d}} \to R^{\mathrm{d}}$, If d or di is sufficiently large to make $V_{\mathrm{n+1}} = F(V_{\mathrm{n}})$, then the model can be described as:

$$V_{1,n+1} = F_1(V_{\mathrm{n}}), V_{2,n+1} = F_2(V_{\mathrm{n}}), \ldots, V_{M,n+1} = F_M(V_{\mathrm{n}}). \tag{4}$$

After establishing the prediction model, there comes out a phase space after reconstruction of multi-dimensional chaotic time series. According to the embedding theorems of chaos theory, there is a smooth mapping function F which can reflect the law of phase space, and the smooth mapping function F can establish a dynamic mathematical model to fit it to predict the motion trace of phase space, which can get the forecasting results of highway network traffic flow. After finishing the reconstruction of phase space through multidimensional time series, we adopt prediction methods like global method, local method and maximum Lyapunov exponent to establish the mathematical model to approximate the mapping function F. In this paper, we use the weighted zero order local forecasting method to make calculation, using V_n as the center point to predict V_{n+1}.

3.2 STEPS OF WEIGHTED ZERO ORDER LOCAL PREDICTION CALCULATION

Steps of weighted zero order local prediction calculation are as following based on reconstruction of phase space:

First, calculating K adjacent points $V_{n,j}^{\min}$, which is the minimum Euclidean distance with V_n in the phase space after reconstruction, the distance of each point are respectively R_j, j = l, 2, …, K, and let $R_{\min}$ be the minimum value. Thus, each adjacent point's weight can be defined as:

$$P_j = e^{-l(R_j - R_{\min})} / \sum_{j=1}^{K} e^{-l(R_j - R_{\min})}. \tag{5}$$

Secondly, V_{n+1} is calculated according to the following equation:

$$V_{n+1} = \sum_{j=1}^{K} P_j V_{n,j}^{\min} \ . \tag{6}$$

Lastly, according to:

$$\begin{aligned} V_{n+1} = (&x_{1,n+1}, x_{1,n+1-r_1}, \cdots, x_{1,n+1-(d_1-1)r_1}, \\ &x_{2,n+1}, x_{2,n+1-r_2}, \cdots, x_{1,n+1-(d_1-1)r_2}, \\ &\cdots, \\ &x_{M,n+1}, x_{M,n+1-r_m}, \cdots, x_{M,n+1-(d_1-1)r_m}), \end{aligned} \tag{7}$$

The V_{n+1} can be isolated from the final traffic flow prediction results as:

$$\left(x_{1,n+1}, x_{2,n+1}, ..., x_{M,n+1}\right). \tag{8}$$

There are two parts of the parameters in the above model. One part is the time σ_i and embedding dimension d_i of phase space reconstruction. The other part is the neighbor number K and the calculation of adjacent weight parameter l in the weighted zero order local forecasting method. In the theory of chaotic time series analysis, there are many ways to determine the time delay σ_i and embedding dimension d_i, such as the auto-correlation method, mutual information method and C-C method. Methods to determine the embedding dimension are G-P algorithm, the false neighbor method C -C method and Cao method. This paper use C-C method to determine the time delay and embedded dimension. The parameters K represents the number of neighboring points. Too few adjacent points leads to low prediction accuracy, while too many neighbor points is unnecessary because it may make the computational complexity increase and affect the prediction effect; the parameter l represents the weights of adjacent points, its weight is determined according to the proximity to the center point of the Euclidean distance in the forecast. In certain Euclidean distance, less l means greater proportion. In this paper use the particle swarm optimization algorithm to select the parameters of K and l, because the particle swarm algorithm is simple and effective and the solving process is easier.

4 Empirical studies

4.1 SOURCES OF DATA

In order to test the model and the method, the traffic flow data comes from Jinggangao Highway, Hurong Highway, Huyu Highway and Yinfu Highway Yuenan Station in the Hubei highway network from 00:00 in September 15 to 23:59 in September 15th, 2013. Through the correlation data pre-processing and completed analysis, the traffic flow of Section 1, Section 2,… Section 6 are sources of data. The flow curves are shown in Figure 1 to 3 (Section 1 for example). The time interval Δ_t is respectively 5 min, 10min and 15min. Then the prediction results and the measured results are compared based on established model, parameters optimization and prediction.

4.2 STEPS OF THE EMPIRICAL STUDIES

Firstly, to make pre-treatment of the selected section traffic flow data and make correlation analysis with the completed data.

Secondly, calculating the time delay embedding dimension and Lyapunov index of each time series using the C-C method shown in Table 1. The lyapnuvo index is positive shown in Table 1, which suggests that the 6 time series are all chaotic time series so we can use the chaos theory to analyze and predict.

TABLE 1 Result of parameters calculated under 6 time series and three intervals

5min	Section 1	Section 2	Section 3	Section 4	Section 5	Section 6
Time delay	1	1	2	2	1	2
Embedding dimension	3	10	4	4	7	3
Lyapunov	1.1605	1.0095	0.712	0.8286	1.2915	0.2845
10min	Section1	Section2	Section3	Section4	Section5	Section6
Time delay	1	1	3	3	1	2
Embedding dimension	2	4	3	2	5	3
Lyapnuvo	0.3905	0.5632	0.2636	0.2486	0.8283	0.2744
15 min	Section1	Section2	Section3	Section4	Section5	Section6
Time delay	1	1	3	2	1	2
Embedding dimension	1.75	2	3	3	5	3
Lyapnuvo	0.379	0.381	0.31	0.4836	0.4541	0.3756

4.3 COMPUTATIONAL PREDICTION AND ERROR

According to the time delay and embedding dimension, using the multi section model of chaotic prediction and the use of particle swarm optimization algorithm of K and optimization of l parameters, the calculation results of traffic flow of each cross section. The prediction results see Figures 1-3 (in Section 1 as an example), the solid

part is the prediction of traffic flow, the dashed part for the measured traffic flow, a unit for every 5 minutes, every 10 minutes and a bus every 15 minutes.

At different time intervals, each section of traffic flow forecast value and the actual value of average error, see Table 2.

TABLE 2 Different time intervals of the cross-section traffic predicted and actual values of average error table

average error	Section1	Section2	Section3	Section4	Section5	Section6
5min	0.026541	0.031395	0.031081	0.018483	0.099386	0.129176
10min	0.01886	0.021353	0.018794	0.006982	0.04686	0.055504
15min	0.018788	0.018246	0.012589	0.004737	0.035359	0.047582

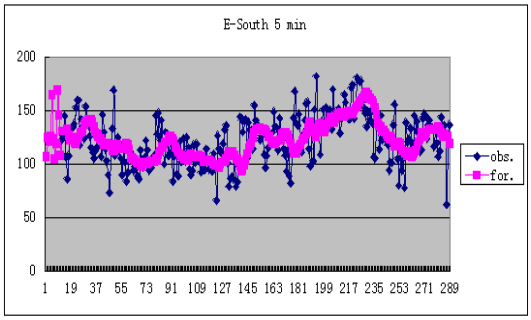


FIGURE 1 Under section 1 in 5min-interval actual and predicted traffic volume comparison chart

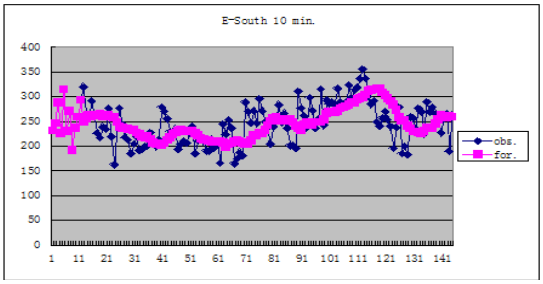


FIGURE 2 Under section 1 in 10min-interval actual and predicted traffic volume comparison chart

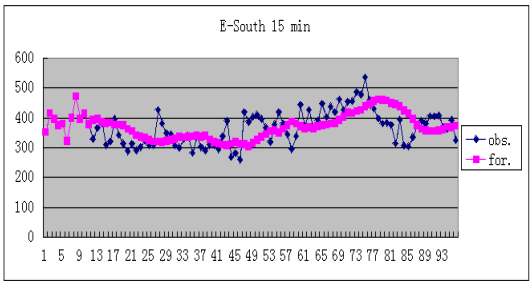


FIGURE 3 Under section 1 in 15min-interval actual and predicted traffic volume comparison chart

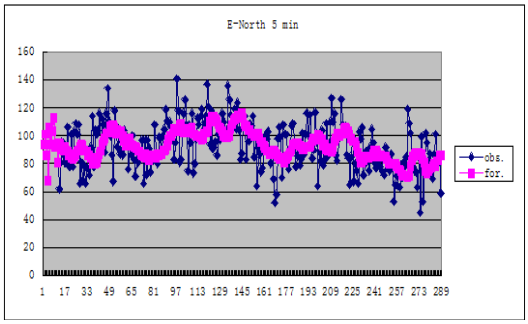


FIGURE 4 Under section 3 in 5min-interval actual and predicted traffic volume comparison chart

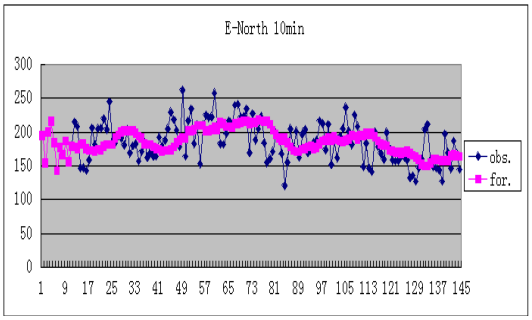


FIGURE 5 Under section 3 in 10 min-interval actual and predicted traffic volume comparison chart

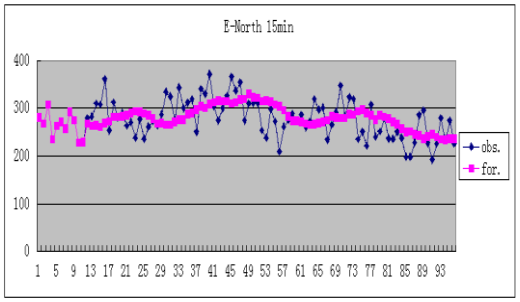


FIGURE 6 Under section 3 in 15 min-interval actual and predicted traffic volume comparison chart

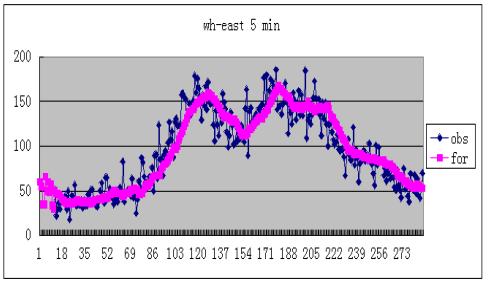


FIGURE 7 Under section 4 in 5min-interval actual and predicted traffic volume comparison chart

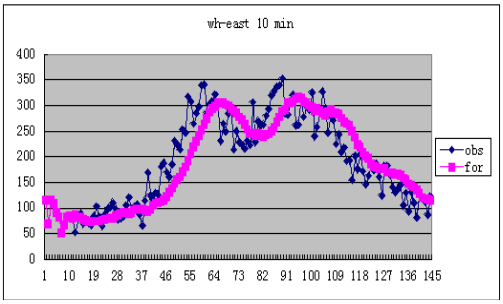


FIGURE 8 Under section 4 in 10 min-interval actual and predicted traffic volume comparison chart

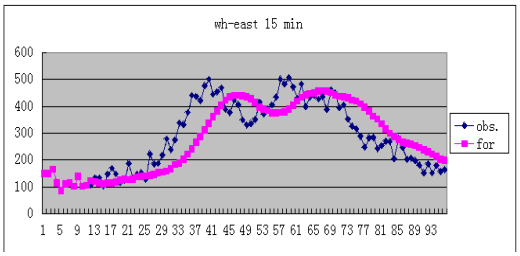


FIGURE 9 Under section 4 in 15 min-interval actual and predicted traffic volume comparison chart

5 Conclusions

Compared the actual and predicted values of traffic flow in section 1 and section3, it shows that the average prediction error tends to decrease with the increasing time intervals. Meanwhile, compared with chaotic time series prediction values in a separate and single section, the prediction accuracy of multiple-sections forecasting method is higher than single-section forecasting methods. Overall, the multiple-sections traffic flow forecasting method in road network based on multi-dimensional chaotic time series prediction is feasible, and the predictive effect is better than of the single-section chaotic time series prediction method. On the contrary to the short-term traffic flow forecasting in urban road network, see Figure 7, Figure 8 and Figure 9 in Section4 located at the intersection of the urban road network and the highway network, if the time intervals relatively increase in short-term traffic flow prediction of highway network, its predictive effect would be better.

References

[1] Hao B L 1989 Elementary Symbolic Dynamics and Chaos in Dissipative Systems *World Scientific Publishing Company*
[2] Okotani I, Stephanedes Y J 1984 Dynamic Prediction ofTraffic Volume Through Kalman Filtering Theory *Transportation Research Part B: Methodological* **18**(1) 1-11
[3] Guo M, Lan J, Xiao X, Lu H 2010 Forecasting Short-time Traffic Flow for Beijing 2nd Ring Road Using Chaos Theory *Journal of Transportation Systems Engineering and Information Technology* **10**(2) 106-11
[4] Liu J 2007 Study on Forecasting of Short-term Traffic Flow Based on Fractal Theory *Doctoral Dissertation at Beijing Jiaotong University (in Chinese)*
[5] Zhu X 2007 Research on a Practical Method of Freeway Forecasting Distribution *Central South Highway Engineering* **32**(3) 161-4 *(in Chinese)*
[6] Zhang Y, Qu S, Wen K 2009 Chaotic Property Analysis and Prediction Model Study for Traffic Flow Time Series *China Civil Engineering Journal* **42**(1) 119-23*(in Chinese)*
[7] Yue Y, Han W 2005 Determination of Parameters in the Phase-space Reconstruction of Multivariate Time Series *Control and Decision* **20**(3) 290-3
[8] Ghosh B, Basu B, O'Mahony M 2009 *IEEE Transactions on Intelligent Transportation Systems* **10**(2) 246-54
[9] Tchrakian T T, Basu B, O'Mahony M 2012 *IEEE Transactions on Intelligent Transportation Systems* **13**(2) 519-26
[10] Javad A, Behzad M, Baher A, Ali K S 2012 Forecasting of short-term traffic-flow based on improved neurofuzzy models via emotional temporal difference learning algorithm *Engineering Applications of Artificial Intelligence* **25**(5) 1022-42
[11] Kit Y C, Khadem S, Dillon T S, Palade V 2012 *IEEE Transactions on Industrial Informatics* **8**(2) 255-66
[12] Kit Y C, Dillon T S, Chang E 2013 *IEEE Transactions on Industrial Electronics* **60**(10) 4114-25
[13] Lippi M, Bertini M, Frasconi P 2013 *IEEE Transactions on Intelligent Transportation Systems* **14**(2) 871-82
[14] Kit Y C, Dillon T S, Singh J, Chang E *IEEE Transactions on Intelligent Transportation Systems* **13**(2) 644-54
[15] Zhang X, Onieva E, Perallos A, Osaba E, Lee V C S 2014 Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction *Transportation Research Part C: Emerging Technologies* **43**(1) 127-142
[16] Hosseini S H, Moshiri B, Rahimi-Kian A, Araabi B N 2012 Short-term traffic flow forecasting by mutual information and artificial neural networks *2012 IEEE International Conference on Industrial Technology (ICIT)* 1136-41
[17] Huang Z, Ouyang H, Tian Y 2011 Short-Term Traffic Flow Combined Forecasting Based on Nonparametric Regression Information Technology *International Conference on Computer Engineering and Management Sciences (ICM)* **1** 316-9
[18] Kaya S, Kilic N, Kocak T, Gungor C 2014 From Asia to Europe: Short-term traffic flow prediction between continents *International Conference on Telecommunications (ICT)* 277-82

Author

Hao Zhang, born in December, 1984, Wuhan city, China

Current position, grades: PhD Candidate in Department industrial engineering, Tsinghua University.
University studies: MS, Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook NY USA, 2009. Area of Study: Statistics and Quantitative Finance. BS, National Mathematics Base Class, Wuhan University, Wuhan China, 2007. Area of Study: Mathematics.
Scientific interests: operations research and logistics management.
Publications: 5 papers.

Optimization of industrial structure configuration based on fruit fly optimization algorithm

Sujiao Liu*

School of Philosophy and Public Administration, Henan University, Kaifeng, Henan, 475000, China

Received 1October 2014, www.cmnt.lv

Abstract

Establishing the mathematical optimization model of Shangqiu's agricultural structure from three aspects - economy, ecology and society, the agricultural industry structure in Shangqiu is regarded as the research object, utilizing fruit fly optimization algorithm to solve the mathematical model. Simulation results show that fruit fly optimization algorithm can solve the optimal solution for Shangqiu's agricultural industry structure to achieve the maximum benefits, thus providing the decision-making basis for adjustment and development of Shangqiu's agricultural structure.

Keywords: fruit fly optimization algorithm, industrial structure, optimization model, population size, iterations

1 Introduction

Shangqiu has a large population, little land, fragile ecological environment, farmers with low cultural quality, poor economy and improper industry positioning which lead to the unfavorable development of poverty reduction in the region. There are a large proportion of absolute poor population in Shangqiu, and backward is still the reality of this region. Based on 2012-2020 poverty reduction program of Henan Province, Shangqiu, for example, will build mathematical model for the agricultural structure in the region from aspects of economy, ecology and society, utilize fruit fly optimization algorithm for multi-objective optimization solution and achieve the optimal allocation in Shangqiu's agricultural structure, thereby providing decision-making basis for the development of agricultural industry structure in Shangqiu area [1-7].

2 Fruit fly optimization algorithm

2.1 OVERVIEW OF FRUIT FLY OPTIMIZATION ALGORITHM

Fruit fly optimization algorithm is a new group intelligence algorithm proposed by Pan Chao [8] in 2011. Fruit flies can detect food source on the air through their olfactory and visual superiority, and after detecting the food, they use acute vision to locate food and companions, and fly toward the food. The method is to deduce a new way to realize global optimization based on the foraging behavior of Fruit flies. Fruit fly optimization process is shown in Figure 1.

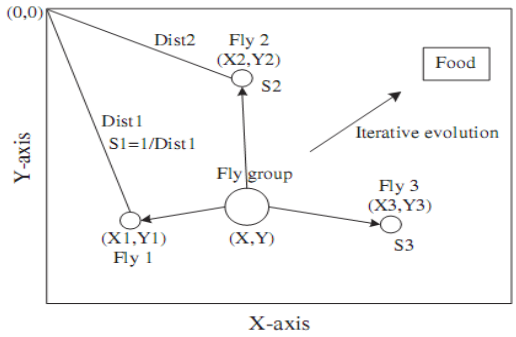


FIGURE 1 Optimization roadmap of fruit fly group

2.2 STEPS OF FRUIT FLY OPTIMIZATION ALGORITHM

Fruit fly optimization algorithms can be divided into seven steps as follows.

1) In Figure 1, the position initialization result of fruit fly group is $Init\ X\_axis$ and $Init\ Y\_axis$.

2) After setting the search direction RV_x and RV_y, random search distance of single fruit fly can be obtained by the following Equation:

$$Yi = Init\ Y\_axis + RV_y . \quad (1)$$

3) The distance from the origin to the current position of fruit fly is estimated a $Xi = Init\ X\_axis + RV_x$ a $Disti$, and the smell concentration is estimated according to Equation (2), which is equal to the reciprocal of the distance.

* ***Corresponding author's*** e-mail: kfliusujiao@126.com

$$Disti = \sqrt{Xi^2 + Yi^2}$$
$$Si = 1/Disti \quad (2)$$

4) The result of Equation (2) is substituted into Equation (3), the determination function of smell concentration, thus calculating the smell concentration value corresponding to the current position of individual fly.

$$Smelli = Function(Si) . \quad (3)$$

5) Best smell concentration of fruit fly group can be obtained by the following Equation

$$[bestSmell\ bestIndex] = \max(Smelli) . \quad (4)$$

6) Best smell concentration of fly group and the corresponding *x* and *y* coordinates are retained, and the fly group flies toward the food source through visual positioning.

$$Smellbest = bestSmell$$
$$X\_axis = X(bestIndex) \quad (5)$$
$$Y\_axis = Y(bestIndex)$$

7) Steps from 2) to 5) are repeated, and the smell concentration is judged. If a better smell concentration is obtained, go to Step 6); otherwise, iterative optimization will be executed.

3 Modeling

3.1 BOUNDARY DETERMINATION

The agricultural industry structure in Shangqiu is selected as the study object with 2012 as the planning year, and the study mainly focuses on the relationship between agriculture, forestry, animal husbandry and fishery, thus conducting the simultaneous layout optimization on the internal structure of crop farming and animal husbandry and achieving coordinated and sustainable development of different industries. Due to complex industrial structure in the region, different industries cannot be studied through a careful division, so a lot of data in the model are replaced by the average values [9,10].

3.2 DECISION VARIABLES

The planting area of selected major crops and cash crops in Shangqiu is X_{1j} (Ten thousand hm2), where rice X_{11}, wheat X_{12}, corn X_{13}, soybeans X_{14}, millet X_{15}, sorghum X_{16}, potato X_{11}, oil plant X_{18}, beet X_{19}, flax X_{110}, tobacco X_{111}, vegetable X_{112}, melon and fruit X_{113}, forage grass X_{114} and forage crop X_{115}. Forage grass and forage crop are regarded as a study entirety. In animal husbandry, the decision variable X_{2j} includes five variables: beef X_{21}, dairy cattle X_{22}, pig X_{23}, sheep X_{24} and poultry X_{25}. Forestry and aquaculture industry are regarded as a study entirety, where woodland area is X_3 (Ten thousand hm2) and the total aquaculture area is X_4 (Ten thousand hm2).

3.3 OBJECTIVE FUNCTION

Objective function $E(x)$ indicates the sum net income of various industries. And in crop farming, net income is the product of net output per unit area and total acreages; the expression of net income in animal husbandry can be obtained through the method of best square approximation [11-13]. The relationship between net income and feeding amount can be expressed by the linear weighted sum of logarithmic function $y = a\ln(1+bx)$ and logistic curve $y = \frac{a^*}{1+b^*e^{-cx}}$, where x indicates the feeding amount, and y indicates the net income; net incomes of forestry and aquaculture farming are similar to that of crop farming, which can be approximated as a linear function. Therefore, the objective function can be expressed as follows.

$$\begin{aligned}\max\ E(x) &= 5547.8X_{11} + 1505X_{12} + 1622X_{13} + 2359X_{14} + 1287.2X_{15} + \\ &1248.4X_{16} + 1197.7X_{17} + 1220X_{18} + 1116.3X_{19} + 4078.8X_{110} + 3173.7X_{111} + \\ &8464.8X_{112} + 18176X_{113} + 2665X_{114} + 1472X_{115} + \frac{442850}{1+65.4e^{-0.0121X_{21}}} \times 0.63 + \\ &572404\ln(1+0.0017X_{21}) \times 0.37 + \frac{588940}{1+162.2e^{-0.077X_{22}}} \times 0.61 + 6829556\ln(1+0.0006X_{22}) \times 0.39 \\ &+ \frac{269450}{1+78.9e^{-0.0049X_{23}}} \times 0.89 + 3601713\ln(1+0.0002X_{23}) \times 0.11 + \frac{82100}{1+19.6e^{-0.0071X_{24}}} \times 0.81 + 99598\ln(1+0.0007X_{24}) \times 0.19 \\ &+ \frac{473800}{1+17.6e^{-0.0004X_{25}}} \times 0.92 + 1564320\ln(1+0.000019X_{25}) \times 0.08 + 67X_3 + 4152X_4\end{aligned} \quad . (6)$$

Due to the fundamental status of soil in agricultural production, soil nutrient and moisture have direct impact on crop growth, animal husbandry development and the expansion of aquaculture area, so the content of organic

matter in soil is regarded as an important decision variable. In this research, the difference $\delta(x)$ between the current content and balance content of organic matter in the soil is an important indicator of soil fertility, while influencing factors including the status of soil, water conservation and vegetation cover are regarded as constraints. The organic matter content in the soil is predicted and evaluated through Jenny mathematical model, wherein the topsoil depth is 0.2mm [14-16].

$$\min\ \delta(x)=0.04-\frac{1}{0.0564\times 225\times(\sum_{j=1}^{15}X_{1j}+X_3)}([0.6\times(2.25(X_{21}+X_{22})+1.5X_{23}+0.52X_{24}+0.015X_{25})+3908.8\times 0.108\times 0.5]\times 0.512+[0.3\times(9X_{11}+4.19X_{12}+11.14X_{13}+2.36X_{14}+3.08X_{15}+8.24X_{16})+2.175X_{11}+0.838X_{12}+0.502X_{13}+0.41X_{14}+0.533X_{15}+0.412X_{16}]\times 0.203)\times(1-e^{-0.0564\times 6})-0.0317e^{-0.0564\times 6} \quad (7)$$

$$\min\ D_\delta(x)=([3200-0.9\times(7.5X_{11}+2.99X_{12}+5.57X_{13}+2.05X_{15}+4.12X_{16})+650-0.9\times 2.05X_{14}]+1800-30.38X_{112})\times 0.72+[450-(0.198X_{21}+0.089X_{23}+0.015X_{24}+0.0019X_{25})+1250-3.8X_{22}+160-0.0044X_{25}+75-0.98X_4]\times 0.28 \quad (8)$$

3.4 CONSTRAINTS

1) Natural resource constraints. The constraint of total available land area:

$$\sum_{j=1}^{15}X_{1j}+X_3+X_4\le 3758.7. \quad (9)$$

The constraint of arable land resource:

$$\sum_{j=1}^{15}X_{1j}\le 1180. \quad (10)$$

2) Acreage constraint. Aquaculture area constraint:

$$X_4\ge 40. \quad (11)$$

Economic crop acreage constraint:

$$\sum_{j=1}^{13}X_{1j}\ge 220. \quad (12)$$

Forage crop acreage constraint:

$$X_{114}+X_{115}\ge 30. \quad (13)$$

3) Ecological constraint. Constraint of organic fertilizer amount: the total amount of required fertilizer for various crops and waste fertilizer should not exceed the amount of provided fertilizer.

$$4.875X_{11}+2.792X_{12}+4.137X_{13}+1.822X_{14}+1.936X_{15}+2.725X_{16}+1.439X_{17}+3.12\sum_{j=1}^{15}X_{1j}+8.274X_{112}+2.17X_{113}+1.19X_3\le 2.25(X_{21}+X_{22})+1.5X_{23}+0.52X_{24}+0.015X_{25}+3908.8\times 0.108+9X_{11}+4.19X_{12}+11.14X_{13}+2.26X_{14}+3.08X_{15}+8.24X_{16}+2.175X_{11}+0.838X_{12}+0.502X_{13}+0.41X_{14}+0.533X_{15}+0.412X_{16}. \quad (14)$$

Green coverage constraint:

$$X_3\ge 0.42\times 4546. \quad (15)$$

4) Demand constraint. Grain output constraint:

$$7.50X_{11}+2.99X_{12}+5.57X_{13}+2.05X_{14}+2.05X_{15}+4.12X_{16}\ge 300036. \quad (16)$$

Meat output constraint:

$$0.198X_{21}+0.089X_{23}+0.015X_{24}+0.0063X_{25}\ge 400. \quad (17)$$

Vegetable output constraint:

$$30.38X_{112}\ge 1200. \quad (18)$$

Constraint of required roughage for the development of animal husbandry:

$$(9X_{11}+4.19X_{12}+11.14X_{13}+2.26X_{14}+3.08X_{15}+8.24X_{16})\times 0.23+4.12X_{114}+600\times 1.41\ge 1.51X_{21}+3.42X_{22}+0.054X_{24}. \quad (19)$$

Constraint of concentrated feed demand:

$$(0.128X_{21}+1.065X_{22}+0.19X_{23}+0.022X_{24}+0.015X_{25}+0.432X_4)\times 0.5\le 0.1\times(7.50X_{11}+2.99X_{12}+2.05X_{14}+2.05X_{15}+4.12X_{16}+3.19X_{17})+0.4\times 5.57X_{13}+4.54X_{115}. \quad (20)$$

Minimum quantity of dairy cattle:

$$X_{22}\ge 260. \quad (21)$$

Non-negative constraint of decision variables:

$$X_{ij}\ge 0\ \ (i=1,2,3,4). \quad (22)$$

4 Solutions for these problems

MATLAB software is utilized as research platform for solving the simulation model of agricultural structure in Shangqiu, China, with the results shown in Table 1, Figure 2, Figure 3 and Figure 4.

TABLE 1 "Optimal solution" of agricultural structure optimization

Variable	Optimization value
X_{11}	154.66
X_{14}	295.04
X_{17}	28.617
X_{110}	42.299
X_{113}	26.356
$X2_1$	500
X_{24}	1212.8
X_4	50.609
$D_\delta(X)$	855.71
X_{12}	40.00
X_{15}	8.238
X_{18}	68.483
X_{111}	25.948
X_{22}	320.5
$E(X)$	604.32
X_{13}	198.56
X_{16}	12.972
X_{19}	23.077
X_{121}	57.509
X_{115}	29.452
X_{23}	2898.5
X_3	2378.4
$\delta(X)$	0.0008

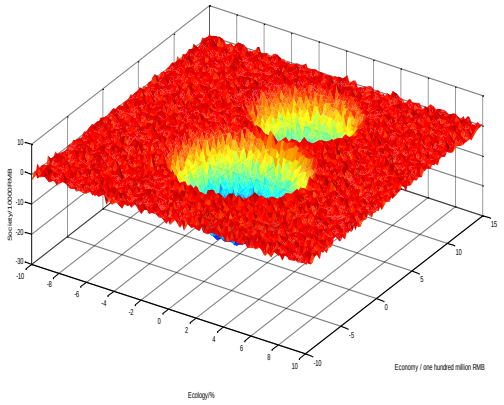


FIGURE 2 Pareto frontier of agricultural structure model

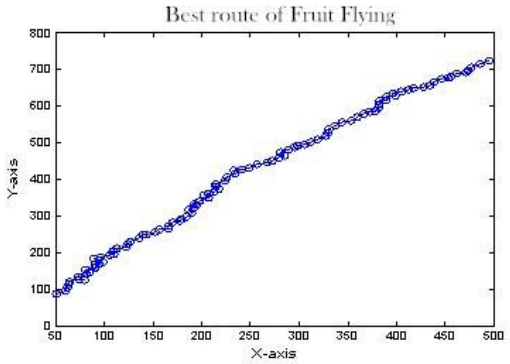


FIGURE 3 Optimization roadmap of fruit fly optimization algorithm

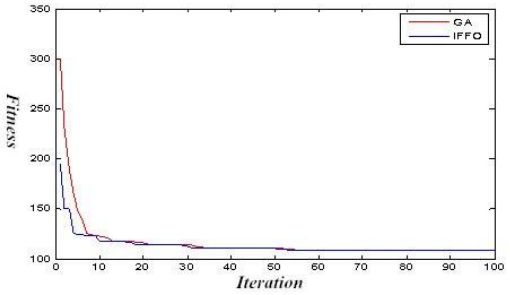


FIGURE 4 Results comparison between fruit fly optimization algorithm (IFFO) and genetic algorithm (GA)

Figure 4 indicates that fruit fly optimization algorithm has a faster convergence and better optimization ability than that of genetic algorithm, which proves the superiority of the proposed algorithm.

5 Conclusions

In this research, based on the coordinating research in three aspects – economy, ecology and society, multi-objective optimization model of agricultural industry structure in Shangqiu is established, and fruit fly optimization algorithm is utilized for solving the multi-objective optimization model, thus optimizing the objective functions in different degrees; meanwhile, multi-index evaluation is carried out on the optimization results, which greatly enhances and improves the coordinated and sustainable development of Shangqiu's agricultural industry structure through fruit fly optimization algorithm.

Researches on industrial structure allocation through fruit fly optimization algorithm have important theoretical and practical significance on agricultural structure adjustment and allocation in Shangqiu, China. And some competitive industries can be developed based on the optimization results, such as increasing the acreage of rice and soybean; economic and forage crops should be mainly developed on the basis of ensuring food security; livestock and dairy products, returning rate of organic fertilizer and soil fertility should be enhanced; besides, local government needs to vigorously develop green food and explore potentials in agricultural industry.

References

[1] Gu X 2003 Multi-objective process system optimization based on fruit fly optimization algorithm *Qingdao University of Science and Technology* (2) 33-6

[2] Wang M 2003 A new multi-objective genetic optimization algorithm and its application *Computing Technology and Automation* (6) 5 -7

[3] Hu T 2009 Neural network method of multi-objective dynamic programming *Journal of Electronics* (10) 70-2

[4] Lin Y 2009 Multi-objective optimization fruit fly optimization algorithm based on fuzzy optimization *Systems Engineering Theory and Practice* (12) 31-6

[5] Wen T 2012 A new fruit fly optimization algorithm: taking the financial distressmodel as an example *Knowledge-Based Systems* (26) 69-74

[6] Li J, Yang A, Dai W 2007 Modeling mechanism of grey neural network and its application *Proceedings of 2007 IEEE International Conference on Grey Systems and Intelligent Services, Nanjin* 404-8

[7] Xiao X, Xiao D, Lin J 2011 Research overview of multi-objective optimization problem. application research of computers **28**(3) 805-8.

[8] Wen T 2012 A new fruit fly optimization algorithm: taking the financial distress-model as an example *Knowledge-Based Systems* (26) 69-74

[9] Shi L 2010 Multi-objective evolutionary optimization algorithm and its application *Nanning: Guangxi Normal University Master's Degree Thesis*

[10] Chun L, Wang Y-P 2008 Dynamic multi-objective optimization evolutionary algorithm based on a new model *Computer Research and Development* **45**(4) 603 -11

[11] Wang Z 2009 Study on model-based uncertainty design optimization *Chengdu Doctoral Dissertation of University of Electronic Science and Technology*

[12] Liu N 2010 Multi-objective algorithm optimization based on evolutionary algorithm and its application *Nanjing University of Aeronautics and Astronautics Master Thesis*

Author

Sujiao Liu, born in Decmber 1977, Huaxian, Henan Province, China

Current position, grades: lecturer of Henan University, China.
University studies: PhD of Henan University, China.
Scientific interest: public administration, industry structure planning.
Publications: 15 papers.

Investigation of the strategic alignment in public sector organisations using knowledge based strategy

Shuwen Ma*

Hangzhou College of Commerce, Zhejiang Gongshang University, 310012, Hangzhou, China

Received 1 May 2014, www.cmnt.lv

Abstract

Ageing workforce is one of the critical challenges a public sector organisation is facing and will face more terribly. The cost of knowledge loss can have huge impact on the bottom line of business. A major issue facing Public Sector Organisations (PSOs) in recent times has been the increasing pressure to demonstrate value from investments in Information Technology (IT). One omnipresent but often overlooked solution is Strategic Alignment (SA) between corporate business objectives and IT initiatives as achieving SA remains one of the more enduring challenges for organisations. Moreover, deriving value from IT investments through SA requires both IT and business executives to foster synergies between their respective areas, which contributes to increased organisational performance. Conversely, business-IT misalignment may eventually lead to failure in achieving business goals. Shared Domain Knowledge (SDK) is a key factor within the social dimension of SA and is concerned with the level of knowledge business and IT have of each other's to each other's missions, objectives, and plans. This paper presents findings from a study which investigated the influence of SDK on SA within organisations in the Australian public sector. The developed research model examined SDK between business and IT professionals as a factor that would potentially influence SA. The findings suggest that increased levels of SDK between professionals from the business and IT domains leads to more efficient SA in PSOs.

Keywords: shared domain knowledge, strategic alignment, public sector, organisational performance

1 Introduction

Ageing workforce is one of the critical challenges a public sector organisation is facing and will face more terribly. The speculation of the mass retirement of the baby boomers generation will leave an organisation with a big challenge of retaining knowledge of experience employees, storing this knowledge and transferring knowledge to the new or existing employees in understandable manner. The cost of knowledge loss can have huge impact on the bottom line of business. A major issue facing public sector organisations (PSOs) is the increasing pressure to demonstrate the value for business from Information Technology (IT) investment which has been more often than not intangible. In order to reap the full benefits of IT, decision-makers are encouraged to integrate IT missions, objectives, and plans into various aspects of business functions within the organisation [1]. In an attempt to boost operational efficiency, organisations have been making continuous efforts to take advantage of IT investments to improve business performance and customer service [2]. Improving the communications, shared domain knowledge, relationship between business and IT executives have been proposed to meet the challenges of value realisation and obtaining maximum benefit from IT investments [3].

A strategic alignment (SA) between business and IT has been long sought by public and private sector organisations for many years [4]. Not only it is hard to achieve, but this type of alignment is also hard to maintain [5]. SA refers to the "congruence between an organisation's business and IS/IT strategies" [6] and is found to have a direct impact on firm performance because of its role as a mediator between IT investments and business objectives [7]. Not only that, but SA is also found to positively influence the effectiveness of IT and contributes to maximising business profits [8]. On the other hand, failure to leverage existing IT investments (i.e., not simply just investing more in IT) is expected to get in the way of organisations performance and viability [9].

Some of the issues that obstruct achieving the social dimension revolve around shared domain knowledge (SDK) between the business and IT domains [10] as SDK is major contributor to the successful creation and execution of strategic plans [11,12]. Consequently, IT managers need to have knowledge about business goals and objectives prior to developing IT strategic plans. For example, it has been observed that CIO's with good business knowledge enjoy greater participation in top management decision-ma king [13], whereas the lack of such knowledge hinders the organisation's ability to employ IT strategically and thus stands in the way of SA [14]. Business executives also need to be aware of IT initiatives so that they are able to objectively make investment decisions, which enables a more realistic expectations of the organisation's IT capabilities [14]. Therefore, exploring effective ways for achieving and sustaining SA in PSOs through SDK remains a challenge requiring more research to address what is still conside-

* ***Corresponding author's*** e-mail: 1622913173@qq.com

red a major concern for executives [15]. While many relevant factors are important in achieving high levels of SA, the focus of the study reported here is solely on understanding the influence of shared domain knowledge on strategic alignment.

2 Related works

Given the emphasis in today's environment on customer focus, stake holder interests, public sector organisational performance and other methods of assessment are employed to address issues in the new public management and prevailing managerialism in measurement of public sector organisations around the world. Therefore, many public sector organisations have been encouraged to implement benchmarking as one way of satisfying the government's requirement that public organisations provide best value services. SA between business and IT is usually described in terms of strategy (i.e., business strategy, IT strategy). In literature, business strategy is considered the driver of an organisation through a set of actions and decision -ma king relating to long-term direction and the scope of activities, addressing changes in the environment to gain advantage over competitors, and building on resources to meet stakeholders' expectations [16,17]. Within the same context, IT strategy often refers to the identification of innovative solutions that provide a competitive edge for businesses and optimise the return of investments [18]. A well-crafted IT strategy is expected to be based on a clear business strategy, and covers technology, information systems, and governance [19].

There are a number of different definitions of SA that describe how business and IT should be integrated, coherent and in sync. SA is defined as "the continuous process, involving management and design subprocesses, of consciously and coherently inter relating all components of the business-IT relationship in order to contribute to the organisation's performance over time." [20] It is no longer considered as an end-state, but rather a continuous and dynamic process that evolves into a relationship between IT and business functions and requires to be actively maintained [21,22]. Further, SA is not an isolated management action as "…no single activity will enable a firm to attain and sustain alignment." [23], but rather a "…continuous and dynamic synchronization of the capabilities inherent in the information infrastructure and the demands of strategy" [24]. SA can be seen as "… applying IT in an appropriate and timely way, in harmony with business strategies, goals and needs…" [25]. In a public sector context, SA is expressed as "the degree to which the IT goals support the strategic goals of a public agency, and to which administration and IT stakeholders are committed to support these goals." [26]

The motivation for SA emerged from a focus on strategic business planning and long-term IT decision-ma king. SA is a major concern for business executives and is ranked among the most important issues faced by IT managers [27]. The most proclaimed advantage in which SA can be beneficial to organisations is the capability to achieve a more focused and strategic use of IT resources leading to improved business and organisational performance [28]. Financial benefits are no longer the only factor used to measure the value added by IT investments. Instead, business executives are now considering non-financial benefits (e.g., the ability to react to new possibilities) as an indication of sound IT investments [29]. Although, the importance of SA has been widely recognised and well documented, many organisations today are still misaligned and this has led to undesirable outcomes including dissatisfied stakeholders, cancelled or failed projects, amplified costs of IT, and systems that do not meet the needs of the business [30]. Thus, the effect of misalignment between business and IT is expected to lead to degraded business performance [31].

A wide range of factors/dimensions that influence SA exist in literature, such as structural, strategic/intellectual, social, and cultural. Although more attention is given to the strategic/intellectual and structural dimensions, SA is also contingent on many of the social and cultural aspects of an organisation [32]. The social dimension is defined as "the state in which business and IT executives within an organisational unit understand and are committed to the business and IT mission, objectives and plans" [33]. The main social dimension factors that have the potential to influence SA are: connections between business and IT planning, communication between business and IT executives, previous imp le mentation of IT plans/projects, Shared Domain Knowledge (SDK), and external influences [11].

SDK is defined as "the knowledge that the IT manager possesses about the business process, the knowledge that the business manager possesses about the potential opportunities to apply IT to improve business process, and the common understanding between the IT and the line manager regarding how IT can be used to improve business process performance." [34] Consequently, two dimensions of SDK have been identified, namely: IT managers' knowledge of business and business executives' knowledge of IT [12].

SDK improves communication [35], increases innovation [36], and enhances IT performance and the use of IT resources [11]. The business and IT executives' knowledge of IT and business strategy results in developing a shared perception and vision [37], and achieves better linkages between objectives and actions [11]. In contrast, business managers' lack of IT knowledge hinders SA [38].

Emerging from the literature thus far, previous research focused on SA as an outcome of integration between business and IT [39], while others have concentrated on the relationship between SA and IT performance [40], or between SDK and IT performance [41]. However, an empirical investigation of the relationship between SDK and SA in PSOs has not been conducted. This paper endeavours to address this limitation by

investigating the influence of SDK on SA within the Australian public sector.

3 The proposed method

Presents, from a systemic perspective, an examination and discussion of performance measurement, performance indicators and associated improvement initiatives, as typically applied in public sector organisations. Such mechanisms are usually implemented as a causal loop which is established between perceived performance and resulting actions, thereby constituting a form of feedback control. Within this context a two dimensional matrix model is postulated in which the independent dimensions are the source of control and the nature of the resultant control action. As PSOs are large consumers of information technology [42], this paper aims at exploring the potential influence of SDK on SA within these organisations. The research methodology, which is exploratory in nature, draws on a survey of business and IT/IS professionals. This paper utilised key informants from the Australian public sector to provide information about PSOs' level of IT and business knowledge, and SA levels. Officers with at least 3-5 years of experience in IT or business and public administration were chosen as key informants. The selection of experienced officers reduces the number of informants required as they tend to be a more reliable source of information; however, this approach provides the potential for common method bias [43]. In addition, this paper endeavoured to control other external and IT characteristic factors which may influence SDK or SA by collecting data from organisations within the same environment/industry. As well as choosing a sample from organisations where IT was regarded of high strategic value, as measured by IT budget and the appointment of a dedicated CIO or senior IT executive.

3.1 RESEARCH MODEL

The model presented below illustrates the conceptual hypotheses and hierarchical relationships of three constructs examined in this paper (i.e., Business objectives knowledge, IT processes knowledge, and SA). The theoretical model depicts the relationships that this paper aims to investigate (i.e., SDK association with SA). The lines connecting constructs symbolise the hypotheses that will be statistically tested. Each of the constructs correlated to items on the data collection instrument (i.e., questionnaire). The model has been designed to examine two hypotheses; business objectives knowledge correlates with the SA (H1), IT processes knowledge correlates with the SA (H2).

3.2 HYPOTHESIS DEVELOPMENT

The available literature contains definitions and measurable indicators (or variables) of the constructs but the operation allocation of these indicators was not specific to the SDK-SA relationship examined in this paper. Rather than adopt these broader indicators, this study used field-driven measures based on existing theory. In total, 22 items were used to measure three constructs. The measurable indicators of each research construct are discussed below.

The first construct, business objectives knowledge, was conceptualised as the aggregate of two variables: experience in public administration (V1) and familiarity with business objectives within the surveyed organisation (V2). Both variables were perceived as important to the ability to participate in strategic decisions within PSOs. A prerequisite of 3-5 years of experience was established for the first variable (V1), whereas the second variable (V2) was assessed on a five point Likert-type scale. The second construct, IT processes knowledge, was operationalised using two variables: experience in IT governance (V3) and familiarity with IT processes within the surveyed organisation (V4). Potentially, both of these variables would lead to identifying opportunities for utilising IT to support business goals. A prerequisite of 3-5 years of experience in IT governance was established for the first variable (V3), whereas the second variable (V4) was assessed on a five point Likert-type scale. As two variables (i.e., V1 and V3) were established prior to data collection and therefore did not require measuring, only two variables (i.e., V2 and V4) were measured using the data collecting instrument to indicate the level of business objectives and IT processes knowledge of each respondent. The third construct, strategic alignment, has been frequently measured using financial and perceptual indicators in previous research (e.g., Kearns and Sabherwal [12]; Reich and Benbasat [35]; and Winkler [26]). Perceptual indicators were selected for this study through utilising high-level processes from the COBIT 5 frame work due to its worldwide application (e.g., private and public industries, governments, accounting and audit firms) as a reliable source for assessing IT governance and business/IT alignment [44]. The latest version, COBIT 5, divides the governance of IT into five domains: Evaluate, Direct and Monitor (EDM) ; Align, Plan and Organise (APO); Build, Acquire and Implement (BAI); Deliver, Service and Support (DSS); and Monitor, Evaluate and Assess (MEA), which are broken into 37 high-level processes [45]. As a result, and judging by the nature of each of the COBIT 5 domains, the EDM (V5-V9) and APO (V10-V22) do ma ins were used as measurable indicators of strategic alignment.

3.3 DATA SETS

Data were drawn from a cross-sectional field study from the public sector population. A pilot test of the questionnaire was administered to five thought leaders from the Queensland public sector with at least ten years experience in the field. Based on their feedback, no further amendments were required to the developed instrument. The survey included participants drawn from

three different representative groups to limit any sample frame bias. Moreover, to reduce the possibility of single-source bias that might result from exaggeration or self-pro motion and to encourage participation, the respondents were assured that the results would be completely anonymous and that they would receive a summary of the study findings.

From a total of 112 invitations, 56 usable responses were available which provided an unadjusted response rate of 50%. Analysis of nonresponse bias was performed by comparing early and late responses [46]. Early respondents are those who respond to the initial email, while late respondents are those who respond only after a reminder is sent. T-tests of the mean differences for each of the constructs and number of respondents failed to demonstrate any significant differences ($p<0.05$, two-tailed), suggesting that nonresponse bias was not an issue in this study.

4 Experiments

Specific statistical approaches were used to assess the model and establish the results. The primary method of assessing the correlation between the knowledge areas was Spearman's Rank Correlation Coefficient. The tables below (see Tables 1-2) detail some descriptive statistics that characterise the data collected. Analysis of descriptive statistics was undertaken to search for possible affects or bias resulting from certain patterns in the sample data. The results of the descriptive statistics for constructs are displayed in Table 2 and show no sign of misrepresentation as all responses seem to follow an approximately normal distribution (i.e. all data-points fall within two standard deviations).

TABLE 1 Role distribution

Role	Count
Executive/Manager	26
Junior/Operational	30
Total	56

TABLE 2 Descriptive statistics

Construct	Total	Average	StdDev
Business objective knowledge	222	4	0.94
IT Processes knowledge	216	4	0.92
Strategic alignment	177	5	1.3

For internal consistency analyses of the Likert-type scales, Cronbach's Alpha reliability coefficient was used as it relates to the measurement of the internal consistency and homogeneity of items in a scale [47]. The results were significant for all items in the questionnaire at Cronbach's alpha >0.7. Construct validity was measured using the Kaiser Meyer Olkin (KMO) and Barlett tests. The KMO test assesses the adequacy of the sample magnitude for factor analysis, while the Barlett test is used to determine whether the data come from multivariate and normal distribution. The KMO's value ranges between 0 to 1 and is expected to be over 0.60 while 0.80-0.90 is considered a very good range [48]. For this study, the KMO value was 0.89 indicating sampling adequacy, and the Bartlett's test result was significant at 1256.7 ($P<0.05$). Other forms of data validation (i.e., confirmatory factor analysis) for convergent and divergent validity could not be performed on such low sample size as the minimum sample recommended for conducting such analysis should be at least 100 [49]. Descriptive statistics for variables (V5–V22) is expressed.

Variable correlation was conducted using the Spearman's rank correlation coefficient, which is a non-parametric correlation suited for small sample sizes where a normal distribution is difficult to be assumed [48]. The Spearman's Rank Correlations values vary from -1 to 1, where a high positive value denotes a stronger relationship between the variables and vice versa. Based on the p-value no null hypotheses can be rejected, therefore no results are significant. Hypothesis one has moderately a stronger positive correlation based on the Spearman's rank ($R=0.21$) whereas hypothesis two has a weaker positive correlation ($R=0.19$).

Despite the low significance of results an attempt will be made to draw implications of the relationship between SDK and SA from the Spearman's rank correlation. The results indicate that the relationship between SDK and SA is not as clearly defined as previous research indicated. Although the initial expectation was that there is a strong direct relation between SDK and SA in PSOs, the results suggest that the level of SA may depend on other factors as well, perhaps more than SDK. For instance, knowledge of IT Governance was found to improve the perceived level of SA.

Looking back at the theoretical frame work, results of the construct measurements (i.e., business objectives knowledge, IT processes knowledge, and SA) suggest that mutual understanding between the business and IT domains may create imp roved SA as both correlations of SDK (i.e., business objectives knowledge and IT processes knowledge) were relatively positive.

Based on the number of items where the average score was greater than the mean score, we note that participants with senior roles placed higher emphasis on stakeholder management, while operational areas (e.g., security) were perceived less important. Similarly, participants with junior roles perceived risk management as more important than benefits delivery and managing innovation. This supports the assumption that junior roles have a great understanding of the use of IT processes and risk management in PSOs while executives have a better understanding of enabling business objectives to meet stakeholders' needs.

FIGURE 1 SDK and SA breakdown based on role

Furthermore, the business objectives knowledge rating for senior and executive participants was higher than their rating of IT processes knowledge as illustrated in Figure 1. In contrast, IT processes knowledge rating for junior participants was higher than their rating of business objectives knowledge.

Moreover, IT knowledgeable business executives and business savvy operational/IT staff are expected to optimise SA in PSOs. Presumably, regular communication is intended to bridge this identified knowledge gap between executives and operational staff. The results also demonstrated that the rating of SDK has a positive relation with the rating of SA. For instance, the SDK rating for junior staff was higher than the one for executives, and as a result, the SA rating for junior staff was also higher. Thus, staff at all levels should invest in identifying capabilities for business and IT to interconnect, which consequently improves SDK and SA. Academics on the other hand are urged to extend the understanding of SDK by testing the individual constructs identified in this paper using a larger sample sizes to refine the exploratory results; explore the applicability of results on non-public sector organisations; and explore alternative measurable indicators/variables for SA and SDK.

5 Conclusions

This research contributes to an overall conceptual understanding of SDK as a factor fo r build ing organisational and Information Systems performance. SDK is concerned with reaching a state by which IT managers grow into business savvies and business executives develop IT governance/processes knowledge. As PSOs are forced to operate in an environment where demands grow and resources are reduced, improving business-IT synergies through SDK becomes critical. Based on the results, it appears as though, achieving high levels of SDK within PSOs result in optimising SA between business – IT. Consequently, in order to improve SA, IT and business staff at different levels should have adequate knowledge about business objectives and IT processes. In order to create an environment for SDK to develop, actions such as physically moving IT staff into business units, planning shared workshops and brainstorming sessions, and sending IT staff on regular visits to frontline offices may be required. Other methods may include rotating business managers through IT roles to reinforce the message that IT is an integral part of the business. In line with previous research [e.g.,50] external factors (e.g., IT processes/governance) have less effect on SA than internal factors (e.g., business objectives/governance). Thus, a possible direction for this research stream is to evaluate if SDK is only necessary for internal evaluations. Research points to the conclusion that SA is contingent on the existence of SDK between the business and IT domains. Improving SDK within PSOs like any other core competency takes time to develop, therefore; IT managers and business executives need a clear roadmap to build and maintain these capabilities.

This study suggests several implications for future research. First, this paper supports the critical role of SA in deriving value from IT investments and highlights the mediating role of SDK in PSOs. Future research could examine if SDK indirectly improves other aspects of IT. Second, future studies could include a more comprehensive conceptualisation and operationalisation of SA that better reflects its multi-dimensional nature. It is suggested that further research investigate whether the results are reproduced using other methods and in other contexts and industries. Moreover, it is important to study other factors (e.g., changing political environments) that might facilitate or inhibit SDK and subsequently influence SA.

References

[1] Vogt M, Hales K 2010 Strategic alignment of ICT projects with community values in local government *43rd Hawaii International Conference on System Sciences (HICSS) 142445509X* 1-10
[2] Bocij P, Hickie S 2008 Business information systems: technology, development and management for the e-business *Prentice Hall*
[3] Mohdzain M B, Ward J.M 2007 A study of subsidiaries' views of information systems strategic planning in multinational organisations *The Journal of Strategic Information Systems* **16**(4) 324-52
[4] Silvius A G, De Waal B M, Smit J 2009 Business and IT alignment; answers and remaining questions *PACIS 2009 Proceedings* 16
[5] Avison D 2004 Using and validating the strategic alignment model *The Journal of Strategic Information Systems* **13**(3) 223-46
[6] Preston D, Karahanna E 2009 How to develop a shared vision: the key to IS strategic alignment *MIS Quarterly Executive* **8**(1) 1-8
[7] Byrd AT, Lewis B R, Bryan R W 2006 The leveraging influence of strategic alignment on IT investment: an empirical examination," information & management **43**(3) 308-21
[8] Skerlavaj M 2011 The organisational learning culture and organisational performance in macedonian companies *European Journal of International Management* **5**(6) 574-607
[9] Noce I 2011 Strategic alignment through organizational modeling: a case study in a public institution *in ENTERprise Information Systems M Cru z-Cunha, et al., Editors Springer Berlin Heidelberg* 129-38
[10] Baker J, Jones D 2008 A theoretical frame work for sustained strategic alignment and an agenda for research *All Sprouts Content* **8**(16)
[11] Reich B H, Benbasat I 2000 Factors that influence the social dimension of alignment between business and information technology objectives *Management Information Systems Quarterly* **24**(1) 81-114

[12] Kearns G S, Sabherwal R 2007 Strategic alignment between business and information technology: a knowledge-based view of behaviors, outcome, and consequences *Journal of Management Information Systems* **23**(3) 129-62
[13] Lv R-j 2013 Shared knowledge and shared understanding between cio and top management team: a literature review, in the 19th international conference on industrial engineering and engineering management *E. Qi, J. Shen, R. Dou, Editors. Springer: Berlin Heidelberg* 759-66
[14] Teo T S, Ang J S 1999 Critical success factors in the alignment of IS plans with business plans *International Journal of Information Management* **19**(2) 173-85
[15] Luftman J, Kempaiah R, Nash E 2006 Key issues for IT executives 2005 *MIS Quarterly Executive***5**(2) 81-101
[16] Johnson G, Scholes K, Whittington R 2005 Exploring corporate strategy: Tet & cases *Prentice Hall*
[17] Huang D, Hu Q 2007 Achieving IT-business strategic alignment via enterprise-wide implementation of balanced scorecards *Information Systems Management* **24**(2) 173-84
[18] Buchta D, Eul M, Schulte-Croonenberg H 2009 Strategic IT-management: increase value, control performance, reduce costs *GablerVerlag*
[19] Willcocks L, Petherbridge P, Olson N 2002 Making IT count *Butterworth-Heinemann*
[20] Maes R 2000 Redefining business-IT alignment through a unified framework *in PrimaVera Working Paper Universiteitvan Amsterdam* 1-25
[21] Chan Y E, Reich B H 2007 IT Alignment: what have we learned? *Journal of Information Technology* **22**(4) 297-315
[22] Luftman J 2000 Assessing business-IT alignment maturity *Communications of the Association for Information Systems* **4**(14) 1-51
[23] Luftman J 2003 Assessing IT/Business alignment *Information Systems Management* **20**(4) 9-15
[24] Prahalad C, Krishnan M the dynamic synchronization of strategy and information technology *Sloan Management Review, available at: http://sloanreview.mit.edu/article/the-dynamic-synchronization-of-strategy-and-informat ion-technology/, retrieved on 03 June*
[25] Luftman J, Brier T 1999 Achieving and sustaining business-IT alignment *California Management Review* **42**(1) 109-22
[26] Winkler T J 2013 IT governance mechanisms and administration/it alignment in the public sector: a conceptual model and case validation *11th International Conference on Wirtschafts informatik* 831-45
[27] Nadali A 2011 Maturity assessment of business/IT *in Digital Enterprise and Information Systems E. Ariwa, E. El-Qawasmeh Editors Springer Berlin* 724-38
[28] Chan Y E, Sabherwal R, Thatcher J B *IEEE Transactions on Engineering Management* **53**(1) 27-47
[29] Tarafdar M, Qrunfleh S 2009 IT-business alignment: a two-level analysis *Information Systems Management* **26**(4) 338-49
[30] Chen L 2010 Business–IT alignment maturity of companies in China *Information & Management* **47**(1) 9-16
[31] Singh S N, Woo C 2009 Investigating business-IT alignment through multi-disciplinary goal concepts *Requirements Engineering* **14**(3) 177-207
[32] Chan Y E 2001 Information systems strategy, structure and alignment in strategic information technology: opportunities for competitive advantage *R. Papp Editor Idea Group Publishing Hershey PA* 56-81
[33] Chan Y E Why Haven't We Mastered Alignment? The Importance of the Informal Organization Structure *MIS Quarterly Executive* **1**(2) 97-112
[34] Ray G, Barney J B, Muhanna WA 2004 Capabilities, business processes, and competitive advantage: choosing the dependent variable in empirical tests of the resource-based view *Strategic Management Journal* **25**(1) 23-37
[35] Reich B H, Benbasat I 2003 Measuring the information systems–business strategy relationship, in strategic information management: challenges and strategies in managing information systems *R.D. Galliers and D.E. Leidner, Editors Butterworth-Heinemann: Oxford UK* 265-310
[36] Jorfi S, Nor K M, Najjar L 2011 Assessing the impact of IT connectivity and IT capability on IT-business strategic alignment: an empirical study *Computer and Information Science* **4**(3) 76-87
[37] Enns H G, Huff S L, Golden BR How CIOs Obtain Peer Commitment to Strategic IS Proposals: Barriers and Facilitators *The Journal of Strategic Information Systems* **10**(1) 3-14
[38] Liu Q, Zhang L, Ni M 2010 *IEEE Transactions on Parallel and Distributed Systems* **21**(3) 405-16
[39] Menzel M, Ranjan R, Wang L, Khan S, Chen J 2014 CloudGenius: a hybrid decision support method for automating the migration of web application clusters to public clouds *IEEE Transactions on Computers, in press*
[40] Hao F, Min G, Chen J, Wang F, Lin M, Luo C, Yang L T 2014 An optimized computational model for task-oriented multi-community-cloud social collaboration *IEEE Transactions on Services Computing, in press*
[41] Qi L, Dou W, Chen J 2014 Weighted principal component analysis-based service selection method for multimedia services in cloud computing *Computing, Springer in pres*.
[42] Wang L, Tao J 2013 G-Hadoop: MapReduce across distributed data centers for data-intensive computing *Future Generation Computer Systems* **29**(3) 739-50
[43] Xu Z 2014 Knowle: a semantic link network based system for organizing large scale online news events *Future Generation Computer Systems 10.1016/j.future.2014.04.002*
[44] Xu Z, Luo X, Zhang S, Wei X, Mei L, Hu C 2013 Mining temporal explicit and implicit semantic relations between entities using web search engines *Future Generation Computer Systems. DOI:10.1016/J.future.2013.9.027*
[45] Liu Y, Ni L M, Hu C 2012 *IEEE Journal on Selected Areas in Communications* **30**(9) 1780-8
[46] Luo X, Xu Z, Yu J, Chen X 2011 *IEEE transactions on automation science and engineering* **8**(3) 482-94
[47] Hu C, Xu Z 2014 Semantic link network based model for organizing multimedia big data *IEEE Transactions on Emerging Topics in Computing, 10.1109/TETC.2014.2316525*.
[48] Liu Y, Zhu Y, Ni L M, Xue G. 2011 *IEEE Transactions on Parallel and Distributed Systems* **22**(12) 2100-7

Author

Shuwen Ma, 1968.02, Jilin, China.

Current position, grades: an associate professor in Hangzhou College of Commerce, Zhejiang Gongshang University.
Scientific interest: information management and innovation management
Publications: 20 papers.

Evaluation of e-commerce website based on fruit fly algorithm optimization RBF algorithm

Gang Lu*

[1]*School of Management Science and Engineering, Shijiazhuang University of Economics, Shijiazhuang 050031, China*

Received 1 May 2014, www.cmnt.lv

Abstract

The feature and various index properties of E-commerce website are considered as a whole by applying the expert grading method, with the construction of the multi-index hierarchical structure of an E-commerce website competitiveness index evaluation and the establishment of an E-commerce website competiveness index evaluation index system. The competiveness level of the website is quantified after calculating the competiveness index of the E-commerce website. On this basis, this work adopted Radial Basis Function (RBF) neural network algorithm to perform evaluation research on the competiveness index of E-commerce website. Aiming at the problems exist in the evaluation research, this work tried to use Fruit Fly Optimization Algorithm (FOA) to perform improvement on the RBF neural network algorithm. Through the simulation and comparison of practical examples, FOA-RBF algorithm is obviously better than RBF neural network algorithm when the E-commerce website competiveness index is calculated and evaluated, thus the validity and reliability of calculating method presented in this work are verified.

Keywords: e-commerce, neural network, radial basis function, fruit fly optimization algorithm, expert grading method

1 Introduction

In recent years E-commerce has obtained fast development, and competition is also becoming increasingly fierce. For E-commerce website, its own competitive power is sufficiently evaluated and understood, which is favourable to strengthen website construction and improve the quality of the website; the premise for enterprises to improve competiveness level and obtain competitive advantage; a problem [1] which needs to be solved urgently.

2 Evaluation research on the competitiveness index of E-commerce website

2.1 TYPES OF E-COMMERCE WEBSITE

The E-commerce website can be divided into different types [2] by different classification methods.

1) Classified according to commercial purpose and business function

- Basic commerce website.

For the basic commerce website, the company publicity and customer service are performed by means of basic measures in network media and E-commerce. This kind of website is appropriate for small enterprises with weak power of professionals and in need of E-commerce service. Its characteristics are cheap price of website construction, and high performance cost ratio.

- Publicity commerce website.

Website can be served as the important window of enterprises' public relationship, which is used to publicize the latest news and the business condition of the enterprises. This kind of website is mainly served for some listed companies at home and abroad, on the official websites of which the columns of company news and introduction of the investors are established, becoming the official channel for enterprises to publish news and source rule. The publicity commerce website is appropriate for all kinds of enterprises, especially for the foreign trade enterprises.

- Customer-oriented service website.

This kind of website can provide the inquiry of the after-sale service and the dynamic service status, satisfying the customer demands at a higher level.

- Complete E-commerce operation website.

Complete E-commerce operation website refers to various aspects of E-commerce, such as the distribution management, online shopping and online recruitment. This kind of website can be more clearly described as a set of business management system software.

2) Classified according to the principal part of constructing a website

- Industry E-commerce website.

Industry E-commerce website means a large-scale E-commerce website constructed with the industry organization as the principal part, providing a platform of information distribution, commodity order and customer communication for enterprises and departments in the Industry to perform E-commerce.

- Enterprise E-commerce.

Enterprise E-commerce means that the website is constructed with the enterprises as the principal part to implement E-commerce activities, and can be further divided

* ***Corresponding author's*** e-mail: sjzlugang@126.com

into variously different types of websites according to the leading products of enterprises and the different services provided.

- Government E-commerce website.

Government E-commerce website means the website is constructed with the government body as the principal part to implement E-commerce activities.

- Service organization E-commerce website.

Service organization E-commerce website means the website is constructed with the service organization as the principal part to implement E-commerce activities.

3) Classified according to the responsibilities of the website owners

- Production commerce website.
- Circulation commerce website.

4) Classified according to E-commerce model.

- B2B commerce website.
- B2C commerce website.
- C2C commerce website.
- G2C commerce website.

2.2 COMPETIVENESS INDEX OF E-COMMERCE WEBSITE

Whatever kinds of E-commerce websites, the dominant factors affecting its operation effects and competitive power, are basically the same. The overall level of the E-commerce website is decided by the combined effect of these factors, producing great influence on the function realization of the website and making the E-commerce website present different competitive power. Therefore, the competiveness index of E-commerce website is proposed to measure E-commerce website in competition, reflecting the level of the website in design operation [3].

Aiming at the development status and characteristics of E-commerce website, this work performed research on the competiveness index of E-commerce website according to the principles of comprehensiveness, scientific, operability, industry representativeness and taking content as the core. On basis of performing in-depth investigation and research on the currently famous E-commerce websites as well as learning the research achievements of other scholars and some organizations [4], the more optimized algorithm is used to simulate and calculate the competiveness index of an E-commerce website.

2.3 INDEX SYSTEM OF E-COMMERCE WEBSITE COMPETIVENESS INDEX MEASUREMENT

Through relatively specific literature research, the evaluation index system of E-commerce website competiveness index is established by adopting comprehensive evaluation method during the process of performing measurement on E-commerce website competiveness index. Twenty representative secondary indexes are finally selected as the evaluation object after performing investtigation analysis and literature learning on lots of E-commerce websites. The finally-designed index system includes the first class index and the secondary index, therein to, the first class index includes five indexes - website content, user service, usability, website technology and website function, and the five first class indexes are subdivided on this basis to constitute the secondary evaluation index [5].

3 RBF neural network algorithm

3.1 CONCEPT OF RBF NEURAL NETWORK ALGORITHM

Artificial neural network is applied successfully in many fields by means of its unique information processing capacity. It not only possesses powerful nonlinear mapping capacity, but also obtains the characteristics of self-adaption, self-learning and fault tolerance, clustering and learning from lots of historical data so as to find the change law of some behaviour. Radial basis function (RBF) is a kind of fresh and valid forward feedback neural network [6], with the property of the best approximation and global optimum. Meanwhile, the training method is quickly and easily to be implemented, and there is no problem of local optimum. These advantages make RBF network broadly applied in nonlinear time series prediction. In 1985 Powell presented the Radial-Basis Function (RBF) method of multivariate interpolation [7]. In 1988 Broomhead and Lowe firstly applied RBF to neural website design, constituting the RBF neural network [8].

3.2 COMPOSITION OF RBF NEURAL NETWORK

The most basic composition of RBF neural network includes three layers – input layer, hidden layer (middle layer) and output layer. Therein to, the input layer is composed of some source points (perception units), which connect the network with the external environment, with the transmit function of data information and no transformation on the input information; the kernel function (or action function) of the hidden layer neuron is adopted as the RBF, which performs nonlinear transformation between the input information and the space of the hidden layer, and the higher dimensionality is usually obtained; the output layer is linear provides response for the activation pattern of the input layer.

Setting the neurons of the hidden layer and the output layer as M and Q, respectively, the input pattern is noted as X, and $X=[x_1,x_2,\cdots,x_R]^T$; the output pattern is noted as Y, and $Y=[y_1,y_2,\cdots,y_Q]^T$. The RBF in this work is Gauss function, so the output of the hidden unit is [9]:

$$z_j=\exp\left(-\left\|\frac{X-C_j}{\sigma_j}\right\|^2\right),\ j=1,2,\cdots,M\ , \tag{1}$$

where z_j is the output value of the j neuron in hidden layer; C_j is the center of the j neuron of the hidden layer, composed by the j neuron of the hidden layer corresponding to the center component of all neurons in the input layer, and $C_j = \left[c_{j1}, c_{j2}, \cdots, c_{jR}\right]^T$; σ_j is the width of the j neuron of the hidden layer, corresponding to C_j; $\|\bullet\|$ is Euclid Norm.

The in-out relational expression of the neurons in output layer is:

$$y_k = \sum_{j=1}^{M} w_{kj} z_j \qquad k = 1, 2, \cdots, Q, \tag{2}$$

where y_k is the output value of the j neuron of the output layer; w_{kj} is the weight between the k neuron of the output layer and the j neuron of the hidden layer.

The parameter of the RBF neuron network hereby mainly means the center, width of the network and the adjust weight [10].

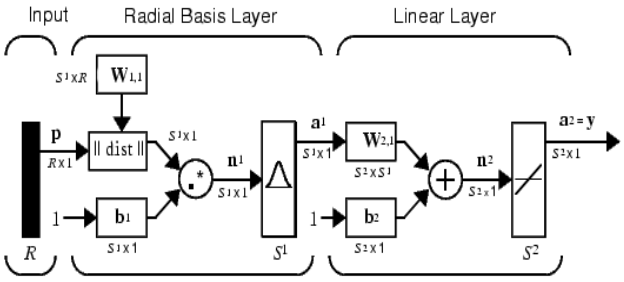


FIGURE 1 RBF neuron network structure

4 Competitiveness index evaluation of E-commerce website based on RBF neuron network algorithm

4.1 ANALYTIC HIERARCHY PROCESS

Analytic Hierarchy Process (AHP) is a kind of hierarchical and structuralized decision analysis method of qualitative and quantitative combination, proposed by American operational research expert T.L.Saaty in 1970s [11]. This method is the most representative comprehensive evaluation method where the factor aggregation is hierarchical, the hierarchical structure model composed, and then the analysis performed by layers according to the interrelationship between factors on the basis of the nature of problems and the component elements of the goal decomposition problems; the weight of various factors is established according to the comparison between factors, appropriate for the decision problems with multiple targets and attribution. It's concise and effective, with extensive practicability.

4.2 DETERMINATION OF RBF NEURAL NETWORK INPUT AND OUTPUT

The method of RBF neural network is adopted to perform measurement and evaluation on the competitiveness index of E-commerce website, and the determination of target output is critical for RBF neural network. Because RBF neural network belongs to supervised learning, a group of expected output data is required as the reference for learning. Therefore, various indexes of E-commerce website competitiveness index should be scored if the target output would be determined.

Expert grading method is a common method, analytic hierarchy process is applied in this work to calculate the weight of various indexes by means of expert grading method, and then the evaluation index system of E-commerce website competitiveness index is used to perform measurement and evaluation on the E-commerce website competitiveness index, evaluating the quality and competitiveness level of E-commerce website [12].

The method of combining expert interview with questionnaire survey is adopted in expert grading method, which can effectively guarantee the objectivity and validity of the obtained data.

The marking rules are as follows: the 20 secondary indexes in the evaluation index system of E-commerce service website are served as the RBF neural network input to perform evaluation on the website (Table 1). The index of each website is scored by experts, and then the quantitative analysis is performed on the results to quantize each evaluation index [13].

TABLE 1 Score list of website evaluation index system

The first class index	The secondary index	Better (score)	Medium (score)	Worse (score)	Scoring methods
A1 Website content	B1 Timeliness	6	4	2	Monitoring method
	B2 Accuracy	6	4	2	Investigation method
	B3 Professional	5	3	1	Investigation method
	B4 Authority	5	3	1	Investigation method
A2 User service	B5 Personalized service	6	4	2	Investigation method
	B6 Protection of user privacy	5	3	1	Investigation method
	B7 Credit monitoring of transaction norm	4	2	0	Investigation method

	B8 Expert technical support	4	2	0	Investigation method
	B9 Website customer service	5	3	1	Investigation method
	B10 User interaction	4	2	0	Investigation method
A3 Usability	B11 Site map	5	3	1	Investigation method
	B12 Website design	5	3	1	Investigation method
	B13 Classification of website information	5	3	1	Investigation method
A4 Web technology	B14 Search function	5	3	1	Investigation method
	B15 Validity of links	5	3	1	Test method
	B16 Browser compatibility	5	3	1	Test method
	B17Website security	5	3	1	Test method
A5 Website operation	B18 Website traffic	5	3	1	Monitoring method
	B19 Quantity of the linked	5	3	1	Monitoring method
	B20 User number of the website	5	3	1	Monitoring method
	Total	100	60	20	

4.3 APPLICATION OF ANALYTIC HIERARCHY PROCESS IN INDEX SYSTEM

What we should urgently solve and determine is that the "contribution degree"---weight of various secondary indexes in the website competitiveness evaluation after the evaluated first class index and various secondary indexes are determined. The analytical hierarch process can be adopted to perform index weight evaluation on the first class index and secondary index in order to determine the weight. And the specific operation steps are that the user survey and data analysis should be firstly performed, the judgment matrix of different levels of indexes are established after the indexes at the same layer are compared pairwise (the comparison of importance degree and non-importance degree), the weight of various indexes and their characteristic root can be obtained after calculation by means of AHP method, the consistency test should be performed on the above results, and the satisfying consistency can be recognized until CR<0.10. The judgment matrix at each layer of index is established in this work according to the steps of analytic hierarchy process, the relative weight value of each layer index could be obtainned by calculation, and then the weight of various indexes of E-commerce website competitiveness evaluation index are obtained by consistency test, thus a complete evaluation index system [14] which is appropriate for E-commerce website competitiveness is finally constructed.

4.4 DETERMINATION STEPS OF EVALUATION INDEX SYSTEM WEIGHT

4.4.1 Establishment of judgment matrix

Generally speaking, users express different attention on the characteristics at various aspects of the website, someone pay more attention to the content elements of the website, and someone place extra emphasis on the obtained service and feelings. For this purpose, the result of expert grading is applied to establish the judgment matrix. The specific operation are as follows: aiming at each index (including the first class index and secondary index), five importance grades are established, which respectively are the most important, very important, important, more important and generally important. The score for the most important is $K_1 = 5$, followed by $K_2 = 4$, $K_2 = 4$, $K_4 = 2$ and $K_4 = 2$. Thus the calculation Equation of the importance score for each index is [15]:

$$M = \sum_{n=1}^{5} P_n g K_n ,$$

where P_1, P_2, P_3, P_4 and P_5 respectively represent the ratio in the overall number of people, aiming at each index and the people who choose different importance grades.

4.4.2 The judgment matrix is determined by scores

Calculation steps are as follows:

1) Calculating the product of the elements in each row of judgment matrix, the Equation is:

$$L_i = \prod_{j=1}^{n} a_{ij} \qquad (i = 1, 2, ..., n)$$

2) Calculating L_i to the N[th] root value in each row, the j is:

$$K_i = \sqrt[n]{L_i} .$$

3) The vector $K = (K_1, K_2, K_3, K_4, K_5)^T$ is performed by normalization processing, $w = K_i \Big/ \sum_{\wedge j=1}^{n} K_j$ (w is the index weight at each level).

Accordingly, the weight of each secondary index in the affiliated level can be calculated. Table 2 shows the calculation result.

TABLE2 Weight of secondary index in the affiliated level

Secondary index	*w*	Secondary index	*w*	Secondary index	*w*
B1	0.4	B8	0.107	B15	0.177
B2	0.4	B9	0.055	B16	0.315
B3	0.2	B10	0.328	B17	0.061
B4	0.247	B11	0.150	B18	0.168
B5	0.113	B12	0.068	B19	0.397
B6	0.247	B13	0.454	B20	0.051
B7	0.231	B14	0.447	-	-

The website with higher scores indicates that the comprehensive quality of the website is better. It's divided into 5 grades which are excellent, good, better, worse and very bad. The score explanation at each grade refers to Table 3.

TABLE 3 Score explanation of website evaluation grades

Grade	Score
Excellent	0.9score and above
Good	0. 75-0. 89 score
Better	0. 65-0. 74 score
Worse	0. 55-0. 64 score
Very bad	0.54score and below

4.5 RBF NEURAL NETWORK ALGORITHM ADOPTED TO CALCULATE THE COMPETITIVENESS INDEX OF E-COMMERCE WEBSITE

The order of adopting RBF neural network to perform evaluation on the website is: firstly, the data of evaluation index system of E-commerce competitiveness index which have been established serve as the input sample, and a RBF neural network model is established to performing training on internet; secondly, test is performed on the established evaluation network model according to the results of training and learning; finally, conclusion and analysis should be performed on the results of website learning and training. From March to June in 2013, the competitiveness index of 100 designated E-commerce websites is scored by means of expert grading method, and the uniform treatment is performed on the obtained data. The front 70 groups of data are used to test, and test is performed on the rear 30 groups of data. The parameter of the RBF neural network is set as: the error goal=0.0001, spread=1, the maximum quantity of the neurons at the hidden layer is mn=20, and the displayed frequency in the training process df=1. The training results are shown in Figure 1:

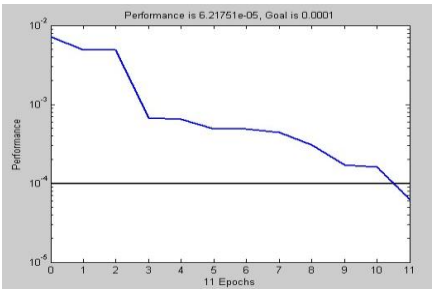


FIGURE 1 Training process figure of RBF neural network

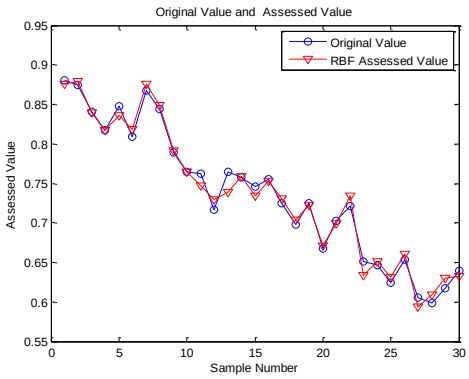


FIGURE 2 Comparison figure of the RBF predicted result and the original value

It can be found that the E-commerce evaluation result of applying RBF neural network is better, and the evaluated absolute error figure and relative error figure are shown in Figure 3 and Figure 4, respectively.

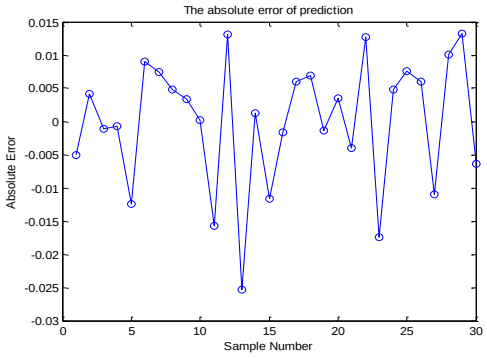


FIGURE 3 Absolute error figure predicted by RBF

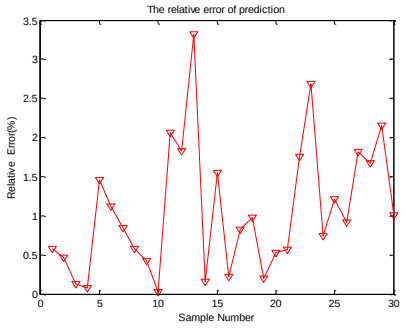


FIGURE 4 Relative error figure predicted by RBF

Figure 4 shows the average value of relative error in E-commerce website evaluation is at 2.5% by adopting

RBF neural network, with the better effect; but the problems of long time evaluation and excessive error in some evaluation points exist

5 Fruit fly algorithm optimization RBF neural network model

5.1 BRIEF INTRODUCTION OF FRUIT FLY OPTIMIZATION ALGORITHM

Fruit Fly Optimization Algorithm (FOA) was a kind of brand new evolutionary computation method presented by Taiwan young teacher Pan Wenchao [5-8] in 2011. Because the fruit fly is superior in the sense of smell and sight, which is shown in Figure 5. Fruit fly searches the food source in air; its keen sense of sight is used to find food and the position where its companions gather after finding food, and finally fly to the position. Therefore, this method considers that fruit fly searches food, it firstly searches the approximate position of the food by smell and then confirms the correct position of the food by sight. This method is a new method of seeking global optimization based on the foraging behavior of fruit fly.

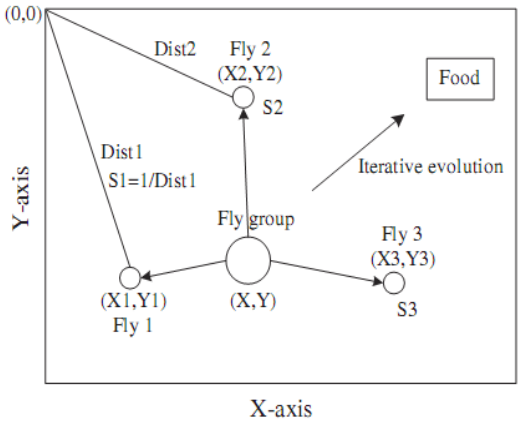


FIGURE 5 Iteration food search figure of fruit fly group

5.2 STEPS OF FRUIT FLY OPTIMIZATION ALGORITHM

The fruit fly optimization algorithm can be divided into seven steps as follows [9]:

1) Figure 5 shows the initialization of the fruit fly group position, and the result of initialization is Init X_axis; Init Y_axis;

2) When the search direction RV_x and RV_x is set, random search distance of the individual fruit fly can be obtained by the following formula:

$$\begin{aligned} Xi &= Init\ X\_axis + RV_x \\ Yi &= Init\ Y\_axis + RV_y \end{aligned} . \tag{3}$$

3) Because the position of the food is unknown, therefore the distance Disti between the current position of individual fruit fly and the original point needs to be evaluated. The decision value Si of taste and consistency can be calculated after that, the decision value of taste and consistency equals to the reciprocal of the distance.

$$\begin{aligned} Disti &= \sqrt{Xi^2 + Yi^2} \\ Si &= 1/Disti \end{aligned} . \tag{4}$$

4) Taste and consistency decision value is substituted into taste consistency to judge the function and calculate the taste consistency of the individual fly's current position.

$$Smelli = Function(Si) . \tag{5}$$

5) The best taste consistency in fruit fly group can be obtained by the following formula:

$$[bestSmell\ bestIndex] = \max(Smelli) .$$

6) Preserve the best taste consistency value of fruit fly group and its corresponding x-coordinate and y-coordinate, then the fruit fly group performs positioning on the food source by means of personal sight, and fly to the position of the food source after that.

$$\begin{aligned} Smellbest &= bestSmell \\ X\_axis &= X(bestIndex) . \\ Y\_axis &= Y(bestIndex) \end{aligned}$$

7) Entering iterative optimization, repeat the iteration steps 2-5, and meanwhile judge if the taste consistency is better than the previous iterative taste and consistency; if it's established, the step 6 shall be performed.

5.3 MODEL OF FRUIT FLY ALGORITHM RBF NEURAL NETWORK

This work uses MATLAB neural network tool box RBF neural network function to establish improved RBF neural evaluation model [16] of fruit fly optimization algorithm. The spreading parameter Spread of RBF neural network is mainly optimized, the bigger the Spread is, the more smooth the function matching is. But the excessive Spread means to require lots of neurons to adapt to the rapid change of function. If the Spread is undersize, which means more neurons are required to be fit to the slow change of the function, thus the website performance is worse.

In the past different Spread values are tried to determine the optimum value during the design process of RBF, thus lots of time will be spent, and meanwhile the best of the determined Spread value can't be guaranteed.

FOA algorithm in this work is adopted to search the optimum value within the whole situation, the predicted error sum of squares and served as the taste judgment function to determine the optimum RBF spreading parameter Spread. The specific steps of algorithm are as follows:

1) Confirming the number of individuals and the maximum iteration times in the group, and meanwhile the initial position of the fruit fly can be generated randomly;
2) Endowing the random flight direction and distance section [-1 1] for fruit fly individual to search food.
3) Estimating the distance between the original point, calculating the taste and consistency judge value, which is the spreading parameter Spread, if the spreading parameter Spread <0.01, thus the Spread=1.
4) Taking Spread into the RBF to perform network training and simulation, the function shall be judged by the predicted error squares and as the taste, the taste and consistency of the fruit fly position can be obtained, which is the error sum of squares.
5) The optimum taste and consistency in the fruit fly group, which means, the fruit fly with the lowest taste consistency make the error sum of squares the lowest.
6) Preserving the optimum spread parameter Spread and its corresponding x-coordinate and y-coordinate, meanwhile the fruit fly group performs positioning on food source by means of own sight, and then fly to the position where the food source exists.
7) Entering the iterative optimization, repeating iterative steps 2-5, meanwhile judging if the taste consistency is better than the previous iterative taste consistency; if it's successful, the step 6 will be performed.

6 E-commerce website competitiveness index is calculated based on the fruit fly optimization RBF neural network algorithm

Set the parameter of fruit fly optimization algorithm, the number of iterations is 100, the population size is 30. MATLAB is applied to solve by programming, and the solved result are as follows:

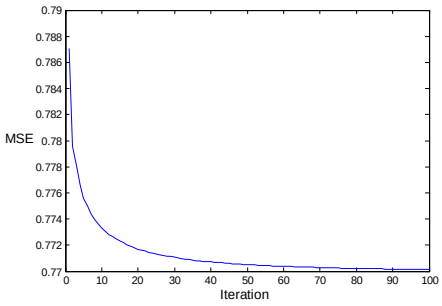


FIGURE 6 The square error convergence map of fruit fly optimization algorithm RBF

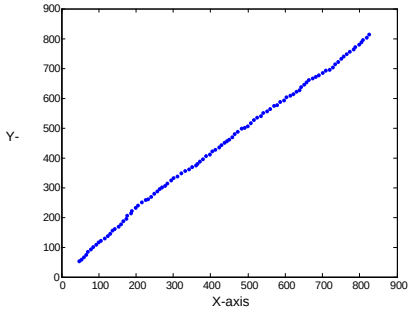


FIGURE 7 Optimization path of fruit fly algorithm

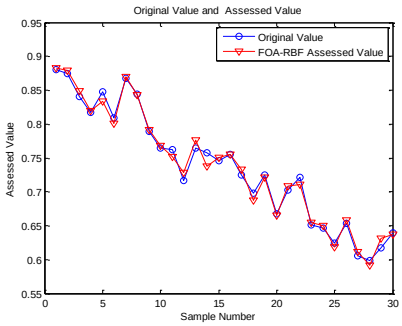


FIGURE 8 Comparison figure of FOA-RBF predicted result and the original value

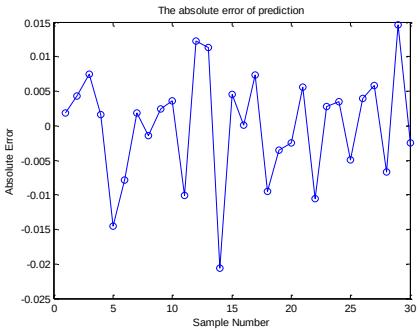


FIGURE 9 FOA-RBF predicted absolute error figure

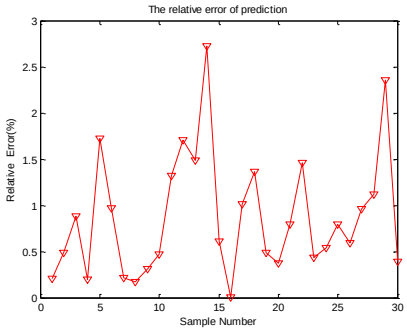


FIGURE 10 FOA-RBF predicted relative error figure

The simulated result of FOA-RBF algorithm shows that the prediction accuracy of FOA-RBF algorithm is higher than the RBF neural network algorithm, and the expected absolute error and relative error are shown in Figure 9 and Figure 10. The optimizing rate of convergence is faster in FOA-RBF algorithm, with good convergence. The characte-

ristic of convergence is shown in Figure 6, and the optimazing path of fruit fly in 2-D space in Figure 7.

7 Conclusions

The expert grading method is used to construct a multi-index hierarchical structure of E-commerce website competitiveness evaluation, with the establishment of an E-commerce website competitiveness evaluation index system, and thus quantizing the competitive level of the website. After that, RBF neural network algorithm is used to perform evaluation research on E-commerce website, aiming at the problems exist in evaluation research, fruit fly optimization algorithm is applied to perform improvement on RBF neural network algorithm. It can be seen by means of simulated comparison of the practical examples, FOA-RBF algorithm is obviously better than RBF neural network algorithm on evaluating the accuracy and processing time of E-commerce website, thus the validity and reliability of the algorithm in this work are verified, and this method can be promoted to other fields so as to easily solve other similar problems.

References

[1] Kennedy J, Mendes R 2003 Neighbourhood topologies in fully-informed and best-of-neighbourhood particles swarms *Proceedings of the 2003 IEEE International Workshop on Soft Computing in Industrial Applications 2003(SMCia/03)* 45-50

[2] Lotfi K G, Sherif A M, Ashraf O N 2008 A particle swarm-based genetic algorithm for scheduling in an agile environment *Com. & Indu. Engineering* 55(3) 707-20

[3] Shi X H, Liang Y C, Lu C, Wang L M 2005 An improved GA and a novel PSO-GA-based hybrid algorithm *Inform Process Lett* 255-61

[4] Xu R, Venayagamoorthy G K, Wunsch D C 2007 Modeling of gene regulatory networks with hybrid differential evolutionary and particle swarm optimization *Neural Networks* 266(1-3) 917-27

[5] Pan Wen-Tsao 2012 A new fruit fly optimization algorithm: Taking the financial distressmodel as an example *Knowledge-Based Systems*(26) 69-74

[6] Li J, Yang A, Dai W 2007 Modeling mechanism of grey neural network and its application *Proceedings of 2007 IEEE International Conference on Grey Systems and Intelligent Services Nanjin* 404-8

[7] Ratnaweera A, Halgamuge S K, Watson H C 2004 Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients *IEEE Transactions on Evolutionary Computation* 8(3) 240-55

[8] Higashi N, Iba H 2003 Particle swarm optimization with Gaussian mutation *Proceedings of IEEE Swarm Intelligence Symposium (SIS)* 71-9

[9] Kennedy J 2006 In Search of the Essential Particle Swarm *IEEE Congress on Evolutionary Computation, Canada: IEEE Press* 1(1) 1694-701

[10] Stacey A, Jancic M, Grundy I 2003 Particle swarm optimization with mutation *Proceedings of IEEE International Conference on Evolutionary Computation* 2 1425-30

[11] Wang H, Liu Y 2007 A Hybrid Particle Swarm Algorithm with Cauchy Mutation *in Proc IEEE Int. Conf. Swarm Intelligence Symposium (SIS)* 356-60

[12] Wang H, Liu Y 2008 An improved particle swarm optimization with adaptive jumps *In Proceedings of IEEE International Conference on Evolutionary Computation* 392-7

[13] Lum P Y, Singh G, Lehman A, et al. 2013 Extracting insights from the shape of complex data using topology *Mathematical Problems in Engineering* 3 article 1236

[14] Herrero J, Valencia A, Dopazo J 2001 A hierarchical unsupervised growing neural network for clustering gene expression patterns Mathematical *Problems in Engineering* 17(2) 126–36

[15] Walczak S 2012 Methodological triangulation using neural networks for business research *Mathematical Problems in Engineering Article ID 517234 12 pages*

[16] Nicolau M, Levine A J, Carlsson G 2011 Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival *Mathematical Problems in Engineering* 108(17) 7265–70

Author

Gang Lu, born in October, 1969, Shanghai, China.

Current position, grades: an associate professor in School of Management Science and Engineering, Shijiazhuang University of Economics, China.
Scientific interest: electronic commerce research.
Publications: 20 papers.

Web search engine based trend analysis of electronic commerce market

Zihui Yang*

School of Traffic & Transportation Engineering, Central South University, Changsha Normal University, China

Received 1 May 2014, www.cmnt.lv

Abstract

Google Trends offer weekly free information about internet searches. Users can see and download search volume patterns for the search term and the information is available by the category and by the location of those making the search. Also, Google provides "Hot searches" and "Top charts" including top and rising searches that include the search term. All these information is up to date as it provides weekly figures for a period up to and including the current week. In here, the proposed methods present a predictive model for Electronic Commerce market using the searched data in Google search engine (Google Trend data). Through the predictive model for the market and analysis of the Google Trend data, the proposed methods can get an efficient and meaningful result for the Electronic Commerce market, also we can enforce the marketing on highly ranked countries and cities. Those are very useful information for the Electronic Commerce manufacturers.

Keywords: electronic commerce, search engine, trend analysis

1 Introduction

Electronic commerce, commonly known as E-commerce, is trading in products or services using computer networks, such as the Internet. Electronic commerce draws on technologies such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems. Modern electronic commerce typically uses the World Wide Web for at least one part of the transaction's life cycle, although it may also use other technologies such as e-mail, mobile devices, social media, and telephones.

Electronic commerce is generally considered to be the sales aspect of e-business. It also consists of the exchange of data to facilitate the financing and payment aspects of business transactions. This is an effective and efficient way of communicating within an organization and one of the most effective and useful ways of conducting business. It is a market entry strategy where the company may or may not have a physical presence. There are already a range of studies about how the data can be used to monitor economic trends because Google data is extremely up to date. Those are called as nowcasting [2]. We can predict a near future based on Google Trends data. Google Trends coupled with an explanation of its potential pitfalls and other caveats, and then we compare patterns of search data from Google Trends with the real statistics from the local information, providing a justification for the possible utility of Google Trends.

In Google Trends, the data are measures of the likelihood of searches, which indicates the likelihood of a random user searching for a particular term from a certain location at a certain time. Repeated queries from a single user over a short period of time are filtered out, so that the level of interest isn't impacted. Also, data are displayed on a scale of 0 to 100 after normalization, hence if a term was most searched for in the first week of the month, the annual chart of such searches would have a peak of 100 in that week and all other weeks would be displayed as proportions of the volumes of searches for the peak week. The "Search Volume Index" is displayed graphically on screen but the underlying data can be downloaded as a CSV file. The series started from the beginning of 2004 and are provided on a weekly basis. Results for combinations of search terms can be generated, and the values obtained are scaled and normalized figures.

Government data are often released with a lag of months or more, which is a delay in assessing the current economic conditions. The term frequency of online search engine like Google provides a highly accurate and up to date but simple way to predict the future business. We can uncover sales trends before government publish the data by recognizing consumers' interests through their online behaviours analysis. As the search engine becomes pervasive, more shoppers are using the Web to collect the information about the product to purchase and narrow down the number of selections, especially for expensive products. There are many examples to show the correlation between search and actual data. For example, Google flu trend showed strong correlation between the search and real flu patient data [7].

With the help of cloud computing [16-19], internet of things [20-23], and Big Data [24-26] Advances in search engine technology, big data analytics technologies, etc.

* ***Corresponding author's*** e-mail: 12498303@qq.com

provide remarkable detailed information of human behaviours. With these technologies, we now can take advantage of (almost) real-time data.

Nowadays, companies have already made predictions, such as consumer preferences, supplies and demands for goods as well as the number of inventory level and turnover rate. A lot of companies such as Amazon, Coca-Cola, and Volvo are making a decision about their business strategies and to achieve tremendous profits from the market.

2 Related works

In the United States some electronic commerce activities are regulated by the Federal Trade Commission (FTC). These activities include the use of commercial e-mails, online advertising and consumer privacy. The CAN-SPAM Act of 2003 establishes national standards for direct marketing over e-mail. The Federal Trade Commission Act regulates all forms of advertising, including online advertising, and states that advertising must be truthful and non-deceptive. Using its authority under Section 5 of the FTC Act, which prohibits unfair or deceptive practices, the FTC has brought a number of cases to enforce the promises in corporate privacy statements, including promises about the security of consumers' personal information. As result, any corporate privacy policy related to e-commerce activity may be subject to enforcement by the FTC.

The Ryan Haight Online Pharmacy Consumer Protection Act of 2008, which came into law in 2008, amends the Controlled Substances Act to address online pharmacies.

Internationally there is the International Consumer Protection and Enforcement Network (ICPEN), which was formed in 1991 from an informal network of government customer fair trade organisations. The purpose was stated as being to find ways of co-operating on tackling consumer problems connected with cross-border transactions in both goods and services, and to help ensure exchanges of information among the participants for mutual benefit and understanding. From this came Econsumer.gov, an ICPEN initiative since April 2001. It is a portal to report complaints about online and related transactions with foreign companies.

There is also Asia Pacific Economic Cooperation (APEC) was established in 1989 with the vision of achieving stability, security and prosperity for the region through free and open trade and investment. APEC has an Electronic Commerce Steering Group as well as working on common privacy regulations throughout the APEC region.

In Australia Trade is covered under Australian Treasury Guidelines for electronic commerce, and the Australian Competition and Consumer Commission regulates and offers advice on how to deal with businesses online, and offers specific advice on what happens if things go wrong. Also Australian government e-commerce website provides information on e-commerce in Australia.

In the United Kingdom the FSA (Financial Services Authority) is the competent authority for most aspects of the Payment Services Directive (PSD). The UK implemented the PSD through the Payment Services Regulations 2009 (PSRs), which came into effect on 1 November 2009. The PSR affects firms providing payment services and their customers. These firms include banks, non-bank credit card issuers and non-bank merchant acquirers, e-money issuers, etc. The PSRs created a new class of regulated firms known as payment institutions (PIs), who are subject to prudential requirements. Article 87 of the PSD requires the European Commission to report on the implementation and impact of the PSD by 1 November 2012.

Moe et al. showed that online behaviours can be used to show consumers' interest and predict purchase outcomes [10]. Ginsberg et al. accurately estimated the current level of weekly influenza activity in each region of the United States. There were highly correlation between the relative frequency of certain queries and the percentage of physician visits in which a patient presents with influenza-like symptoms. Through their work, we could detect influenza epidemics in areas with a large population of web search users using search queries [7]. Hand et al. investigated and showed that forecasts of cinema admissions based on seasonal patterns in the data could be improved using Google Trend data [9]. Lui et al. tested predictive power against the US congressional elections (2008 and 2010). Based on their investigation, Google Trend data was not a good predictor in this case and they explained why this may be the case [10]. Choi et al. Showed how Google Trend data may help predict initial claims for unemployment benefits in the United States. And they found that Google Trend data played a key role to improve forecasting accuracy [4,5]. Wu et al. predicted housing prices and sales with Google Trend data and found that housing search index is strongly predictive of the future housing market sales and prices. Also, they suggested how those data can be used in other markets [15]. Contreras et al. provided a method to predict next-day electricity prices based on the ARIMA methodology. They showed the result of Spain and California market [6]. With the ability to collect search queries over time, Google trends data is useful to capture the decision makers' intention. Google trends data provides unprecedented opportunities to make predictions in electronics markets. Google search frequencies can be used as a reliable predictor for the underlying electronics market trends both in the present and in the near future, which is called "nowcast" using very simple regression model [2].

3 Google trends

In 2010 the United Kingdom had the biggest e-commerce market in the world when measured by the amount spent per capita. The Czech Republic is the European country where ecommerce delivers the biggest contribution to the enterprises´ total revenue. Almost a quarter (24%) of the country's total turnover is generated via the online channel.

Among emerging economies, China's e-commerce presence continues to expand every year. With 384 million

internet users, China's online shopping sales rose to $36.6 billion in 2009 and one of the reasons behind the huge growth has been the improved trust level for shoppers. The Chinese retailers have been able to help consumers feel more comfortable shopping online. China's cross-border e-commerce is also growing rapidly. E-commerce transactions between China and other countries increased 32% to 2.3 trillion yuan ($375.8 billion) in 2012 and accounted for 9.6% of China's total international trade.

Google is dominant in the search engine field. People from all around the world are using Google because Google's search algorithm is superior. Google is processing more than 60% of all the online queries in the world. [10] I got the data from Google Trends, which analyse web searches to compute how many searches have been done for the terms you input, and provide weekly and monthly reports on query statistics. Online queries submitted to the Google search engine since 2004 are captured and categorized them into several predefined categories. Definitely, queries not within the predefined categories are also captured systematically.

A search index from Google Trends doesn't provide the absolute number of counts of queries. Instead, it is calculated as the search volume for each query in a given geographical location divided by the total number of queries, so the index is always from 0 to 100. Google Trends data is easy to access and up to date.

I don't use a predefined category in Google Trends. I checked with tv, smart tv, led tv, hdtv and 3d tv at the Google search engine. As you can see in Figure 1, many people are still searching by just typing 'tv' than smart tv, led tv, hdtv or 3d tv. Google flu estimate showed that Google Trend data have the potential to predict people's interest over time and I believe that the volume of Google search queries can be used to predict future economic indicators.

4 Google flu trends

Google provides weekly influenza activity estimates for the world based on aggregated Google search data. Not every person who searched for "flu is actually sick, but some pattern can be found. Through the comparison between Google's query counts and the number of people actually have flu symptoms, we can find a close relationship [7].

As you can see, there is a positive relationship between Google flue trends estimate and the data of US centres for disease control. In here, we are going to compare TV data from Google search engine. Actually, we compared TV, Smart TV, HDTV, LED TV and 3D TV. As you can see in Figure 1, Search for just 'TV' is dominant. All other terms including Smart tv, HDTV, LED tv and 3D tv are 0 or 1 as a normalized value in most of the period. As you can see in Figure 2, Samsung tv is ranked #1 in from around 2009 to now. LG and Sony are competing for the 2nd place. As you can see in Figure 2, the search of each tv manufacturer has some seasonality, so it showed the local peak in end of each year. It means that end of the year is usually a shopping season.

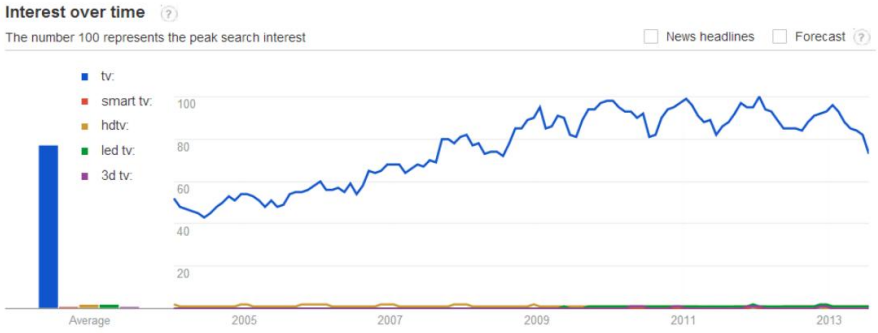


FIGURE 1 Comparison of TV related search terms in Google search engine

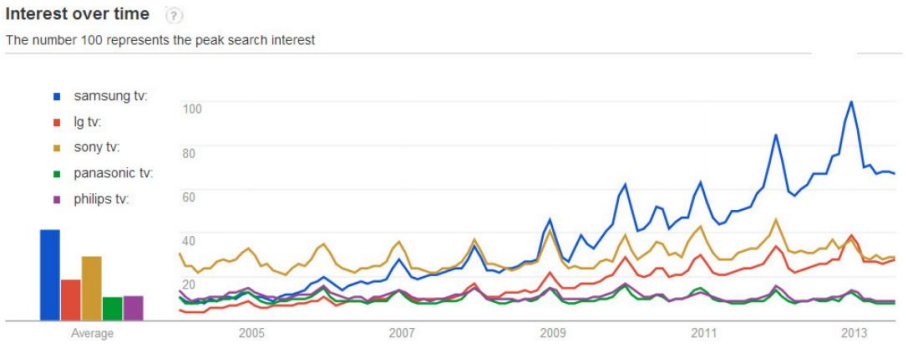


FIGURE 2 Comparison of 5 TV manufacturers in Google search engine

5 Predictive model of TV time series data

I used ARIMA (Auto Regressive Integrated Moving Average) [6] as a statistical model and the model was implemented in R. Exponential smoothing is useful for making forecasts and make no assumptions about the correlations between successive values of the time series data.

ARIMA models include an explicit statistical model for the irregular component of a time series data and it allows for non-zero autocorrelations in the irregular component. ARIMA models are defined for stationary time series. Hence, your time series data is a non-stationary, first you have to difference a time series data to obtain a stationary time series. Typically, the ARIMA model can be written as ARIMA (p,d,q), where p is the number of autoregressive terms, d is the order of differencing and q is the number of moving average terms. For example, ARIMA (1,1,0) is a first-order AR model with one order of differencing. In most cases, the best model turns out a model that uses either only AR terms or only MA terms.

Economists have theorized that e-commerce ought to lead to intensified price competition, as it increases consumers' ability to gather information about products and prices. Research by four economists at the University of Chicago has found that the growth of online shopping has also affected industry structure in two areas that have seen significant growth in e-commerce, bookshops and travel agencies. Generally, larger firms are able to use economies of scale and offer lower prices. The lone exception to this pattern has been the very smallest category of bookseller, shops with between one and four employees, which appear to have withstood the trend.

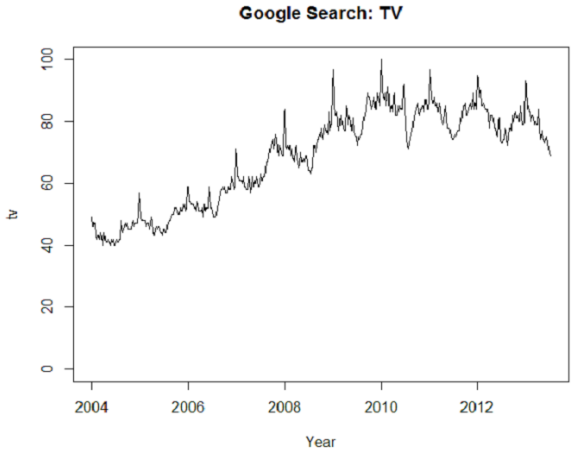


FIGURE 3 TV search result in Google search engine

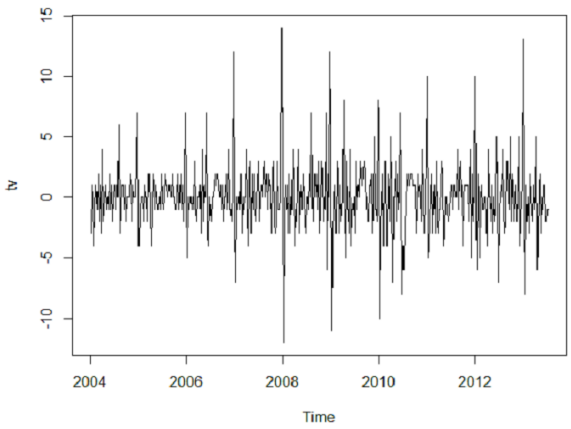


FIGURE 4 Time series data of first difference

The mean forecast error is 0.048. The time series forecast errors have a positive mean rather than a zero mean. Figure 5 showed the results of the prediction model using an ARIMA model. Blue line represents the nowcast using the prediction model. Red line represents the result with 5% significance level.

Individual or business involved in e-commerce whether buyers or sellers rely on Internet-based technology in order to accomplish their transactions. E-commerce is recognized for its ability to allow business to communicate and to form transaction anytime and anyplace. Whether an individual is in the US or overseas, business can be conducted through the internet. The power of e-commerce allows geophysical barriers to disappear, making all consumers and businesses on earth potential customers and suppliers. E-bay is a good example of e-commerce business individuals and businesses are able to post their items and sell them around the Globe.

6 Experiment and results

In case of 'tv' search on Google search engine, local maxima are always January of each year, so January 2013, January 2012, January 2011 was local maxima. Local maxima are August 2012, June 2011, July 2010, and July 2009. Hence, we can see that the search data of 'tv' is seasonal, which is increased on January of each year and decreased on summer (June, July or August) of each year. I depicted the volume of Google search engine for 'tv'. As you can see, the graph is not stationary time series, so I can difference the time series once and plot the difference series in Figure 3. As you can see in Figure 3, the search result for 'tv in Google search engine is somewhat seasonal and not stationary. As you can see in Figure 4, the differenced time series looks stationary in mean.

Based on ACF and PACF, MA(1) model is appropriate for the data. In case of ACF, the value at lag 2 exceeded the significance bounds. In case of PACF, the values at lag 2, 11, 14 and 17 exceeded the significance bounds.

Pakistan, France and Romania are top 3 'tv' search countries on Google search engine. Figure 6 showed the term 'tv' search result based on city. Istanbul, Paris and Ankara are top 3 'tv' search cities on Google search engine. According to the Google Trends data, the search for 'tv' is hot in Pakistan, France, Romania, Turkey (Country) and Istanbul, Paris, Ankara and Warsaw (City),

so manufacturers have to intensify the marketing in these countries and cities right now to increase the revenue. I think that the searched data in Google search engine have a positive relationship with real sales like the Google Flu Trends data. Also, as seen on prediction model in Figure 5, market for 'tv' might be increased smoothly.

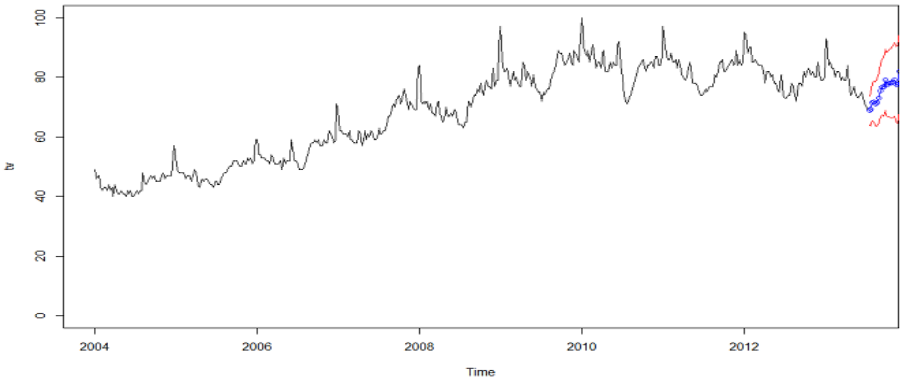


FIGURE 5 Prediction model of Google Trends data

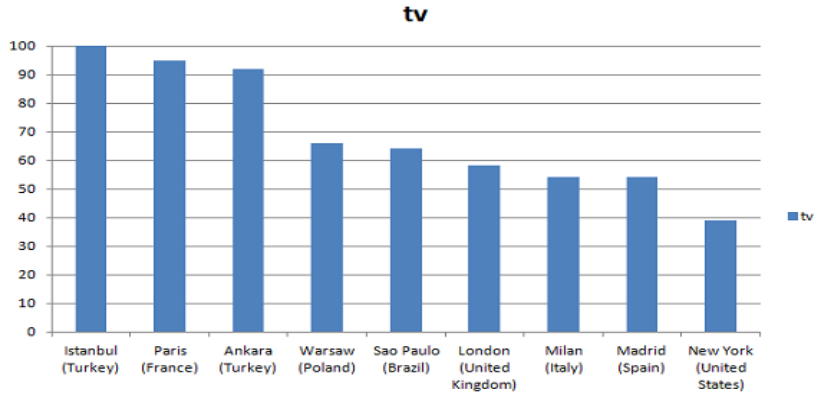


FIGURE 6 Search Results based on the City

I also investigated the 'tv' related terms such as smart tv, hdtv, led tv and 3d tv. For smart tv, Pakistan, Turkey and Ghana are top 3 in search frequency. For hdtv, United States, South Korea and Germany are top 3. For led tv, India, Brazil and Hungary are top 3. For 3d tv, Brazil, United Kingdom and Hungary are top 3. All these results are based on current Google search, so if some issues or policy changes of each country happen, it can be changed quickly. For example, recently BBC announced that they have no more plan for 3D format after the trial period ends, so it would affect the search result in UK sooner or later. However, compared with the 'tv', search result of other terms was so trivial.

7 Conclusions

In this paper I reviewed the TV search data through Google search engine and predicted the TV market based on the searched data and ARIMA model. Also, I could estimate the target market based on the frequencies of the search on the Google search engine. By intensifying the marketing in the regions which showed a high search frequency, the sales of the item (TV) might be increased. The search result of Google Trend will be different as time goes by because the search of people will be different based on various factors.

I've found that we can predict the near future of the market of home electronics (e.g., TV) using Google Trends data. Google Trends data reveals up to date trend of the people.

It might be applied to other products (e.g., Refrigerator, Cell phone, etc.) or areas (e.g., Home sales, diseases, etc.) Through the analysis of the Google Trend data, we can see that people from which country and from which city like the keyword which we've searched in the Google search engine. We can find useful trend or information by watching the Google search trend. For example, it can predict daily price moves in the Dow Jones industrial average. We already saw Google flu trend data, which correlated with the real flu data. Like this, we will see more useful correlation result in the near future.

References

[1] Askitas N, Zimmerman K 2009 Google econometrics and unemployment forecasting *Applied Economics Quarterly 55:107.120* URL: http://ftp.iza.org/dp4201.pdf
[2] Carrière-Swallow Y, Labbé F Nowcasting with google trends in an emerging market http://www.bcentral.cl/estudios/documentos-trabajo/pdf/dtbc588.pdf
[3] Catherine L, Panagiotis T, Mustafaraj E 2011 On the predictability of the U.S. elections through search volume activity http://cs.wellesley.edu/~pmetaxas/e-Society-GTrends-Predictions.pdf
[4] Choi H, Varian H 2012 Predicting the present with Google Trends *The Economic Record* 88 2–9
[5] Choi H, Varian H 2009 Predicting the present with Google Trends *Google Research Blog Thursday April 2 2009* http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html.
[6] Contrera J, Espinola R, Nogales F, Conejo A 2003 ARIMA models to predict next-day electricity prices *IEEE Transactions on Power Systems* **18**(3) 1014-20 DOI: 10.1109/TPWRS.2002.804943
[7] Ginsberg J, Mohebbi M, Patel R, Brammer L, Smolinski M, Brilliant L 2009 Detecting influenza epidemics using search engine query data *Nature* 457 doi:10.1038/nature07634
[8] Google, How does Google Trends Work, Official Site, 2010, http://www.google.com/intl/en/trends/about.html.
[9] Hand C, Judge G 2012 Searching for the picture: forecasting UK cinema admissions using Google Trends data *Applied Economics Letters* 19 1051-5 DOI:10.1080/13504851.2011.61374
[10] Lui C, Mustafaraj E 2011 On the predictability of the U.S. Elections through Search Volume Activity *e-Society Conference Avila Spain*
[11] Moe W, Fader P 2008 Dynamic conversion behavior at e-commerce sites *Management Science* **50**(3) 326-35
[12] Pentland A "Honest Signals: How They Shape Our World" The MIT Press: London 2008
[13] Nielsen Report 2008 http://www.polepositionmarketing.com/emp/august-2008-search-3/
[14] Polgreen P, Chen Y, Pennock D, Forrest N 2008 Using internet searches for influenza surveillance *Clinical Infectious Diseases* 47 1443–8
[15] Preis T, Moat H, Stanley H 2013 Quantifying Traing Behavior in Financial Markets Using Google Trends *Nature, Scientific Reports 3, Article number* 1684 DOI:10.1038/srep01684
[16] Wu L, Brynjolfsson E 2009 The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales *NBER Conference Technological Progress & Productivity Measurement*
[17] Zhang L Q, Ni L M 2010 *IEEE Transactions on Parallel and Distributed Systems* **21**(3) 405-16
[18] Menzel M, Ranjan R, Wang L, Khan S, Chen J 2014 CloudGenius: A Hybrid Decision Support Method for Automating the Migration of Web Application Clusters to Public Clouds *IEEE Transactions on Computers in press*
[19] Hao F, Min G, Chen J, Wang F, Lin M, Luo C, Yang L T 2014 An optimized computational model for task-oriented multi-community-cloud social collaboration *IEEE Transactions on Services Computing, in press*
[20] Qi L, Dou W, Chen J 2014 Weighted Principal component analysis-based service selection method for multimedia services in cloud computing *Computing Springer in press*
[21] Wang L, Tao J, et al. 2013 G-Hadoop: MapReduce across distributed data centers for data-intensive computing *Future Generation Computer Systems* **29**(3) 739-50
[22] Xu Z, et al. 2014 Knowle: a Semantic Link Network based System for Organizing Large Scale Online News Events *Future Generation Computer Systems* 10.1016/j.future.2014.04.002.
[23] Xu Z, Luo X, Zhang S, Wei X, Mei L, Hu C 2013 Mining temporal explicit and implicit semantic relations between entities using web search engines *Future Generation Computer Systems* DOI:10.1016/J.future.2013.9.027.
[24] Liu Y, Ni L M, Hu C 2012 *IEEE Journal on Selected Areas in Communications* **30**(9) 1780-8
[25] Luo X, Xu Z, Yu J, Chen X 2011 *IEEE transactions on automation science and engineering* **8**(3) 482-94
[26] Hu C, Xu Z Semantic link network based model for organizing multimedia big data *IEEE Transactions on Emerging Topics in Computing* 10.1109/TETC.2014.2316525
[27] Liu Y, Zhu Y, Ni L M, Xue G 2011 *IEEE Transactions on Parallel and Distributed Systems* **22**(12) 2100-7

Author

Zihui Yang, borm in December, 1972, Changde, Hunan Province, China.

Current position, grades: PhD candidate of Central South University, an associate professor of Changsha Normal University, China.
Scientific interest: logistics engineering, e-commerce.
Publications: 15 papers.

Impacts of tourist destination image on place attachment and tourist loyalty

Hong-Liang Qiu*

Department of Business Administration, Tourism College of Zhejiang, Hangzhou, Zhejiang, 311231, China

Received 0 October 2014, www.cmnt.lv

Abstract

Although the importance of tourist destination image as a tool to enhance tourist loyalty is commonly acknowledged, prior research on the relationship between tourist destination image and tourist loyalty is not in-depth. Drawing on place attachment theory, a model depicting the relationship among tourist destination image, place attachment and tourist loyalty is constructed. Using the sample of 337 inbound tourists from Japanese and Korean and the structural equation modeling method, the empirical results reveal that: 1) Landscape image, merchandise image and facility image have positive effects on affective image. Landscape image and merchandise image significantly and directly affect place attachment while partially mediating the effect of affective image. Service image has a direct effect on place attachment. Facility image has an indirect effect on place attachment. Facility image significantly and directly affects tourist loyalty while partially mediating the effect of place attachment. 2) Affective image has a direct effect on place attachment and it is an antecedent of tourist loyalty while completely mediating the effect of place attachment.

Keywords: tourist destination image, cognitive image, affective image, place attachment, tourist loyalty

1 Introduction

Since more than 40 years ago, tourist destination image has grown into one of the most pervasive areas in tourism studies [1,2]. Previous researches have sought to emphasize the crucial part it plays in tourism production and consumption [3,4]. Indeed, tourist destination image is understood to play an important role in destination positioning [5], destination branding [6], and tourist decision [7,8]. Therefore, the construction and maintenance of effective tourist destination images has been a major focus of the marketing [9]. In other words, tourist destination image plays an important role in enhancing tourist loyalty. Regardless of significant progress achieved, few studies examine the role of place attachment in the relationship between tourist destination image and tourist loyalty. Due to the importance of place attachment in revealing tourist loyalty [10,11], there is an imperative need to examine the relationship among tourist destination image, place attachment and tourist loyalty.

With the fast development of tourism industry in China, the number of inbound tourists has increased sharply. It is worth mentioning that Japan and Korea have been the main focus of China's inbound tourism marketing in recent years. Hangzhou is famous for a beautiful international tourist destination. Despite all this, Hangzhou's tourist destinations have long suffered a critical problem. An important of manifestation of the problem reveals low revisit to a particular place. Hence, it is urgent that destination marketing organizations find the effective solutions for cultivating positive tourist destination images without compromising the viability of natural and cultural resources. Based on the above consideration, in order to solve this problem, the main purpose of this article is to contribute to this study of tourist destination image, by examining the relationship among tourist destination image, place attachment and tourist loyalty.

2 Literature review and hypothesis development

2.1 DEFINITION OF RELATED CONCEPTS

Tourist destination image studies have proliferated as a major focus of tourism research over the past four decades [12]. Tourist destination image is defined as the sum of the beliefs and impressions that an individual has of a tourist destination [13]. Phelps (1986) defines tourist destination image as an individual's overall perception of a tourist destination [14]. Fakeye & Crompton (1991) defines tourist destination image as an individual's mental representation of knowledge and overall perception of a particular tourist destination [15]. It is referred to the sum of perceptual beliefs and impressions based on information processed from a number of sources over time [16]. It is viewed as an individual's mental representation of knowledge and global impressions about a tourist destination [7]. So most related academic scholars commonly agree that it could be conceptualized from a psychological perspective.

Regardless of this dispute, it is commonly accepted that tourist destination image as a multidimensional concept. Prior researchers describe tourist destination image could

* ***Corresponding author's*** e-mail: qiuhongliang1127@163.com

be divided into several sub-constructs from the cognitive image perspective [5,17]. With the deepening of tourist destination image studies, it is widely understood that tourist destination image could be divided into two separate dimensions of cognitive image and affective image [18-25]. Therefore, tourist destination image could be conceptualized as the sum of perceptual beliefs and affective feeling that a tourist has about a tourist destination, which could be divided into cognitive image and affective image. In order to explore what constitutes the multi factors of tourist destination cognitive image, this article uses the focus group discussion to ascertain it. Four experts are chosen to conduct this survey. It is widely accepted that tourist destination cognitive image of landscape image, merchandise image, facility image, service image and entertainment image. On the basis of the previous studies and the focus group discussion, tourist destination cognitive image could be divided into landscape image, merchandise image, facility image, service image and entertainment image. Therefore, the following hypothesis is developed.

Tourist destination image can be divided into landscape image, merchandise image, facility image, service image, entertainment image and affective image (H1).

It is commonly acknowledged that the notion of place attachment is traceable to attachment theory [26]. Precisely, the concept of place attachment originates from interpersonal attachment theory, which is first reported by Bowlby (1979) [27]. It's further developed by several scholars [28,29]. Hence, attachment originally is defined as the emotional and psychological bonds of interpersonal relationships. Interpersonal attachment theory is found extendable beyond interpersonal relationship context to person-to-object context [30,31]. For tourism marketing, place attachment is capable of bonding the tourist emotionally and psychologically with the tourist destination [32].

Williams et al. (1992) defines place attachment as the emotional bond between an individual and a place [33]. Guiliani & Feldman (1993) defines place attachment as the bonding between individuals and places [34]. Mazumdar (2005) describes place attachment as the emotional bond between individuals and places [35]. Regardless of this dispute, it is widely acknowledged that place attachment is described as a multidimensional notion. It is commonly understood that place attachment could be divided into two separate dimensions of place dependence and place identity [33, 36-38]. Therefore, place attachment is viewed as a bond with a particular tourist destination, which could be divided into place dependence and place identity. Place dependence is viewed as tourists' functional attachment to a specific destination [33]. Place identity refers to a far-reaching connection between a place and an individual's personal identity [39].

2.2 RELATIONSHIP AMONG TOURIST DESTINATION IMAGE, PLACE ATTACHMENT AND TOURIST LOYALTY

Prior academic researchers have examined the relationship among tourist destination image, place attachment, tourist loyalty [40]. However, few studies examine the role of tourist destination image factors in the relationship. At the same time, the relationship between cognitive image and affective image is tested by most scholars. That is, cognitive image is an antecedent variable of affective image. So there is an imperative need to examine the effect of each factor of tourist destination image of place attachment and tourist loyalty. Taking account into the preceding discussion, the following hypotheses are developed.

Landscape image, merchandise image, facility image, service image and entertainment image relate positively to affective image respectively (H2a-H2e).

Landscape image, merchandise image, facility image, service image and entertainment image relate positively to place attachment respectively (H3a-H3e).

Landscape image, merchandise image, facility image, service image and entertainment image relate positively to tourist loyalty respectively (H4a-H4e).

Affective image relate positively to place attachment (H5).

Affective image relate positively to tourist loyalty (H6).

Place attachment relate positively to tourist loyalty (H7).

2.3 THE HYPOTHETICAL MODEL

Taking account into the preceding discussion, the six-factor of tourist destination image measurement model (M1) is proposed, which consists of landscape image, merchandise image, facility image, service image and entertainment image and affective image. Moreover, the model depicting the relationship among tourist destination image, place attachment and tourist loyalty (M2) is constructed, which is presented in Figure 1.

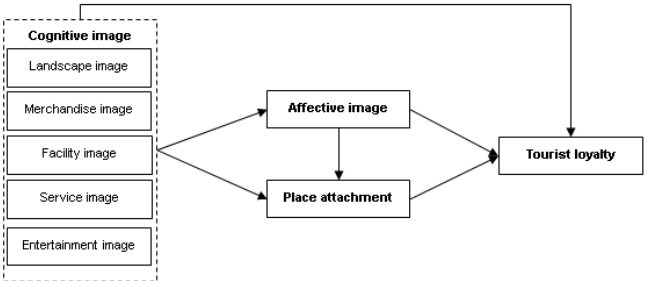


FIGURE 1 Research theory model

3 Methods

3.1 MEASURES

Measurements of the variables in this article are drawn from the prior literature and the focus group discussion. These constructs in this article are measured with multiple items. Each factor of tourist destination image is measured with 19 items by a combination of previous studies and focus group discussions. Landscape image is measured with four items (e.g., "I think Hangzhou is a beautiful international tourist destination"). Merchandise image is measured with three items (e.g., "I think Hangzhou has unique local specialties and souvenirs"). Facility image is measured with three items (e.g., "As an international tourist destination, I think internal traffic of Hangzhou is convenient"). Service image is measured with three items (e.g., "I think Hangzhou has a high service quality of catering"). Entertainment image is measured with two items (e.g., "I think leisure activities of Hangzhou are rich and colourful"). Affective image is measured with four items (e.g., "As an international tourist destination, I think visiting Hangzhou is excited"). Place attachment with four items is adapted from Williams et al. (1992) [33]. Tourist loyalty with 2 items is modified from Bao & Hu (2008) [41].

3.2 PARTICIPANT AND PROCEDURE

The sample is collected in Hangzhou tourist destination which locates at Zhejiang province. Facing the visiting inbound tourists from Japan and Korea, the 400 questionnaires are distributed. This study obtained 337 completed valid questionnaires, for an effective response rate of 84.25%. The socio-demographic and tourist profiles of the respondents are summarized as follows. Briefly, 55.2% are female and 44.8% are male; 34.5% are 25-34 years old and 25.9% are 35-44 years old; 80.7% have a university education and 15.3% have a high school education; 66.3% are repeated visit and 33.7% are first visit.

4 Tables

4.1 DIMENSIONALITY OF TOURIST DESTINATION IMAGE

Exploratory factor analysis is carried out using the principle component method with VARIMAX rotation to examine the dimensionality of tourist destination image [42]. As recommended by Hair et al. (2010), factor loadings greater than 0.50 and eigenvalues greater than 1.0 indicate significance [43]. For the tourist destination image scale, none items do not meet the criteria, and thus none items are eliminated from further analyses, yielding a six-factor model with 19 items (See Table 1). The factor solution accounts for approximately 68.704% of the total variance extracted. Therefore, all initial indicators of tourist destination image are preserved for reliability and validity testing, which is shown in Table 1.

TABLE 1 Exploratory factor analysis, reliability and validity

Variable	Eigenvalues	Item coding	Factor loadings(EFA)	Cronbach α	Factor loadings (CFA)	C.R.	AVE
Landscape image	1.931	A1 A2 A3 A4	0.778 0.752 0.795 0.649	0.777	0.714 0.696 0.748 0.586	9.262 9.135 9.478 -	0.556
Merchandise image	1.282	B1 B2 B3	0.732 0.787 0.729	0.699	0.554 0.744 0.697	8.064 9.317 -	0.562
Facility image	1.240	C1 C2 C3	0.646 0.811 0.710	0.702	0.512 0.787 0.657	7.594 9.582 -	0.526
Service image	1.783	D1 D2 D3	0.695 0.788 0.826	0.746	0.573 0.780 0.774	- 9.201 9.193	0.595
Entertainment image	1.117	E1 E2	0.751 0.812	0.619	0.829 0.544	5.146 -	0.612
Affective image	5.702	F1 F2 F3 F4	0.840 0.800 0.769 0.831	0.888	0.887 0.805 0.735 0.837	15.882 14.487 - 15.067	0.657
KMO sample measurement and Bartlett test: KMO=0.846 x^2=2447.791 df=171 Sig.= 0.000							
Overall fit : $\chi 2$/df=2.183 RMR=0.029 RMSEA=0.059 GFI=0.913 TLI=0.912 NFI=0.878							

Reliability and validity testing

Table 1 indicates that the Cronbach's α of all factors of tourist destination image reach 0.7, satisfying the specified standard of internal consistency [44]. So tourist destination image scale is considered acceptable as a good indication of reliability. In addition, as recommended by Song et al. (2014), each construct's Average variance extracted (AVE) above 0.5 is treated as indications

of convergent validity [45]. The AVE for the scale ranges from 0.526 to 0.657, indicating an acceptable level of convergent validity. According to the criteria by Lee et al. (2012), if the average variance value extracted for each construct is greater than the squared correlation coefficient for corresponding inter-constructs, the scale is considered acceptable as a good indication of discriminant validity [46]. Based on the above criteria, the scale of tourist destination image has acceptable levels of discriminant validity (see Table 2). Moreover, according to the criteria by Hung & Petrick (2012), the overall fit measures indicate that tourist destination image scale is a good representation of the structures underlying the observed data [47]. Therefore, H1 generally receive support from the empirical findings.

TABLE 2 Discriminant validity analysis

Variable	1	2	3	4	5	6
1.Landscape image	(0.746)					
2.Merchandise image	0.337	(0.750)				
3.Facility image	0.553	0.380	(0.725)			
4.Service image	0.553	0.369	0.471	(0.771)		
5.Entertainment image	0.268	0.275	0.341	0.482	(0.782)	
6.Affective image	0.467	0.562	0.532	0.364	0.289	(0.811)

4.3 STRUCTURAL MODEL AND HYPOTHESES TESTING

According to the criteria by Hung & Petrick (2012), the overall fit measures indicate that M2 has acceptable levels of overall fit (see Table 3). On this basis, this study builds CM2 through eliminating no significant relationship on M2. In Table 3, CM2 is more reasonable for examining the relationship among each factor of tourist destination image, place attachment and tourist loyalty. In the light of the output results of CM2, this study sorts out the relevant information, this is shown in Table 4. According to the output results of CM2, this study draws the final relationship figure (see Figure 2).

TABLE 3 M2 /CM2 fit index

Category	X^2/DF	RMR	RMSEA	GFI	IFI	CFI	TLI	NFI
M2	2.159	0.029	0.059	0.887	0.913	0.911	0.892	0.849
CM2	2.035	0.027	0.056	0.898	0.927	0.926	0.913	0.866

TABLE 4 Test results of CM2

Hypothesis	Estimate	C.R.	P-value	Result
Landscape image→ affective image	0.154	2.087	0.037	Supported
Merchandise image→ affective image	0.398	5.522	***	Supported
Facility image→ affective image	0.309	3.790	***	Supported
Landscape image→ place attachment	0.434	5.963	***	Supported
Merchandise image→ place attachment	0.176	2.497	0.013	Supported
Service image→ place attachment	0.173	2.984	0.003	Supported
Affective image→ place attachment	0.350	4.954	***	Supported
Facility image→ tourist loyalty	0.245	2.731	0.006	Supported
Place attachment→ tourist loyalty	0.846	7.511	***	Supported

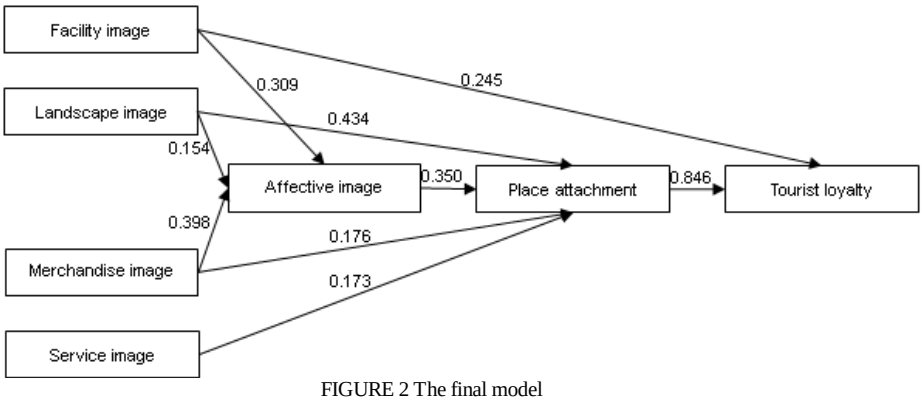


FIGURE 2 The final model

In the light of the relationship described by Figure 2, this study figures up direct effects and indirect effects (see Table 5). Table 5 indicates that:

1) Among the factors of tourist destination cognitive image, merchandise image is the important antecedent variable of affective image.
2) Merchandise image is the important antecedent variable of place attachment.
3) Facility image is the important antecedent variable of tourist loyalty.

TABLE 5 Influential effects

Path	Standardized estimate		
	Direct effects	Indirect effects	Indirect effects
landscape image→ affective image	0.154	0	0.154
Merchandise image → affective image	0.398	0	0.398
facility image → affective image	0.309	0	0.309
landscape image→ place attachment	0	0.054	0.054
Merchandise image → place attachment	0.176	0.139	0.315
facility image → place attachment	0	0.108	0.108
service image → place attachment	0.173	0	0.173
landscape image→ tourist loyalty	0	0.046	0.046
Merchandise image → tourist loyalty	0	0.266	0.266
facility image → tourist loyalty	0.245	0.091	0.336
service image → tourist loyalty	0	0.146	0.146

5 Discussion and conclusion

This article examines the relationship among each factor of tourist destination image, place attachment and tourist loyalty. The results of the structural relationship analysis reveal that:

1) Landscape image, merchandise image and facility image have positive effects on affective image.
2) Landscape image and merchandise image significantly and directly affect place attachment while partially mediating the effect of affective image.
3) Service image has a direct effect on place attachment.
4) Facility image has an indirect effect on place attachment.
5) Facility image significantly and directly affects tourist loyalty while partially mediating the effect of place attachment. Therefore, each factor of cognitive image has different effects on the relationship among affective image, place attachment and tourist loyalty, indicating that each factor of cognitive image should be treated differently.

Another finding of this article is that Affective image has a direct effect on place attachment. Not only that, but affective image is an antecedent variable of tourist loyalty while completely mediating the effect of place attachment. In other words, affective image and place attachment are two important antecedent variables of tourist loyalty. These results reveal that:

1) It is imperative to foster affective image in order to improve place attachment.

2) It is urgent that international tourist destination management should cultivate place attachment in order to enhance tourist loyalty.

Acknowledgments

This study is supported by China National Committee on Ageing (No.QLB2014B009), Educational Commission of Zhejiang Province (No.Y201328759), Key Social Science Foundation of Hangzhou City (No.A13FX04) and Zhejiang Province Federation of Social Sciences project (No. 2014Z063).

References

[1] Pike T S 200 Destination image analysis - A review of 142 papers from 1973 to 2000 *Tourism Management* **23**(5) 541–9
[2] Lai K, Li Y 2012 Core-periphery structure of destination image *Annals of Tourism Research* **39**(3) 1359–79
[3] Jenkins O 1999 Understanding and measuring tourism destination images *International Journal of Tourism Research* **1**(1) 1–15
[4] Sonmez S, Sirakaya E 2002 A distorted destination image: The case of Turkey *Journal of Travel Research* **41**(2) 185–96
[5] Pike S, Ryan C 2004 Destination positioning analysis through a comparison of cognitive, affective, and conative perceptions *Journal of Travel Research* **42**(4) 333–42
[6] Selby M & Morgan N 1996 Reconstruing place image: A case study of its role in destination market research *Tourism Management* **17**(4) 287–94
[7] Baloglu S, McCleary K W 1999 A model of destination image formation *Annals of Tourism Research* **26**(4) 868–97
[8] Beerli A, Martin J 2004 Factors influencing destination image *Annals of Tourism Research* **31**(3) 657–81
[9] Garrod B, Kosowska A 2012 Destination image consistency and dissonance: A content analysis of goa's destination image in brochures and guidebooks *Tourism Analysis* **17**(2) 167–80

[10] Mechinda P, Seriat S, Gulid N 2009 An examination of tourists attitudinal and behavioral loyalty: Comparison between domestic and international tourists *Journal of Vacation Marketing* **15**(2) 129–48
[11] Yuksel A, Yuksel F, Bilim Y 2010 Destination attachment: Effects on customer satisfaction and cognitive, affective and conative loyalty *Tourism Management* **31**(2) 274–84
[12] Kim D, Perdue R R 1979 The influence of image on destination attractiveness *Journal of Travel & Tourism Marketing* **28**(3) 225–39
[13] Crompton J L 1979 An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image *Journal of Travel Research* **17**(4) 18–24
[14] Phelps A 1986 Holiday destination image – the problem of assessment: an example developed in Menorca. Tourism Management **7**(3) 168–80
[15] Fakeye P C, Crompton J L 1991 Image difference between prospective, first-time and repeat visitors to the Lower Rio Grande Valley *Journal of Travel Research* **30**(2) 10–6
[16] MacKay K J, Fesenmaier D R 1997 Pictorial element of destination in image formation *Annals of Tourism Research* **24**(3) 537–65
[17] Sirakaya E, Sonmez S F, Choi H 2001 Do destination images really matter? Predicting destination choices of student travelers *Journal of Vacation Marketing* **7**(2) 125–42
[18] Bigne J E, Sanchez M I, Sanchez J 2001 Tourism image，evaluation variables and after purchase behavior: Inter-relationship *Tourism Management* **22**(6) 607-16
[19] Baloglu S, Mangaloglu M 2001 Tourism destination images of Turkey, Egypt, Greece, and Italy as perceived by US-based tour operators and travel agents *Tourism Management* **22**(1) 1–9
[20] Kim H, Richardson S L 2003 Motion picture impacts on destination images *Annals of Tourism Research* **30**(1) 216–37
[21] Son A 2005 The measurement of tourist destination image: Applying a sketch map technique *International Journal of Tourism Research* **7**(4-5) 279–94
[22] Ekinci Y, Hosany S 2006 Destination personality: An application of brand personality ot tourism destinations *Journal of Travel Research* **45**(2) 127–39
[23] Hosany S, Ekinci Y, Uysal M 2006 Destination image and destination personality: An application of branding theories to tourism places *Journal of business Research* **59**(5) 638–42
[24] Lin C, Morais D B, Kerstetter D L, Hou J 2007 Examining the role of cognitive and affective image in predicting choice across natural, developed, and theme-park destinations *Journal of Travel Research* **46**(2) 183–94
[25] Martin H S, Bosque I A R 2008 Exploring the cognitive-affective nature of destination image and the role of psychological factors in its formation *Tourism Management* **29**(2) 263–77
[26] Bowlby J 1969 Attachment and loss: Attachment *London: Hogarth Press*
[27] Bowlby J 1979 The making and breaking of affection bonds *London: Tavistock Publications*
[28] Hazan C, Shaver P R 1994 Attachment as an organizational framework for research on close relationships *Psychological Inquiry* **5**(1) 1–22
[29] Gillath O, Shaver P R, Baek J M, Chun D S 2008 Genetic correlates of adult attachment style *Personality and Social Psychology Bulletin* **34**(10) 1396–405
[30] Belk R W 1988 Possessions and the extended self *Journal of Consumer Research* **15**(2) 139–68
[31] Milligan M J 1998 Interactional past and potential: The social construction of place attachment *Symbolic Interaction* **21**(1) 1–33
[32] Tsai S 2012 Place attachment and tourism marketing: Investigating international tourists in Singapore *International journal of tourism research* 14(2) 139–52
[33] Williams D R, Patterson M E, Roggenbuck J W 1992 Beyond the commodity metaphor: Examining emotional and symbolic attachment to place *Leisure Sciences* **14**(1) 29–46
[34] Guiliani M, Feldman R 1993 Place attachment in a developmental and cultural context *Journal of Environmental Psychology* **13**(3) 267–74
[35] Mazumdar S 2005 Religious place attachment, squatting and "qualitative" research: A commentary *Journal of Environmental Psychology* **25**(1) 87–95
[36] Moore R L, Graefe A R 1994 Attachments to Recreation settings: The case of Rail-trail users *Leisure Science* **16**(1) 17–31
[37] Kyle G, Graefe A, Manning R, Bacon J 2003 An examination of relationship between leisure activity involvement and place attachment among hikers along the Appalachian Trail *Journal of Leisure Research* **35**(3) 249–73
[38] Gross M J, Brown G 2008 An empirical structural model of tourists and places: Progressing involvement and place attachment into tourism *Tourism Management* **29**(6) 1141–51
[39] Prohansky H M 1978 The city and self-identity *Environment and Behavior* **10**(2) 147–69
[40] Prayag G, Ryan C 2012 Antecedents of tourists' loyalty to mauritius: The role and influence of destination image, place attachment, personal involvement, and satisfaction *Journal of Travel Research* **51**(3) 342–56
[41] Bao G M, Hu F S 2008 A study on the impact of destination tourism image on tourists' after-sale behavior *Tourism Tribune* **23**(10) 40-6
[42] Chen C F, Phou S 2013 A closer look at destination: Image, personality, relationship and loyalty *Tourism Management* 36(June) 269-78
[43] Hair J F, Black W C, Babin B J, Anderson R E 2010 Multivariate data analysis(7th ed.) *New Jersey Pearson Prentice Hall*
[44] Cheng T M, Wu H C, Huang L M 2013 The influence of place attachment on the relationship between destination attrativeness and environmentally responsible behavior for island tourism in Penghu, Taiwan *Journal of Sustainable Tourism* **21**(8) 1166-87
[45] Song H, You G J, Reisinger Y, Lee C K, Lee S K 2014 Behavioral intention of visitors to an Oriental medicine festival: Anextended model of goal direced behavior *Tourism Management* 42(June) 101-13
[46] Lee C K, Song H J, Bendle L J, Kim M J, Han H 2012 The impact of non-pharmaceutical interventions for 2009 H1N1 influenza on travel intentions: A model of goal-directed behavior *Tourism Management* **33**(1) 89-99
[47] Hung K, Petrick J F 2012 Testing the effects of congruity, travel constraints, and self-efficacy on travel intentions: An alternative decision-making model *Tourism Management* **33**(4) 855-67

Author

Hong-Liang Qiu, China.

Current position, grades: a lecturer in the Department of Business Administration, Tourism College of Zhejiang, China.
Scientific interest: tourist destination image, place attachment, tourist behavior, and pro-environmental behavior, on the relationship among tourist destination image, place attachment and tourist loyalty.

Deployment algorithm based on dynamic multi-populations particle swarm optimization for wireless sensor networks

Lei Hong*

School of Information Technology, Jinling Institute of Technology, Nanjing 211169, China

Received 1 March 2014, www.cmnt.lv

Abstract

Aiming at improving coverage rate and reducing coverage holes of wireless sensor networks, this paper proposes a deployment algorithm based on dynamic multi-populations particle swarm optimization. K-Means clustering algorithm is employed to divide the network into several sub-populations dynamically, which could weaken particles on the pursuit of local optima, realize the improvement of basic PSO (Particle Swarm Optimization) algorithm, and solve the "premature" problem of basic PSO algorithm effectively. In addition, it also accelerates the convergence of the algorithm. Simulation results show that this deployment algorithm can improve the network coverage rate effectively. Comparing with the conventional particle swarm optimization algorithm, its coverage rate is increased by 3.66%.

Keywords: deployment, particle swarm optimization, *k*-means, wireless sensor networks

1 Introduction

Wireless sensor network is constituted in ways of self-organization and multi-hop by large volumes of sensor nodes with communication and computation capability. Nodes in the network are able to collaboratively perceive, collect, process, and transmit the information of perceived objects within the coverage area of the network, as well as to report the information to users [1,2]. Wireless sensor network has great potential application value in military, transportation, medical care, and environment monitoring [3,4]. Node deployment is a key issue in wireless sensor network, which reflects the quality of awareness service provided by the network.

Currently, there are mainly two ways to deploy nodes for wireless sensor network: deterministic deployment and random deployment [5]. Deterministic deployment refers to the approach that, when the status of wireless sensor network is relatively fixed, or the size of deployment area is determined, network topology structure may be determined according to pre-set node position, or sensor node density in key areas may be increased. However, in practical natural environment, the environment of monitored area may be quite severe, or the network status is not determined in advance. As for this, random deployment may be adopted. However, this deployment method may easily lead to dead zones. Under certain deployment density, it may be hard to reach the required coverage rate. Hereby, how to reduce dead zones and to improve network coverage rate is an important problem urgently needing to be solved by wireless sensor network.

Based on parallel optimization and fast convergence of fish swarm algorithm, Literature [6] proposed WSNs coverage optimization strategy based on fish swarm algorithm. Literature [7] put forward a mobile sensor deployment algorithm based on virtual force. The algorithm integrates the concepts of field potential and disc packet. Yet, VFA didn't solve the problem of "dead zone". Literature [8] proposed a sensor node deployment algorithm based on virtual rhombic grid. The algorithm integrates deterministic deployment and self-organizing deployment into a unified platform. Particle Swarm Optimization (PSO) [9] is an intelligent optimization algorithm firstly proposed by Professor Kennedy and Eberhard from the US. Its ideology derived from artificial life and evolutionary computation theory, which was mainly inspired by birds flock's behaviour of looking for food. Literature [10] and Literature [11] presented a wireless sensor network deployment optimization method based on particle sward optimization algorithm. Although particle sward optimization algorithm is proved with the ability to optimize wireless sensor network deployment, in spatial searching, standard particle sward optimization algorithm is easy to take on "precocity" phenomenon, which limits the searching range of particle. In allusion to the problem of regional optimization that may be resulted by precocity of PSO algorithm, Literature [12] introduced the concept of disturbance factor into the basic particle sward optimization algorithm, and also applied disturbance factor in mixed wireless sensor networks. In Literature [13], Lin, et al. put forward multi-populations PSO algorithm, which was able to improve the optimazation performance of particles. It's running efficiency and precision was both superior to mono-population PSO algorithm. However, the literature divided sub-population by randomly selecting particles in populations. Such sub-population division strategy was blind to some extent.

* ***Corresponding author's*** e-mail: honglei@jit.edu.cn

With respect to the above issues, the paper puts forward a wireless sensor network node deployment optimization strategy based on dynamic multi-populations particle sward optimization (DMPSO) algorithm. In the optimization process, *k*-means clustering algorithm is introduced to divide the population into several sub-populations, so that optimization may be performed independently on these sub-populations. On the other hand, in order to enhance the information exchange between sub-populations, sub-populations are dynamically re-grouped, so as to reduce parties' pursuit for regional optimal point, and to effectively avoid particle "precocity" in basic PSO algorithm. This improves the coverage rate of network.

The rest part of this paper is organized as follows: in Section 2, we propose the deployment model for wireless sensor networks. In Section 3, we propose the DMPSO algorithm. In Section 4, we simulate the DMPSO optimazation scheme by using computer software and evaluate its performance. Finally, in Section 5, we reach the main conclusions.

2 Deployment model

2.1 PARTICLE SWARM OPTIMIZATION MODEL

Standard particle sward optimization algorithm [14] takes individuals as particles without weight and volume in *N*-dimensional space, which flies at a certain velocity in the searching space. The flying velocity is dynamically adjusted according to individual and population flying experience.

Assuming that $X_i=(X_{i1}, X_{i2}, \ldots, X_{iN})$ is present position of Particle *i*; $V_i=(V_{i1}, V_{i2}, \ldots, V_{iN})$ is present flying velocity of Particle *i*; $P_i=(P_{i1}, P_{i2}, \ldots, P_{iN})$ is the best position experienced by Particle *i*, i.e. the position experienced by Particle *i* with the best adaptive value, which is denoted as P_{best}, also referred to as individual best position. Assuming $f(x)$ as minimized fitness function, so that the best position of Particle *i* may be determined by the below equation:

$$P_i(t+1)=\begin{cases} P_i(t) & f(X_i(t+1)) \geq f(P_i(t)) \\ X_i(t+1) & f(X_i(t+1)) < f(P_i(t)) \end{cases}. \quad (1)$$

Assuming the number of particles in the population is *s*, and the best position experienced by all particles in the population as P_g, also referred to as global best position g_{best}, so that:

$$P_g(t) \in \{P_0(t), P_1(t), \ldots, P_s(t)\} \mid f(P_g(t)) = \min\{f(P_0(t)), f(P_1(t)), \ldots, f(P_s(t))\}. \quad (2)$$

For the *t*-th iteration, motion of Particle *i* in *d*-dimensional space ($1 \leq d \leq D$) follows the below Equation:

$$v_{id}(t+1)=\omega \cdot v_{id}(t)+c_1 \cdot rand() \cdot (P_i - x_{id}(t)) \quad c_2 \cdot rand() \cdot (P_g - x_{id}(t)), \quad (3)$$

$$x_{id}(t+1)=x_{id}(t)+v_{id}(t+1), \quad (4)$$

where, ω is inertia coefficient, which endows particles with the tendency to expand the searching space, so as to search new areas; c_1 and c_2 are acceleration constant; rand() refers to the random values within [0,1].

2.2 OPTIMIZATION MODEL FOR DMPSO

Particle swarm optimization algorithm based on dynamic multi-populations (DMPSO) is an improved multi-populations PSO algorithm. Basic ideology of DMPSO: In the algorithm, population is divided into several sub-populations by *k*-means clustering algorithm, so that sub-populations may be optimized independently. Meanwhile, after several generations of iteration, sub-populations are then re-divided to form new sub-populations, so as to enhance information exchange between sub-populations.

Assuming that there are m particles in Population *M*, the m particles are then divided into *K* sub-populations, which may be described as $M=(M_1, M_2, \ldots, M_k)$. Each sub-population may also be represented as $M_i=(X_1, X_2 \ldots X_p)$, ($i=1, 2 \ldots K$; $0 \leq p \leq m$), and elements in sub-populations stand for individual particles, and the total number of particles in all sub-populations is *m*. For each Particle *i* in sub-populations, X_i is employed to indicate its position $X_i=(X_{i1}^1, X_{i1}^2, X_{i2}^1, X_{i2}^2, \ldots, X_{iN}^1, X_{iN}^2)$, where elements in it separately refer to horizontal coordinate and vertical coordinate of sensor nodes in the particle. The flying velocity of Particle *i* is described as $V_i=(V_{i1}^1, V_{i1}^2, V_{i2}^1, V_{i2}^2, \ldots, V_{iN}^1, V_{iN}^2)$, where elements in it the component velocity of sensor nodes in the particle along X axis and Y axis. Clustering center is taken as the best position passed by all particles in each sub-population, represented by P_{lg}, also referred to as Lg_{best}. P_{lg} is used to replace population global best position P_g in Equation (1). After the improvement, velocity evolution formula of Particle *i* in *d*-dimensional space ($1 \leq d \leq D$) is shown below:

$$v_{id}(t+1)=\omega \cdot v_{id}(t)+c_1 \cdot rand() \cdot (P_i - x_{id}(t)) \quad c_2 \cdot rand() \cdot (P_{lg} - x_{id}(t)). \quad (5)$$

In each sub-population, particles often perform searching with clustering center as the best position of the sub-population. If clustering center of the sub-population is coincidentally located at the position of regional optimal solution, the sub-population may take on "precocity" convergence. Here, after *R* generations' iteration, all particles in population will be considered as a whole, and will be divided into new sub-generations by *k*-means clustering algorithm. Dynamic re-combination of sub-population is designed to improve information exchange between sub-populations, so as to avoid "precocity" of particles.

Moreover, *k*-means clustering algorithm is adopted to partition the data of all node coordinates into *K* areas, i.e. the number of sub-populations. The number of sub-populations determines social information sharing degree between particles in each sub-population. If parties are lack

of social information sharing, network coverage may be significantly reduced.

In order to enhance global searching performance of particles in searching process, inertia coefficient factor in velocity evolution Equation (5) will be appropriately adjusted:

$$\omega(\mathrm{t}) = 0.9 - \frac{t}{\max Iterations} \cdot 0.5, \tag{6}$$

where, t refers to present iteration generation of particle, and maxIterations stands for the maximum iteration generation in the algorithm. It may be seen from Equation (6) that ω reduces linearly with the iteration generation increases. As for this, the algorithm is endowed with strong global searching ability in the beginning and strong regionnal searching ability in later phase.

3 DMPSO deployment algorithms

In the paper, random deployment method is employed. In the beginning when nodes are scattered into the monitoring area, dead zones may be easily caused in the network. As for this, secondary deployment shall be performed in allusion to such "problematic" area. In order to simplify the network model, assuming that Monitoring Area A is a two-dimensional plane, while N sensor nodes with the same parameters are scattered in the area. The perception radius of each node is denoted as R_s, and communication radius as R_C. In order to keep the connectivity of network, communication radius is set no less than two times of the perception radius [15], i.e. $R_c \geq 2R_s$. Sensor node set is denoted as $S=\{S_1, S_2, \ldots, S_N\}$, where, $S_i = \{x_i, y_i, R_s\}$, $i \in (1, 2, \ldots, N)$. The coverage model of each node may be considered as a circle with coordinates of the node as the center, and R_s as the radius.

Digitally discretizing Monitoring Area A as $\mathrm{m} \times \mathrm{n}$ pixels, with coordinates of pixels as (x,y), and the distance between sensor node $\mathrm{S_i}$ and a certain pixel p shall be:

$$d(\mathrm{S}_i,\mathrm{p}) = \sqrt{(x_i - x)^2 + (y_i - y)^2}. \tag{7}$$

Here, node Boolean coverage model, i.e. 0-1 coverage model, is employed. Assuming the event of a certain pixel's being covered by sensor nodes in the monitoring area as $\mathrm{t_i}$, if the event is true, $p(\mathrm{t_i})=1$; or else $p(\mathrm{t_i})=0$. Being described in Equation:

$$p(\mathrm{t}_i) = \begin{cases} 1, & d(S_i, p) \leq R_s \\ 0, & d(S_i, p) > R_s \end{cases}. \tag{8}$$

Monitoring Area A is a $\mathrm{m} \times \mathrm{n}$ rectangle, which is divided into $\mathrm{m} \times \mathrm{n}$ pixels of equal size, with area as 1. The discrete precision is 1. For pixel (x,y), as long as there is 1 node in Node Set S covers the pixel, the pixel shall be considered being covered by Node Set S. Otherwise, the pixel (x,y) shall be considered uncovered. Denoting the rate of pixel (x,y)'s being covered by Node Set S as p(x,y,S), so that:

$$p(\mathrm{x},\mathrm{y},\mathrm{S}) = \mathrm{p}(\bigcup_{i=1}^{N} t_i) = 1 - \prod_{i=1}^{N} (1 - \mathrm{p}(\mathrm{t}_i)). \tag{9}$$

As for this, the total area covered by Sensor Node Set S is just the union set of all pixels covered by all nodes in the node set, denoted as S_{area}, then:

$$S_{area} = \int_0^m \int_0^n p(\mathrm{x},\mathrm{y},\mathrm{S})\,\mathrm{dxdy}. \tag{10}$$

The optimization goal of wireless sensor network node deployment is to maximum the coverage rate of the network. Here, coverage rate refers to the specific value between coverage area of Node Set S and the total area of the monitoring area, i.e.:

$$\sigma = \frac{S_{area}}{\mathrm{m} \times n}. \tag{11}$$

Denoting the fitness function as follows:

$$f(\mathrm{X}) = \max \sigma = \max(\frac{S_{area}}{m \times n}). \tag{12}$$

When f(X) obtains it maximum value, node position information is the best deployment position of node in wireless sensor network.

Each particle in monitoring area represents one sensor node deployment method. When Equation (9) is taken as the fitness function, node deployment optimization algorithm based on DMPSO is shown as Algorithm 1:

ALGORITHM 1: DMPSO algorithm

Algorithm: DMPSO

a) Initializing m particles, i.e. randomly generating Position X_i and Velocity V_i of each particle;
b) Dividing the population into K sub-populations with k-means clustering algorithm;
c) Updating the velocity and position of each particle in each sub-population with Formula (2)~(4);
d) Calculating the coverage rate of each particle according to fitness function;
e) Comparing the coverage rate of particle with its best position P_{best}; if the result is better, re-setting P_{best};
f) Comparing the coverage rate of each particle in each sub-population with the best position of the sub-population Lg_{best}; if the result is better, re-setting Lg_{best};
g) Judging if R iteration generations have been reach; if so, returning to b); or else, executing h);
h) If the pre-set maximum iteration generation or satisfactory coverage rate has been reached, stopping the process; optimal individual position X_i of the population shall be taken as the result; or else, returning to c) and repeating the steps.

4 Experiments and result analysis

4.1 COVERAGE AND UNIFORMITY

Generally, coverage can be considered as the measure of quality of service of a sensor network. Gage invented the concept of coverage in the research of multi-robot systems [16]. We define it as the ratio between sum of the coverage area of all the nodes and the area of the entire target region,

shown in Equation (13). The definition of sum of the coverage area is taken from the concept of union in the Set Theory, thus the coverage is usually less than or equal to 1.

$$Coverage = \frac{\bigcup_{i=1,\ldots,N} A_i}{A}. \quad (13)$$

The uniformity of coverage is a well-defined standard to measure the service life of a network. Article [17] describes the concept as the standard deviation of distance between nodes. Smaller standard deviation means better coverage uniformity of the network, as shown in Equation (14).

$$\begin{cases} U_i = \left[\frac{1}{n}\sum_{j=1}^{n}(d_{i,j} - \bar{d})^2\right]^{\frac{1}{2}}, \\ U = \frac{1}{N}\sum_{i=1}^{N} U_i \end{cases} \quad (14)$$

where in Equation (14), U is the Uniformity, N is the total number of nodes, U_i is the standard deviation of distance between the i-th node and its adjacent nodes, n is the number of neighbors of the ith node, $d_{i,j}$ is the distance between i-th and j-th nodes, $\bar{d}$ is the mean of internal distances between the ith node and its neighbors. So far, we have discussed the relation between communication and coverage. Article [16] has proved that when the communication range of node is twice or larger than the sensing range, coverage will contain pure connections. In practical deployment, we only have to consider the coverage so as to ensure the connection. At the moment, coverage contains connection problems.

4.2 SIMULATION

Assuming that the monitoring area of wireless sensor network is a 50m×50m square, perception radius of each sensor node R_s=5m, and communication radius R_c=2R_s=10m. 35 nodes are randomly deployed in the monitoring area. As particles are diversified, when the number of particles is larger, distribution of nodes may be relatively even. However, with the number of particles increases, the calculation duration may increase exponenttially, largely reducing the calculation speed. Taking into consideration the above factors, the number of particles in the population is assumed to be 30. The flying velocity of particles is limited within -3-3m/s, parameter $c1$=0.9, $c2$=0.9, maximum iteration generation I=500.

The number of sub-populations divided with k-means clustering algorithm may eventually affect the coverage rate of the monitoring area. 50 experiments are conducted under situations with K=2 to 30. On this basis, the average value is figured out, leading to the experimental data shown in Figure 1.

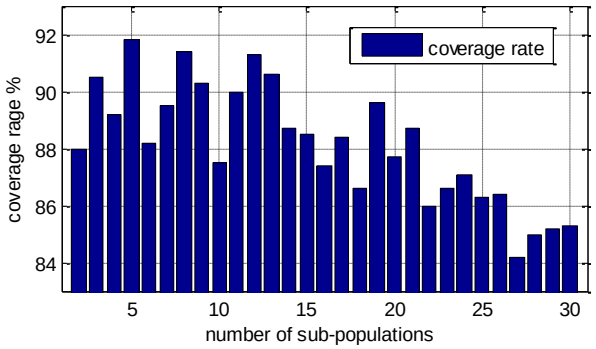


FIGURE 1 Coverage vs. various sub-populations

It can be seen from Figure 1 that, when the number of sub-populations is less than half of the population scale, the coverage rate is generally higher than that when the number of sub-populations is more than half of the population scale. As for this, the number of sub-populations is controlled within 3-8, and the effect would be better. The reason is that, when the number of sub-populations is relatively less, a certain amount of particles for each subpopulation could be guaranteed. In essence, PSO algorithm is a sort of swarm intelligence algorithm. The velocity of particle is co-determined by its flying experience and companions' flying experience. Population diversifycation of each sub-population shall be guaranteed, so that particles in each sub-population may be able to interact and to exchange information. On the contrary, when the number of divided sub-populations is larger, there might be fewer particles for each sub-population. As for this, particles may be lack of social information sharing, and the probability of reaching optimal solution would be lower. According to the experiment, when the population is divided into 5 sub-populations, wireless sensor network deployment optimization effect is the best, and the coverage rate at the moment is 91.89%.

In order to test the effectiveness of DMPSO algorithm, simulation experiment based on two models is performed. In other words, standard PSO algorithm and improved DMPSO algorithm are separately applied to optimize node deployment for wireless sensor network. Figure 2a shows the initial status of node random deployment; Figure 2b is wireless sensor network deployment result based on the optimized PSO algorithm; Figure 3c is wireless sensor network deployment result based on the improved DMPSO algorithm.

In the Figure 2, solid black dots indicate the position of sensor nodes, while circles refer to the perception range of nodes. After 500 times' iteration, the coverage rate of wireless sensor network deployed with standard PSO algorithm is 88.23%; the coverage rate of wireless sensor network deployed with DMPSO algorithm is 91.89%. Thus, it can be seen that, the coverage rate of the improved algorithm is increased by 3.66%, reaching the effect of deployment optimization. Shown by Figure 2, node distribution in (c) is more even than that in (b), so that the rate of "dead zone" and repeat coverage is relatively lower. DMPSO algorithm divides particles into

several sub-populations via clustering. On this basis, particles in each sub-population will be able to adjust the flying direction according to their respective flying experience and the "global optimum" of their sub-population. When the frequency of worst fitness of a certain particle reaches the pre-set value, the particle will consequently be regarded as being unfit for present searching environment, which needs to be optimized. As for this, the particle will be removed from regional optimal value for optimal solution. In this way, particles may be able to get rid of regional optimal value, to expand the searching range, and to effectively solve the problem of "precocity".

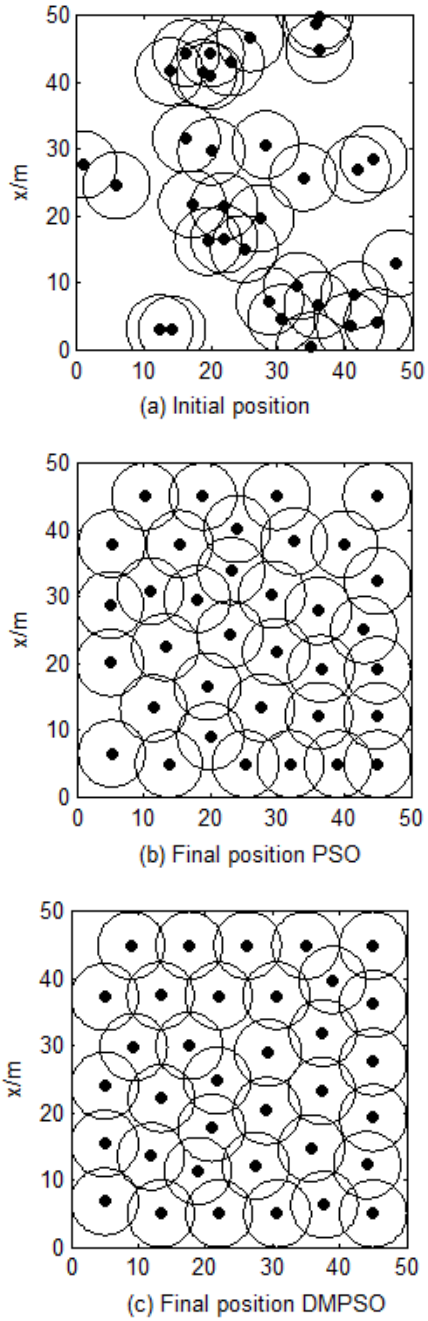


FIGURE 2 Deployment of sensor nodes

Figure 3 shows coverage and uniformity changing curve of DMPSO algorithm. According to the figure, the initial coverage rate of network is approximately 35%. After 500 iterations, the coverage rate is significantly improved. Seeing from the growth slope of the curve, in the first 100 generations of iteration, the slope of curve is relatively higher, and the coverage rate increases sharply. After 150 iterations, the slope reduces obviously, and the coverage rate grows gently, eventually being stabilized to a constant. The reason for this is that, under the influence of inertia coefficient, the algorithm is easy to converge to a global best position in the early stage. With the number of iteration increases, particles begin to oscillate around the best position, so that the result obtained tends to be stabilized to the best result 91.89%. Compared with the initial coverage rate, the coverage rate at the moment is improved by approximately 55%. Thus, it may be seen that, the algorithm shows more obvious effect in improving network coverage rate.

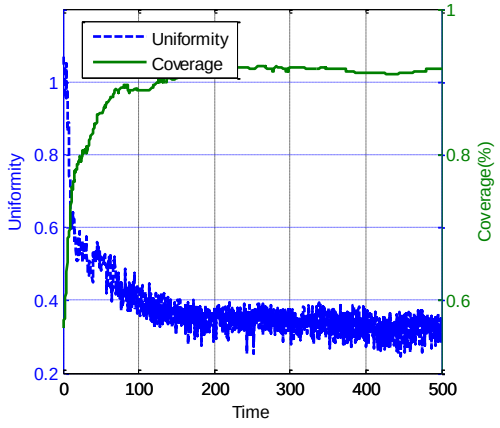


FIGURE 3 Coverage and Uniformity of DMPSO

Figure 4 shows the coverage rate of standard PSO algorithm and DMPSO algorithm. According to Figure 4, the number of iteration of standard PSO algorithm when reaching convergence is around 350, and for DMPSO algorithm, the number is around 300. The convergence speed of the improved algorithm is improved by approximately 14%.

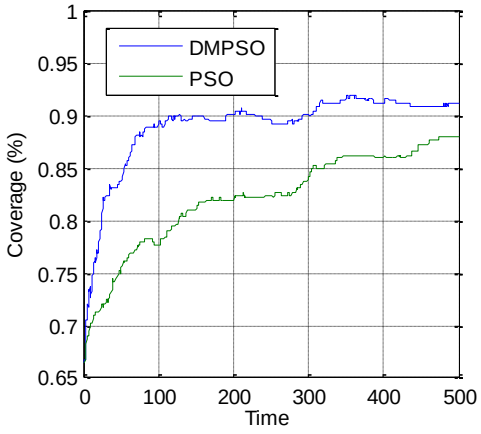


FIGURE 4 Comparison of PSO and DMPSO coverage

Thus, it may be seen that, in wireless sensor network deployment optimization, DMPSO algorithm has better convergence performance than that of PSO algorithm.

In order to further test the feasibility of DMPSO algorithm, common genetic algorithm, and bee colony algorithm are employed to compare with DMPSO algorithm proposed in this paper. The number of iteration of all the three algorithms is 500. Table 1 shows simulation result comparison of the 3 algorithms after 50 times of statistics.

TABLE 1 Comparison of various algorithms

Algorithm	Coverage %	Iteration
Genetic Algorithm	80.3	485
Bee Colony Algorithm	85.2	426
DMPSO	91.89	308

Shown by Table 1, compared with genetic algorithm and bee colony algorithm, DMPSO algorithm improves network coverage rate separately by 11.59% and 6.65%. Moreover, DMPSO algorithm reaches the best solution within less generation. As for this, it is feasible to apply DMPSO algorithm to optimize wireless sensor network node deployment.

4 Conclusions

In this paper k-means clustering method is introduced to propose a particle swarm model for dynamically dividing sub-populations. The algorithm effectively solves the "precocity" problem of standard PSO algorithm. Moreover, the simulation result also shows that, the proposed DMPSO algorithm is able to optimize the deployment of WSN node, and to improve the network coverage. How to reduce the repeat coverage of node, so as to further improve network coverage rate is a key problem to be further studied in future.

References

[1] Losilla F, Garcia-Sanchez A J, Garcia-Sanchez F, et al.2011 A comprehensive approach to WSN-based ITS applications: A survey *Sensors* **11**(11) 10220-65

[2] Kulkarni R V, Venayagamoorthy G K. 2011 Particle swarm optimization in wireless-sensor networks: A brief survey *Systems, Man, and Cybernetics Part C: Applications and Reviews IEEE Transactions on* **41**(2) 262-7

[3] Alippi C, Camplani R, Galperti C, et al. 2011 *Sensors Journal IEEE* **11**(1) 45-55

[4] Zhang J, Varadharajan V 2010 Wireless sensor network key management survey and taxonomy *Journal of Network and Computer Applications* **33**(2) 63-75

[5] Ren Y, Zhang SD, Zhang HK 2006 Theories and algorithms of coverage control for wireless sensor networks *Journal of Software* **17**(3) 422-33

[6] Yiyue W, Hongmei L, Hengyang H 2012 Wireless sensor network deployment using an optimized artificial fish swarm algorithm *Computer Science and Electronics Engineering (ICCSEE) International Conference on. IEEE* 2 90-4

[7] Zou Y, Chakrabarty K 2003 Sensor deployment and target localization based on virtual forces *Proceedings of the 22ndAnnual Joint Conference of the IEEE Computer and Communications SanFrancisco USA IEEE* 1293-303

[8] Wang XQ, Yang YT 2006 Sensor deployment algorithm based on virtual rhomb grid *Computer Applications* **26**(7) 1554-6

[9] Kennedy J, Eberhar R 1995 Particle swarm optimization *in: Proceedings of the 4thIEEE International Conference on Neural Networks Piscatawya:IEEE Service Center* 1942-8

[10] Burne R A,Buczak A L,JIN Yao-chu 1999 Self-organization cooperative sensor network for remote surveillance: Current result

[11] Wang X, Wang S, Ma J J 2007 Parallel Particle Swarm Optimization Based Mobile Sensor Node Deployment in Wireless Sensor Networks **30**(4) 563-9

[12] Xiang XX, Huang HG, Li YD 2010 Hybrid sensor networks coverage-enhancing approach based on particle swarm optimization *Application Research of Computers* **27**(6) 2273-5

[13] Lin ZL, Feng YJ 2009 Research on strategy for optimizing coverage of WSNs based on multi-particle PSO *Application Research of Computers* **26**(12) 4701-3

[14] Clerc M 2010 Particle swarm optimization *John Wiley & Sons*

[15] Zhang H-h, Hou J C 2005 Maintaining sensing coverage and connectivity in large sensor networks *Wireless Ad Hoc and Sensor Networks* **1**(1) 89-124

[16] Gage, Douglas W 1992 Command control for many-robot systems. *Naval command control and ocean surveillance center rdt and e div San Diego CA*

[17] Nojeong H, Varshney P K 2003 A distributed self spreading algorithm for mobile wireless sensor networks *Wireless Communications and Networking WCNC IEEE* 3

Author

Lei Hong, born in Mart, 1976, Jiangsu, China.

Current position: a lecturer at the School of Information Technology, Jinling Institute of Technology, China.
University studies: Bachelor Degree from Anhui Normal University in 1998, the Master. Degree from Hohai University in 2005.
Scientific interest: software engineering and wireless sensor network.
Publications: 3 papers.

Financial contagion dynamics and fragility assessment of industrial complex network

Bao Wu*

[1]*China Institute for Small and Medium Enterprises, Zhejiang University of Technology, Chaowang Rd. 18, Hangzhou, Zhejiang, P.R. China*

[2]*Zhejiang Economic &Trade Polytechnic, Xueling st.108, Hangzhou, Zhejiang, P.R. China*

Abstract

The paper proposed a mathematical modelling of financial contagion dynamics that is tightly linked to systematic risk of industrial complex network. And the paper provides a practical method to assess fragility of industrial complex network in the context of financial contagion. To examine its function, an experimental analysis based on real data set of a Chinese textile industrial network is conducted. The experimental analysis shows that the method proposed in the paper is effective ad reliable, and is capable to assess fragility of industrial complex network in the context of financial contagion.

Keywords: financial contagion, fragility, industrial complex network, assessment method

1 Introduction

Industrial complex network, such as Industrial clusters, are complex network of lots of interconnected companies in a specific industry. These companies are bonded together with highly complicated business, technological and financial relationships. In this way, these industrial systems are essentially complex networks. And these industrial complex networks play significant roles in the process of regional development in European countries, Asian countries and other areas. In rural area of emerging countries, many township economies are essentially mixtures of several industrial clusters that are typical industrial complex networks. The performance of these industrial clusters is crucial for economic development and social stability of these developing areas.

However, these industrial complex networks are readily destructed by market fluctuation and economic crisis because their interior complex network imposed risk interdependency on each individual organization. "Robust yet fragile" is a characteristic of complex network. Idiosyncratic risk of bankruptcy of an important company in the network would be likely to trigger systematic risk by inter-company financial contagion in context of interdependent industrial network. Domino-style bankruptcy in such interdependent business cycle occurs more frequently in recent years, and receives concerns of policy-makers. The argument that inter-company financial contagion is a critical step leading to systematic crisis [1] is supported by empirical evidences and case studies in recent literatures. With financial contagion and systematic risk are widely talk about in literatures, fragility of industrial complex network in the context of financial contagion became an important topic. Recent work on physical networks, such as internet and power grid, has addressed the resilience of these networks to idiosyncratic risk. Scientific works in complex network-related field has provides ideas about how to measure fragility of physical complex network. And recent financial literatures offer in-depth insights about financial contagion [2-7]. Though similar topics are discussed by scholars from disciplines, more efforts are needed to shed some light on financial contagion and its impact on fragility of industrial complex network. In previous works, the author proposed a framework for testing stability of regional industrial cluster considering inter-company financial contagion [9], and developed a network model for inter-company financial contagion [10]. In the paper, the author tries further to propose a new method for evaluating fragility of industrial complex network in the context of financial contagion.

The paper is organized as follows. The paper discusses contagious effect of idiosyncratic risk in the complex network of an industrial network in the second section. Then the paper introduces a new method for measuring fragility of industrial complex network in third section. In fourth section, the paper would illustrate an experimental analysis in the background of a Chinese industrial cluster. Its fragility would be discussed in details in this section, along with validation of this new measuring method.

2 Modelling financial contagion dynamics

In this section, the paper derives a basic mathematical framework for modelling dynamic process of financial contagion triggered by bankruptcy of one or several companies in context of industrial complex network. The paper considers mechanisms of risk interdependence and trend reinforcement in an inter-company network in the model. The paper describes an industrial complex network as a graph (donated by $G=(N,R)$), in which a set of N companies are connected with others via financial contacts.

* ***Corresponding author's*** e-mail: jamesbewood@126.com

The network is also associated with an adjacency matrix $M=\left[M_{ij}\right]$, where $M_{ij}=0$ means there is not financial contract from company *i* to company *j*, and $M_{ij}=1$ represents there is some kind of financial contract from company *i* to company *j*, i.e. company *i* invested some money or lend some money to company *j*. And the out-degree of company *i* in this network, k_i represents the number of counterparties to which company *i* is exposed. And a weight matrix $H=\left[H_{ij}\right]$ is introduced. In which, the weight H_{ij} represents the exposure of *i* to *j* relative to its portfolio of all exposures. So, the matrix H is defined as a row-stochastic matrix, and subjects to $\sum_j H_{ij}=1$. Denote M_{ij} must equal to 1, if H_{ij} is more than zero.

According to literatures about financial contagion, the dynamic law of financial contagion process can be modelled as the two mechanisms of interdependence and trend reinforcement (or financial accelerator). The paper employs stochastic differential equation (SDE) to include mechanisms of interdependence and trend reinforcement in our framework, and to investigate their interplay in the contagious process. Denote a financial robustness indicator β_i for each node *i*, to measure its distance from actual default or to indicate its state of liability reputation. If $\beta_i=0$, the node *i* has been on the situation of default or bankruptcy. A drift term that depends on the past realizations is introduced in the SDE for representing the mechanism of financial accelerator. The paper restricts the dependence at time t on the past only to $t'=t-T$. This is, β_i at t time is dependent on the realization of β_i at time $t'=t-T$. So, the paper models the law of motion of robustness by the following time-delayed SDE.

$$d\beta_i=L\left(\beta(t),\beta(t')\right)dt+\sigma d\varepsilon_i, \tag{1}$$

where $L\left(\beta(t),\beta(t')\right)$ is not more than 0, σ is the variance of idiosyncratic shocks that hit company *i*, and $d\varepsilon_i$ denotes the Wiener stochastic process.

Then, the expected fist passage time T_f is $T_f\sim 1/\sigma^2$, if financial accelerator is absent ($L\equiv 0$). While, the negative drift would make the expected fist passage time shorter, i.e. $T_f\left(L\prec 0\right)\prec T_f\left(L=0\right)$. In this way, financial accelerator would increase the probability of default or bankruptcy. And the paper restricts interdependence in the context of complex network. Denote C_{ij} as the asset related to company *j* and held by company, such as a liability of company *j* to company *i*. Then the value of asset C_{ij} depends on financial robustness of company *j*, i.e. its ability to meet the obligation to company *i*. To be more specific, the paper assumes the value of asset C_{ij} is proportional to the financial robustness of β_i, i.e. $C_{ij}=C_{ij}^0\times\beta_j$, where C_{ij}^0 is nominal value of the asset. Then the total assets of company *i* at a time *t* is determined based on the estimate of financial robustness of its counterparties at time $t-1$.

$$C_i(t)=\sum_j C_{ij}(t-1)=C_i^0\sum_j H_{ij}\beta_j(t-1), \tag{2}$$

where C_{ij}^0 is the nominal value of investments or obligations invested or lent by company *i*, H_{ij} is fraction of investments or obligations of company *i* that is related to company *j*. Thus, H_{ij} is a reasonable proxy of relative impact on total asset of company *i* due to a change in financial robustness of company *j*. In this way, the interdependence mechanism is modelled in the context of a network of financial contacts.

The paper assumes financial robustness is linearly dependent on its counterparties in absence of bankruptcies:

$$\beta_i(t)=\sum_j H_{ij}\left(\beta_j(t-1)+\sigma\varepsilon_i(t-1)\right). \tag{3}$$

And assume financial robustness is subject to independently and identically normal distribution. If company *i* is hit by idiosyncratic shock, its own loss is only proportionally to the term H_{ij} due to this shock hit. Other loss would be shared among its counterparties. And company *i* also would be proportionally affected by the shocks hitting its counterparties. Then, the change of financial robustness is express by following equation:

$$\beta_i(t)-\beta_i(t-1)=\sum_j H_{ij}\left(\beta_j(t-1)-\beta_i(t-1)\right)+\sigma\sum_j H_{ij}\varepsilon_j(t-1). \tag{4}$$

Passing to the limit of continuous time and introducing Equation (1) into Equation (4) to capture the financial accelerator, then:

$$d\beta_i=\left[\sum_j H_{ij}\beta_j(t)-\beta_i(t)+L\left(\beta(t),\beta(t')\right)\right]dt+\sigma\sum_j H_{ij}d\varepsilon_j, \tag{5}$$

where $d\varepsilon_j$ is an independent Wiener stochastic process. Since $d\varepsilon_j$ is independent, its linear combination is still a Wiener stochastic process. So, denoted as $d\tau_i=\sum_j H_{ij}d\varepsilon_j$, Equation (5) could rewrite as following:

$$d\beta_i=\left[\sum_j H_{ij}\beta_j(t)-\beta_i(t)+L\left(\beta(t),\beta(t')\right)\right]dt+\sigma d\tau_i. \tag{6}$$

A company goes bankrupt when its robustness falls below a given threshold. If the bottom threshold is hit then the robustness value is initiated to Zero. Also, for the sake of simplicity, the paper assumes the robustness cannot exceed an upper barrier at $\beta_i=1$. Then the dynamics of financial contagion process in the context of industrial complex network can be captured by the theoretical model.

3 Assessing fragility of industrial complex network

To simulate financial contagion process as described in the above theoretical model, the paper formulated a simulation procedure as illustrated in Figure 1. The simulation procedure is realized in MATLAB circumstance, operated with the help of software MATLAB 2010(a). Please see Appendix A for details of the simulation procedure.

Recent literatures of scale-free and small-world complex network research have deepened understanding of vulnerability of real-world networks [10,11]. In these researches, fragility of a complex network is generally evaluated by damage of the complex network after removing some most important nodes. To measure fragility of an industrial complex network, the paper simulates a deliberate attack at initial stage, pick up a number of top systematically-weighted nodes as bankrupts, and removes these nodes simultaneously from the network at initial stage. Then a contagion process triggered by this deliberate attack is simulated according to above mentioned procedure. Top 1, top 2 and top 3 systematically weighted nodes are respectively removed in specific simulation circumstances. For simplicity, systemically importance is evaluated by node degree. And the fragility of industrial complex network in such a deliberate attack is measured by the ratio of nodes going bankrupt and total node number.

$$F_T = B/N \,, \tag{7}$$

where F denotes fragility of industrial cluster, T represents simulation circumstance in which top 1, top 2 or top 3 systematically-weighted important nodes are removed by risk shock. B is the number of companys in the industrial complex network that go bankrupt after the ending of contagion process. N is the total number of companys at initial stage.

4 Experimental analysis

4.1 EXPERIMENTAL DATA

To examine the effectiveness of the method for assessing fragility of industrial complex network, this section evaluates fragility of an industrial complex network in Huzhou City of P.R. China. Huzhou city is located in the centre of Yangtze River Delta Economic Area. The city is composed of 6 districts with a total area of 5817 square Kilometres and a population of 2.54 million.

In its history, the city is known for its silk products and other textile products. A network dataset of Huzhou textile industrial cluster is employed in the experimental analysis. And visualization of the industrial complex network is illustrated in Figure 1.

All 27 nodes represent 27 textile companies in the city, which is interconnected with financial bonds. In here, financial bonds include share-holding relationship, debtor-creditor relationship and mutually credit guarantee. Financial information is collected from annual financial reports of these 27 companies, and their capacity to resist risk impact and financial robustness are evaluated based on their asset data and financial indicators.

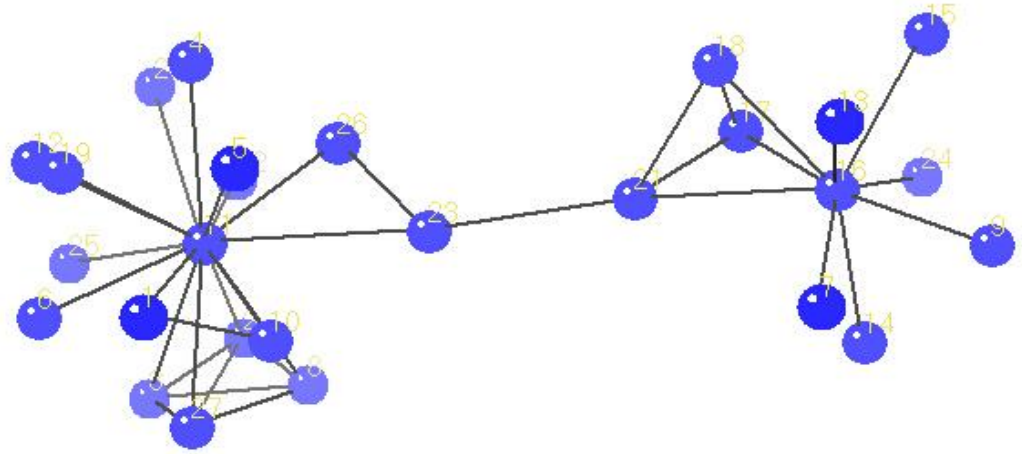


FIGURE 1 Visualization of the industrial complex network in Huzhou City

The strengths of financial bonds are calculated based on their financial releases about investments, liabilities and credit guarantees. The structure characteristics of the network of financial bonds are analyzed by UCINET 6.0, please see Table 1 for details.

TABLE 1 Network characteristics of the industrial complex network in Huzhou city

Network Characteristics	Value
Network centralization	55.08%
Network scale	27 (nodes)
Financial Bonds	73 (ties)
Average Degree	2.71
Network Density	0.1054
Clustering Coefficient	0.370

In the network, node code# 11 and node code# 9 is listed as systematically important nodes. Their share of degree is respectively 21.6% and 12.2% (see Appendix A). And there are many nodes are connected with the two nodes. If the two nodes are removed, these nodes will be disconnected.

4.2 RESULTS

In the experimental analysis, financial contagion process is simulated in a given experimental environment that is similar to real conditions in the Huzhou textile industrial cluster. Contagion effects trigged by each nodes are calculated (Table 2).

In the experimental analysis, the companies that are infected in the contagious process and finally go to bankrupt in the process are counted as infected number. These infected companies that have direct financial contact with the impacted node that is knock out by risk shock at initial stage are listed as number of directly infected. And in the experimental analysis, the industrial complex network is fragile, because 62.96% of nodes in the industrial cluster will go bankrupt in the contagion process if the node code # 11 is removed by risk shock. Furthermore, the whole industrial cluster will be crashed down if only the two nodes (code #11 and 16) are removed by risk shocks. In other aspect, the network is also robust because any other individual bankruptcy, excluding the two nodes (code #11 and 16), will not cause significant impact. Their *F* values are lower than 20%. Only deliberately risk attacks to node code #11 and 16 would cause serious financial panics leading to collapse of the whole network (Table 3).

TABLE 2 Simulation results of contagion process triggered by single bankruptcy

impacted node #	Infected number	Directly infected	Indirectly infected	*F* value	impacted node #	Infected number	Directly infected	Indirectly infected	*F* value
11	16	16	0	62.96%	6	0	0	0	3.70%
16	11	9	2	44.44%	12	0	0	0	3.70%
18	4	3	1	18.52%	9	0	0	0	3.70%
17	4	3	1	18.52%	7	0	0	0	3.70%
27	3	3	0	14.81%	5	0	0	0	3.70%
3	3	3	0	14.81%	19	0	0	0	3.70%
2	3	3	0	14.81%	20	0	0	0	3.70%
8	3	3	0	14.81%	15	0	0	0	3.70%
21	2	2	0	11.11%	22	0	0	0	3.70%
26	1	1	0	7.41%	4	0	0	0	3.70%
10	1	1	0	7.41%	24	0	0	0	3.70%
23	1	1	0	7.41%	25	0	0	0	3.70%
1	1	1	0	7.41%	13	0	0	0	3.70%
14	0	0	0	3.70%					

5 Conclusions

The paper has proposed a methodology to evaluate fragility of industrial cluster in the context of financial contagion. The evaluation is based on calculation of contagious effects triggered by bankruptcy of one or several companies and simulation of the contagion process. Huzhou textile industrial cluster is analyzed using the methodology proposed by the paper. The paper concludes that the industrial cluster is fragile, because deliberate attacks targeting only two specific companies will lead to collapse of the whole industrial cluster in the contagion process. And this experimental result supports existing evidence about risk evaluation of similar industrial clusters [6,11,12]. So, the methodology proposed in the paper is an effective way to evaluate fragility of industrial cluster in the context of financial contagion.

TABLE 3 Fragility assessments of deliberate attacks

Deliberate Attacks	Infected nodes (Bankruptcy)	F value Fragility Assessment
Top 1	16 nodes +1 nodes	62.96%
Top 2	25 nodes+2 nodes	100%
Top3	24 nodes+3nodes	100%

In the experimental analysis, all nodes are given an initial financial robustness value ($\beta_i(0)=1$ for all nodes). The financial robustness will be changed rapidly in the dynamic process of financial contagion. The simulation shows that 25 nodes out of total 27 nodes will not trigger substantial financial contagion. This is, the bankruptcy of these companies is not likely to cause financial panics in the industrial complex network, and lead to systematic risk.

In additional tests, the author found that in more severe circumstance, which means more average lower financial robustness, these nodes is still less likely to cause systematic risk. While, the two systematically important nodes (code No.11 and 16) is still more likely to cause large-scale financial contagion and lead to systematic risk. So, in the case of the experimental analysis, network structure is more important factor in determining the fragility of an industrial complex network. This argument is supported by many evidences in related literatures. In the experimental analysis, the paper ranks systematic importance of nodes according its position, especially its centrality degree. The experimental result shows this method is effective. And the logic underlying the

method is also connected with above argument, and supported by literatures.

In the experimental analysis, direct contagion is major effect in the contagion process. Only 3 nodes are cause indirect contagion. However, the author thinks indirect contagion is restricted by its star-centred structure in the experimental analysis. The trend reinforcement effect or financial accelerator plays important role in the contagion dynamics simulated in the experimental analysis. Indirect effects should not be undervalued.

Acknowledgements

This work is supported by The National Social Science Fund of China (Grant No. 14CSH071). The research leading to this paper is also supported by grant from of MOE (Ministry of Education in China) Humanities and Social Sciences Project (Grant No. 12YJC630229). This material is also based upon work funded by Zhejiang Provincial Natural Science Foundation of China under Grant No. LQ12G03003.

Appendix A Simulation procedure based on theoretical model (Figure 2)

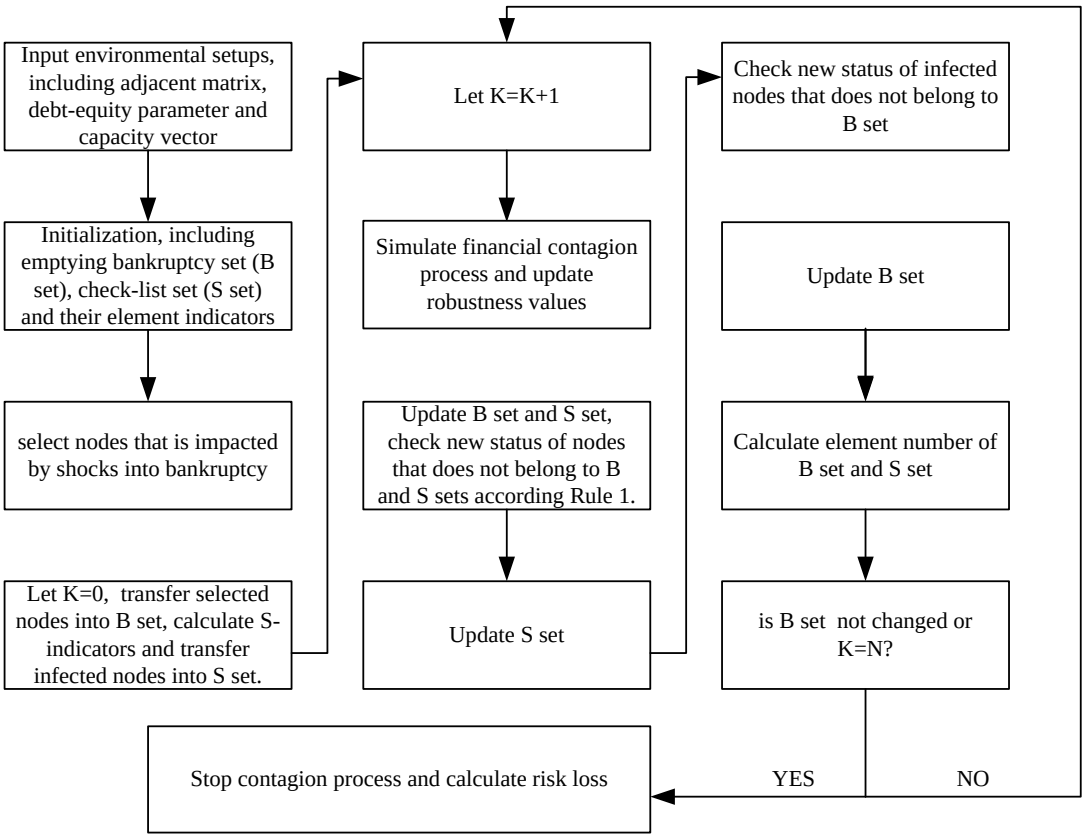


FIGURE 2 A simulation procedure based on theoretical model

Appendix B Degree distribution in the industrial complex network (Table 4)

Table 4 Degree ranking in the network of financial bonds in Huzhou city

Rank	Node Code	Degree	share	Rank	Node Code	Degree	share
1	11	16	0.216	15	12	1	0.014
2	16	9	0.122	16	9	1	0.014
3	27	4	0.054	17	7	1	0.014
4	3	4	0.054	18	5	1	0.014
5	2	4	0.054	19	19	1	0.014
6	21	4	0.054	20	20	1	0.014
7	8	4	0.054	21	15	1	0.014
8	17	3	0.041	22	22	1	0.014
9	23	3	0.041	23	4	1	0.014
10	18	3	0.041	24	24	1	0.014
11	26	2	0.027	25	25	1	0.014
12	1	2	0.027	26	13	1	0.014
13	10	2	0.027	27	14	1	0.014
14	6	1	0.014				

Appendix C Degree distribution in the industrial complex network (Table 5)

Table 5 Financial capacity and financial robustness used in the experimental analysis

Rank	Node Code	Financial capacity	Initial financial robustness	Rank	Node Code	Financial capacity	Initial financial robustness
1	16	20	1.00	15	9	1.25	1.00
2	11	11.25	1.00	16	17	1.25	1.00
3	12	5	1.00	17	23	1.25	1.00
4	18	5	1.00	18	7	1.25	1.00
5	6	5	1.00	19	15	1.25	1.00
6	1	5	1.00	20	4	1.25	1.00
7	22	5	1.00	21	24	1.25	1.00
8	5	5	1.00	22	25	1.25	1.00
9	2	3.75	1.00	23	13	1.25	1.00
10	19	3.75	1.00	24	14	1.25	1.00
11	20	3.75	1.00	25	26	1.25	1.00
12	3	2.5	1.00	26	27	1.25	1.00
13	21	2.5	1.00	27	10	1.25	1.00
14	8	1.25	1.00				

Note: Financial capacity is calculated and normalized based on financial information, including total asset and net asset of every nodes in the industrial complex network

References

[1] Allen F, Gale D 2001 *Journal of Political Economy* **108** 1-33
[2] Battiston S, Gatti D D, Gallegati M, Bruce C N, Greenwald J E Stiglitz 2009 *NBER Working Paper Series* w15611
[3] Lagunoff R, Schreft S L 2001 *Journal of Economic Theory* **99** 220-64
[4] Carlson J M, Doyle J 2000 *Physical Review Letter* **84**(11) 2529-32
[5] Carlson J M, Doyle J 2002 *PNAS* **99**(1) 2539-45
[6] Chi R, Wu B 2011 *JCIT* **6**(6) 221-30
[7] Upper C, Worms A 2004 *European Economic Review* **48**(4) 827-49
[8] Gatti D D, Gallegati M, Bruce C N, Greenwald A R, Stiglitz J E 2006 *Physica A* **370** 68-74
[9] Freixas X, Parigi B M, Rochet J C 2000 *Journal of Money, Credit and Banking* **32**(3) 611-38
[10] Wetter M 2011 *Journal of Building Performance Simulation* **4**(3) 185-203
[11] Guan H 2010 *Journal of Convergence Information Technology* **5**(7) 148-54
[12] Rodriguez R I, Jia Y 2011 *International Journal on Smart Sensing and Intelligent Systems* **4**(2) 325-37

Author

Bao Wu, September 1979, Dongyang, Zhejiang Province, P.R. China.

Current position, grades: associate professor in China Institute for Small and Medium Enterprises, Zhejiang University of Technology. Research fellow at Zhejiang Economic &Trade Polytechnic.
University studies: PhD degree at Zhejiang University of Technology.
Scientific interests: risk analysis, social complex network and system.

Study of collaborative relations through "Daily Maersk" service in China

Han Jingwei[1,2*], Owen Liefung Yue[3]

[1]*Beijing Jiaotong University, No. 3, Shangyuan Village Haidian District, Beijing, China*

[2]*Waterborne Transport Research Institute of Ministry of Transport of PRC, No 8,Xitucheng Road,Haidian District, Beijing, China*

[3]*Hong Kong Logistics Association, LG1, HKPC Building 78 Tat Chee Avenue, Kowloon, Hong Kong*

Received 1 October 2014, www.cmnt.lv

Abstract

Maersk Line has recently introduced a daily service in Asia-Europe route. This new service will bring in a revolutionary impact on the current fixed schedule shipping service in China. There are different views from China market on such daily service since the daily Asia-Europe service route calls mainly China's ports. Through the study of the 'daily Maersk' from 5 perspectives, namely shipping market demand, the required basic port infrastructure, relevant Government policies and regulations, operational aspects and low carbon emission, it is possible to reveal a kind of 'collaborative relations' in the China fixed schedule shipping market. The analysis adopts a qualitative approach via a questionnaire survey of selected experts. With the support of a theoretical mathematical model, it is possible to quantify the collaborative relations in different development phases in China. The study concludes that there are 3 broad phases of the development of collaborative relations and the current daily service emerges in the 'new development era' phase. It is because there are sufficient freight and adequate port infrastructure. The daily service outperforms the general industry practice under the "Rotterdam Rules" on carrier's responsibility and liability and injects contemporary logistics management concepts in the industry. Moreover, it is in the same pace of current China's requirements on energy saving and carbon emission reduction targets.

Keywords: 'Maersk', collaboration, collaborative relations, Asia-Europe

1 Introduction

Maersk Line has introduced a new 'daily Maersk' service for the Asia-Europe route starting in Shanghai [1] from November 2011. The daily service borrows the idea of fixed schedule service from air transport service. It allows shippers to have a better control on their delivery schedules. However, there are some concerns about the provision of daily service within the China shipping industry. The port conditions and level of shipping service have not developed evenly in China to support seamless freight flow for foreign trade. Can the Maersk's daily service in China survive in the short and long run? Through the study of the collaborative relations, the feasibility of providing daily service will be assessed by looking into 5 perspectives, namely shipping market demand, calling port infrastructure, Government policies and regulations, shipping operational factors and carbon emission. A mathematical model has been established to assist in the analysis of different phases of collaborative relations development in China. Finally it is possible to conclude that there are general 3 broad phases in the development of collaborative relations in China shipping industry. The introduction of the 'daily Maersk' is subsumed in the third phase and it should be able to survive. Some enhancements for such daily service will be identified and recommended at the end of this study.

2 Concerns about daily service

This daily service involves the 4 calling ports in Asia with three ports in China (Ningbo, Shanghai, Yantian) and one in Malaysia (Tanjung Pelepas) whereas the European ports at Philip Ke Situo (UK), Rotterdam (Netherlands), and Bremen (Germany). For shipment with Maersk Line scheduled service, it is expected to be on-time but the current 50% on-time is common in the China shipping market. The poor service level forces the shippers to adjust their production schedule as well as its supply chain. The daily service changes the operation of the shipping industry and the shippers' inventory planning. It creates a new standard and marks a milestone. Shippers can now benefit from better inventory control and bank finance. At the moment of writing, the booking of the daily service is getting momentum and growing gradually.

However, the concerns are merger of the 'big' players in order to share their resources to strike against this new challenge. For the small and medium shipping companies, they would easily go broke due to high cost of running daily service and possible demurrage. Subsequently, a new wave of merger and takeover could eliminate some existing players but Maersk Line will secure its monopolistic position in the China shipping industry. On the other hand,

[*] ***Corresponding author's e-mail:*** 552186316@qq.com

the Chinese shipping industry is still developing and the level of ancillary service and conditions may not be able to match with Maersk Line's expectation. Can the Maersk's daily survive in the short and long run? The development of Chinese ports and the supporting service could be a critical factor.

3 Influential factors

3.1 SHIPPING MARKET DEMAND IN CHINA

Would there be sufficient volume of freight to support the development of the Maersk's daily service? The financial tsunami in 2008 had significant impact on Sino-Europe trade. However, after 3 years of recovery in world economy, the trade volume has rebounded and exceeded the level in 2008. According to the 2008 and 2010 China Customs statistics [3] and the internal data analysis (FOXPRO program) of the China Waterborne Transport Research Institute, the container trade between China and the Europe are summarised as follows:

1. In 2010, the trading level reached the value of US$573.1 billion with the export value of US$355.2 billion and import value US$217.9. When comparing with the 2008 level, the 2010 level increased by 12.2%, 3.6% and 29.6% respectively.
2. The demand for container trade is about a total of 11 million TEUs and export from China to Europe with 8.3 million TEUs and 3.2 million TEUs vice versa. It is a huge imbalance of trade. The major sources of TEUs for the Sino-Europe route are coming from 7 provinces in China (Guangdong, Zhejiang, Jiangsu, Shandong, Fujian, Shanghai and Hebei) counting towards 87% of the total (see Table 1 below).
3. In the Asia-Europe route, Maersk will put in 70 ships in the "daily Maersk' service. The capacity is around 8,500 TEUs to 15,000 TEUs. It carries about 57,000 TEUs per week or about 3 million TEUs per year. According to the above Table 1, the annual demand for container transportation will be about 7.7 million TEUs from China to Europe. The Maersk is sharing about 40 percent of Sino-Europe trade so there should be sufficient freight for 'daily Maersk'. Maersk is taking a strategic move in the vulnerable shipping market (after the financial tsunami) in order to scramble for more market share. Other Chinese shipping companies need not be panic and it is just to make sure that their service level could be improved and reliable, rather than simply competing on price in the sea freight.

TABLE 1 Sources of freight (TEUs) in Sino-Europe container trade

Province	**%**	**Province**	**%**
Guangdong	27%		
Zhejiang	19%		
Jiangsu	16%		
Shandong	9%		
Fujian	7%		
Shanghai	5%		
Hebei	4%		
Sub-total	87%	Others	13%

Source: estimates from 2010 China Customs statistics

4. The unit value of the container trade between China and Europe after financial tsunami has increased as China has undergone structural change in the industry although the container trade volume is decreased (-12.8%) from year 2008 of 225.2TEU / billion U.S. dollars down to year 2010 of 199.6TEU / billion U.S. dollars shown in Table 2. Comparing the year 2010 and 2008, Guangdong, Jiangsu, Zhejiang and Shanghai (municipal) forerun in the adjustment of industrial structure. As long as the world economy is not fluctuated significantly, the demand for container transportation between China and Europe is increasing.

TABLE 2 Sino-Europe container trade in 2008 and 2010

Year	**2008**			**2010**		
	Exp	*Imp*	*Exp-Imp*	*Exp*	*Imp*	*Exp-Imp*
China	2587	1570	2252	2331	1450	1996
Jiangsu	1977	1872	1954	1673	1372	1595
Zhejiang	3406	3023	3354	2540	2034	2459
Shanghai	1290	802	1084	984	534	763
Guangdong	3268	2547	3081	3874	2625	3103
Total	**12528**	**9814**	**11725**	**11402**	**8015**	**9916**

Source: projection from the China Customs data unit: TEU/ Billion US$

3.2 BASIC INFRASTRUCTURE

3.2.1 Shanghai Port

Shanghai Port refers to container terminals located in Yangshan, Waigaoqiao, and Wusong involving a total of 46 container berths, 155 container quay cranes and 463 gantry cranes, and the yard area of 6.34 million m2. All these container terminals are included as the Shanghai Port integrated by sea and land shuttle service and 'daily Maersk' uses Yangshan Container Terminal (Table 3).

TABLE 3 Yangshan container terminal

	Mingdong	Shengdong	Guandong	Total
Quay Length(m)	1110	3000	2600	6710
No. of berths	4	9	7	20
Water depth(m)	–12.8	–16	–17.5	
No. of cranes	16	34	26	76
No. of RTG	48	108	82	238
No. of Stackers	2	5	6	13
Yard area (million m^2)	0.82	1.49	1.42	3.73
Warehouse space (10,000 m^2)	0.95	n.a.	n.a.	0.95
Annual capacity (10,000 TEUs)	70	430	500	1000

Source: website of port of Shanghai at www.portshanghai.com.cn [4]

3.2.2 Ningbo Port

Ningbo Port refers to Beilun, Chuanshan and Daxie container terminals with total 14 berths and quay length of 4,733 meters. The basic infrastructure for these container terminals are summarised in Table 4.

TABLE 4 Ningbo Port

Container terminal	Berth capacity	Quay length	Berth	Water depth
	(10,000 tonnes)	(m)	(no.)	(m)
Beilun	5~10	2138	7	13.5~15
Chuanshan	10	1785	5	17
Daxie	10	810	2	17.5

Source: website of Ningbo Port at www.nbport.com.cn [5]

3.2.3 Shenzhen Port

Shenzhen Port refers to Shekou, Chiwan and Yantian container terminals with total 44 container berths and a capacity of 18.91 million TEUs.

3.3 GOVERNMENT POLICY AND REGULATIONS

Maersk Line has been keen to maintain close relationship with China Customs to ensure the smooth operation. China Customs is also willing to support Maersk Line via Daxie's Customs which has set up an innovative high-tech regulatory model. The monitoring process is through online data exchange between the Customs and Maersk. Provided that the information is correct, the rebate for the export tax is fast tracked and customs clearance is greatly improved. This is one of the reasons that Maersk Line strategically positions its daily Maersk service in Ningbo Port.

A major possible obstacle in hindering 'daily Maersk' service is the collaboration with different local governments along their voyage when ship replacement is needed. On the other hand, the international regulation "Rotterdam Rules" may have impact on China shipping industry as China has not signed in yet. This does not affect Maersk Line as they have already fixed the demurrage of US$100 per container for delay of first 3 days and US$300 per container for the fourth days or more.

3.4 OPERATIONAL ASPECTS

The ocean container transportation has significantly changed the pattern of world trade. In the past decades, the development of China shipping industry has however no major breakthrough. The unreliable shipment has been a common criticism in the China shipping industry and only 50 percent arriving on time. Every year, China has about a total of 11 million TEUs 'in' and 'out' from/to Europe excluding the transfer of the empty container which adds up to around 15 million TEUs. Maersk Line has offered a fixed schedule daily service in the China shipping market with fixed demurrage. The daily service does not just relieve the pressure of the shippers' inventory control but also enhances the competitiveness of the manufacturing companies in China. This service will change the weakest link in the supply chain into a powerful link. The recent observation is about 95% on-time which projects an average saving between US$200 to US$400 per container.

3.5 ENERGY SAVING AND REDUCTION OF CARBON EMISSION

The trend to change the economic development structure so as to achieve resources saving and carbon emission reduction has been stated in the national policy level since the '11th Five-Year Plan'. The energy consumption in the operation unit of the ship transportation and port throughput fell by 7% and 4% respectively and achieved the related reduction target in energy-saving and carbon emission. In the '12th Five-Year Plan' i.e. the period up to 2015, the proposed energy consumption for the unit of the ship transportation and port throughput to be decreased by 15% and 8% respectively compared to 2005, carbon dioxide emissions would be decreased by 16% and 10% respectively [6].

'Daily Maersk' service is at 13% carbon dioxide emissions lower than the industry average per container with its newly investment in the twenty '3E' class ships to run on the Asia-North Europe route would become the world's most energy-efficient route.

4 Phases of Collaborative Relations

'Collaboration' concept in this study refers to the balancing various needs in the shipping industry via positive and constructive relations. It is a dynamic, interactive and complicated process in order to materialize collaborative relations. The 5 perspectives elaborated in the previous sections would be quantified in a mathematical model below to differentiate the development of collaboration in the China shipping industry at different stage if any. The collaborative relation is therefore quantified through a theoretical mathematical model as follows [7]:

$$F_k = A_k C_k^k \left(X_1,\ X_2,\ \ldots,\ X_i\right) + b, \tag{1}$$

where F_k is the collaborative relations 'k' phase. Collaborative relations phase F_k is directly proportional to the collaborative degree A, which increases as the F_k increases:

$$\begin{cases} F_1 = A_1 C_1 (X_1,\ X_2,\ \ldots, X_i) + b \\ F_2 = F_1 + A_2 C_2^2 (X_1,\ X_2,\ \ldots, X_i) \\ F_3 = F_2 + A_3 C_3^3 (X_1,\ X_2,\ \ldots, X_i) \\ F_4 = F_3 + A_4 C_4^4 (X_1,\ X_2,\ \ldots, X_i) \end{cases}, \tag{2}$$

If $F_1 \leqq F_k < F_2$: the collaborative relations are under an incubation stage;

If $F_2 \leqq F_k < F_3$: collaborative relations in the developing phase;

If $F_3 \leqq F_k < F_4$: collaborative relations in a new era stage where a new business environment emerges:

1) 'A' represents the degree of collaboration subject to factors X_1, X_2, …, X_i (see section 3 above) i.e. different phases when 'A' has developed into different state, then the collaborative relations has evolved into a new environment and new collaborative relation phase; and the value of A_1, A_2, …, A_k can be obtained through a "power function" method:

$$d_i = e^{-e^{\frac{x_i - s}{s-n}}}.$$

When x_i reaches the threshold between the uncoordinated and basic coordination, the A_2 is: d_2=0.3679. When x_i reaches the critical value between the coordination and basic coordination, the A_3 is: d_3=0.6922. So the boundary value of A, that is 0, 0.3679, 0.6922 and 1 respectively, assessments using the Equation (2), calculating the boundary value of F_k when A_i at various degree (A_1=0, A_2=0.3679, A_3=0.6922, A_4=1), 'C' value is 1; and boundary value with no randomness, 'b' = '0' with 'i' factors, therefore:

$$\begin{cases} F_1 = 0 \\ F_2 = F_1 + 0.3679(4i) = 1.4716i \\ F_3 = 1.4716i + 0.6922(8i) = 7.0092i \\ F_4 = 7.0092i + 10i = 17.0092i \end{cases}.$$

2) 'C' represents the degree of collaboration in different phase;

3) 'B' represents the random impact factor; this is to cancel out the random effect of Xi. It can be resolved by curve fitting and i and n are natural number. In order to limit the calibration in this study, the values of C_1, C_2, …, C_k are assumed to be '1' for equal weighting of importance.

5 Expert Assessment on Collaborative Relations

5.1 EXPERTS AND QUESTIONNAIRE

A questionnaire survey to assess the rating of the 5 selected factors was issued to the experts from China Waterborne Transport Research Institute (WTI), China Ocean Shipping Company (COSCO), Tianjin Port (TJ), Shanghai Customs (S Customs), China Maritime Safety Administration (China MSA), Shanghai Maritime University (SMU). The rating provided by the experts were separated in 3 periods within the last 20 years: years 1990~2000 (S1), 2001~2005 (S2), and 2006~2010 (S3); And the feedbacks were summarized in Table 5 below:

TABLE 5 Rating of Experts on the 3 Collaborative Relations

Experts	X_1			X_2			X_3			X_4			X_5		
	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_1	S_2	S_1	S_2
WTI	2	4	9	1	5	10	0	5	8	0	5	0	5	0	5
COSCO	1	5	8	1	5	8	1	6	8	1	4	1	4	1	4
TJ	2	4	8	1	4	8	2	5	8	1	6	1	6	1	6
S Customs	3	5	10	1	4	9	1	4	8	0	5	0	5	0	5
China MSA	3	5	9	1	4	9	0	4	8	0	4	0	4	0	4
SMU	1	4	8	1	4	9	1	5	8	0	3	0	3	0	3

In Table 5, the experts rated the 5 factors including market demand, basic infrastructure, Government policies and regulation, corporate operational aspect and reduction in carbon emission in the range of '0' to '10' where '10' the highest rating and '0' the lowest relations. In this study, there are 5 factors so i=5, therefore:

$$\begin{aligned} F_1 &= 0 \\ F_2 &= 1.4716 \cdot 5 = 7.358 \\ F_3 &= 7.0092 \cdot 5 = 35.046 \\ F_4 &= 17.0092 \cdot 5 = 85.046 \end{aligned}.$$

The theoretical value of the collaborative relations is calculated by averaging the total expert rating in a period for each factor (Table 6).

$$S_{tvi} = \frac{1}{n}\sum_{k=1}^{n}\frac{\overline{s_{tvki}}}{n}. \quad (3)$$

S_1 period: *WTI* (X_1, X_2, X_3, X_4, X_5)=(2, 1, 0, 0, 0), $S_{tv1} = X_1 + X_2 + X_3 + X_4 + X_5 = 3$, and so on as in Equation (3): S_{tv1} = theoretical value = (3+4+6+5+4+3)/6=4.2

TABLE 6 Theoretical Value of Collaborative Relations at each period

Experts	S_{tvi}		
	S_{tv1}	S_{tv2}	S_{tv3}
WTI	3.0	24.0	43.0
COSCO	4.0	26.0	40.0
TJ	6.0	23.0	40.0
S Customs	5.0	24.0	44.0
China MSA	4.0	22.0	41.0
SMU	3.0	20.0	41.0
Theoretical value	4.2	23.2	41.5

5.2 THREE PHASES OF COLLABORATIVE RELATIONS

1. In 1990~2000, S_{tv1}=4.2, $F_1 < S_{tv1} < F_2$, the collaborative relations is under an incubation stage. It is the beginning of the container transport market and just receiving attention. The market demand is not strong whereas the container transport infrastructure was poor. The Government policies and regulations did not provide support and there was little effective operation concept. The requirement was basically 'shipped' service since China was in the economic development stage and had little attention on energy saving and environmental protection. It is therefore the expert assessment were in the lowest range '0' for all factors X_1, X_2, …, X_i.
2. In 2001~2005, S_{tv2}=23.2, $F_2 < S_{tv2} < F_3$ collaborative relations in the developing phase together with the rapid development in China's economy and foreign trade (import and export), and the improvement in container transport security, the popularity of using container transport was gaining momentum. Subsequently container transport facilities and infrastructures as well as the Government policies and regulation were gradually matching with the development of container transportation. Furthermore, the introduction of contemporary logistics concepts, national concerns in energy saving and environmental protecttion etc, has put container transport under the spot light. The shippers required more and requested 'shipped well' service. In this phase, the experts rated the factors, X_1, X_2, …, X_i around the mid way 5 points. In this stage, the shipping companies began with strategic alliances to share the capacity among themselves and offer new services in order to reduce the risk. Therefore, this is the developing stage.
3. In 2006~2010, S_{tv3}=41.5, $F_3 < S_{tv3} < F_4$, collaborative relations in a new development era phase where a new business environment emerges. Under this phase, the industrial structure has been adjusted and the economy has recovered from the financial turmoil. There is growing demand in container transport in the foreign trade so the port facilities have also been improved accordingly in particular the improvement in the efficiency of the Customs so as to support the foreign trade. The "Rotterdam rules" gradually attracted the worldwide attention even the shippers require not only "shipped" but "shipped well" so as to achieve environmental benefits. The International Maritime Organization (IMO) adopted the MARPOL 73/78 Convention, marking the official commencement of the low-carbon era. It is therefore the experts rated the factors X_1, X_2, …, Xi for 8 points and more in most cases.

6 Conclusions and Suggestions

6.1 CONCLUSIONS

The three phases of collaborative relations were identified through a mathematical model. It highlighted that the 'Daily Maserk' would definitely bring in a breakthrough in the China fixed schedule shipping service. It is certainly a new standard of service for the shipping industry and raises the level of service to record high. According to the analysis of the collaborative relations, it is concluded that Maersk has launched the 'daily Maersk' service in China's third phase of collaborative relations (new development era phase) so:

1) It has sufficient freight from China calling ports.
2) It has some calling ports with adequate port infrastructure.
3) It can outperform China's requirement in the carrier's responsibility and liability ("Rotterdam Rules").
4) It can inject contemporary logistics concepts into the China ship management practice.
5) It can keep up with the same pace of China's energy saving and reduction of carbon emission.

6.2 SUGGESTIONS

The study reveals that the demand for Sino-Europe route is 11.5 million TEUs in 2010 in which 8.3 million TEUs from China to Europe and 3.2 million vice versa. This serious imbalance generates five million empty containers to be reshuffled. The Maersk operation management should pay attention to this imbalance.

'Daily Maersk' can handle about 3 million TEUs in the Asia-Europe route each year. There is a strong demand on road distribution which may bring in a heavy traffic pressure on the calling ports and creating pollution problem therein. Although a carbon emission of "Daily Maersk" service at its sea link is low, there are more carbon emissions in the upstream of the road transportation chain. Would it be beneficial to use more river fee-

der service? On the other hand, it is recommended that 'daily Maersk' service could be extended to the Taicang Port and Nanjing Port so as to reduce urban traffic and exhaust pollution in Shanghai, as well as to provide shippers in Jiangsu Province with direct service.

References

[1] Daily Maersk service commenced on 10.11.2011 at http://www.100allin.com/topic/maersk20111011/ *(in Chinese)*
[2] http://baike.baidu.com/view/1646035.htm *(in Chinese)*
[3] Internal raw data for 2008 and 2010 Chinese Customs statistics
[4] http://www.portshanghai.com.cn *(in Chinese)*
[5] http://www.nbport.com.cn/portal/wps/portal/!ut/p/c5/04_SB8K8xLL *(in Chinese)*
[6] http://www.wti.ac.cn/article_video.aspx?menuid=1401&tab=tab_video&tabid=19 *(in Chinese)*
[7] Gaoaiying 2009 Research on the Key Issues of Enterprise Purchasing in Supply Chain Environment – Doctoral Dissertation **4** 55-6 Beijing Jiaotong University *(in Chinese)*

Authors

Jingwei Han, born in November 1979, China.

Current position, grades: associate researcher.
University studies: Beijing Jiaotong University.
Scientific interests: water transportation strategy and regulation; water transportation; modern logistics.
Publications: 15.
Experience: Waterborne Transport Research Institute of Ministry of Transport of PRC since 2001.

Owen Liefung Yue, born in November 1956, Australia.

Current position: Chief Town Planner, PhD, MSc., MTCP, MMgt, FHKIP, FHKLA, MHKIE.
Scientific interests: land use and transport modelling; supply chain.
Publications: 9.
Scientific interests: land use and transport planning, supply chain management consultancy.
Experience: Planning Department, HKSAR Government (1998-2013). Henderson Land Development Company Limited since 2014.

Empirical analysis on influencing factors of capital structure of China's real estate listed company: evidence from Chinese listed company

Zhuo Yang[1*], Hong-liang Qiu[2]

[1]*School of Business Administration, Zhejiang Gongshang University, Hangzhou, China*

[2]*Department of Business Administration, Tourism College of Zhejiang, Hangzhou, China*

Received 1 March 2014, www.cmnt.lv

Abstract

Capital, the guarantee for normal operation of an enterprise, is crucial to production. Therefore, optimizing the capital structure has been an important task in the development of enterprises. Combining with VAR model, this paper selects 7 factors, which are profitability, operation ability, the current debt servicing ability, development ability, tax, strategic position and the assets structure, so as to make an exploratory study on the capital structure of Chinese real estate industry, based on the data from June 2002 to December 2012. The study shows that the Liquid ratio and the gross profit are of great effects on the capital structure of a listed company.

Keywords: capital structure, listed company, the real estate, impulse responses

1 Introduction

Capital structure refers to the composition and relationship of capital. Research on the capital structure of the real estate industry mainly concentrates in the issue of financing and capital structure. Gau and Wang (1990) found that the liability of real estate Company, which has a positive effect on investment cost, has a negative effect on the expected cost of non-debt tax shield, market interest rates and financial crisis. Chen and Xia (2006) studied on the financing behaviour and financing options of the real estate company. They found that asset-liability ratio, which has a positive effect on firm size, has a negative effect on operation ability and ownership concentration. In addition, asset-liability ratio does not have a significant relationship with the profitability, solvency, and non-debt tax shields.

2 Data processing

Existing studies on influencing factors of capital structure contain the industry ones. However, there is not a confirmative relationship between capital structure and industry. Many scholars have found that capital structure has significant industry difference (Scott, 1972; Scott and Martin, 1975; Bradley, Jarrel and Kim, 1984; Liu, 2003; He, 2005; Tan, 2005). However, some studies have quite contrary conclusions (Hong and Shen, 2000). This study makes an analysis on the influencing factors in Chinese listed Company of real estate industry only.

2.1 THE SELECTED VARIABLES

We design dependent variable and independent variable as follow:

1) The dependent variable.

We adopt asset-liability ratio (total liability/ total assets), equity ratio (total liability/ Owners' Equity) and $\frac{\text{current liabilities}}{\text{total assets}}$ to measure company's capital structure.

2) The independent variable.

This paper has selected 7 factors, which is profitability, operation ability, the current debt servicing ability, development ability, tax, strategic position and the assets structure (Feng, Wu and Liu, 2000; Xiao and Wu, 2002; Zhou and Xu, 2012), shown as Table 2.1.

[*] ***Corresponding author's*** e-mail: extraterrestrial@163.com

TABLE 2.1 Indicators and illustration

Influencing factor	Index	Simbol	Illustration
Profitability	Gross profit	XSMLL	Profit from operation/ Operating income
Operation ability	Turnover rate of return on common stockholders' equity	GDSYZZL	Revenue / Net assets
The current debt servicing ability	Liquid ratio	LDB	Current Assets / Current liability
Development ability	The growth rate of net profit	JLRZZL	(Current net profit –The net profit for last period)/ The net profit for last period
Tax	The actual tax rate	SJSL	The actual expense of tax /Total profit before tax
Strategic position	Enterprise size	QYGM	ln(Revenue)
The assets structure	The percentage from fixed assets to the total assets	GDZCB	Fixed assets at the end of the period /Total assets at the end of the period

2.2. DATA SCREENING OF THE SAMPLE

According to <the classification guide of listed Company>, which is issued by the National Commission in April 2001, this paper selects the whole data of Shanghai and Shenzhen listed Company of the real estate industry, dating from June 2002 to December 2012. This paper excludes ST companies and incomplete ones, getting 42 companies. After standardization, we make descriptive statistics of variables shown as Table 2.2:

TABLE 2.2 Descriptive statistics

Variable	Mean	Standard error	Median	Minimum	Maximum
XSMLL	0.038252588	0.228387053	0.145544823	-0.514002733	0.214036792
GDSYZZL	0.589755935	0.255617372	0.56523868	0.278723053	1.042955974
LDB	2.152296439	0.255165836	2.152734428	1.785741754	2.7967487
JLRZZL	1.06783448	3.231546326	-0.255036358	-1.036384764	13.22693968
QYGM	20.22273148	0.729246252	20.22796994	18.86080145	21.43912173
GDZCB	0.078978064	0.051225492	0.078812145	0.021866854	0.177129085
SJSL	0.263949435	0.041329917	0.26732886	0.180939192	0.341066238

(Date from: annual report of SSE and SZSE listed companies)

3 Processing and analysis of data

Making a second order autoregressive analysis from every influencing factors to asset-liability ratio, equity ratio and, we can obtain a VAR estimation result, shown as Table 3.1.

TABLE 3.1The fitting coefficients of VAR regression

	Modified coefficient of determination of the asset-liability ratio	Modified coefficient of determination of the equity ratio	Modified coefficient of determination of $\frac{\text{current liabilities}}{\text{total assets}}$
XSMLL	0.832302	0.512510	0.390131
GDSYZZL	0.923190	0.910093	0.904657
LDB	0.187130	0.313559	0.235956
JLRZZL	0.834355	0.125683	-0.087677
QYGM	0.846181	0.923855	0.935781
GDZCB	0.845998	0.859906	0.874597

It can be seen from Table 3.1 that, the fitting to capital structure from liquid ratio and the growth rate of net profit is low, while other influencing factors have a significant fitting result.

3.1 ANALYSIS OF IMPULSE RESPONSES

Based on VAR estimation, we can make a static simulate shown as Figure 3.1:

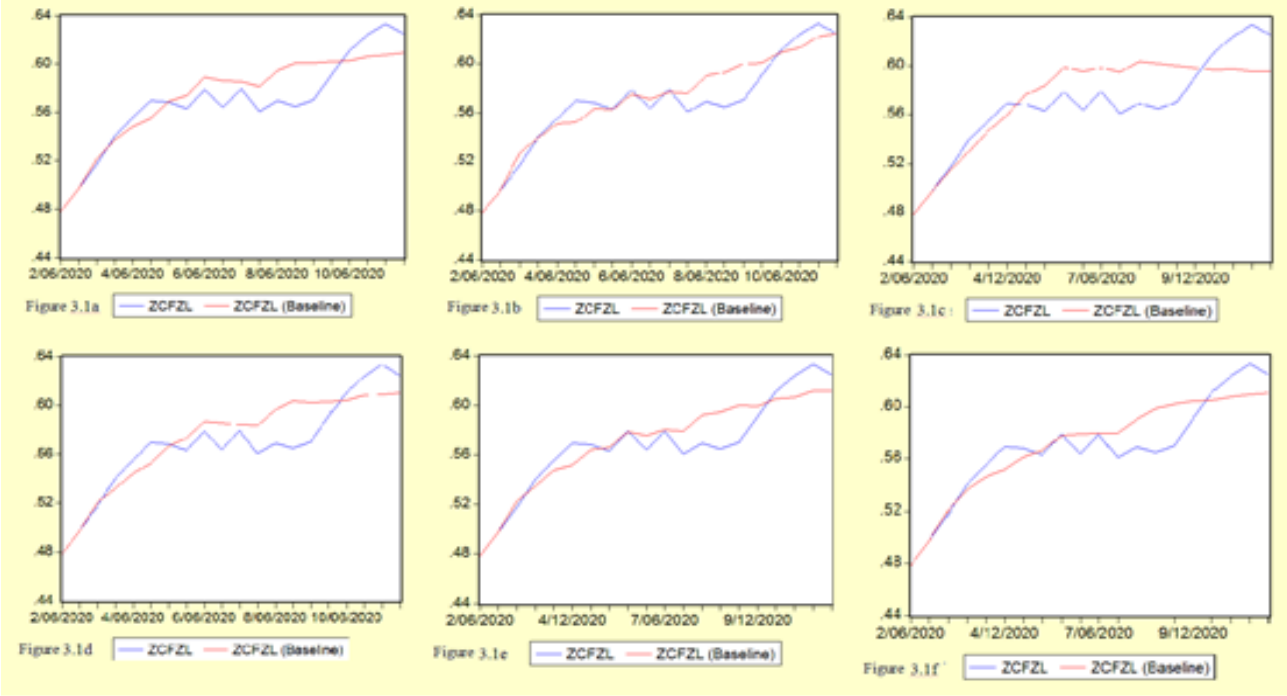


FIGURE 3.1 Static simulation from every influencing factor to asset-liability ratio

It can be seen from Figure 3.1 that, the simulation result is fine. Based on VAR model, we make an analysis of impulse response function of every factor. After giving a unit of residual shock to each factor, we could get an impulse response result shown as Figure 3.2. The horizontal axis represents the period of impulse response, while the vertical axis represents a unit of response to a residual shock. The solid line, which represents the function of impulse response, is on behalf of the response for every factor to a shock of asset-liability ratio residual.

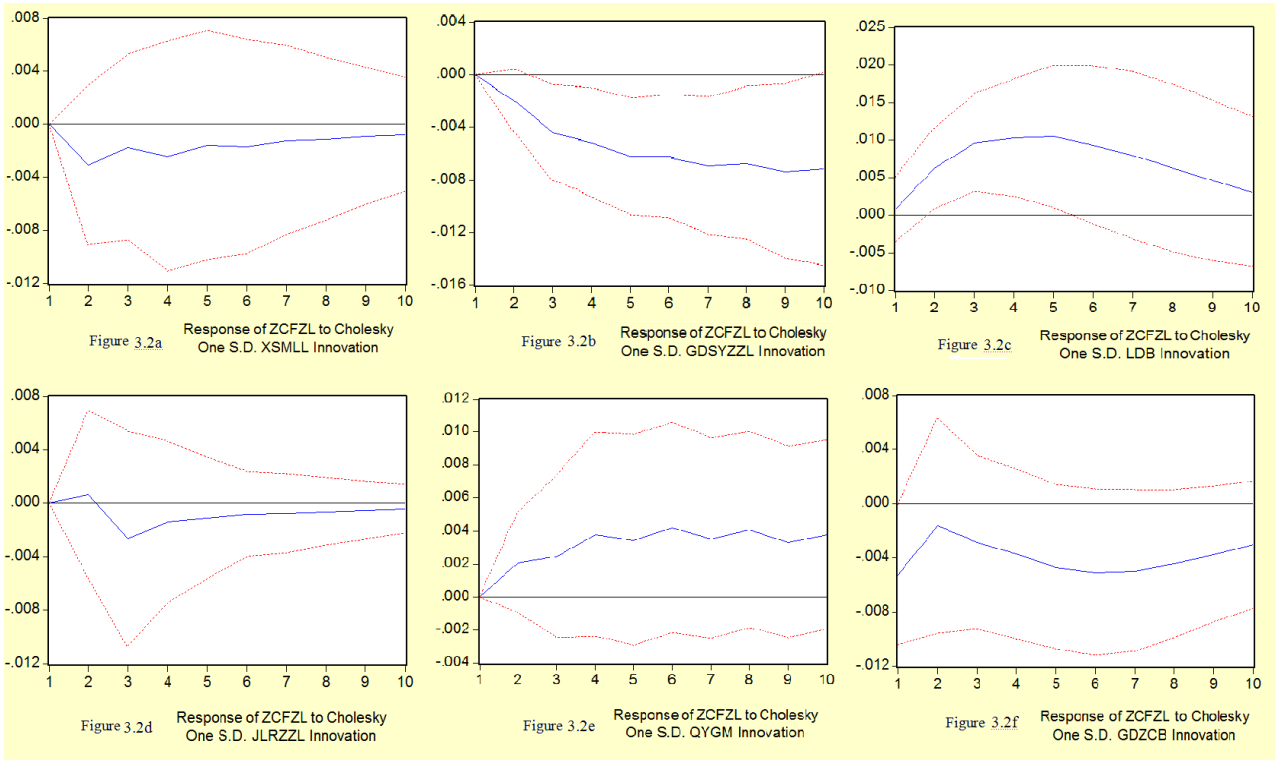


FIGURE 3.2 Impulse responses analysis from the influencing factors to asset-liability ratio

When it gives a residential shock to gross profit or the growth rate of net profit, asset-liability ratio will soon achieve convergence. Meanwhile, the shock from turn-over rate of return on common stockholders' equity, liquid ratio, enterprise size, and the percentage from fixed assets to the total assets to asset-liability ratio will not achieve convergence in the short period. We could see that in Figure 3.2.

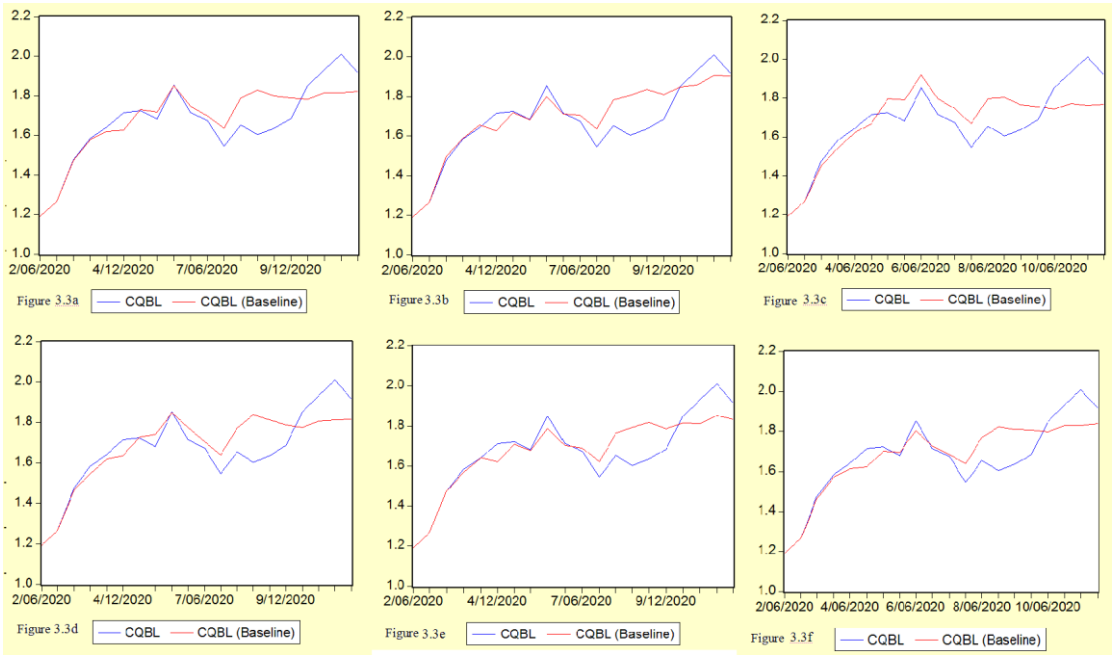


FIGURE 3.3 Static simulation from every factor to the equity ratio

In VAR simulation, every factor fits well to the equity ratio.

When it gives a residential shock to gross profit, the growth rate of net profit, or the percentage from fixed assets to total assets, the equity ratio will soon achieve convergence in current period. Meanwhile, other factors cannot achieve convergence in short period. We could see that phenomenon in Figure 3.4 as follow:

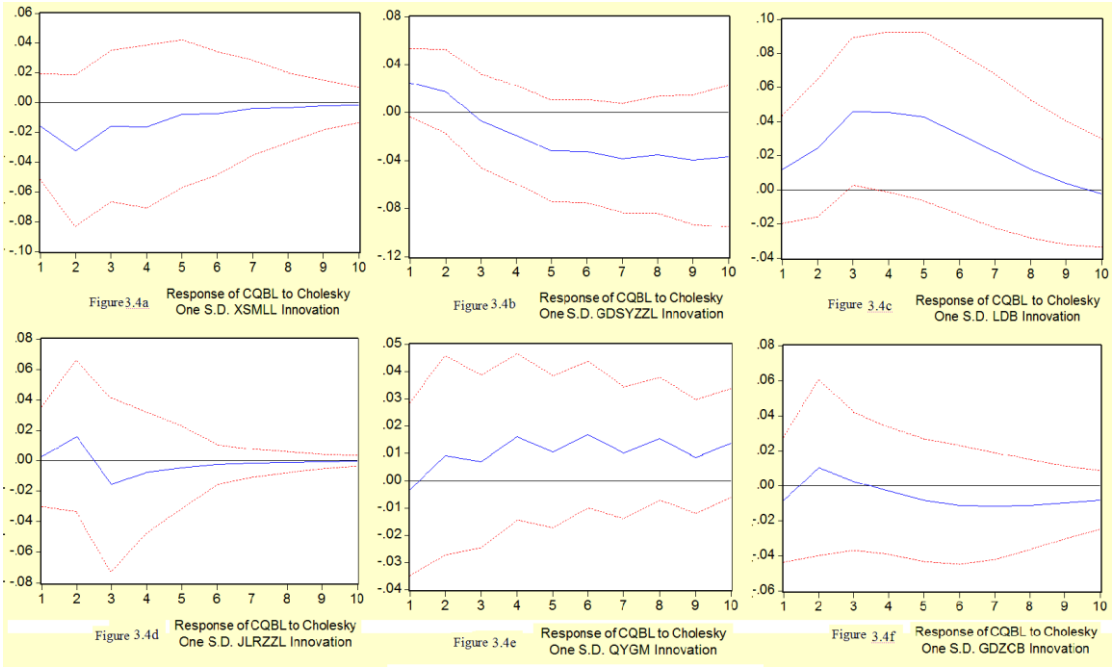


FIGURE 3.4 Impulse responses analysis from the influencing factors to equity ratio

After the VAR analysis from each influencing factor to $\frac{\text{current liabilities}}{\text{total assets}}$, we can make a static simulation. We find that every factor fits good, shown as Figure 3.5:

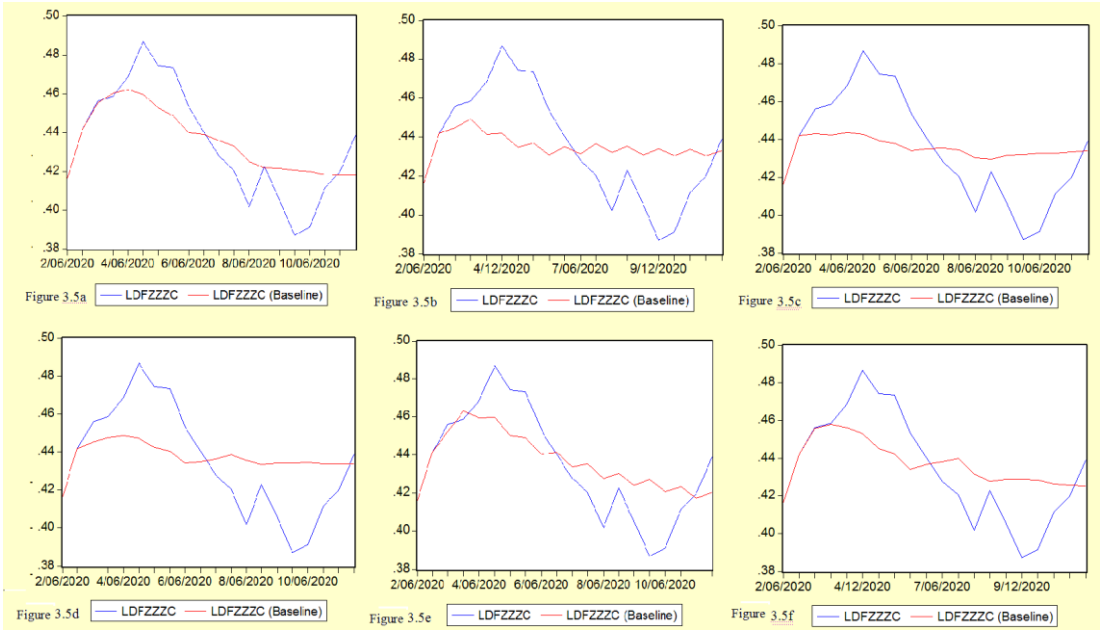


FIGURE 3.5 Static simulation from every factor to $\frac{\text{current liabilities}}{\text{total assets}}$

We make an analysis of impulse responses for every influencing factor to $\frac{\text{current liabilities}}{\text{total assets}}$, shown as Figure 3.6:

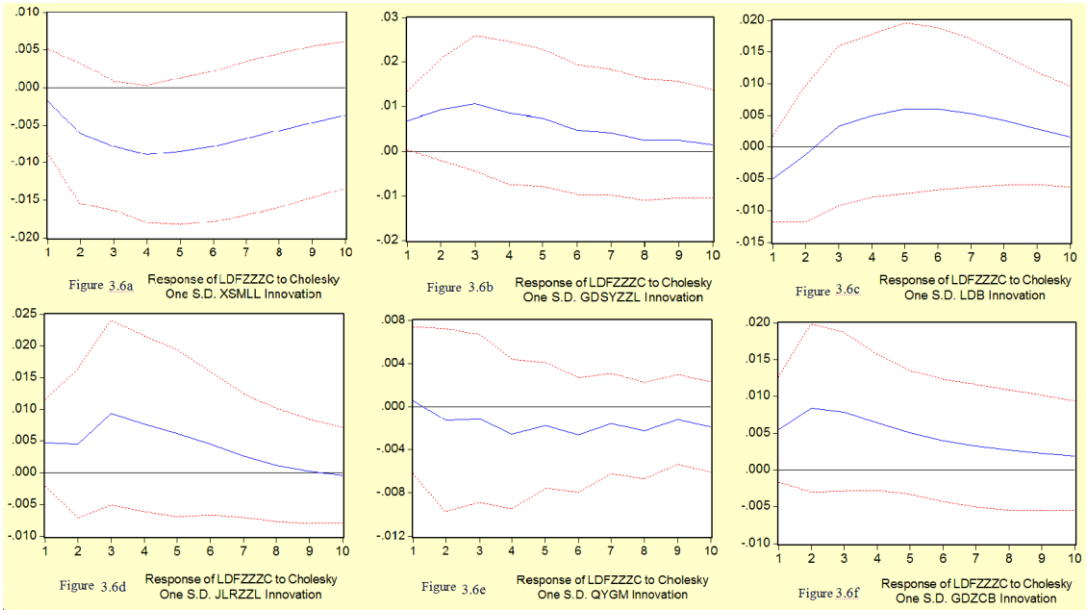


FIGURE 3.6 Impulse responses analysis from the influencing factors to $\frac{\text{current liabilities}}{\text{total assets}}$

When it gives a shock to influencing factors, $\frac{\text{current liabilities}}{\text{total assets}}$ will not achieve convergence in a short period, shown as Figure 3.6.

Therefore, we can see from the analysis of impulse responses that turnover rate of return on common stockholders 'equity, enterprise size, and the percentage

from fixed assets to the total assets could impact capital structure largely, while gross profit can just impact capital structure in a short time.

3.2 VARIANCE DECOMPOSITION

We decompose variance, shown as in Table 3.2 as follow:

TABLE 3.2 the results of variance decomposition

Responsing variable	periods	Impulsing variable					
		XSMLL	*GDSYZZL*	*LDB*	*JLRZZL*	*QYGM*	*GDZCB*
ZCFZL	1	0.000000	22.224570	0.840752	0.000000	0.000000	20.498520
	2	3.125139	16.144570	27.979700	0.144896	1.558124	13.097470
	3	2.782470	16.123960	53.527540	1.682919	2.931824	12.807550
	4	3.319327	21.732530	67.078600	1.822697	6.227982	15.699750
	5	3.310592	30.848870	74.765300	1.876871	8.709116	20.793870
	6	3.463694	38.473050	78.712390	1.878024	12.275810	26.095500
	7	3.484419	45.705750	80.842440	1.896125	14.641180	30.342760
	8	3.537469	51.068540	81.831640	1.906555	17.603180	33.291010
	9	3.552802	55.936660	82.246960	1.913696	19.463710	35.222580
	10	3.573449	59.644260	82.335160	1.917958	21.745740	36.455110
CQBL	1	4.456686	15.497560	3.066314	0.145480	0.242774	1.157865
	2	11.494300	12.168540	9.636969	2.594981	0.958399	1.679596
	3	10.758400	10.370880	26.308690	3.493269	1.142577	1.410912
	4	11.336660	12.769120	37.437450	3.551348	2.977420	1.407895
	5	11.155590	20.135610	45.103720	3.547287	3.750793	1.938934
	6	11.226450	26.498850	48.599240	3.531348	5.696850	2.857597
	7	11.176420	33.646910	49.789530	3.527947	6.349977	3.873392
	8	11.183860	38.602930	49.730650	3.525803	7.858826	4.756385
	9	11.169760	43.658490	49.368730	3.525076	8.306747	5.433732
	10	11.169720	47.256250	49.123970	3.524727	9.477962	5.919936
LDFZZZC	1	1.545402	21.821640	11.499150	9.952159	0.153286	12.176240
	2	9.835713	23.222100	5.836232	7.915500	0.401072	17.987530
	3	18.512280	27.750330	5.399360	15.037020	0.471048	21.219260
	4	27.737700	29.904490	6.869310	17.227640	1.230231	23.164540
	5	34.827280	32.093780	9.423416	18.586280	1.562264	24.327240
	6	39.932240	32.935580	11.883680	19.447150	2.361749	25.028250
	7	43.167160	33.687120	13.853390	19.764220	2.641021	25.463650
	8	45.143310	33.900400	15.107020	19.849610	3.224893	25.744370
	9	46.263580	34.167900	15.677590	19.840780	3.383532	25.931560
	10	46.873460	34.214470	15.768450	19.820790	3.788013	26.059520

It can be seen from Figure 3.2 that, the liquid ratio and turnover rate of return on common stockholders' equity have a larger impact on the equity ratio and the asset-liability ratio, while is largely impacted by gross profit.

4 Conclusions and suggestions

4.1 THE CONCLUSIONS OF EMPIRICAL ANALYSIS

Through VAR analysis of influencing factors and measure variables of capital structure, we find that, gross profit and the turnover rate of return on common stockholders' equity impact capital structure most.

Therefore, in the short period, when it comes to the real estate company, we should be mainly concern about profitability and operation ability. In the long term, we should take more attention to the stability of operation ability.

4.2 SUGGESTIONS

The influencing factors of capital structure can reflect the preferences of the company's financing choice. According to the research, we take some advice as follow:

1) Focus on the profitability of the company

How profitable a company's operation is a crucial target. In the corporative governance, we should pay attention to how to maintain it profitable. A company, which is in a profitable operation condition, can not only provides funds to it, but also maintain its capital structure stable in a short period.

2) The company should ensure the stability of operation ability

To ensure the funds fluid is an important guarantee for the sustainable operation of a company. Enough turnover rate of return on common stockholders' equity can not only makes the funds fluid, but also maintain shareholders rights, so as to maintain the stability of capital structure. Therefore, a certain level of turnover rate of

return on common stockholders' equity is a protection for a long term operation.

Hence, whether from the point of profitability or the stability of capital structure, an enterprise shall keep the earning capacity and the operation capacity in balance.

Acknowledgments

This study is supported by China National Committee on Ageing (No.QLB2014B009), Educational Commission of Zhejiang Province (No.Y201328759), Key Social Science Foundation of Hangzhou City (No.A13FX04) and Zhejiang Province Federation of Social Sciences project (No. 2014Z063).

References

[1] Gau G W, Wang,K 1990 Capital structure decisions in real estate investment *AREUEA Journal* 501-21
[2] Chen H H, Xia H S 2006 Demonstration analysis of China real estate industry listed company's capital structure influence factor *Special Zone Economy* 131-2
[3] Scott D F J 1972 Evidence on the importance of financial structure *Financial Management* 45-50
[4] Scott D F J, Martin J D 1975 Industry influence on financial structure *Financial Management* 67-79
[5] Liu Z B, Jiang F X 2003 Capital structure and competition in product market *Economic Research Journal* 60-8
[6] He G J, Yan Y Y 2005 Analysis on industrial differences of capital structure in Chinese listed companies *Journal of Hunan Financial and Economic College* 31-4
[7] Tan K 2005 Capital structure, ownership structure and performance: A study of Chinese listed companies *Industrial Economics Research* 23-9
[8] Hong X X, Shen Y F 2000 Our state's appeared on market companies' assets structure influence the factors of evidence analysis *Journal of Xiamen University(Arts & Social Sciences)* 114-20
[9] Feng G F, Wu L J 2000 An analysis of the influential factors affecting the formation of capital structure in Chinese listed companies *Economist* 59-66
[10] Xiao Z P, Wu S N 2002 Empirical research on the influencing factors of capital structure in Chinese listed company *Securities Market Herald* 39-44
[11] Zhou K G, Xu Y H 2012 How stable is capital structure in China listed company *The Journal of World Economy* 106-20
[12] Yang C H, Liu X M, Wei Z X, Geng Z Z 2011 Empirical research of pricing efficiency of Chinese futures market based on a VAR model *Journal of Applied Statistics and Management* 330-8

Authors

Zhuo Yang, China.

Current position, grades: PhD candidate of business management in the School of Business Administration, Zhejiang Gongshang University, China.
Scientific interest: business model and the strategy analysis tools.

Hong-Liang Qiu, China.

Current position, grades: a lecturer in the Department of Business Administration, Tourism College of Zhejiang, China.

Research on the performance measurement model of knowledge management based on Grey relational analysis

Liming Yang, Yanwen Wang*, Xiuju Gao

College of Finance, Hebei Normal University of Science & Technology, Qinhuangdao, Hebei, P.R. China

Received 1 October 2014, www.cmnt.lv

Abstract

Performance measurement of knowledge management is a decision-making analysis project that involves multiple complex factors, levels and fuzzy uncertain information. On the basis of analysis of influence factors of knowledge management performance measurement, the study established an evaluation index system of enterprise knowledge management performance. Meanwhile, by combining grey relational analysis method and Euclidean distance measurement, a performance measurement model of knowledge measurement was established. Via standardization of different types of evaluation indexes of knowledge management performance, Euclidean distances of standardized evaluation indexes of knowledge management performance and the grey relational coefficients based on Euclidean distances were established respectively. Then the weighted grey correlations of evaluation indexes of knowledge management performance were obtained. In this way, evaluation analysis of enterprise knowledge management performance was realized. Finally, the model and algorithm was tested with a case study. The result proves that the method of combining grey relational analysis and Euclidean distance is efficient and has its application value in performance evaluation of knowledge management.

Keywords: knowledge management, performance measurement, Grey relational analysis, euclidean distance, model

1 Introduction

In the era of knowledge economy, knowledge has become key resource in economic growth, social development and enterprise development. Efficient knowledge management is important for an enterprise to acquire and maintain advantages in competition. Thus, knowledge management is increasingly important in enterprise management. As a key section in knowledge management, performance evaluation of knowledge management can evaluate the level and capacity of enterprise knowledge management performance in an effective, accurate and objective way. It is beneficial to the enterprise to control its knowledge management and development changes and can help with the evaluation of enterprise development. What's more, enterprise can find problems in knowledge management in the process of performance measurement of knowledge management and make plans to deal with these problems, which is an important way for the enterprise to evaluate its capacity of knowledge management [1-3]. Thus, finding the key influence factors in the improvement of performance and taking effective measures on time are of theoretical importance and engineering application value. Knowledge management combines multiple disciplines and methods. Enterprise managers improve the overall organizing efficiency of enterprise system, Reaction capacity of enterprise business operation, enterprise marketing competitiveness, innovation capacity in production, enterprise capital appreciation, etc. Thus, the performance measurement of enterprise knowledge management usually concentrates on these issues. By far, there have been some studies direct on this problem and have gained corresponding achievements [4-8]. However, different scholars and specialists have different perspectives in performance measurement of knowledge management. Thus, there are no unified evaluation indexes of knowledge management performance and elements in knowledge management cannot be reflected. Besides, some methods of performance measurement of knowledge management are not operable and performable enough. In a word, the current methods of performance measurement of knowledge management have some limitations. Thus, this study, based on existing studies and researches, tentatively puts forward a performance measurement model of knowledge management based on grey relational analysis [8-11]. Via a case study, the feasibility and operability of the model were tested.

2 The evaluation index system of performance measurement of enterprise knowledge management

2.1 PRINCIPLE OF THE CHOICE OF EVALUATION INDEXES OF PERFORMANCE MEASUREMENT OF KNOWLEDGE MANAGEMENT

There are complex dynamic factors of multiple levels and perspectives that influence the result of performance measurement of knowledge management, which make the design principle of choosing of the evaluation indexes of knowledge management performance in the evaluation index system of enterprise knowledge management performance. Whether the evaluation indexes of knowledge

* ***Corresponding author's*** e-mail: 125721815@qq.com

management performance is valid or not concerns the reliability and validity of the whole evaluation index system of knowledge management performance. There is no unified design principle in choosing evaluation indexes of knowledge management performance in the field of designation of evaluation system of knowledge management performance, and different scholars and specialists have different perspectives in principles to follow in the designation of evaluation indexes of knowledge management performance. The author of this paper deems that the choice of evaluation indexes of knowledge management performance should guarantee the objecttiveness, accurateness and effectiveness of the performance measurement result of knowledge management, and the performance measurement of knowledge management should be conducted from multiple perspectives and levels. Thus, in the establishment of evaluation indexes of knowledge management performance, the following principles should be followed:

1) Scientific: in choosing evaluation indexes of knowledge management performance, elements in enterprise knowledge management and the rationality of the overall structure of evaluation indexes should be considered in the first place. The indexes should analyse the condition of enterprise knowledge management from a scientific perspective. Thus, the reliability and representativeness of chosen evaluation indexes can be guaranteed.
2) Comprehensiveness: the evaluation index system of enterprise knowledge management should comprehensively reflect the overall condition of enterprise knowledge management system. It should be able to analyse the condition of knowledge management from different perspectives. Thus, the choice of evaluation indexes of knowledge management performance should consider factors in multiple aspects.
3) Integral: the evaluation index system of enterprise knowledge management is a complex decision-making system. Thus, the choice of evaluation index of knowledge management performance should keep the integral and internal relations among each element and avoid imperfect and omission of index element.
4) Objectiveness: the choice of evaluation indexes of knowledge management performance should avoid the influence of subjective factors to the greatest extent. Each evaluation index should be able to reflect the actual condition of enterprise knowledge management.
5) Consistency: different evaluation indexes of knowledge management performance should apply consistent standard in assignment.
6) Hierarchy: the levels of evaluation index system of knowledge management performance and the membership between each two levels should be clear.
7) Operability: the evaluation indexes of knowledge management performance should be operable enough so as to conduct quantitative or qualitative measurement.

2.2 ESTABLISHMENT OF EVALUATION INDEX SYSTEM OF ENTERPRISE KNOWLEDGE MANAGEMENT PERFORMANCE

Based on the above-mentioned design principles, an evaluation index system of knowledge management performance was established in this paper, as presented in Table 1. The evaluation index system contains five 1st level indexes, namely the capacity of information management, the marketing capacity, the level of knowledge stock, the maturity of learning organization and the transfer ability of knowledge using.

Table1 The evaluation index system of knowledge management performance

Objective level	1st level index	2nd level index
Evaluation index system of knowledge management performance U	The capacity of information management U_1	Information level of the enterprise u_{11} Information communication level among employees u_{12} Information communication level among departments u_{13} Information communication level between the enterprise and the customers u_{14} Information support level of cooperation in production u_{15}
	The marketing capacity U_2	Customer satisfaction u_{21} Customer profitability u_{22} Market retention rate u_{23} Market share u_{24} Quick reaction capacity in marketing u_{25}
	The level of knowledge stock U_3	The ratio of technical personnel u_{31} The holding quantity of technological achievements u_{32} The conservation rate of technical personnel u_{33} The average level of education of technical personnel u_{34}
	The maturity of learning organization U_4	The learning competence in external communication u_{41} The learning competence in internal training u_{42} The improvement level of incentive mechanism u_{43} The ability of the knowledge managers u_{44}
	The transfer ability of knowledge using U_5	The knowledge acquisition capacity u_{51} The knowledge innovation capacity u_{52} The knowledge transformation capacity u_{53} The knowledge learning capacity u_{54}

3 Performance measurement model of knowledge management based on grey relational analysis

3.1 THE SET OF EVALUATION PLANS AND THE SET OF EVALUATION INDEXES

Assume that there are *m* enterprises that conduct evaluation and analysis of knowledge management performance. Thus, the set of evaluation plans will be formed as $C=\{C_1,C_2,...,C_n\}$. In the set, C_i is the evaluation plan of knowledge management performance of one of those different enterprises? Meanwhile, based on the structure and content of Table 1, the set evaluation indexes of knowledge management performance *U* with multiple levels can be determined. In this model $U=\{U_1,U_2,U_3,U_4,U_5\}$ and $\underset{i\neq j}{\forall}(U_i\cap U_j)=O,1\le i,j\le 5$. In the set $U_1=\{u_{11},u_{12},u_{13},u_{14},u_{15}\}$, $U_2=\{u_{21},u_{22},u_{23},u_{24},u_{25}\}$, $U_3=\{u_{31},u_{32},u_{33},u_{34}\}$, $U_4=\{u_{41},u_{42},u_{43},u_{44}\}$, $U_5=\{u_{51},u_{52},u_{53},u_{54}\}$.

3.2 STANDARDIZATION OF EVALUATION INDEXES

According to the content of evaluation indexes of knowledge management performance, different evaluation indexes can have different types of value of information. On the one hand, some values of evaluation indexes are accurate, while those of others are fuzzy and uncertain. On the other hand, some evaluation indexes have positive effect for the level of enterprise knowledge management performance and are positive evaluation indexes, and some have negative effect for the level of enterprise knowledge management performance and are negative evaluation indexes. Thus, in order to conduct effectively the performance measurement of knowledge management, different types of evaluation indexes need to be unified and standardized.

In order to keep the generality of statement, assume that the value of enterprise *i* on performance evaluation index *j* is $V_{ij}=[v_{ij}^{\min},v_{ij}^{\max}]$, $v_{ij}^{\min}<v_{ij}^{\max}$.

If the evaluation index is a positive evaluation index, its standardized value $\bar{V}_{ij}=\left[v_{ij}^{-\min},v_{ij}^{-\max}\right]$ of enterprise *i* on performance evaluation index *j* is:

$$\bar{V}_{ij}=\left[v_{ij}^{-\min},v_{ij}^{-\max}\right]=\left[\frac{v_{ij}^{-\min}-\min\limits_{1\le i\le m}(v_{ij}^{-\min})}{\max\limits_{1\le i\le m}(v_{ij}^{-\max})-\min\limits_{1\le i\le m}(v_{ij}^{-\min})},\frac{v_{ij}^{-\max}-\min\limits_{1\le i\le m}(v_{ij}^{-\min})}{\max\limits_{1\le i\le m}(a_{ij})-\min\limits_{1\le i\le m}(a_{ij})}\right]. \quad (1)$$

If the evaluation index *j* is a negative evaluation index, its standardized value $\bar{V}_{ij}=\left[v_{ij}^{-\min},v_{ij}^{-\max}\right]$ of enterprise *i* on performance evaluation index *j* is:

$$\bar{V}_{ij}=\left[v_{ij}^{-\min},v_{ij}^{-\max}\right]=\left[\frac{\max(v_{ij}^{-\max})-v_{ij}^{-\min}}{\max\limits_{1\le i\le m}(v_{ij}^{-\max})-\min\limits_{1\le i\le m}(v_{ij}^{-\min})},\frac{\max(v_{ij}^{-\max})-v_{ij}^{-\max}}{\max\limits_{1\le i\le m}(a_{ij})-\min\limits_{1\le i\le m}(a_{ij})}\right]. \quad (2)$$

If the evaluation index *j* is a moderate evaluation index, its standardized value $\bar{V}_{ij}=\left[v_{ij}^{-\min},v_{ij}^{-\max}\right]$ of enterprise *i* on performance evaluation index *j* is:

$$\bar{V}_{ij}=\left[v_{ij}^{-\min},v_{ij}^{-\max}\right]=\left[\frac{\frac{\max(v_{ij}^{-\max})+\min\limits_{1\le i\le m}(v_{ij}^{-\min})}{2}-v_{ij}^{-\min}}{\frac{\max\limits_{1\le i\le m}(v_{ij}^{-\max})-\min\limits_{1\le i\le m}(v_{ij}^{-\min})}{2}},\frac{\frac{\max(v_{ij}^{-\max})+\min\limits_{1\le i\le m}(v_{ij}^{-\min})}{2}-v_{ij}^{-\max}}{\frac{\max\limits_{1\le i\le m}(v_{ij}^{-\max})-\min\limits_{1\le i\le m}(v_{ij}^{-\min})}{2}}\right]. \quad (3)$$

It can be observed that those standardized evaluation indexes of knowledge management performance all satisfy $0\le v_{ij}^{-\min}\le 1$, $0\le v_{ij}^{-\max}\le 1$. Then the differentiation among those evaluation indexes is removed and measure standard of all the evaluation indexes is unified, which is beneficial to the accuracy of performance measurement of knowledge management.

3.3 THE GREY RELATIONAL DEGREE OF PERFORMANCE MEASUREMENT OF ENTERPRISE KNOWLEDGE MANAGEMENT

Grey relational analysis is a decision-making analysis method that analyzes and determines the degree of influence between two systems by analyzing the reference sequence and comparing the proximity of the geometrical shape of sequences to judge the proximity of changing tendency based on the analysis of the geometrical proximity of the data sequence of the system. It measures the relational degree between systems or between factors based on the grey relational degree, and describes the relative changes of factors in the development of the system. In the development process of the system, if the consistency of change trends of two factors is high, the grey relational degree between the two factors is high; otherwise the grey relational degree between the two factors is low. The grey relational analysis analyzes the development trend of the system, so it does not require strictly the sample size and typical regulations of distribution. Thus, it can be widely applied. However, when grey relational analysis is conducted to fuzzy and uncertain values of evaluation indexes, the classical grey

relational analysis method needs to be improved. Thus, this study introduces Euclidean distance in analysis.

Assume that all the evaluation indexes of knowledge management performances have been standardized, and the standardized indexes are all positive indexes. Thus, the optimal evaluation index sequence V^* of the set of evaluation indexes of knowledge management performance can be established as:

$$V^* = \left(\bar{V}_1^*, \bar{V}_2^*, ..., \bar{V}_n^*\right) = \left(\left[v_1^{-\min*}, v_1^{-\max*}\right], \left[v_2^{-\min*}, v_2^{-\max*}\right], ..., \left[v_n^{-\min*}, v_n^{-\max*}\right]\right). \quad (4)$$

In the sequence:

$$\bar{V}_j^* = [\max_{1\le i\le m}(v_{ij}^{-\min}), \max_{1\le i\le m}(v_{ij}^{-\max})]. \quad (5)$$

Thus, the Euclidean distance $D_{ij}(\bar{V}_{ij} \to \bar{V}_j^*)$ of the enterprise i on the evaluation index of knowledge management performance j and the ideal optimal sequence of evaluation indexes V^* is:

$$D_{ij}\left(\bar{V}_{ij} \to \bar{V}_j^*\right) = \frac{\left[\left|v_{ij}^{-\min} - \max_{1\le i\le m}\left(v_{ij}^{-\min}\right)\right|^2 + \left|v_{ij}^{-\max} - \max_{1\le i\le m}\left(v_{ij}^{-\max}\right)\right|^2\right]^{\frac{1}{2}}}{\sqrt{2}}. \quad (6)$$

According to the relative theories of grey relational analysis method, the relational coefficient ξ_{ij} of enterprise i on the evaluation index of knowledge management performance j and the ideal optimal evaluation index sequence V^* is:

$$\xi_{ij} = \frac{\min\limits_{1\le i\le m}\min\limits_{1\le i\le n} D_{ij}(\bar{V}_{ij} \to \bar{V}_j^*) + \rho \max\limits_{1\le i\le m}\max\limits_{1\le i\le n} D_{ij}(\bar{V}_{ij} \to \bar{V}_j^*)}{D_{ij}(\bar{V}_{ij} \to \bar{V}_j^*) + \rho \max\limits_{1\le i\le m}\max\limits_{1\le i\le n} D_{ij}(\bar{V}_{ij} \to \bar{V}_j^*)}, \quad (7)$$

where $i = 1, 2, ..., m$; $j = 1, 2, ..., n$; ρ is the resolution ratio and $\rho \in (0,1)$. Generally, $\rho = 0.5$.

Considering the weights w_{ij}^2 of different 2nd level evaluation indexes are different, the weighted grey relational degree φ_i^2 of enterprise i on the 2nd level evaluation index on knowledge management performance is:

$$\varphi_i^2 = \sum_{j=1}^{n^2}(w_{ij}^2 \xi_{ij}). \quad (8)$$

Combining the weights w_{ij}^1 of the 1st level of evaluation indexes, the weighted grey relational degree ψ_i of enterprise i on the evaluation index set of knowledge management performance is:

$$\psi_i = \sum_{j=1}^{n^1}(w_{ij}^1 \varphi_i^2). \quad (9)$$

Thus, according to the principle of selecting the closest in the analysis of grey relational analysis with multiple attributes based on grey relational degree of comprehensive weight, if:

$$\psi_i = \max\{\psi_1, \psi_2, \cdots, \psi_m\}. \quad (10)$$

The level of knowledge management performance of the enterprise i is the closest to the level of knowledge management performance correspondent to the ideal optimal evaluation index sequence.

3.4 REALIZATION OF THE MODEL AND ALGORITHM OF PERFORMANCE MEASUREMENT OF KNOWLEDGE MANAGEMENT BASED ON GREY RELATIONAL ANALYSIS METHOD

Based on what mentioned above, the algorithm implementation of performance measurement model of knowledge management based on grey relational analysis method can be described as below:

Step 1: Determine the basic principles in choosing evaluation indexes of enterprise knowledge management performance directing at the complex dynamic influence factors of multiple levels and aspects in the process of enterprise knowledge management.

Step 2: Establish the evaluation index system of knowledge management performance under the guidance of principles of choosing evaluation indexes, and form evaluation plan and evaluation index set of knowledge management performance based on this evaluation index system.

Step 3: Determine the types of evaluation indexes in the above-mentioned evaluation index set, and obtain the values of evaluation indexes of knowledge management performance plans of different enterprises.

Step 4: Standardize positive indexes, negative indexes and moderate indexes respectively according to Equations (1)-(3) and unify the measure standard of each type of evaluation indexes.

Step 5: Construct the optimal evaluation index sequence of evaluation index set of knowledge management performance with Equations (4) and (5).

Step 6: Obtain the Euclidean distance between the plan of knowledge management performance and the ideal optimal evaluation index sequence with Equation (6).

Step 7: Obtain the grey relational coefficient between the plan of knowledge management performance and the ideal optimal evaluation index sequence with Equation (7).

Step 8: consider the weights of different levels of evaluation indexes of knowledge management performance and obtain the comprehensive weighted grey relational degree with Equations (8) and (9).

Step 9: According to the value of comprehensive weighted grey relational degree obtains the optimal plan of knowledge management performance with Equation (10).

4 Verification of the model and the algorithm

This paper set the evaluation and analysis of knowledge management performance of three enterprises in the economic and technological development zone of the new city zone of a provincial capital as the object for the case study, in order to prove the algorithm feasible and practicable. The set of analysis object of performance measurement of knowledge management is $C=\{C_1,C_2,C_3\}$. C_1, C_2 and C_3 respectively stand for the three enterprises. The weight values were obtained from specialists in domain design with comprehensive understanding of the mechanism of the three enterprises. They combined the investigation and analysis of the relative materials of the three enterprises, their own knowledge, experience and personal preference and analysed the weights of the 1st level indexes and the 2nd level indexes of the above-mentioned evaluation index system of knowledge management performance with comprehensive judgment approach. The corresponding weighted values are presented in Table 2. Meanwhile, via investigation, statistics and analysis, the information of the knowledge management performance of the three enterprises was obtained, as also presented in Table 2.

If a table is too long to fit onto one page, the table number and headings should be repeated on the next page before the table is continued.

Alternatively, the table can be spread over two consecutive pages (first on even-numbered, then on odd-numbered page).

For a wide table you can use 1-column section (Table 1), for a small standard table 2-column section is used (Table 2).

Table 2 Information of evaluation index of knowledge management performance

1st level index	Weight	2nd level index	Weight	Information of evaluation index		
				Enterprise 1	Enterprise 2	Enterprise 3
The capacity of information management U_1	0.21	Information level of the enterprise u_{11}	0.21	0.83–0.85	0.80–0.83	0.67–0.70
		Information communication level among employees u_{12}	0.19	0.68–0.73	0.75–0.80	0.78–0.83
		Information communication level among departments u_{13}	0.22	0.75–0.78	0.75–0.80	0.75–0.78
		Information communication level between the enterprise and the customers u_{14}	0.18	0.80–0.85	0.75–0.80	0.75–0.80
		Information support level of cooperation in production u_{15}	0.20	0.76–0.80	0.70–0.74	0.67–0.70
The marketing capacity U_2	0.15	Customer satisfaction u_{21}	0.18	0.80–0.85	0.60–0.65	0.80–0.85
		Customer profitability u_{22}	0.17	1.35	2.32	0.98
		Market retention rate u_{23}	0.20	85–90	80–85	85–90
		Market share u_{24}	0.25	0.016–0.018	0.013–0.015	0.012–0.015
		Quick reaction capacity in marketing u_{25}	0.20	0.60–0.70	0.75–0.85	0.65–0.75
The level of knowledge stock U_3	0.19	The ratio of technical personnel u_{31}	0.20	0.73	0.58	0.69
		The holding quantity of technological achievements u_{32}	0.30	15	22	18
		The conservation rate of technical personnel u_{33}	0.28	0.75–0.80	0.85–0.90	0.75–0.80
		The average level of education of technical personnel u_{34}	0.22	1.34	1.50	2.01
The maturity of learning organization U_4	0.22	The learning competence in external communication u_{41}	0.25	2.50–2.80	1.60–1.80	2.00–2.20
		The learning competence in internal training u_{42}	0.25	20.00–25.00	12.00–14.00	16.00–18.00
		The improvement level of incentive mechanism u_{43}	0.26	0.77–0.80	0.68–0.75	0.54–0.60
		The ability of the knowledge managers u_{44}	0.24	0.80–0.85	0.75–0.80	0.70–0.75
The transfer ability of knowledge using U_5	0.23	The knowledge acquisition capacity u_{51}	0.30	90–95	80–85	85–90
		The knowledge innovation capacity u_{52}	0.25	80–85	85–90	85–90
		The knowledge transformation capacity u_{53}	0.25	80–85	70–75	80–85
		The knowledge learning capacity u_{54}	0.20	70–75	80–85	80–85

Based on the type of evaluation indexes, applying the method of standardization of evaluation index in the paper, the standardized information of different evaluation indexes can be obtained. The values are presented in Table 3.

Table 3 The standardized value of evaluation indexes of knowledge management performance

2nd level index	Information of evaluation index		
	Enterprise 1	**Enterprise 2**	**Enterprise 3**
Information level of the enterprise u_{11}	0.976–1.000	0.941–0.976	0.788–0.824
Information communication level among employees u_{12}	0.819–0.880	0.904–0.964	0.940–1.000
Information communication level among departments u_{13}	0.938–0.975	0. 938–1.000	0.938–0.975
Information communication level between the enterprise and the customers u_{14}	0.941–1.000	0.882–0. 941	0.882–0.941
Information support level of cooperation in production u_{15}	0.950–1.000	0.875–0.925	0.838–0.875
Customer satisfaction u_{21}	0.941–1.000	0.706–0.765	0.941–1.000
Customer profitability u_{22}	0.582	1.000	0.422
Market retention rate u_{23}	0.944–1.000	0.889–0.944	0.944–1.000
Market share u_{24}	0.889–1.000	0.722–0.833	0.667–0.833
Quick reaction capacity in marketing u_{25}	0.706–0.824	0.882–1.000	0.765–0.882
The ratio of technical personnel u_{31}	1.000	0.795	0.945
The holding quantity of technological achievements u_{32}	0.682	1.000	0.818
The conservation rate of technical personnel u_{33}	0.833–0.889	0.944–1.000	0.833–0.889
The average level of education of technical personnel u_{34}	0.667	0.746	1.000
The learning competence in external communication u_{41}	0.893–1.000	0.571–0.643	0.714–0.786
The learning competence in internal training u_{42}	0.800–1.000	0.480–0.560	0.640–0.720
The improvement level of incentive mechanism u_{43}	0.963–1.000	0.850–0.938	0.675–0.750
The ability of the knowledge managers u_{44}	0.941–1.000	0.882–0.941	0.823–0.882
The knowledge acquisition capacity u_{51}	0.947–1.000	0.842–0.895	0.895–0.947
The knowledge innovation capacity u_{52}	0.889–0.944	0.944–1.000	0.944–1.000
The knowledge transformation capacity u_{53}	0.941–1.000	0.824–0.882	0.941–1.000
The knowledge learning capacity u_{54}	0.824–0.882	0.941–1.000	0.941–1.000

By applying the calculation formula of Euclidean distance in the paper, the matrix of Euclidean distance of performance measurement of knowledge management can be obtained:

$$\mathbf{D}=\begin{bmatrix}0.000 & 0.030 & 0.182\\ 0.170 & 0.051 & 0.000\\ 0.018 & 0.000 & 0.018\\ 0.000 & 0.059 & 0.059\\ 0.000 & 0.075 & 0.119\\ 0.000 & 0.235 & 0.000\\ 0.418 & 0.000 & 0.578\\ 0.000 & 0.056 & 0.000\\ 0.000 & 0.167 & 0.279\\ 0.176 & 0.000 & 0.166\\ 0.000 & 0.205 & 0.055\\ 0.318 & 0.000 & 0.182\\ 0.111 & 0.000 & 0.111\\ 0.333 & 0.254 & 0.000\\ 0.000 & 0.340 & 0.197\\ 0.000 & 0.385 & 0.228\\ 0.000 & 0.091 & 0.230\\ 0.000 & 0.059 & 0.118\\ 0.000 & 0.105 & 0.053\\ 0.056 & 0.000 & 0.000\\ 0.000 & 0.118 & 0.000\\ 0.118 & 0.000 & 0.000\end{bmatrix}_{22x3}$$

The matrix of grey relational coefficient of 2nd level indexes of performance measurement of knowledge management can be obtained based on the matrix of Euclidean distance:

$$\mathbf{H}=\begin{bmatrix}1.000 & 0.906 & 0.614\\ 0.630 & 0.850 & 1.000\\ 0.941 & 1.000 & 0.941\\ 1.000 & 0.830 & 0.830\\ 1.000 & 0.794 & 0.708\\ 1.000 & 0.552 & 1.000\\ 0.409 & 1.000 & 0.333\\ 1.000 & 0.838 & 1.000\\ 1.000 & 0.634 & 0.509\\ 0.622 & 1.000 & 0.635\\ 1.000 & 0.637 & 0.840\\ 0.476 & 1.000 & 0.614\\ 0.723 & 1.000 & 0.723\\ 0.465 & 0.532 & 1.000\\ 1.000 & 0.459 & 0.595\\ 1.000 & 0.429 & 0.559\\ 1.000 & 0.761 & 0.557\\ 1.000 & 0.830 & 0.710\\ 1.000 & 0.734 & 0.845\\ 0.838 & 1.000 & 1.000\\ 1.000 & 0.710 & 1.000\\ 0.710 & 1.000 & 1.000\end{bmatrix}_{22x3}$$

The sequence of comprehensive weighted grey relational degree can be obtained via considering comprehensively the weight of 1st level evaluation indexes and 2nd level evaluation indexes. According to the principle of selecting the closest of the comprehensive weight grey relational degree, the knowledge management performance of Enterprise 1 is the optimal in the three, which consists with the actual condition.

5 Conclusion

Directing at the problems in the process of performance measurement of knowledge management, this paper puts forward a performance measurement model of knowledge management based on grey relational analysis. By establishing evaluation index system of knowledge management performance, standardization of different types of evaluation indexes is realized and an improved grey relational coefficient calculation model of evaluation indexes of knowledge management performance based on Euclidean distance is put forward. Thus, the weighted grey relational degree of performance measurement of knowledge management can be obtained, and evaluation and analysis of the knowledge management performance can be realized. Then by a case study, this model proves feasible and practicable, and can provide decision basis to the guidance of knowledge management development, which can help to improve the level of knowledge management and strengthen the competitiveness of the enterprise. What's more, it can also verify the regulations in studies of knowledge management and find new problems in knowledge management of enterprise, which is beneficial to the development of the discipline of enterprise knowledge management.

Acknowledgment

The Research Project on social scientific development of Hebei province in 2014. (Project number: 2014041408): Study on Transformation and Upgrading of Rural Tourism Industry in Beijing-Tianjin-Hebei and around Bohai Sea Regional Integration. The supervisor of project: Yang Liming.

References

[1] Hult G T M, Ketchen D J, Cavusgil S T, Calantone R J 2006 Knowledge as a strategic resource in supply chains *Journal of Operations Management* 24 458-75
[2] Lee K C, Lee S, Kang I W 2005 KMPI: measuring knowledge management performance *Information and Management* 42 469-82
[3] Wang J, Guan J 2002 The application of composite DEA method in measuring the performance of knowledge management of enterprises Studies in Science of Science 20(1) 84-8
[4] Fairchild A M 2002 Knowledge management metrics via a balanced scorecard methodology *Proceedings of the 35th Hawaii International Conference on System Science* USA 2002
[5] Li S, Chang Li, Zou S 2001 Grey relational assessment of know ledge stocks of enterprises *Science Research Management* 22(3) 73-8 *(in Chinese)*
[6] Zhu Q H, Zhang G 2003 Appraisal model of enterprises know ledge management based on artificial neural network *Science of Science and Management of S&T* (8) 32-4.
[7] Tseng S-M 2009Knowledge management performance measure index *Expert systems with Applications* (34) 734-45
[8] Wang T, Yang A, Bu L 2013 Mechanism scheme design based on multi-attribute extension gray relevant optimized decision-making model *Systems Engineering – Theory & Practice* 33(9) 2321-9 *(in Chinese)*
[9] Liu S, Fang Z, Yang Y, et al 2012 General grey numbers and its operations *Grey Systems: Theory and Application* 2012 2(3) 4-15
[10] Xiao X P, Wen J H, Xie M 2010 Grey relational analysis and forecast of demand for scrap steel *The Journal of Grey System* 22(1) 73-80
[11] Xie N M, Liu S F 2009 Research on evaluations of several grey relational models adapt to grey relational axioms *Journal of Systems Engineering and Electronics* 20(2) 304-9

Authors

Liming Yang, born in November, 1978, Changli, Hebei, China.

Current position, grades: experimentalist at Hebei Normal University of Science & Technology, Qinhuangdao, Hebei.
University studies: bachelor's degree in economics at Hebei Normal University of Science & Technology, Qinhuangdao, Hebei in 2008.
Scientific interests: economics.
Publications: 1.

Yanwen Wang, born in July 1977, Changli, Hebei, China.

Current position, grades: lecturer at Hebei Normal University of Science & Technology, Qinhuangdao, Hebei.
University studies: Master's degree of economics in University of International Business and Economics (2010).
Scientific interests: Economics, trade, and education.
Publications: 1.

Xiuju Gao, December 1971, Changli, Hebei, China.

Current position, grades: associate professor at Hebei Normal University of Science & Technology, Qinhuangdao, Hebei.
University studies: Master's degree of economics in Hebei University of Economics and Business, Shijiazhuang, Hebei (2012).
Scientific interests: financial accounting.
Publications: 1.

Corporate growth, liquidity assets value and financing decision

Yaqin Lu[1], Cunzhi Tian[2], Yi Wu[2*]

[1]*Economic Research Institute, Yunnan University of Finance and Economics, 650221, Kunming, Yunnan, China*

[2]*Faculties of Management and Economics, Kunming University of Science and Technology, 650093, Kunming, Yunnan, China*

Received 1 October 2014, www.cmnt.lv

Abstract

Based on the external financing analysis framework under asymmetric information, this paper analyzes the influence of corporate growth on liquid assets value and financing decision. Both of the theory and numerical calculations show that liquid assets value would increase with the increasing in corporate growth. For the reinvestment decision in the case of higher reinvestment demand, if the price of liquidity assets is higher than value, it is optimal not to reinvest; if the price is lower than value, it is optimal to reinvest; if the price is equal to value, there is no difference found in reinvesting or not.

Keywords: liquidity shocks, liquid assets value, corporate growth, financing decision

1 Introduction

Shortage of liquidity has been considered to be the main reason of financial crisis sweeping global. As an important part of the companies' operation process, holding of liquidity asset may influence its operating performance, and directly impact on its management, investment and financing activities. Why do companies hold large pools of liquidity? What will holding liquidity do for the companies' value? How to determine the value of liquidity assets? What factors will affect the value of liquid assets? Such problems have been hot issues of the academic research.

As far as factors affecting the value of liquid assets is concerned. Harford (1999) proves that mergers and acquisitions will damage the companies' cash value by studying the relationship between policy of mergers and acquisitions and value of holding liquid assets. Faulkender and Wang (2006) et al. consider that the marginal value of holding liquid assets increases with the increasing in financing constraints. Denis and Sibilkov (2007) deem that holding liquid assets can help the enterprise get projects with higher yields, and so the enterprise with higher financing constraints may own higher value of holding liquid assets. Pinkowitz and Williamson (2007), Kalcheva and Lins (2007) make a comparative study for multinational corporations and find that, compared to countries with poor shareholders' protection, investors contain higher liquidity assets valuation in those with better protection. Tong (2009) believes the enterprise's diversification can increase agency cost and reduce the value of liquid assets. Drobetz et al. (2010) think the entrepreneur's moral hazard could reduce the value of liquid assets on account of asymmetric information.

In point of research on the growth and value of liquidity assets, Myers and Majluf (1984) believe that, compared with the enterprise with lower growth, the enterprise with higher growth may own higher cost of underinvestment and more serious problem of asymmetric information, so the cash value is relatively higher held by the enterprise with higher growth. Mikkelson and Partch (2003) find that companies holding more liquidity assets usually own more investment, more research and development expenses and bigger expansion of asset scale. Saddour (2006) deems the companies' market value is positively correlated with holding level of liquid assets, and compared with the enterprise with lower growth; this kind of positive correlation is more significant for the enterprise with higher growth. Through the empirical analysis, Pinkowitz and Williamson (2007) think that for the company in high-speed growth stage, its market value of holding liquidity assets would be higher.

Based on the external financing analysis framework under asymmetric information built by Tirole (2006), this article investigates the influence of corporate growth on liquidity assets value, and then makes a deep exploration on the entrepreneur's optimal financing decision in the condition of certain liquidity asset price. Basis of the study is that the enterprise can not meet his reinvestment demand by create enough internal liquidity relying on its future earnings.

2 Assumptions

The basic assumptions will be given in the following: three periods: $t=0$ represents ex enter period; $t=1$ represents intermediate period; $t=2$ represents ex post period.

Participants: an entrepreneur and investors.

At date $t=0$, the entrepreneur has "assets" A and a project requiring variable investment $I \in [0,\infty)$. To imp-

[*]***Corresponding author's*** e-mail: wuyi0871@126.com

lement the project the entrepreneur must borrow $I-A$ from investors.

If the project can get financing, the company meets a reinvestment opportunity requiring an amount ρI at date $t=1$, where ρ means the reinvestment amount required by one unit of initial investment for reinvesting. ρ is unknown at date $t=1$ and subjects to the following two-point distribution:

per unit of reinvestment	ρ_H	ρ_L
probability	$1-\lambda$	λ

Average reinvestment demand for per unit of initial investment is:

$$\bar{\rho} \equiv (1-\lambda)\rho_L + \lambda\rho_H .$$

Whether the enterprise reinvests or not, the company can continue to operate.

At date $t=2$, the project either succeeds, that is, yields verifiable income RI, or fails, that is, yields no income, where R is the yield of initial investment in the case of success?

The probability of success would be influenced by the effort degree of the entrepreneur and reinvestment opportunities.

At date $t=1$, if the company does not reinvest, the probability of success would be p; on the contrary, if the company reinvest, the probability of success would be $p+\tau$. Where $\tau>0$ indicates the size of growth opportunities. The bigger τ, the greater growth opportunities.

p is affected by the effort degree of the entrepreneur, but it is unobservable. Behaving yields probability $p=p_H$ of success and no private benefit to the entrepreneur, and misbehaving results in probability $p=p_L<p_H$ of success and private benefit $BI>0$. Let $\Delta p=p_H-p_L$.

Per unit investment has positive NPV if the entrepreneur behaves at date $t=2$, but negative NPV even if one includes the entrepreneur's private benefit, if he does not. In other words, the initial contract need to motivate the entrepreneur to behave at date $t=2$.

Reinvestment is also optimal for the society, even if the company requires reinvesting $\rho_H I$ for growth opportunities. That is:

$$(p_H+\tau)RI-\rho_H I-I>p_H RI-I \Rightarrow \tau R>\rho_H . \quad (1)$$

At date $t=1$, the entrepreneur can raise internal liquidity $\tau(R-B/\Delta p)I$ when facing reinvestment demand. Let:

$$\rho_L I<\tau(R-B/\Delta p)I<\rho_H I . \quad (2)$$

This means that, when faced with reinvestment needs $\rho_L I$, the enterprise could meet his reinvestment needs through internal liquidity; otherwise, when faced with reinvestment needs $\rho_H I$, he could not meet his reinvestment needs through internal liquidity.

There exist in the economy liquid assets. That is, 1 unit invested at $t=0$ delivers a return of 1 unit at $t=1$.

The price of liquid assets is q at $t=0$, where $q\geq 1$.

In order to ensure the threshold of liquidity assets price is not less than 1, let:

$$\rho_H+\tau'[(1-\lambda)(\rho_H-\rho_L)-1]\leq 0, \quad (3)$$

where $\tau'=\tau/p_H$. In fact, τ' reflects the increasing proportion of the success probability due to growth opportunities, in the case of behaving; and it could be used to measure the marginal productivity of growth opportunities.

In order to ensure the investment scale of equilibrium is positive, let:

$$1+(1-\lambda)\rho_L-\rho_0-(1-\lambda)\tau'\rho_0>0, \quad (4)$$

where $\rho_0=p_H(R-B/\Delta p)I$ denotes the expected pledge able income per unit of investment at $t=2$ without reinvestment.

Both the entrepreneur and investors are risk neutral.

Both the entrepreneur and investors have not time preference; the riskless rate is taken to be 0.

Investors behave competitively in the sense that the loan, if any, makes zero profit.

The entrepreneur has bargaining power, and he puts forward a financing contract for investors "either accept or reject", he is also protected by limited liability.

The timing could be summarized in Figure 1.

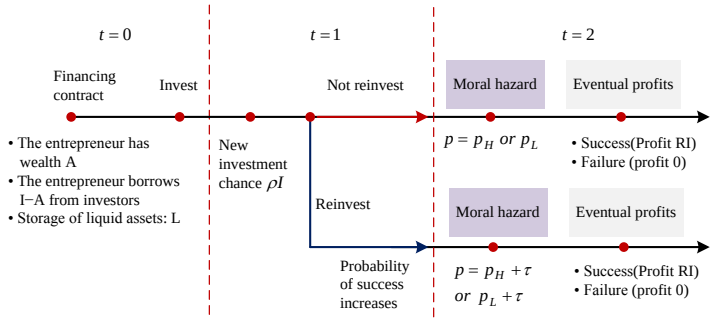


FIGURE 1 Figure of the timing

3 Optimal models

Suppose that the financing contract signed at date $t=0$ between the entrepreneur and investors takes the following state-contingent form:

$$\{I, L; (1, x); (R_b, 0), (RI - R_b, 0)\},$$

where the contract specifies an initial investment level I.

The entrepreneur can purchase L units of liquidity assets with the price q at date $t=0$.

At date $t=1$, if reinvestment level is $\rho_L I$, the entrepreneur always reinvests; contrarily, if reinvestment level is $\rho_H I$, he could only reinvests with probability x.

At date $t=2$, if the project succeed, the entrepreneur and investors get R_b and $RI-R_b$, respectively, if the project fail, both of them get 0.

The enterprise's optimal model under this kind of contract form is like that:

$$\begin{cases} \max\limits_{R_b, I, x} \{(1-\lambda)(p_H+\tau)+\lambda[x(p_H+\tau)+(1-x)p_H]\}R_b - A \\ s.t.(a_1)(p_H+\tau)R_b \geq (p_L+\tau)R_b + BI \\ \quad (a_2)\, p_H R_b \geq p_L R_b + BI \\ \quad (a_3)\, L\times 1+\tau(RI-R_b) \geq \rho_L I \\ \quad (a_4)\, L\times 1+x\tau(RI-R_b) \geq x\rho_H I \\ \quad (a_5)\lambda x[(p_H+\tau)(RI-R_b)-\rho_H I]+\lambda(1-x)p_H(RI-R_b) \\ \qquad +(1-\lambda)[(p_H+\tau)(RI-R_b)-\rho_L I] \geq I - A + (q-1)L \end{cases},$$

where the objective function is the entrepreneur's net utility. When reinvestment level is $\rho_L I$, the entrepreneur always reinvests, and the probability of success is $p_H+\tau$. When reinvestment level is $\rho_H I$, the entrepreneur could only reinvest with probability x (the success probability is $p_H+\tau$), and could not reinvest with probability $1-x$ (the success probability is p_H).

The constraint (a_1) is the entrepreneur's incentive-compatibility constraint with reinvestment. This constraint ensures the entrepreneur behaving at date $t=2$ if he reinvests at date $t=1$. Where the left side of the inequality is the expected profit if he behaves, the right side is the expected profit if he misbehaves, and (a_1) could be simplified as:

$$R_b \geq BI/\Delta p. \tag{5}$$

The constraint (a_2) is the entrepreneur's incentive-compatibility constraint without reinvestment. This constraint ensures the entrepreneur behaving at date $t=2$ if he does not reinvest at date $t=1$. Where the left side of the inequality is the expected profit if he behaves, the right side is the expected profit if he misbehaves, and (a_2) could be simplified as Equation (5). Therefore, the entrepreneur's incentive-compatibility constraint will not be affected by reinvestment. In fact, the moral hazard happens after reinvestment, so the incentive-compatibility constraint will be the same whether the project gets growth opportunity or not.

The constraint (a_3) ensures that the entrepreneur can reinvest if reinvestment level is $\rho_L I$; (a_4) ensures that the entrepreneur can reinvest if reinvestment level is $\rho_H I$. In fact, the enterprise can meet his reinvestment needs by internal liquidity if reinvestment level is $\rho_L I$, that means (a_3) is totally unnecessary. In addition, because storage of liquid assets may generate liquidity premium $q-1$ and too many storage of liquidity assets is irrational, (a_4) would hold with equality for maximizing the entrepreneur's utility, that is:

$$L = x[\rho_H I - \tau(RI - R_b)].$$

The constraint (a_5) is the investors' individual-rationality constraint. Specifically, if reinvestment level is $\rho_L I$, the investors pay reinvestment $\rho_L I$ and can get expected return $(p_H+\tau)(RI-R_b)$ at the same time. Contrarily, if reinvestment level is $\rho_H I$, the expected probability for investors getting $RI-R_b$ is $x(p_H+\tau)+(1-x)p_H$, and the investors' expected reinvestment is $x\rho_H+(1-x)\cdot 0$, $I-A$ is their initial invest at date $t=0$, $(q-1)L$ is the liquidity premium generated by L units of liquidity assets. In fact, the investors' individual-rationality constraint can be simplified as:

$$R_b \leq \frac{H\cdot I + A}{p_H + [(1-\lambda)+\lambda x]\tau + (q-1)x\tau}, \tag{6}$$

where:

$$H = p_H R + [(1-\lambda)+\lambda x]\tau R - (q-1)x(\rho_H - \tau R) - 1 - (1-\lambda)\rho_L - \lambda x\rho_H.$$

4 Optimal contracts

The optimization problem can be solved in three steps.

Step 1: Solve the optimal R_b^* and $I^*(x)$ for a given x.

In fact, all the solutions satisfied the constraint Equations (5) and (6) are called the "feasible contract set" S. The optimization problem shows that the entrepreneur needs to find one financing contract to maximize his own profit from the feasible set S. Figure 2 illustrates the "feasible contract set" of the optimization problem.

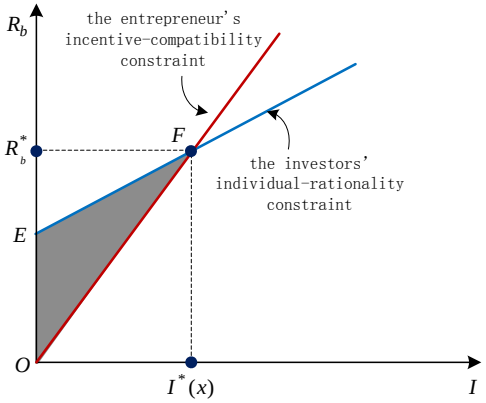


FIGURE 2 The "feasible contract set"

We can get that the "feasible contract set "from a geometric point of view, the "feasible contract set" is constituted by the shaded area OEF surrounded by the incentive-compatibility constraint, individual-rationality constraint and non-negative constraint. The objective function shows that it is the bigger the better for R_b. Therefore, the point F constitutes the optimal contract, and then:

$$R_b = \frac{BI}{\Delta p}$$

and

$$I^*(x) = A \cdot \{1-\rho_0 - [(1-\lambda)+\lambda x]\tau'\rho_0 + (q-1)x(\rho_H - \tau'\rho_0) + (1-\lambda)\rho_L + \lambda x\rho_H\}^{-1}, \quad (7)$$

where

$$\rho_0 = p_H(R - B/\Delta p), \rho_1 = p_H R, \tau' = \tau / p_H .$$

In Equation (7), $(1-\lambda)+\lambda x$ represents the probability of reinvesting or getting growth opportunity at date $t=1$, $\rho_0 + [(1-\lambda)+\lambda x]\tau'\rho_0$ expresses the expected pledge able income generated by per unit of initial investment; $(1-\lambda)\rho_L + \lambda x\rho_H$ shows the expected reinvestment cost of per unit of initial investment; $x(\rho_H - \tau'\rho_0)$ is the holding amount of liquid assets for one unit initial investment; $q-1$ is the liquidity premium.

Step 2: Consider the optimal x^* by taking R_b^* and I^* into the optimization problem.

In conclusion, the optimization problem can be expressed as:

$$\max_x \{p_H + [(1-\lambda)+\lambda x]p_H\tau'\}\frac{B}{\Delta p}I^* - A, \quad (8)$$

take I^* into the optimization Equation (8), and then it can be further simplified as:

$$\max_x \frac{\rho_1 - c(x,q)}{c(x,q)-\rho_0}A,$$

where:

$$c(x,q) = \frac{(q-1)x(\rho_H - \tau'\rho_0) + (1-\lambda)\rho_L + \lambda x\rho_H + 1}{1+(1-\lambda+\lambda x)\tau'},$$

as:

$$\max_x \frac{\rho_1 - c(x,q)}{c(x,q)-\rho_0}A \Leftrightarrow \min_x c(x,q),$$

this means the entrepreneur could get maximum utility when $c(x,q)$ takes the minimum.

Suggestion 1: When the entrepreneur needs to reinvest $\rho_H I$ because of getting the growth opportunity, his optimal strategy is closely related to the liquidity assets q it displays as that:

if $q \in [1, \bar{q})$, it is optimal to reinvest with probability $x^* = 1$;

if $q \in (\bar{q}, +\infty)$, it is optimal to reinvest with probability $x^* = 0$;

if $q = \bar{q}$, it is optimal to reinvest with probability $x^* = [0,1]$.

where $\bar{q} = 1 + \lambda\frac{\tau'[1-(1-\lambda)(\rho_H - \rho_L)] - \rho_H}{(\rho_H - \tau'\rho_0)[1+(1-\lambda)\tau']}$ represents the value of liquid assets.

Proof: Due to $\frac{\partial c(x,q)}{\partial x} = \frac{E}{F}$, where:

$$E = (q-1)(\rho_H - \tau'\rho_0)[1+(1-\lambda)\tau'] + \lambda\rho_H + \lambda\tau'[(1-\lambda)(\rho_H - \rho_L) - 1],$$
$$F = [1+(1-\lambda+\lambda x)\tau']^2$$

let:

$$\bar{q} = 1 + \lambda\frac{\tau'[1-(1-\lambda)(\rho_H - \rho_L)] - \rho_H}{(\rho_H - \tau'\rho_0)[1+(1-\lambda)\tau']},$$

which leads to:

if $q > \bar{q}$, $\partial c(x,q)/\partial x > 0 \Rightarrow x^* = 0$;

if $q < \bar{q}$, $\partial c(x,q)/\partial x < 0 \Rightarrow x^* = 1$;

if $q = \bar{q}$, $\partial c(x,q)/\partial x = 0 \Rightarrow x^* \in [0,1]$.

In fact, $\bar{q}$ is the highest liquid assets price which can be accepted by the entrepreneur?

Suggestion 1 shows three kinds of phenomena. Firstly, if the price of liquid assets is relatively higher ($q > \bar{q}$), it is the optimal strategy for the entrepreneur not to store liquid assets. In fact, holding one unit of liquid assets could generate expected return $\bar{q}$ for the entrepre-

neur that means the cost of holding liquid assets is bigger than expected earnings. Therefore, it is irrational for the entrepreneur to hold liquid assets. Secondly, if the price of liquid assets is relatively lower ($q < \bar{q}$), it is optimal to hold $\rho_H I - \tau'\rho_0 I$ units of liquid assets, in order to ensure the reinvestment in any case by getting plenty of liquidity. Thirdly, if the price of liquid assets is equal to the entrepreneur's expected return $\bar{q}$ generated by one unit of liquid assets, there is no difference whether to store liquid assets or not for the entrepreneur.

Step 3: Find the optimal investment level I^*.

Suggestion 1 and Equation (7) indicate that the optimal investment level is closely related to the liquid assets price q, specifically as follows:

if $q < \bar{q}$, the optimal investment level would be

$$I^* = \frac{1}{1-\rho_0-\tau'\rho_0+(q-1)(\rho_H-\tau'\rho_0)+\bar{\rho}} A;$$

if $q > \bar{q}$, the optimal investment level would be

$$I^* = \frac{1}{1-\rho_0-(1-\lambda)\tau'\rho_0+(1-\lambda)\rho_L} A;$$

if $q = \bar{q}$, the optimal investment level would be

$$I^*(x) = A\cdot\{1-\rho_0-[(1-\lambda)+\lambda x]\tau'\rho_0+ (\bar{q}-1)x(\rho_H-\tau'\rho_0)+(1-\lambda)\rho_L+\lambda x\rho_H\}^{-1}.$$

5 Comparative static analyses

The relationship between corporate growth and liquidity assets value will be given next.

Suggestion 2: The better the corporate growth is, the higher the enterprise can get from one unit of liquid assets, that means, liquid assets value $\bar{q}$ and corporate growth τ are positively correlated.

Proof: Proposition 1 indicates that the value of one unit of liquid assets for the enterprise is:

$$\bar{q} = 1+\lambda\frac{\tau'[1-(1-\lambda)(\rho_H-\rho_L)]-\rho_H}{(\rho_H-\tau'\rho_0)[1+(1-\lambda)\tau']}.$$

Solving the partial derivative of $\bar{q}$ with respect to τ':

$$\frac{d\bar{q}}{d\tau'} = -\lambda\frac{G}{J},$$

where:

$$J = \{(\rho_H-\tau'\rho_0)[1+(1-\lambda)\tau']\}^2,$$
$$G = [(1-\lambda)(\rho_H-\rho_L)-1](\rho_H-\tau'\rho_0)[1+(1-\lambda)\tau']+ \{\rho_H+\tau'[(1-\lambda)(\rho_H-\rho_L)-1]\}\rho_0[1+(1-\lambda)\tau']- \{\rho_H+\tau'[(1-\lambda)(\rho_H-\rho_L)-1]\}(\rho_H-\tau'\rho_0)(1-\lambda).$$

G can be simplified as:

$$G = [(1-\lambda)(\rho_H-\rho_L)-1-\rho_H(1-\lambda)](\rho_H-\tau'\rho_0)+ \{\rho_H+\tau'[(1-\lambda)(\rho_H-\rho_L)-1]\}\rho_0[1+(1-\lambda)\tau'],$$

based on Equation (3):

$$\rho_H+\tau'[(1-\lambda)(\rho_H-\rho_L)-1] \le 0,$$

it leads to:

$$(1-\lambda)(\rho_H-\rho_L)-1 \le 0$$

and as $\rho_H-\tau'\rho_0 > 0$, $G < 0$, then $d\bar{q}/d\tau' > 0$. Due to $\tau' = \tau/p_H$, $\frac{d\bar{q}}{d\tau} > 0$.

This means the better the corporate growth, the higher the liquid assets value. In other words, the highest liquid assets price $\bar{q}$ which can be accepted by the enterprise would increase with the increasing in corporate growth τ.

6 Numerical simulations

Now, some numerical calculations would be made for the theoretical results. Table 1 indicates the influence of corporate growth on liquid assets value $\bar{q}$, and then the optimal financing contract could be given in the condition of a certain liquid assets price. Where basic parameters are $q=1.5$, $A=1$, $p_H=0.4$, $p_H=0.2$, $B=0.4$, $R=3$, $\rho_H=0.6$, $\rho_L=0.3$, $\lambda=0.5$. Table 1 shows that, both of the entrepreneur's expected net utility U_b and liquid assets value $\bar{q}$ increase with the increasing in corporate growth τ. If the price of liquidity assets is higher than value, the probability of reinvestment $x=0$ in the case of higher reinvestment demand. The number of holding liquid assets $L=0$; if the price is lower than value, the enterprise will store enough liquid assets to meet his reinvestment need, and the number of holding liquid assets would decrease with increasing in corporate growth.

TABLE 1 The influence of corporate growth on liquid assets value and financing decision

τ	$\bar{q}$	x^*	L	I^*	R_b^*	U_b
0.30	1.045	0	0.000	1.667	3.333	0.833
0.32	1.102	0	0.000	1.695	3.390	0.898
0.34	1.165	0	0.000	1.724	3.448	0.966
0.36	1.237	0	0.000	1.754	3.509	1.035
0.38	1.320	0	0.000	1.786	3.571	1.107
0.40	1.417	0	0.000	1.818	3.636	1.182
0.42	1.533	1	0.250	1.389	2.778	1.278
0.44	1.675	1	0.232	1.449	2.899	1.435
0.46	1.856	1	0.212	1.515	3.030	1.606
0.48	2.094	1	0.190	1.587	3.175	1.794
0.50	2.423	1	0.167	1.667	3.333	2.000
0.52	2.913	1	0.140	1.754	3.509	2.228

7 Conclusions

Based on the external financing analysis framework under asymmetric information, this paper analyzes the influence of corporate growth on liquid assets value, and then discusses the enterprise's optimal financing decision in the condition of a given liquid assets price. The study shows that liquid assets value would increase with the increasing in corporate growth, in other words, the enterprise with higher growth is willing to pay higher price to store liquid assets. If the price of liquidity assets is higher than value, the enterprise would not hold liquidity assets, although this may make him unable to seize the growth opportunity in the case of higher reinvestment demand; if the price is lower than value, the enterprise will store enough liquid assets to meet his reinvestment need in any case for getting the growth opportunity; if the price is equal to value, there is no difference found in reinvesting or not.

Acknowledgments

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper. The authors acknowledge the financial support of the project “strategic trading behaviour of institutional investors and stock price volatility”, project No. 2011FZ016.

References

[1] Harford J 1999 Corporate Cash Reserves and Acquisitions *The Journal of Finance* 54(6) 1969-1997
[2] Faulkender M, Wang R 2006 Corporate Financial Policy and the Value of Cash *Journal of Finance* 61(4) 1957-90
[3] Denis D J, Sibilkov V 2007 Financial Constraints, Investment, and the Value of Cash Holdings *Review of Financial Studies* 23(1) 247-69
[4] Pinkowitz L, Williamson R 2007 What is the Market Value of a Dollar of Corporate Cash? *Journal of Applied Corporate Finance* 19(3) 74-81
[5] Kalcheva I, Lins K V 2007 International Evidence on Cash Holdings and Expected Managerial Agency Problems *Review of Financial Studies* 20(4) 1087-1112
[6] Tong Z X 2011 Company Diversification and the Value of Corporate Cash Holdings *Journal of Corporate Finance* 17(3) 741-58
[7] Drobetz W, Matthias C, Simone H 2010 Information Asymmetry and the Value of Cash *Journal of Banking and Finance* 34(9) 2168-84
[8] Myers S C and Majluf N S 1984 Corporate Financing and Investment Decisions When Companies Have Information that Investors Do Not Have *Journal of Financial Economics* 13(2) 187-221
[9] Mikkelson W H, Partch M M 2003 Do Persistent Large Cash Reserves Hinder Performance? *Journal of Financial and Quantitative Analysis* 38(2) 275-94
[10] Saddour K 2006 The Determinants and the Value of Cash Holdings: Evidence from French Companies *CEREG* 1-33
[11] Tirole J 2006 The Theory of Corporate Finance *Princeton University Press*: Princeton

Authors

Yaqin Lu, 1972, born in Dayao, Yunnan, China.

Current position, grades: associate professor in Economic Research Institute at Yunnan University of Finance and Economics.
Scientific interests: international trade theory and inter-regional trade theory.

Cunzhi Tian, born in 1969, Heqing, Yunnan, China

Current position, grades: professor at the Faculty of Management and Economics at Kunming University of Science and Technology of China.
University studies: PhD degree in Economics at Nankai University in 2001, Tianjin, China.
Scientific interests: securities market microstructure theory, financial engineering and corporate finance.
Experience: Postdoctoral position at Shanghai University of Finance and Economics, China.

Yi Wu, born in 1986, Wuhan, Hubei, China.

Current position, grades: lecturer at the Faculty of Management and Economics at Kunming University of Science and Technology in China.
Scientific interests: corporate finance and mathematical economics.

Identification and application of investors' risk appetite-based on the analysis of risk allocation of China multi-layer capital market system

Xiaoyuan Geng[1], Yongde Wang[1*]

[1]*School of Accountancy, HeiLongjiang Bayi Agricultural University, Daqing, 163319, China*

Received 1 October 2014, www.cmnt.lv

Abstract

In order to measure investors' risk appetite more accurately, from the focus on the investors' demand for the capital market, this article deduces the utility level of investors in the capital market by the inverse of the investor demand (only when the demand function satisfies integrablility, then it will be deduced the utility function inversely), and thus measure and identify the investors' risk appetite. While based on this theory approach, the paper empirically analyses the risk allocation of China multi-layer capital market, and the results show that: risk allocation of China multi-layer capital market system is non Pareto efficient, the risk allocation of each market does not meet the structure of multi-layer capital market established, but these problems can be improved by adjusting the market trading mechanisms.

Keywords: Stochastic demand; Integrability problem; Risk appetite; Risk allocation

1 Introduction

By correctly understanding the investors' risk appetite in the multi-layer capital market, it can be provided a strong basis for accurately formulating the policy for multi-layer capital market, formulated more targeted policy, security policy role to achieve the desired effect. October 30, 2009 28 SME GEM official visit in China, which means China multi-layer capital market system, has taken shape. The multi-layer capital market is needed to establish by reforming and developing capital market in China, because the single capital market structure is caused by inefficient capital markets, speculative prominent, regulatory costs and other problems of the institutional reasons [1], only through effectively operating the multi- layer capital market system can reduce the high transaction costs due to a single hierarchy lead, narrow the scope of the transaction costs incurred [2]. The purpose of the establishment of multi-layer capital market is not only to optimize the allocation of resources, but also to optimize the allocation of risk. Optimal allocation of resources refers to it provides a direct financing platform for the different types of companies which they can maximally overcome the obstacles to credit constraints, the optimal allocation of risk refers to it provides investment environment for the investors with different characteristics of risk preferences which consistent with their risk appetite, and launch of the GEM adapts to the demand of risk lovers. That is to say, the establishment of multi-layer capital market is not only to improve the efficiency of resource allocation, but also to improve risk allocation efficiency by establishing different levels of market risk, namely the risk allocation should achieve Pareto efficiency, the various risks fall on the risk appetite of investors with such features [3]. Current research focuses on the study of efficiency of resource allocation, the study of risk allocation efficiency is relatively less, but this aspect of the research for the development of the market is how important it is.

Therefore, the focus of this paper is allocation efficiency of multi-market risk capital, combining the purpose of multi-purpose capital market system established, the paper aims to measure the allocation efficiency of risk by identifying the investors' risk appetite in the market. Identification of investors' risk appetite can be attributed to its utility function for the study, while investors' utility function can be obtained by studying its demand function, which is the integrable problems, and the starting point of this study from the risk appetite is not common. That is to say, based on the analysis of investors' market demand function, if the demand function satisfies the integrability, namely the demand function has a symmetrical, semi-negative definite substitution matrix, then we can get the investors' utility function which consistent with their demand function, and identify the investors' risk appetite, thereby determine whether the investors' performance in the market consistent with the purpose of the establishment of multi-layer capital market, namely whether it is Pareto efficient.

Visibly, the establishment of investors' demand function for the market does not like the conventionally established model, namely under the given investors' utility function achieving the maximum utility for solving the utility function. Since the time of establishment of multi-layer capital market system is short, not mature enough, so to some extent, the current capital market is a speculative market [4]. When making investments, its purpose is not to become a corporate shareholder, with part of its title, but trying to take advantage of price movements to get the profits arising from bid-ask spread, when making investment decisions, investors don't consider there is any

** Corresponding author's* e-mail: byndwyd@sina.com

essential difference among the listed companies, but unsteadily they see them as symbol or code, they randomly select these symbol or code according to their own subjective and objective constraints. Therefore, according to investors' performance in the market, we construct the expected demand function of random variables. If this function is able to satisfy the integrability, we can get the investors' utility function. Utility function is an important economic analytical tool, which is real, and can be proved by various methods [5][6]. At the same time, it is also been widely used in the financial field[7]: used the utility function to improve Markowitz mean – variance portfolio model, resulting in asset allocation to meet the different investors' risk preferences[8]; also based on investors' appetite for risk and return, established investors' utility function, and thus used the utility function in the insurance industry[7]; certainly, also built the utility function by direct to measure the investors' risk attitude[9]. It can be seen, in the use of the utility function to solve problems, these analysis carried out under the prerequisite condition of known investors' risk appetite, but this article is based on the premise of investors' risk appetite unknown, by the utility function to identify investors' risk appetite ; although the article also build utility function, the utility function is constructed based on the investors' real need performance in the multi-layer capital market in this paper, rather than built directly.

Construct the investors' risk appetite measurement model in accordance with their utility unction in this article. In this regard, the paper does not use the standard deviation σ or variance σ^2 proposed by Markowitz (1952) and β coefficient method proposed by William Sharpe $\beta = \frac{\mathrm{COV}_{jM}}{\sigma_M^2}$, in which: σ_M^2 is the variance of the market portfolio. Because of these two methods does not reflect the risk difference of asset purchasing at different prices, for investors' risk are different facing the same kinds of asset purchasing at different prices, the purchasing price more higher the risk more greater. Strictly speaking, by σ and β measuring the risk is the same ideologically, there is no essential difference between them, because β is the variant of σ.

Then, in the second part of this article, we first construct stochastic demand model, and validate the integrability of this model, and then get the random utility model which consistent with investors' demand model; in the third part of this article, we construct risk appetite measurement model, in the fourth part we identify the risk appetite of China multi-layer capital market, determine its validity, that is to say identify the current investors' performance on risk allocation in the multi-layer capital market, whether it is achieved Pareto efficient; in the five part we correspond the policy and recommendations.

2 Construction of utility theory model – demand

Random expectation demand model raised by Becker is the most classic, most influential. He considered individual randomly assigned their wealth among competing goods under linear budget constraints, constructs the probabilistic choice models based on the assumption of uniform distribution of two commodities, non-satiation, and found that this model was still in line with the needs of law [10]. Since then a large number of scholars have conducted research based on Becker's model, Machina and Gul and Pesendorfer proposed, even the model resulting in the non-rational behavior can contain the satisfactory and limited rational behavior depending on the utility function and random selection [11][12], Sanderson reduced the assumptions and constraints of Becker's model, he expanded the Becker's model and used the expanded model to analyze the family output, and produced better results. Visibly random expected demand model have been developed more perfect. In this article, the benchmark model is still Becker's model, but reduces the probability distribution assumptions, while expand the two products to three, because this article studies the main market, small board and GEM of capital market, according to this method it can also be adapted to model of plurality products certainly. Although the article reduces the assumptions, the model still has excellent properties.

2.1 ASSUMPTIONS

Assumption 1: $X \equiv (X_1, X_2, X_3)$ represents the market vector, $X_i = X_1, X_2, X_3, i = 1,2,3$ separately represents the three markets: main market, small board and the GEM market, assuming that all investors who choose to enter the i th market constitute a general, and the overall homogeneity is the overall unit performance has the same preferences, while the differences in the presence among the overall unit makes this general heterogeneity.

Assumption 2: $p \equiv (p_1, p_2, p_3)$ represents the price vector relatively, and w represents the individual's total wealth, individual randomly allocates his total wealth w to three markets, $w = \sum_{i=1}^{3} \pi_i w$, π_i is the probability of the total wealth randomly assigned to the i th market, the wealth allocated to the i th market is w_i, $w_i = \pi_i w$. To avoid degradation, we assume the income and price are strictly positive and finite. So their budget constraint is:

$p_1 X_1 + p_2 X_2 + p_3 X_3 \le w$,that is to say,

$$p \cdot X = \sum_{i=1}^{3} p_i X_i \le w \tag{1}$$

Assumption 3: individuals randomly choose their willingness to participate in the sub-stock market, the cumulative distribution function is:

$F(x_1, x_2, x_3) \equiv \Pr(X_1 \le x_1, X_2 \le x_2, X_3 \le x_3)$

and the probability density function is $f(x_1, x_2, x_3)$.

$x \equiv (x_1, x_2, x_3)$, and $f(x) \equiv dF(x)/dx$.

To ensure its differentiability, we assume $f(x)$ is continuous. $x \in R_+^3$ is a standard set of requirements.

2.2 RANDOM EXPECT DEMAND MODEL OF THE i TH MARKET

Investors' demand for the i th market $(i=1,2,3)$, if $p_1X_1 \le w$, then the choice of X_1 is feasible. That is, if $X_1 \le X_1^{\max}$, $X_1^{\max}$ is the maximum possible demand of X_1, $X_1^{\max} \equiv w/p_1$. In this case, the viable condition of choice X_2 is need to be considered for the part of the wealth to X_1. If X_1 is feasible, $p_2X_2 \le w-p_1X_1$ set up, then X_2 is feasible, the maximum possible amount of consumption is $X_2^{\max} = w-p_1X_1$. Similarly, if X_1 and X_2 are feasible, $p_3X_3 \le w-p_1X_1-p_2X_2$ set up, then X_3 is feasible, the maximum demand is $X_3^{\max} = w-p_1X_1-p_2X_2$, and it can be summarized as $X_i^{\max} = w-\sum_{i=1}^{2} p_iX_i$.

Therefore, we make $X^{\max} \equiv (X_1^{\max}, X_2^{\max}(x_1), X_3^{\max}(x_1,x_2))$ as the largest possible demand vector on the condition of three markets.

Define $F(p,w) \equiv F(X_1^{\max}, X_2^{\max}(x_1), X_3^{\max}(x_1,x_2))$, it ensures the establishment of the budget constraint (1), given:

$$F(p,w) = \Pr(0 \le pX \le w) = \int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} \int_0^{X_3^{\max}(x_1,x_2)} f(x_1,x_2,x_3)dx_3dx_2dx_1 \tag{2}$$

Investors' random expect demand for the i th market to meet the:

$$\bar{x}_i(p,w) \equiv E(X_i \mid 0 \le p\cdot X_i \le w) = \frac{\int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} \int_0^{X_3^{\max}(x_1,x_2)} x_i f(x_1,x_2,x_3)dx_3dx_2dx_1}{F(p,w)} \tag{3}$$

Model (3) is investors' random expect demand model for the i th market. In essence, $\bar{x}_i(p,w)$ is the truncated expectation of random variable X_i, truncation occurs because all of the true value of the vector must be located within the budget set. In other words, $\bar{x}_i(p,w)$ is the expectation of random variable from $f(x)/F(p,w)$. From (2) and (3), it can be concluded that the demand function is zero-order homogeneous for prices and wealth. This result is easily confirmed, for θ >0, $\bar{x}(\theta p,\theta w) = \bar{x}(p,w)$.

Define $\bar{x}(p,w) \equiv (\bar{x}_1(p,w), \bar{x}_2(p,w), \bar{x}_3(p,w))$ as the random expect demand vector.

2.3 VALIDATION OF INTEGRABILITY OF RANDOM EXPECT DEMAND MODEL

When a given set of demand function $x(p,w)$, if it has a negative semi-definite symmetric substitution matrix, then this function can be set to satisfy the integrability, namely we can get the utility function set through the function set which consistent with.

If $S(p,w)$ is Slutsky substitution matrix, Σ is the variance – covariance matrix of $\bar{x}_i(p,w)$.

$S(p,w) = -\Sigma$, and the elements of Σ as follows:

$\Sigma_{ij} = Cov(X_i,X_j) \equiv E[(X_i-\bar{x}_i(p,w))(X_j-\bar{x}_j(p,w)) \mid X_1 \le X_1^{\max}, X_2 \le X_2^{\max}(x_1), X_3 = X_3^{\max}(x_1,x_2)]$, in this case if $i \ne j$, then the elements are the covariance $Cov(X_i,X_j)$, if $i=j$, then the elements are the variance $Var(X_i)$.

So $Q_i(p_i,w_i) = \bar{x}_i(p_i,w_i) = \frac{\int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} \int_0^{X_3^{\max}(x_1,x_2)} x_i f(x_1,x_2,x_3)dx_3dx_2dx_1}{F(p,w)}$, then $\bar{x}_i(p_i,w_i) = \frac{Q_i(p_i,w_i)}{F_i(p_i,w_i)}$, and

$$\left.\frac{\partial Q_i(p,w)}{\partial p_i}\right|_{\bar{w}} = \int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} x_i f(x_1,x_2,\ X_3^{\max}(x_1,x_2))\left(\left.\frac{\partial X_3^{\max}(x_1,x_2)}{\partial p_i}\right|_{\bar{w}}\right)dx_2dx_1 \tag{4}$$

$$\left.\frac{\partial F(p,w)}{\partial p_i}\right|_{\bar{w}} = \int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} f(x_1,x_2,\ X_3^{\max}(x_1,x_2))\left(\left.\frac{\partial X_3^{\max}(x_1,x_2)}{\partial p_i}\right|_{\bar{w}}\right)dx_2dx_1 \tag{5}$$

1) Verification of symmetry

$$\left.\frac{\partial Q_i(p,w)}{\partial p_j}\right|_{\bar{w}} = \int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} x_i,\ f(x_1,x_2,X_3^{\max}(x_{1,}x_2))\left(\left.\frac{\partial X_3^{\max}(x_{1,}x_2)}{\partial p_j}\right|_{\bar{w}}\right)dx_2dx_1,$$

$$\left.\frac{\partial F_i(p,w)}{\partial p_j}\right|_{\bar{w}} = \int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} f(x_1,x_2,X_3^{\max}(x_{1,}x_2))\left(\left.\frac{\partial X_3^{\max}(x_{1,}x_2)}{\partial p_j}\right|_{\bar{w}}\right)dx_2dx_1,$$

after calculation it can be obtained:

$$\left.\frac{\partial \bar{x}_i(p,w)}{\partial p_j}\right|_{\bar{w}} = \int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} (x_i - \bar{x}_i(p,w))\left(\frac{f(x_1,x_2,X_3^{\max}(x_1,x_2)}{F(p,w)} \left.\frac{\partial X_3^{\max}(x_1,x_2)}{\partial p_j}\right|_{\bar{w}}\right) dx_2 dx_1 \tag{6}$$

Among them, $$\left.\frac{\partial X_3^{\max}(x_1,x_2)}{\partial p_j}\right| = \frac{\bar{x}_j(p,w) - x_j}{p_3} \tag{7}$$

Let (7) substitute into (6), we get,

$$\left.\frac{\partial \bar{x}_i(p,w)}{\partial p_j}\right|_{\bar{w}} = -\int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} [x_i - \bar{x}_i(p,w)][x_j - \bar{x}_j(p,w)]\left[\frac{f(x_1,x_2,X_3^{\max}(x_1,x_2))}{F(p,w)} \frac{1}{p_3}\right] dx_2 dx_1 \tag{8}$$

Similarly it can be calculated:

$$\left.\frac{\partial \bar{x}_j(p,w)}{\partial p_i}\right|_{\bar{w}} = -\int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} [x_i - \bar{x}_i(p,w)][x_j - \bar{x}_j(p,w)]\left[\frac{f(x_1,x_2,X_3^{\max}(x_1,x_2))}{F(p,w)} \frac{1}{p_3}\right] dx_2 dx_1 \tag{9}$$

Thus, (8)=(9),that is to say, $\left.\frac{\partial \bar{x}_i(p,w)}{\partial p_j}\right|_{\bar{w}} = \left.\frac{\partial \bar{x}_j(p,w)}{\partial p_i}\right|_{\bar{w}}$,this substitution matrices satisfy the symmetry.

2) Validation of semi-negative definite

For $\left.\frac{\partial \bar{x}_i(p_i,w_i)}{\partial p_i}\right|_{\bar{w}_i} = \left.\frac{\partial Q_i(p_i,w_i)}{\partial p_i}\right|_{\bar{w}_i} \left(\frac{1}{F_i(p_i,w_i)}\right) - \left.\frac{\partial F_i(p_i,w_i)}{\partial p_i}\right|_{\bar{w}_i} \left(\frac{Q_i(p_i,w_i)}{F^2(p_i.w_i)}\right)$, after arrangement we get:

$$\left.\frac{\partial \bar{x}_i(p_i,w_i)}{\partial p_i}\right|_{\bar{w}} = \left(\frac{1}{F_i(p_i,w_i)}\right)\left[\left.\frac{\partial Q_i(p_i,w_i)}{\partial p_i}\right|_{\bar{w}_i} - \bar{x}_i(p_i,w_i)\left.\frac{\partial F_i(p_i,w)}{\partial p_i}\right|_{\bar{w}_i}\right] \tag{10}$$

Let (4) (5) substitute into (10), after arrangement, we get:

$$\left.\frac{\partial \bar{x}_i(p,w)}{\partial p_i}\right|_{\bar{w}} = \int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} [x_i - \bar{x}_i(p,w)]\left[\frac{f(x_1,x_2,X_3^{\max}(x_1,x_2))}{F(p,w)} \left.\frac{\partial X_3^{\max}(x_1,x_2)}{\partial p_i}\right|_{\bar{w}}\right] dx_2 dx_1 \tag{11}$$

$$\left.\frac{\partial X_3^{\max}(x_1,x_2)}{\partial p_i}\right| = \frac{\bar{x}_i(p,w) - x_i}{p_3} \tag{12}$$

Let (12) substitute into (11), we get:

$$\left.\frac{\partial \bar{x}_i(p,w)}{\partial p_i}\right|_{\bar{w}} = -\int_0^{X_1^{\max}} \int_0^{X_2^{\max}(x_1)} [x_i - \bar{x}_i(p,w)]^2\left[\frac{f(x_1,x_2,X_3^{\max}(x_1,x_2))}{F(p,w)} \frac{1}{p_3}\right] dx_2 dx_1 \tag{13}$$

Meanwhile, $Var(X_i) \equiv E[(X_i - \bar{x}_i(p,w))^2 \,|\, X_1 \le X_1^{\max}, X_2 \le X_2^{\max}(x_1), X_3 = X_3^{\max}(x_1,x_2)]$, so

$$\left.\frac{\partial \bar{x}_i(p,w)}{\partial p_i}\right|_{\bar{w}} = -Var(X_i) \tag{14}$$

At this time, the price effect of expect demand is non-positive.

And because $Cov(X_i,X_j) \equiv E[(X_i - \bar{x}_i(p,w))(X_j - \bar{x}_j(p,w)) \,|\, X_1 \le X_1^{\max}, X_2 \le X_2^{\max}(x_1), X_3 = X_3^{\max}(x_1,x_2)]$, $i \ne j$,

$$\left.\frac{\partial \bar{x}_i(p,w)}{\partial p_j}\right|_{\bar{w}} = -Cov(X_i,X_j).$$

Because of the variance – covariance matrix Σ is semi-positive definite, $S(p,w)$ is semi-negative definite.

It can be seen that this model satisfies the integrability, we can get the utility function which consistent with it, that is to say we can obtain three different investtors' utility function from the main market, small board and the GEM market, that is $U(X_i) = \int_p \bar{x}_i(t,w)dt$.

3 Measurement of risk appetite

When the investors make their own choice, and at the same time they know the results they will have to bear in the future, or benefit from this choice, or bear the loss, or neither benefit nor lost, in a risk-free status. When investors choose to enter the main market, small board and GEM market, they tend to be more concerned about their possible gain or loss, if R_i represents the collection which contains all appeared unknown results after investors choose to enter the i th markets, R_i^+ means the benefit collection from investors' choice, R_i^- means loss collection from investors' choice, it is clear that $R_i^+, R_i^- = R$, $R_i^+ \cap R_i^- = \Phi$.

Risk is the possible loss due to the uncertainty to investors, more objective expression is that when investtors choose to fall R^+, investors will benefit, will not bear the loss; If when the investor's choice falls R^-, that is $R^- \neq \Theta$, at this time investors have to face the risk.

Investors bear the risk:

$$E_{R^-}(U_i) = \int_{R^-} U_i dF(U_i) = \int_{R^-} U_i f(U_i) dU_i \text{ , } E_{R^-}(U_i) < 0$$

Investors obtain the benefit:

$$E_{R^+}(U_i) = \int_{R^+} U_i dF(U_i) = \int_{R^+} U_i f(U_i) dU_i \text{ , } E_{R^+}(U_i) > 0.$$

Assuming r is risk-free return, and then we construct the measurement tools of investors' risk appetite:

$$\eta_i = r - E_{R^+}(U_i) - E_{R^-}(U_i) \tag{15}$$

When $\eta_i > 0$, investors are risk lovers, and the value of η_i the larger indicates the degree of risk investtors pursuit the higher. At this point, investors are more willing to make a choice, rather than election of fair game, because the exposure may be $-E_{R^-}(U_i)$, there is loss of revenue r (because if they don't choose, r is the opportunity costs for investors) may outweigh the benefits obtained possibly.

When $\eta_i < 0$, investors are risk averse, and the value of η_i the smaller indicates the degree of risk adverse avoidance the higher. Because the choice may not only bring greater revenue than raised losses, but also greater than the sum of r and $-E_{R^-}(U_i)$, they believes the risk of fair game is too high, so they should be risk averse.

When $\eta_i = 0$, investors consider at this time to make a choice is equal to fair game, then the investors are risk neutral, they does not behave like a risk averse too conservative, and does not behave as risk lovers too optimistic, would be more objective treatment of the problem, when $\eta_i > 0$, he will not select, on the contrary at the time $\eta_i < 0$, they will select, because at this time they believe in their favor.

Through the above theoretical analysis, we will find that: when $\eta_i = 0$ and when $\eta_i < 0$, there is the intersection between the investors' choice. Because under certain conditions, the risk netballers' choice is similar with the risk averse', that is, it is difficult to identify the similar choice whether from the risk averse investors or from the risk neutral investors. Visibly, this approach is not suitable for determining the real investors' specific type of risk appetite.

To this end, we have to be adjusted on the basis of this model, let η_i will be used to be replaced by $\eta_i^{\max}$, which is the measurement maximum of investors withstand, the model is:

$$\eta_i^{\max} = \max[r - E_{R^+}(U_i) - E_{R^-}(U_i)] \tag{16}$$

If and only if

$$E_{R^+}(U_i) = E_{R^+}^{\min}(U_i), E_{R^-}(U_i) = E_{R^-}^{\min}(U_i)$$

According to the assumptions (1) it is available that $E_{R^+}^{\min}(U_i)$ and $E_{R^-}^{\min}(U_i)$. Therefore, the model (16) can be used to identify the investors' risk preferences in different markets.

4 Empirical analysis based on the risk allocation of China multi-layer capital market

On the basis of theoretical analysis framework, this article choose the main market, small board, the GEM market as the research object, using the method of empirical analysis to analyze the investors' risk appetite and allocative efficiency in each market.

4.1 INDICATORS AND DATA SELECTION

According to the foregoing analysis, the model contains three corresponding variables: price, quantity, and has a wealth of purchase. To ensure the comparability, the paper selects the composite index. In the main market, we select the closing price on the Shanghai Composite Index and Shenzhen Composite Index as the price, and their volume (unit: million) as the corresponding quantity demanded; in the small board, select the small plate KLCI closing price, volume (Unit: million) as the corresponding variables; in the GEM, select the GEM KLCI closing price, volume (unit: million) as its variables; meanwhile, we sum the turnover (unit: million) corresponding to the Shanghai Composite Index and Shenzhen Composite Index, the small plate KLCI and GEM KLCI, then obtain the total wealth in demand market which investors can distribute. Since the GEM KLCI launches from August 20, 2010, the data in this paper is August 20, 2010 to November 21, 2013. Theoretically, the Shanghai Composite Index and Shenzhen Composite Index are main market index, should be unified analysis, but found that in the Shanghai Stock Exchange and Shenzhen Stock Exchange market there are significant differences in the degree of risk appetite, therefore, the paper retains the Shanghai and Shenzhen two main markets. The above data is from CSMAR database.

4.2 STATIONARY TEST OF DATA

Stationary test takes ADF test which is the most commonly to be used. To avoid the heteroscedasticity of estimation model, this article deals with the data to be logarithmic treatment, and do the stationary test to these variables and the variables after the first differences, the test results are shown in table 1:

TABLE 1 Stationary test results

Variables	ADF test statistic	$p-value$
ln(Shanghai Composite Index)	-1.523085	0.8198
ln(*volumeof* Shanghai Composite Index( )	-5.92698	0.0000
ln(Shenzhen Composite Index)	-1.474319	0.8364
ln(*volumeof* Sshenzhen Composite Index)	-4.567777	0.0002
ln(Small board inde)	-1.34279	0.8751
ln(volume of Small board index)	-4.878331	0.0004
ln(GEM Index)	-1.852633	0.6764
ln(volume of GEM Index)	-4.563277	0.0014
ln(Total Turnover)	-4.160081	0.0009
Variables after the first differences	ADF test statistic	$p-value$
d ln(Shanghai Composite Index)	-16.90434	0.0000
d ln(*volumeof* Shanghai Composite Index)	-18.11173	0.0000
d ln(Shenzhen Composite Index)	-16.06674	0.0078
d ln(*volumeof*Sshenzhen Composite Index)	-15.20731	0.0000
d ln(Small board index)	-15.58342	0.0000
d ln(volume of Small board index)	-15.2738	0.0000
d ln(GEM Index)	-15.94434	0.0000
d ln(volume of GEM Index)	-16.01485	0.0000
d ln(Total Turnover)	-15.44591	0.0000

The results shows that: the variables after logarithmic treatment, at 10% significance level, the Shanghai Composite Index, Shenzhen Composite Index, the small board index and GEM index are non-stationary sequence, the rest is smooth sequence; after first differential treatment with the logarithmic variables, all variables are stationary sequence.

4.3 SELECTION AND ESTIMATION OF RANDOM EXPECT DEMAND MODEL

To ensure that the estimated demand model satisfies the integrability, based on the results of stationary test, we select the first difference variables as the variables, so the model is double logarithmic first-order differential form. Also, because of the differential model, the model does not contain a constant term. In this paper, the level of significance is $\alpha = 0.05$. The basic model is:

$$d\ln x_i = \sum_{i=1}^{3} \alpha_i d\ln p_i + \beta d\ln w + \varepsilon_i \tag{17}$$

Among them, x_i is the investors' demand for i th market, p_i is the price for i th market, w is the total wealth for investors in the capital market, α, β are the parameters to be estimated.

Meanwhile, from $d\ln x_i = \sum_{i=1}^{3} \alpha_i d\ln p_i + \beta d\ln w$, we can obtain $\ln x_i = \sum_{i=1}^{3} \alpha_i \ln p_i + \beta \ln w + c$, wherein C an arbitrary constant is. To facilitate the analysis, let $c = 0$, then $\ln x_i = \sum_{i=1}^{3} \alpha_i \ln p_i + \beta \ln w$, the deformation can be obtained:

$$x_i = \prod_{i=1}^{3} p_i^{a_i} w^{\beta} \tag{18}$$

That is demand model for i th market.

1) Demand model of main Market

a. Demand model of Shanghai main market

First differential model of Shanghai main market:

$$d\ln x_{Shanghai} = 1.1289 d\ln w$$
$$(22.0965)$$
$$\bar{R}^2_{Shanghai} = 0.6296$$

Demand model: $x_{Shanghai} = w^{1.1290}$

From the above first difference model, we can see that, $\bar{R}^2_{Shanghai} = 0.6296$ indicates that the extent of the first differential model of Shanghai Stock Exchange Main Board fit well, the equation is significant, and only the total wealth variable coefficients is significant, that is to

say changes in investors' demand for the Shanghai Stock Exchange main board market is only the case with investors owned total wealth, and shows the same relationship to changes.

b. Demand model of Shenzhen main market

First differential model of Shenzhen main market:

$$d\ln x_{Shenzhen} = -2.4655 d\ln p_{Shanghai} + 6.0191 d\ln p_{Shenzhen} - 3.4643 d\ln p_{Smallboard} - 0.8143 d\ln p_{GEM} + 0.9695 d\ln w$$

(-4.3874) (3.9071) (-3.0724) (-2.8082) (79.4239)

$$\bar{R}^2 = 0.9564$$

Demand model:

$$x_{Shenzhen} = p_{Shanghai}^{-2.4654} p_{Shenzhen}^{6.0191} p_{Smallboard}^{-3.4643} p_{GEM}^{-0.8143} w^{0.9695}$$

Seen, $\bar{R}^2 = 0.9564$ indicates the model fits well, and the coefficients of prices and total wealth variables of Shanghai Stock Exchange main market, Shenzhen main market small board and the GEM market are significant, namely changes in investors' demand for the Shenzhen main market are related to the changes of price of Shanghai main market, Shenzhen main market, small board market, GEM market and investors' total wealth. And changes in investor demand of Shenzhen main market with its own price changes and changes in total wealth is in the same direction of changes, in relationships with the price changes of other alternative market such as Shanghai Stock Exchange main market, small board and the GEM market inversely to changes in market relations.

2) Demand model of small board market

First difference model of Small board:

$$d\ln x_{smallboard} = 3.1144 d\ln p_{smallboard} + 0.8891 d\ln w$$

(2.0220) ((53.3173)

$$\bar{R}^2 = 0.9077$$

Demand model: $x_{smallboard} = p_{smallboard}^{3.1144} w^{0.8891}$

Seen, $\bar{R}^2 = 0.9077$ indicates the model fits well, while the coefficients of variables of small board market price and total wealth are significant, that is, changes in investors' demand for small board market are related to its own price and investors' total wealth. And changes in investors' demand for small board market with its own price changes and changes in total wealth was the relationship in the same direction.

3) Demand model of GEM market

First differential model of GEM:

$$d\ln x_{GEM} = -7.7118 d\ln p_{Shenzhen} + 9.6736 d\ln p_{GEM} + 0.7874 d\ln w$$

(-2.9680) (8.6117) (16.6508)

$$\bar{R}^2 = 0.5275$$

Demand model: $x_{GEM} = p_{Shenzhen}^{-7.7118} p_{GEM}^{9.6736} w^{0.7874}$

Seen, $\bar{R}^2 = 0.5275$ indicates the model fits better, the coefficients of variables of prices of Shenzhen main market and GEM market and total wealth are significant, changes in investors' demand for the GEM market are related to its own market price changes, price movements of Shenzhen main market and investors' total wealth. And changes in investors' demand for the GEM market with its own price changes and changes in total wealth had a positive relationship, with the price changes of its alternative market Shenzhen main market inversely to changes in the relations.

Through the above analysis, it shows that changes in investors' demand for i th market are related to its own price in the same movement direction(except Shanghai Stock Exchange Main Market), which reflects in the actual market with the increasing price the amount of investors to buy the stock increases, with the decline in stock prices the amount reduces, namely "chase sell"; reverse changes in the relationship with the price of other alternative market(except Shanghai Stock Exchange Main Market), that is, after the prices of alternative markets increasing investors will reduce the demand for changes the original market, in the situation of wealth unchanged, will increase the wealth of assigned to the alternative markets, which is an increase of the purchase of the alternative markets, and conversely it is also be established, which once again confirms the "chase sell " phenomenon; with the same movement relationship to the investors' own overall wealth, that the total wealth increasing investors will increase their demand for the market, the total wealth reducing the investors' demand for the market will be reduced.

4.4 UTILITY FUNCTION AND DISTRIBUTION SET

1) Utility function of each market

According to the definition of the utility function of the foregoing, the utility function can be obtained in each market:

a. Utility function of Main Market

Utility function of Shanghai Main Market:

$$u_{Shanghai} = \int_{p_{Shanghai}^0}^{p_{Shanghai}^1} w^{1.1290} dp_{Shangha}$$

Among them, $p^0_{Shanghai}$ is the base price of the Shanghai main market, $p^1_{Shanghai}$ is the reporting period price of Shanghai main market.

When we get the utility function of Shanghai main market, but also shows that this layer of meaning: when investors in Shanghai main market randomly choose whether to enter the market, or increase or decrease the market demand, actually investors follow their own implied the selecting behavior when the utility function maximized in Shanghai main market, that is, investors' demand for Shanghai main market is guided to conduct under the above utility function essentially.

Utility function of Shenzhen main market:

$$u_{Shenzhen} = \int_{p^0_{Shenzhen}}^{p^1_{Shenzhen}} p_{Shanghai}^{-2.4654} p_{Shenzhen}^{6.0191} p_{Smallboard}^{-3.4643} p_{GEM}^{-0.8143} w^{0.9695} dp_{Shenzhen}$$

Among them, $p^0_{Shenzhen}$ is the base price of Shenzhen main market, $p^1_{Shenzhen}$ is the reporting period prices of Shenzhen main market.

Similarly we can see that the utility function of Shenzhen main market is the basis to guide investors' demand choice behavior in Shenzhen main market, and its conducting code is to maximize the utility function of Shenzhen main market.

Obviously, although the current main market is constitute by Shanghai main market and Shenzhen main market together, there is a big difference for the investors following the conducting code in the two main market, that is, investors follow the two different conducting codes in Shanghai and Shenzhen main markets.

b. Utility function of small board

Utility function of Small board:

$$u_{Smallboard} = \int_{p^0_{Smallboard}}^{p^1_{Smallboard}} p_{Smallboard}^{3.1144} w^{0.8891} dp_{Smallboard}$$

Among them, $p^0_{Smallboard}$ is the base market price of small board, $p^1_{Smallboard}$ is the reporting period price of small board market.

After the above analysis, investors' demand for small board market is completed under the guidance of the utility function of small board.

c. Utility function of GEM

Utility function of GEM:

$$u_{GEM} = \int_{p^0_{GEM}}^{p^1_{GEM}} p_{Shanghai}^{-7.7118} p_{GEM}^{9.67363} w^{0.7874} dp_{GEM}$$

Among them, p^0_{GEM} is the base price of GEM market, p^1_{GEM} is the reporting period price of GEM market.

Similarly, investors get the guidelines of demand of GEM, that is, actually investors' behavior is the performance when the utility function maximize in the GEM.

2) Distribution set of utility function of each market

Based on the above utility function to estimate the utility level in different markets, due to the distribution of the utility level sequence for each market is unknown, the paper carries out the empirical distribution test. The paper uses the Watson test, Cramer-von Mises test, Kolmogorov testing and other testing methods to test the four markets utility level sequences goodness of fit with the theoretical distribution (normal, chi-square distribution, exponential distribution, extreme value distribution, logistic distribution, the Pareto distribution, uniform distribution, etc.) in Shanghai main market, Shenzhen main market, small board and GEM, the results are shown in table 2:

TABLE 2 Empirical distribution test of utility level sequence (original assumptions: utility level obeys to a particular sequence of distribution)

sequence	distribution	Test method	Asymptotic distribution statistics	Limited sample adjustment statistics	$p-value$
$u_{Shanghai}$	exponential	Watson	0.1139	0.1141	0.226
$u_{Shenzhen}$	normal	Watson	0.1179	0.1181	0.239
$u_{Smallboard}$	exponential	Cramer-von Mises	0.1051	0.1051	0.2985
u_{GEM}	uniform	Kolmogorov	0.0033	0.0581	0.9933

Table 2 shows, the utility level sequences for Shanghai main market obeys to the exponential distribution by Watson test, because of $p-value=0.226>\alpha=0.05$; similarly we can see that the utility level sequences for Shenzhen main market obeys to the normal distribution by Watson test, the utility level sequences for small board obeys to the exponential distribution by Cramer-von Mises test, and the utility level sequence GEM obeys to the exponential distribution by Kolmogorov test.

Using the maximum likelihood method to estimate the parameters of each distribution density function, the results are shown in table 3:

TABLE 3 The utility level sequence probability density function and its parameters estimate

sequence	Density functional form	Associated parameter estimates and test values	
$u_{Shanghai}$	$f(u)=\frac{1}{\mu}\exp(-\frac{u-a}{\mu})$	*parameter 1* $a=21041391$ $p-value$ 0.0296	*parameter 2* $\mu=6.10E+09$ $p-value$ 0.0000
$u_{Shenzhen}$	$f(u)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}(u-\mu)^2)$	parameter 1 $\mu=0.0052$ $p-value$ 0.0485	parameter 2 $\sigma=0.3331$ $p-value$ 0.0000
$u_{Smallboard}$	$f(u)=\frac{1}{\mu}\exp(-\frac{u-a}{\mu})$	parameter 1 $a=3.06E+14$ $p-value$ 0.0425	parameter 2 $\mu=2.96E+17$ $p-value$ 0.0000
u_{GEM}	$f(u)=\frac{1}{b-a}$	parameter 1 $a=1.52E+08$ $p-value$ 0.0399	parameter 2 $b=8.72E+10$ $p-value$ 0.0000

In table 3, the second column shows the form of density distribution function of different markets utility level sequence obedience, third and fourth columns are given the estimate of parameters of density function and test conditions. And by determining $p-value$ and the significant level, it shows the estimate of parameters in density functions are significant ($p-value<\alpha$).

4.5 MEASUREMENT OF RISK APPETITE

Based on the each density function we can calculate the risk and gains constructed before, select the minimal risk $E_{R^-}^{\min}(U_i)$ and minimal gains $E_{R^+}^{\min}(U_i)$ and bring them into risk appetite measurement model, because the risk $E_{R^-}(U_i)$ is negative, and measure the degree of risk appetite in different markets, thus identify them. In risk appetite measurement models, this article defines the one-year bank deposit rates 3.5% as the risk free interest rate, and the measurement unit of "Fortune" indicators is million, so to keep the data caliber consistency here, we adjust the risk free rate to 350 (per million). According to the calculation we can separately get $E_{R^+}^{\min}(U_i)$, $E_{R^-}^{\min}(U_i)$ and $\eta_i^{\max}$, and the results are listed in table 4:

TABLE 4 Measurement and identification of risk appetite of multi-layer capital market

Markets	risk free interest rate r	$E_{R^+}^{\min}(U_i)$	$E_{R^-}^{\min}(U_i)$	$\eta_i^{\max}$	risk appetite
Shanghai main market	350	1079370.6029	-238860113539527	238860112460506	risk lovers
Shenzhen main market	350	0.0001	-67184.2497	67534.2496	risk lovers
Small market	350	25458830575936	-267706412381378000	267680953550803000	risk lovers
GEM	350	20301.4482	-14424635394	14424615442.5215	risk lovers

According to table 4, in China current multi-layer capital market, no matter in which capital market, investors are expressed as risk lovers, only the degree of risk appetite is different. By comparison, the degree of risk pursuit of Shenzhen main market is the weakest, followed by the GEM market, the Shanghai main market, the highest level of risk pursuit is small board market, that is to say, $\eta^{\max}_{Smallboard} > \eta^{\max}_{Shanghai} > \eta^{\max}_{GEM} > \eta^{\max}_{Shenzher}$. Although the Shanghai main market and Shenzhen main market together constitute the current overall main market, at the point of view of the degree of investors' risk loving, these two markets and can not be explained as a whole, there is a clear difference between them. The original intention of establishing multi-layer capital market is to distinguish the different levels of the capital market among them, meanwhile the performance of the internal layer of risk should be roughly the same, the performance should be the difference among the risk layer, and the investors' attitude to risk shows a progressive trend, that is, the weakest degree of market risk hobby should be the main market, the degree of small board risk hobby is in the middle, the degree of GEM risk hobby is the strongest. Through the comparison of current realities and theoretical purposes, it is not difficult to find that there is a clear departure between them, not only on the level of the layers the risk away from the hobby, but also on the level of internal layer it clearly does not conform the intention of the theory. The risk appetite of main market is not the weakest, but the situation of Shanghai Stock Exchange main market also exceeds GEM which should have been the strongest risk-loving, and small board shows the strongest degree of risk-loving. This shows that the current multi-layer capital market is not reasonable division for investors with different risk preferences, which shows the current multi-layer capital market in China is non-Pareto efficient on risk allocation, Pareto improvement be needed.

5 Policy and recommendations

We can say that it is a state of disorder China current multi-layer capital market in the risk allocation, the root cause resulting in this non-valid and disorder may be the trading mechanism of the current market. Because in the multi-layer capital market, there is difference on the basis in each market, and there is a hierarchy among market distinction, these indicate that each market has its own characteristics, and in China the basically same transaction mode is still in the use of multi-layer capital market, that is, in the use of uniform, symmetric price limits, which is obviously inconsistent with the characteristics of multi-layer capital market. To change this inconsistent situation, we must distinguish the various markets, with different trading patterns, according to the characteristics of each market to develop and implement the different price limits, such as appropriately rising the price limits of small board and GEM in accordance with their level of risk should bear, or introducing the asymmetric price limits for each market, in order to improve the Pareto efficiency, so that China multi-layer capital market can be improved.

References

[1] Fangsheng Zhou 2003 Constructed pyramid capital market system *China Youth College for Political Sciences* **22**(1)83-6

[2] Dezong Han 2005 Analysis of the stock market to build a multi-level transaction costs *Economic Theory and Business Management* (8) 21-6

[3] Lasheng Li, Guan Minfang, ShenPing 2010 Research of Risk allocation efficiency on China multi-layer capital market system *Guangdong University of Finance Journal* **25**(5) 26-39

[4] Liping Xu, Huang Xiaoqi 2009 Listed companies to introduce strategic investor behavior *Statistics and Decision Research* **288** (12) 144-6

[5] Bingxia Zhang 2008 Existence of utility function *Northern China Water Conservancy and Hydropower College* **29** (1)104-5

[6] Zongqian Liu, Fu Weiqiang, Feng Sufen 2008 Another proof of expected utility function theorem *Capital Normal University (Natural Science)* (2) 6-17

[7] Sufen, Feng 2010 Utility function used in finance[J] B*eijing Polytechnic College* (1)

[8] Lijun Song, Yang Yongyu 2008 Investment portfolio based utility function *Beijing University of Chemical Technology* **9** (2)119-21

[9] Zhengjie Ren, Zhou Feng 2006 EV utility function in the Shanghai Securities Market *Henan Science* **24** (4) 600-3

[10] Becker G S1962 Irrational Behavior and Economic Theory *Journal of Political Economy* **45**(1) 23-37

[11] Machina M J 1985 Stochastic Choice Functions Generated from Deterministic Preference over Lotteries **95**(4)161-80

[12] Gul F Pesendorfer W 2006 Random Expected Theory *Econometrica* **24**(1)82-6

Author

Xiaoyuan Geng, born on October 22, 1981, Heilongjiang province of China

Current position, grades: lecturer in Heilongjiang Bayi Agricultural University; PH.D. candidate in statistics in Tianjin University of Finance and Economics
University studies: Bachelor degree in statistics from Tianjin University of Finance and Economics in 2004; masters degree in statistics from Yunnan University of Finance and Economics in 2007; now PH.D. in statistics candidate in Tianjin University of Finance and Economics
Scientific interest: Mathematical Statistics; Technical analysis of financial risk
Publications: She has published more than 10 papers.

Yongde Wang, born in 1964, Heilongjiang province of China

Current position, grades: Professor; PH.D.
University studies: Bachelor degree in accounting from Heilongjiang Bayi Agricultural University in 1986; PH.D. in management in Northeast Forestry University
Scientific interest: Agricultural trade; Accounting Theory and Methods
Publications: He has published more than 30 papers

The Decision of Scrap Reverse Logistics Operation Mode for Steel Enterprises Base on Evolutionary Game Theory

Yan LI[1], Fuyu Wang[1*]

[1] *Department of Industrial Engineering, Anhui University of technology, Maanshan, China, 243032*

Received 1 March 2014, www.cmnt.lv

Abstract

This paper discusses how steel enterprises choose the appropriate scrap reverse logistics operation mode from self-operation mode and the third party mode. To solve the problem, evolutionary game theory is used to research the game relation of cooperation between steel enterprises and the third -party enterprises. Firstly, the replication dynamic equations of both players are build based on payoff matrix .Secondly, the evolutionary stable strategies are acquired by stability analysis on evolution dynamic process of the game two players. Finally, combined with numerical simulation, some factors that impact stable strategy choice are analyzed such as initial state of system, extra income and risk cost of cooperation, invested initial cost and loss for cooperation. The conclusion provides theoretical reference for steel enterprises selecting scrap reverse logistics mode.

Keywords: scrap reverse logistics mode;evolutionary game theory;replication dynamic equation;steel enterprises

1 Introduction

During the twelfth five-year plan, China sped up the adjustment of its industrial structure and accelerated the transformation of its development mode to develop a low-carbon economy and circular economy. In this context, the steel industry, which is a great energy consumer and a serious polluter, was targeted to take a low carbon path. Therefore, many steel enterprises used a variety of strategies to actively meet the challenge of the low carbon economy. Among them, developing scrap steel industry and increasing scrap supply have a high practical value. Scrap which is different from the other wastes is the main steel-making raw material. Comprehensive energy consumption of steel production, carbon emissions and mineral resources exploittation can be reduced by using scrap in the steel production process [1]. So scrap reverse logistics is one of the effective means to implement a low carbon economy of steel industry. At present, there are self-operation mode, alliance mode and the third party logistics mode for enterprises [2]. This paper only discusses self-operation mode and the third party logistics mode. The former, is that steel enterprises themselves set up reverse logistics system and manage scrap steel reverse logistics activities of enterprise inner such as recycling, transport and reprocessing. The advantage of this mode is mainly that enterprises can control the operations of each link, efficiently recycle resources and reduce raw material cost. It also has disadvantages:transport costs are high, scrap reverse logistics may conflict with forward logistics, technical defects will hinder the development of reverse logistics, etc. The latter is that steel enterprises entrust scrap recycling work to the third party logistics in the form of an agreement and pay some service fees to them. The advantage of this mode is that the cost of building reverse logistics for steel enterprise can be reduced, more money can be saved to develop the main core business, the competitiveness of the enterprises can be improved. However, it also has certain disadvantages, such as product information are easily leaked, the number and time of recycling products are uncertain, the requirements of production cannot be met, the real power of the third party reverse logistics enterprises cannot be handled well enough that the quality of scrap steel may be poor. In a word, the two kinds of reverse logistics operational mode have both advantages and disadvantages. In practice, the decision of scrap reverse logistics operation mode is one of the key issues to consider for steel enterprises.

In recent years, many domestic and foreign scholars studied the decision-making of reverse logistics operation mode from different facets, such as, Ravi evaluated the reverse logistics operational decisions from the customer, internal business, innovation and learning, finance [3]. Ren Mingming used the fuzzy comprehensive evaluation method to establish mathematical model of operation mode selection [4], Chen Zhikui applied network analytic hierarchy process to analyze the decision of reverse logistics operation mode based on setting up evaluated index [5]. Cheng Guangping put forward an operation mode selection evaluation method aiming at the characteristics of domestic reverse logistics by using Fuzzy hierarchy analysis[6] .In the above literature, scholars largely carried out qualitative and quantitative analysis by considering the influencing

* ***Corresponding author***'s e-mail: liyanshangshang@dddd.com

factors of enterprise itself, few scholar considered the impact of decisions of other enterprises and the third party logistics enterprises. In fact, the problem of choosing the self-operation mode or the third party mode in scrap reverse logistics for steel enterprises can be regarded as game analysis between steel enterprises and the third-party logistics enterprises. For instance, Ceng Minggang used the "principal-agent" model to study the reverse logistics operation mode selection[7], however, he simply analyzed the game process under the assumption that two players were completely rational which did not accord with reality. Therefore, evolutionary game theory based on bounded rationality is used to study evolutionary game relation between steel enterprises and the third party logistics enterprises in the paper. The steel production enterprises choose the scrap steel reverse logistics operation mode according to the evolutionary stable strategy.

2 Evolutionary Game Model

2.1 ASSUMPTION

Evolutionary game theory originates from biological evolutionary theory. The theory based on bounded rationality hypothesis is used to analyze strategy adjustment processes, trend and local stability of limited rational players in the process of long-term repeated behavior, study subjects of the theory are groups rather than a single individual behavior[8]. In this paper, the two players are randomly assigned – one from steel enterprises group and one from the third party logistics enterprise groups respectively. The two players are random when playing the game. The game strategies of both players are cooperation and non-cooperation. The cooperation is that enterprises will contract part or all business of scrap reverse logistics out to the third party. The non-cooperation is the enterprises will not reach an agreement with the third party logistics enterprises; the enterprise itself will handle all of the scrap recycling business. Based on reference literature [9-10], one assumes that the benefits and costs of steel enterprises and third-party logistics enterprises under different strategies are as following:

D_m,D_L denote the obtained normal returns of independent operation when steel enterprises and third-party logistics enterprises choose "non-cooperation",

B_m,B_L are generated excess returns when two players cooperate,

F_m,F_L indicate risk costs of two player’s cooperation,

C_m,C_L signify the initial costs which two players invest for co-operation,

V_m is the suffered loss when steel enterprises select cooperation, while the third party logistics enterprises choose non- cooperation,

V_L is the suffered loss when the third party logistics enterprises select co-operation, while steel enterprises choose non- cooperation,

Where $B_m > F_m + C_m$, $B_L > F_L + C_L$

According to the above parameters, the payoff matrix of steel enterprises and third-party logistics enterprises is shown in Table 1.

TABLE 1 Payoff matrix of steel enterprises and third-party logistics enterprises

	The third party logistics enterprises	
	Cooperation(x)	Non- cooperation(1-x)
Cooperation(y)	$D_m + B_m - C_m - F_m$, $D_L + B_L - C_L - F_L$	$D_m - C_m - V_m$, D_L
Non-cooperation(1-y)	D_m, $D_L - C_L - V_m$	D_m, D_L

2.2 THE ESTABLISHMENT OF THE REPLICATED DYNAMIC EQUATION

The expected returns of steel enterprises group selecting cooperation strategy and non-co-operation strategy are as follows:

$$u_{s1} = x(D_m + B_m - C_m - F_m) + (1-x)(D_m - C_m - V_m), \quad (1)$$

$$u_{s2} = xD_m + (1-x)D_m. \quad (2)$$

The average payoff of steel enterprises group is as follows:

$$u_{ss} = yu_{s1} + (1-y)u_{s2}. \quad (3)$$

Similarly, the expected returns of the third party logistics group selecting cooperation strategy and non-cooperation strategy are as follows:

$$u_{x1} = y(D_L + B_L - C_L - F_L) + (1-y)(D_L - C_L - V_L), \quad (4)$$

$$u_{x2} = yD_L + (1-y)D_L. \quad (5)$$

The average payoff of the third party logistics group is as follows:

$$u_x = xu_{x1} + (1-x)u_{x2}. \quad (6)$$

According to evolutionary game theory, if the fitness or payoff of one strategy is higher than the average fitness of the group, the group will trend to adopt this strategy. Namely, the growth rate of this strategy is greater than zero, which can be expressed by the following differential equation [11]:

$$\frac{dx}{dt} = [u(k,s) - u(s,s)]x_k, k = 1,2,\ldots,K \quad . \quad (7)$$

Equation (7) is replicated dynamic equation, where x_k is the proportion of using strategy k among group, $u(k,s)$ is payoff of using strategy k, $u(s,s)$ is the average payoff of the group,k denotes a different strategy, K indicates strategy number of group, s is all strategy set of group.

The replicated dynamic equation of steel enterprises selecting the "cooperation" strategy is as follows:

$$F(y)=\frac{dy}{dt}=y(u_{s1}-u_{ss})=y(1-y)(xB_m+xV_m-xF_m-C_m-V_m)\ . \qquad (8)$$

Let $\frac{dy}{dt}=0$, the stabilized state points of replicated dynamic equation ,which mean the proportion y of steel enterprises selecting co-operation strategy in the dynamic process keep relatively stable equilibrium, are obtained:

$$y=0, y=1,\ \ x^*=\frac{C_m+V_m}{B_m+V_m-F_m}\ .$$

In the same way, the replicated dynamic equation of the third party logistics enterprises selecting the "cooperation" strategy is as follows:

$$F(x)=\frac{dx}{dt}=x(u_{x1}-u_{xx})=x(1-x)(yB_L+yV_L-yF_L-C_L-V_L)\ . \qquad (9)$$

Let $\frac{dy}{dt}=0$, the stabilized state points are obtained:

$$x=0, x=1,\ \ y^*=\frac{C_L+V_L}{B_L+V_L-F_L}$$

3 Stability Analysis Of Model

The above obtained stabilized state points are not absolute stable equilibrium. According to the stability theorem of differential equation and the nature of evolutionary stable strategy[12], the calculated state points are evolutionary stable strategy(ESS) when $\frac{dF(x)}{dx}<0$ and $\frac{dF(y)}{dy}<0$.

For the iron and steel enterprises,

$$F'(y)=(1-2y)(xB_m+xV_m-xF_m-C_m-V_m)\ . \qquad (10)$$

If $x=\frac{C_m+V_m}{B_m+V_m-F_m}$, then $F(y)\equiv 0$. This means all state points y are stable.

If $x\neq\frac{C_m+V_m}{B_m+V_m-F_m}$, $y=0$ and $y=1$ are stable state points, the two cases in this condition are as follows:

When $x>\frac{C_m+V_m}{B_m+V_m-F_m}$, $\frac{dF(y)}{dy}\Big|_{y=1}<0$, $\frac{dF(y)}{dy}\Big|_{y=0}>0$, $y=1$ is ESS;

When $x<\frac{C_m+V_m}{B_m+V_m-F_m}$, $\frac{dF(y)}{dy}\Big|_{y=0}<0$, $\frac{dF(y)}{dy}\Big|_{y=1}>0$, $y=0$ is ESS;

The conclusion can be drawn from above process analysis: when the probability of the third party logistics enterprises group choose co-operation is bigger than $\frac{C_m+V_m}{B_m+V_m-F_m}$, the final strategy of steel enterprises group is co-operation , that is to say, steel enterprises select the third party logistics mode; when the probability is smaller than $\frac{C_m+V_m}{B_m+V_m-F_m}$, the final strategy is non cooperation ,that is, steel enterprises select self-operation mode.

For the third party logistics enterprises,

$$F'(x)=(1-2x)(yB_L+yV_L-yF_L-C_L-V_L)\ . \qquad (11)$$

If $y=\frac{C_L+V_L}{B_L+V_L-F_L}$, then $F(x)\equiv 0$. This means all state points x are stable state points.

If $y\neq\frac{C_L+V_L}{B_L+V_L-F_L}$, $x=0$ and $x=1$ are stable state points, the two cases are analyzed:

when $y>\frac{C_L+V_L}{B_L+V_L-F_L}$, $\frac{dF(x)}{dx}\Big|_{x=1}<0$, $\frac{dF(x)}{dx}\Big|_{x=0}>0$ $x=1$ is ESS, that is to say, when the probability of steel enterprises groups choosing co-operation is bigger than $\frac{C_L+V_L}{B_L+V_L-F_L}$,the final strategy of the third party logistics enterprises group is co-operation;

when $y<\frac{C_L+V_L}{B_L+V_L-F_L}$, $\frac{dF(x)}{dx}\Big|_{x=1}>0$, $\frac{dF(x)}{dx}\Big|_{x=0}<0$, $x=0$ is ESS, that is, the final strategy of steel enterprises group strategy is non co-operation.

The replication dynamic equations of two players are synthetically analyzed. Their respective stable points are placed in coordinate plane $\{(x,y):0\le x,y\le 1\}$ .There are five local equilibrium points in the plane:

$(x,y)=\{(0,0),(0,1),(1,0),(1,1),(x^*,y^*)\}$, where

$$x^*=\frac{C_m+V_m}{B_m+V_m-F_m} \qquad y^*=\frac{C_L+V_L}{B_L+V_L-F_L}$$

The stability of the equilibrium is obtained by local stability analysis on Jacobean matrix of the system [12-13]. According to method of Friedman proposed, the Jacobean matrix, the corresponding determinant and trace respectively are as follows:

$$J=\begin{bmatrix}\frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y}\\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y}\end{bmatrix}=\begin{bmatrix}(1-2x)(yB_L+yV_L-yF_L-C_L-V_L) & x(1-x)(B_L+V_L-F_L)\\ y(1-y)(B_m+V_m-F_m) & (1-2y)(xB_m+xV_m-xF_m-C_m-V_m)\end{bmatrix}, \tag{12}$$

$$De(J)=\begin{vmatrix}\frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y}\\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y}\end{vmatrix}, \tag{13}$$

$$Tr(J)=\frac{\partial F(x)}{\partial x}+\frac{\partial F(y)}{\partial y}. \tag{14}$$

The five local equilibrium points are analyzed by local stability analysis method. The results are shown in Table 2.

It is shown from Table 2 that there are two evolutionary stable strategies, two unstable equilibrium points and a saddle point among the five local equilibrium points. Figure 1 indicates dynamic process of evolutionary game between steel enterprises and the third party logistics enterprises. The broken line-ADB which two unstable equilibrium points A(0,1),B(1,0) and saddle point D(x^*,y^*) link is critical line of system converging to different states. Because o(0,0) and C(1,1) are ESS, all members either choose win-win cooperation strategy or choose independent operation(non cooperation) when the game between steel enterprises and the third party enterprises reach a stable state. Which state the game dynamic process of two players evolve to depends on the size of the divided area ADOB and ADBC by the broken line-ADB. S_{ADBC} denotes the area of zone-ADBC ,S_{ADOB} is the area of zone-ADOB.If $S_{ADBC}>S_{ADOB}$, the system will evolve to the direction of co-operation. If $S_{ADBC}<S_{ADOB}$, the system will evolve to the direction of non-cooperation. If $S_{ADBC}=S_{ADOB}$, the direction of system evolution is uncertain.

4 Analyses of Factors Affecting Evolutionary Stability Strategy

It is known from Figure 1. that the area of zone-ADOB is determined by coordinates of saddle point $D(x^*,y^*)$, so extra incomes B_m,B_L and risk costs F_m,$F_L$of co-operation, invested initial costs C_m,C_L and losses V_m,V_L for co-operation affect the evolution stable results between steel enterprises and the third party logistics enterprises. Besides, the initial state of system also has certain influence.

TABLE 2 The stability analysis result

Equilibrium Point	De(J)	symbol	Tr(J)	symbol	result
O(0,0)	$(C_L+V_L)(C_m+V_m)$	+	$-C_L-V_L-C_m-V_m$	-	ESS
A(0,1)	$(B_L-F_L-C_L)(C_m+V_m)$	+	$B_L-F_L-C_L+C_m+V_m$	+	instability
B(1,0)	$(B_m-F_m-C_m)(C_L+V_L)$	+	$B_m-F_m-C_m+C_L+V_L$	+	instability
C(1,1)	$(B_L-F_L-C_L)(B_m-F_m-C_m)$	+	$-B_L+F_L+C_L-B_m+F_m+C_m$	-	ESS
D(x*,y*)	$-\frac{(C_m+V_m)(B_m-F_m-C_m)(C_L+V_L)(B_L-F_L-C_L)}{(B_m+V_m-F_m)(B_L+V_L-F_m)}$	-	0		saddle point

4.1 INFLUENCE OF THE INITIAL STATE OF SYSTEM ON ESS

When the initial state of steel enterprises group and the third logistics enterprises group is in III-zone, that is, the probability of steel enterprises selecting cooperation is bigger than $\frac{C_L+V_L}{B_L+V_L-F_L}$ and the probability of the third logistics enterprises selecting co-operation is bigger than $\frac{C_m+V_m}{B_m+V_m-F_m}$,the game of system will finally converge to C(1,1), namely, both steel enterprises group and the third logistics enterprises group choose cooperation strategy. When the initial state is in I-zone, the system will finally converge to point o(0,0),that is, both players choose non co-operation, thus,steel enterprises only choose self-operation mode. When the initial state is in II and IV zone, the convergence of system is uncertain, it may evolve to C(1,1) or to o(0,0). If the initial state falls into II areas, the speed of steel enterprises convergence to

y=1 is greater than the speed of the third party logistics convergence to *x=0*, the final stability strategy is C(1,1).If the initial state falls into IV area, the speed of steel enterprises convergence to *y=0* is greater than the speed of the third party logistics convergence to *x=1*,the final stability strategy is o(0,0).

4.2 INFLUENCE OF EXTRA INCOMES ON ESS

It can be seen from the coordinate D(x*,y*) that values of x^* and y^* reduce respectively, the saddle point D moves to lower left direction ,the area of zone ADBC increases , on the contrary ,the area of zone ADBO reduces when extra incomes B_m,B_L of co-operation increase under the premise of other parameters constant, so system will evolve to co-operation direction along DC path. In order to intuitively display the influence of the changed parameter on the evolution strategy, numerical simulation calculation is used in the paper. Suppose that $C_m=C_L=3$, $V_m=V_L=2$,$F_m=F_L=1$ for convenience of calculation .

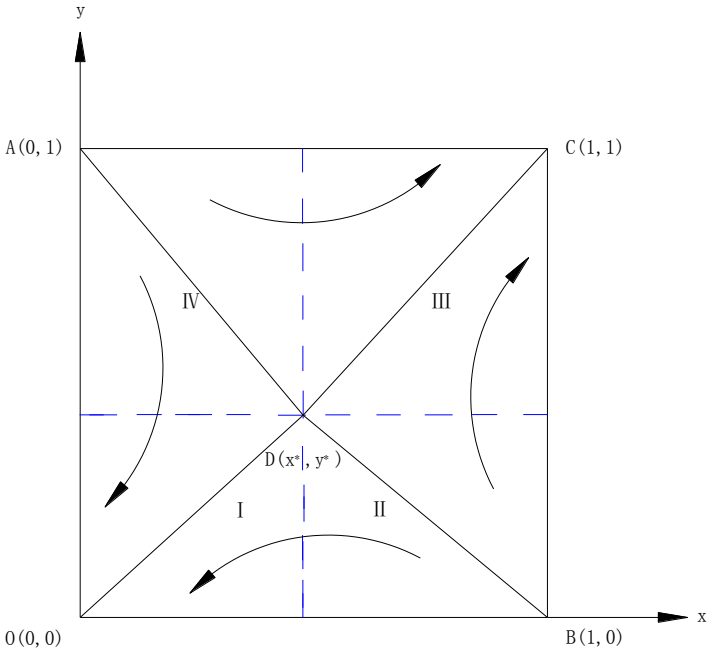


FIGURE 1 Dynamic process figure of evolutionary game between steel enterprises and the third party logistics enterprise

The calculation Process is as shown in Table 3 when the values of B_m and B_L are changed. From Figure 2, the changed path of saddle point D can be seen. The evolution path of saddle point D is from D1 to D2,D3,D4,D5, when the values of B_m and B_L are bigger and bigger.The probability of system converging to c(1,1) is larger and larger with the growing S_{ADBC} in the evolution process, so steel manufacturing enterprises and third-party logistics enterprises will be more and more inclined to choose cooperation strategy. From the above analysis , if the obtained extra incomes of the two players from cooperation become higher and higher , both sides are willing to cooperate actively.

TABLE 3 The coordinates of saddle point under different extra income

B_m=B_L	x*	y*
5	0.83	0.83
14	0.33	0.33
23	0.21	0.21
35	0.14	0.14
43	0.11	0.11

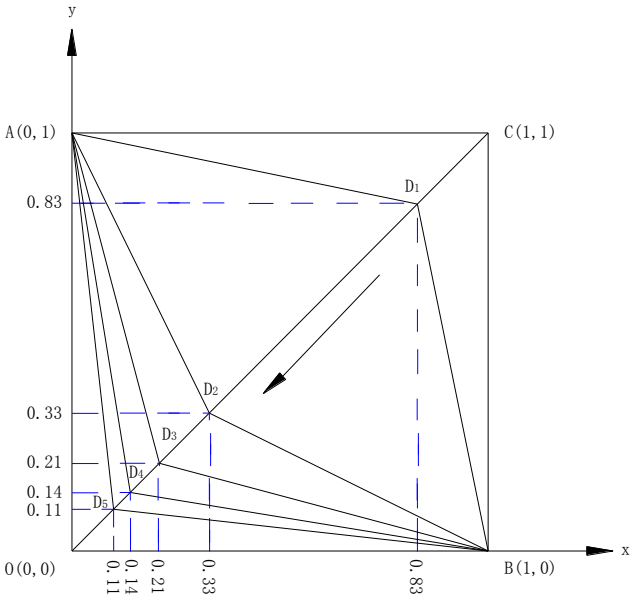


FIGURE 2 The dynamic evolution path of system in the case of extra incomes changed

4.3 INFLUENCE OF RISK COSTS ON ESS

It can be seen that values of $x^*$and y^* increase respectively, the saddle point D moves to upper right direction ,the area of zone ADBO increase , on the contrary ,the area of zone ADBC reduce when risk costs F_m,F_L of cooperation increase under the premise of other parameters constant, so system will evolve to cooperation direction along DC path. Suppose that $C_m=C_L=3$, $V_m=V_L=2$, $B_m=B_L=15$. The calculation process is as shown in Table 4 when the value of F_m and F_L is changed. From Figure.3, the changed path of saddle point D can be seen. The evolution path of saddle point D is from D1 to D2,D3,D4 when the values of F_m and F_L are bigger and bigger. The probability of system converging to o(0,0) is higher and higher with the growing S_{ADBO} in the evolution process, so steel manufacturing enterprises and third-party logistics enterprises will be more and more inclined to choose non cooperation strategy. According to the above analysis, if risk costs of the two players from co-operation are higher, both sides are not willing to co-operate.

TABLE 4 The coordinates of saddle point under different income

$F_m=F_L$	x^*	y^*
1	0.31	0.31
5	0.42	0.42
7	0.50	0.50
9	0.63	0.63

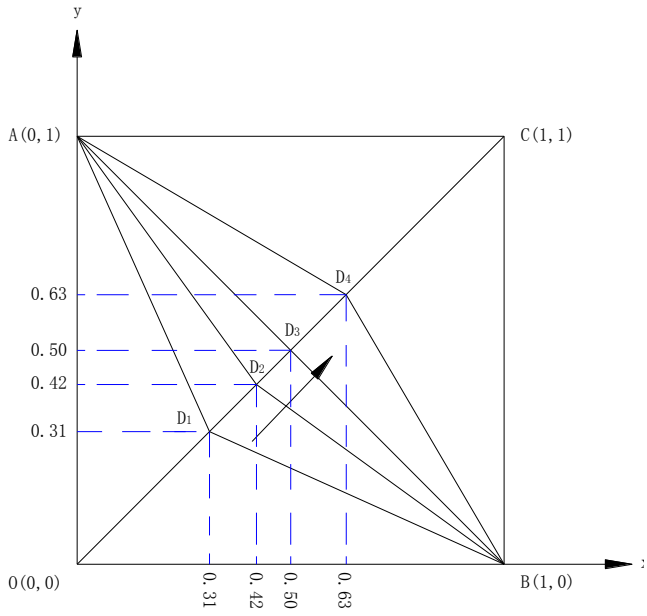


FIGURE 3 The dynamic evolution path of system in the case of risk cost changed

4.4 INFLUENCE OF THE INVESTED INITIAL COSTS ON ESS

Similarly, the values of $x^*$and y^* increase respectively, the saddle point D moves to upper right direction and approaches to C(1,1) when the risk costs C_m,C_L of co-operation increase , so the probability of system converging to o(0,0) is larger and larger, that is, the chances of steel enterprises and third-party logistics enterprises choosing non-co-operation are increasing. According to the above analysis, if the invested risk costs of the two players for cooperation are higher; both sides are not willing to co-operate.

4.5 INFLUENCE OF THE SUFFERED LOSSES ON ESS

The abscissa and ordinate of saddle point D reduce respecttively when the losses V_m,V_L of cooperation are increasing under the premise of other parameters constant. If losses V_m and V_L reduce at the same rate, the abscissa and ordinate of saddle point D keep away from c(1,1)at the same speed and approach to o(0,0) eventually. If the speeds of loss costs V_m andV_L reducing are different ,the magnitudes of saddle point D declining in horizontal and vertical direction are inequality , but it approaches to o(0,0) eventually and the region-ADBC of implement co-operation for the two players are gradually enlarged. According to the above analysis, if the suffered loss of one player choosing cooperation when the other player not cooperate is higher, they are not willing to co-operate, on the contrary, they will choose co-operation.

5 Conclusions

In the paper, the problem of steel enterprise choosing the self-operation mode or the third party mode in scrap iron reverse logistics is regarded as a game problem between steel enterprises and the third party logistics enterprise. Based on evolutionary game theory, co-operation or not between steel production enterprises and the third party logistics enterprises is analyzed. By the evolution game analysis of two players, the two evolutionary stable strategies of both co-operation and non-co-operation are obtained. Finally, the extra income and risk cost from co-operation ,invested initial cost and suffered loss for co-operation which have influence on the evolution stable result are analyzed. If the extra incomes from co-operation are higher, the risk costs from co-operation, the initial costs and losses for cooperation are lower, the two players will eventually choose co-operation development, that is, steel enterprises choose the third party reverse logistics operation mode, on the contrary, they choose the self-management mode.

Acknowledgements

The paper is supported by Humanities and Social Science Foundation of Anhui Province (No.Z12018) and Young Teacher Scientific Research Foundation of Anhui University of Technology (No.QZ201117).

References

[1] Qiping Yan ,Yinxue Li ,Low Carbon 2011 Economy and scrap utilization S*teel Research* **39**(2)1-2.

[2] Xiuyu He Study on Reverse Logistics Operation Model for Steel Enterprise Based on the Concept of Circul Economy. *HuNan XiangTan University*

[3] Ravi V, Shankar R,Tiwari M K 2005 Analyzing alternatives in reverse logisties for end-of-life computers ANP and balanced score-card approach *Computers and Industrial Engineering* **48** 327-56

[4] Mingming Ren, Haolin Tong 2009 The selection of ERP recovery mode based on fuzzy comprehensive evaluation 2009 *Statistics and Decision* **14** 42-44

[5] Zhikui Chen, Chaokai Xue, Fuxing Duan 2013 Reseach on operation models of reverse logistics based on analytic network process 2013 *Journal of Shangdong University of Technology (Natural Science Edition*) **27** 55-61

[6] Guangping Chen ,Wei Liu 2007Comprehensive evaluation research of reverse logistics model for enterprise *Statistics and Decision* **12** 150-2

[7] Mingang Zeng, Yanting Zhou 2009 Study on the Reverse Logistics in Manufacturing Industry Based on Multi-game Model *Industrial Engineering Journal* **12**(6)33-37

[8] Maynard S J. 1974 The theory of games and the evolution of animal conflict *Journal of Theoretical Biology* **47** 209-23

[9] Zhenzhen Wang, Gongyu Chen 2012 Evolutionary game analysis of Joint development between manufacturing and logistics industry E*conomic Issues in China* **2** 86-96

[10] Bin Li, Jing Chen 2011 The evolution game of collaborative relationship among the third party logistics enterprises *Statistics and Decision* **3** 183-5

[11] Taylor P D, Jonker L B 1978 Evolutionarily stable strategy and game dynamics *Math Biosci* **40** 145-56

[12] Herbert Gintis 2000 Game theory evolving *Princeton University*

[13] Shiyu Xie 2001 Economic game theory *ShangHai FuDan University* 233-72

Authors

Yan LI, born on August 15, 1984, Anhui China

Current position, grades: Master of Mechanical and Electrical Engineering,lecturer in Anhui University of Technology
University studies: Industrial Engineering in Anhui University of Technology
Scientific interest: computer simulation ,reverse logistics and intelligent algorithms
Publications: 5 papers
Experience: Master degree on machinery and electronics engineering from Anhui University of Technology,2009. Lecturer at Ahui University of Technology, Ma anshan, China. Mainly focuses on computer simulation , reverse logistics and intelligent algorithms

Fu-yu Wang, born on September 26, 1977, Miyang, China

Current position: Master of Mechanical Engineering, associate professor and master supervisor in Anhui University of Technology
University studies: Bachelor's degree in 1999 and a Master's degree in 2005 Mechanical and Electronic Engineering in China University Of Petroleum, Shandong, China.
Scientific interest: Industrial Engineering and Intelligent Optimization
Publications: 2 Patents, 15 Papers
Experience: His research interests include Industrial Engineering and Intelligent Optimization

A Scientific and Research Performance Evaluation model of Institutes of Higher Learning Based on Multilevel Fuzzy Comprehensive Decision Analysis

Biao Song*

Nanjing University of Information Science & Technology

Abstract

Scientific and research performance is an important part of the evaluation on the overall strength and the ability to achieve sustainable development of institutes of higher learning. With the purpose to deal with the multi-attribute, multilevel and fuzziness, this paper constructs a scientific and research performance evaluation system for institutes of higher learning and proposes an evaluation model based on multilevel fuzzy comprehensive decision analysis. After indicators are standardized, we can get the fuzzy nearness by fuzzy comprehensive decision analysis and Analytical Hierarchy Process for evaluating various schemes. It provides support for the analysis of teaching ability, research ability and the sustainable development of institutes of higher learning.

Keywords: Higher school; scientific and research; fuzzy decision analysis; performance evaluation; model

1 Introduction

Teaching ability and research ability are two integral part of the sustainable development of institutes of higher learning. High teaching ability has much to do with training hi-tech talents, establishing talent pool and sustainable talent seeking, which are the fundamental purposes of institutes of higher learning. High research ability reflects the ability to do scientific research, scientific exploration and technological application. Therefore, it is significant to do scientific and research performance evaluation on institutes of higher learning [1-3]. Given that it is a complicated process, many factors need to be taken into consideration. Many researchers have studied scientific and research performance evaluation and proposed relevant methods with fruitful results [4-8]. However, current evaluation methods are more of analysing sections in the process of implementation and lack the wholeness. Fuzzy information also brings about some limits. Therefore, based on previous researches, this paper intends to study scientific and research performance evaluation for institutes of higher learning by comprehensive fuzzy decision analysis [9-10] and AHP [11-12].

2 Scientific and Research Performance Evaluation System for Institutes of Higher Learning

It is a complicated process to construct scientific and research performance evaluation system. The selection of indicators should comply with the scientific principle, criticality, completeness and objectivity. The weight and data should be reasonable and accurate. The evaluation system is divided into three layers, namely, target layer, criterion layer and indicator layer, as is shown in Table 1.

TABLE 1 Scientific and scientific and research performance evaluation system for institutes of higher learning

Target layer	Criterion layer	Target layer
institutes of higher learning scientific and research performance evaluation system **U**	Teaching ability u_1	Ratio of full-time teachers to students u_{11}
		Proportion of professional teachers u_{12}
		Number of competitive classes u_{13}
		Professional knowledge degree of teaching content u_{14}
		Advanced level of teaching method u_{15}
		Software and hardware u_{16}
	Research ability u_2	The number of senior number reserve u_{21}
		The number of papers issued u_{22}
		The number of patent u_{23}
		Key laboratory above provincial level u_{24}
		The number of research project above provincial level u_{25}
		The number of award above provincial level u_{26}
	Talent training ability u_3	Graduation rate u_{31}
		Unqualified rate u_{32}
		The number of award of students above provincial level u_{33}
		Student's innovation ability u_{34}
		Student's learning ability u_{35}
	Comprehensive service u_4	Turning rate of scientific research u_{41}
		Social satisfaction on student u_{42}
		Scientific service ability u_{43}
		The number of identification of scientific and technological achievements u_{44}
	Potential for development u_5	Teaching input u_{51}
		Scientific input u_{52}
		Overall management quality u_{53}
		Social awareness u_{54}

* ***Corresponding author*** e-mail: biaos@126.com

3 Scientific and Research Performance Evaluation System for Institutes of Higher Learning Based on Multilevel Fuzzy Comprehensive Decision Analysis

3.1 INDICATOR SET AND SCHEME SET OF MULTILEVEL SCIENTIFIC AND RESEARCH PERFORMANCE EVALUATION

The evaluation system is the set of all indicators. It has hierarchy. The criterion layer is the first-class layer of indicators:

$$U=\left(u_1,u_2,u_3,u_4,u_5\right). \tag{1}$$

Under each criterion layer, there is the second-class indicator set,

$$u_{ij}=\left(u_{i1},u_{i2},\cdots,u_{is}\right), \tag{2}$$

i is the NO. of the criterion layer u_{ij}. s is the number of indicators under the criterion layer i and there are $1\le i\le 5$, $1\le s\le 5$.

Suppose there are m institutes of higher learning for evaluation, the scheme set is:

$$C=\left(C_1,\cdots,C_k,\cdots,C_m\right) \tag{3}$$

Each evaluation scheme C_k contains the indicator set of two layers.

3.2 WEIGHT ALLOCATION OF INDICATORS IN SCIENTIFIC AND RESEARCH PERFORMANCE EVALUATION

AHP can ensure the weight is allocated objectively, reliably and reasonably. This paper uses 1-9 ratio scale to measure the weighed allocation matrix $\boldsymbol{P}$:

$$\boldsymbol{P}=\begin{vmatrix} p_{11} & \cdots & p_{1i} & \cdots & p_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ p_{i1} & \cdots & p_{ii} & \cdots & p_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ p_{n1} & \cdots & p_{ni} & \cdots & p_{nn} \end{vmatrix}_{n\times n}. \tag{4}$$

In the expression, p_{ij} is the ratio scale value of indicators. There are

$$1\le p_{ij}=\frac{1}{p_{ji}}\le 9,\qquad 1\le p_{ij}=\frac{1}{p_{ji}}\le 9.$$

The weighed consistency indicator R_{CI} and the consistency rate R_{CR} are:

$$\begin{cases} R_{CI}=\dfrac{max\left(\lambda_t\left(\boldsymbol{P}\right)\right)-n}{n-1} \\ R_{CR}=\dfrac{R_{CI}}{R_{RI}} \end{cases}. \tag{5}$$

In the expression, $\lambda_t\left(\boldsymbol{P}\right)$ refers to the characteristic value of indicator weight-allocation matrix $\boldsymbol{P}$. R_{RI} refers to random consistency indicator.

Under consistency requirement, the weight w_i of indicator i is:

$$w_i=\frac{\sum_{j=1}^{n}p_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}}. \tag{6}$$

3.3 STANDARDIZATION OF SCIENTIFIC AND RESEARCH PERFORMANCE EVALUATION

It is clear that there are two types of indicators, quantitative and qualitative. They have different scales. Some are large and some are small. Therefore, they need to be subject to standardization to get unified indicators.

For quantitative indicators, we use fuzzy membership degree to describe them, as in Table 2.

TABLE 2 Value of indicators based on membership degree

Fuzzy membership degree	Explanation	
	Large indicators	*Small indicators*
1.0	Completely compliance	Not compliance
0.8	Fairly compliance	Poorly compliance
0.6	Compliance	Basically compliance
0.4	Basically compliance	Compliance
0.2	Poorly compliance	Fairly compliance
0	Not compliance	Completely compliance
0.9,0.7,0.5,0.3, 0.1	In between	In between

For quantitative indicators, suppose the i-th scheme about indicator j is $u_{ij}=\left[u_{ij}^L,u_{ij}^R\right],u_{ij}^L\le u_{ij}^R$ and if it is a large indictor, the standardized indicator v_{ij} is:

$$v_{ij}=\left[v_{ij}^L,v_{ij}^R\right]=\left[\frac{u_{ij}^L-\min\limits_{1\le i\le m}\left(u_{ij}^L\right)}{\left\|\Delta_{u_{ij}}\right\|_{max}},\frac{u_{ij}^R-\min\limits_{1\le i\le m}\left(u_{ij}^L\right)}{\left\|\Delta_{u_{ij}}\right\|_{max}}\right], \tag{7}$$

$\left\|\Delta_{u_{ij}}\right\|_{max}$ is the maximum norm in the fuzzy interval, there is: $\left\|\Delta_{u_{ij}}\right\|_{max}=\max\limits_{1\le i\le m}\left(u_{ij}^R\right)-\min\limits_{1\le i\le m}\left(u_{ij}^L\right)$. (8)

If it is a small indictor, the standardized indicator v_{ij} is:

$$v_{ij}=\left[v_{ij}^{L},v_{ij}^{R}\right]=\left[\frac{\max\left(u_{ij}^{R}\right)-u_{ij}^{L}}{\left\|\Delta_{u_{ij}}\right\|_{max}},\frac{\max\left(u_{ij}^{R}\right)-u_{ij}^{R}}{\left\|\Delta_{u_{ij}}\right\|_{max}}\right]. \quad (9)$$

3.4 FUZZY CLEARNESS OF INDICATORS IN SCIENTIFIC AND RESEARCH PERFORMANCE EVALUATION

After standardization, indicators are unified. Suppose there are *m* institutes of higher learning for scientific and research performance evaluation. The i-th scheme about indicator *j* is $v_{ij}=\left[v_{ij}^{L},v_{ij}^{R}\right]$, and fits $0\le v_{ij}^{L}\le v_{ij}^{R}\le 1$, then the maximum fuzzy indicator field $v_{\Delta j}$ of indicator *j* is:

$$v_{\Delta j}=\left[v_{\Delta j}^{L},v_{\Delta j}^{R}\right]=\left[\underset{1\le i\le m}{max}\left(v_{ij}^{L}\right),\underset{1\le i\le m}{max}\left(v_{ij}^{R}\right)\right]. \quad (10)$$

Similarly, the minimum fuzzy indicator field $v_{\Delta j}$ of indicator *j* is:

$$v_{\nabla j}=\left[v_{\nabla j}^{L},v_{\nabla j}^{R}\right]=\left[\underset{1\le i\le m}{min}\left(v_{ij}^{L}\right),\underset{1\le i\le m}{min}\left(v_{ij}^{R}\right)\right]. \quad (11)$$

The fuzzy clearness τ_{ij}^{Δ} between the i-th scheme and the j-th maximum fuzzy indicator field $v_{\Delta j}$ about indicator j is:

$$\tau_{ij}^{\Delta}=\sqrt[p]{\frac{\left|\underset{1\le i\le m}{max}\left(v_{ij}^{L}\right)-v_{ij}^{L}\right|^{T}+\left|\underset{1\le i\le m}{max}\left(v_{ij}^{R}\right)-v_{ij}^{R}\right|^{T}}{2}}. \quad (12)$$

The weighed fuzzy clearness ρ_{i}^{Δ} between the i-th scheme and the j-th maximum fuzzy indicator field $v_{\Delta j}$ about indicator j is:

$$\rho_{i}^{\Delta}=\sum_{j=1}^{n}\left(w_{j}*\tau_{ij}^{\Delta}\right). \quad (13)$$

Similarly, the fuzzy clearness τ_{ij}^{Δ} between the i-th scheme and the j-th minimum fuzzy indicator field $v_{\Delta j}$ about indicator j is:

$$\tau_{ij}^{\nabla}=\sqrt[p]{\frac{\left|\underset{1\le i\le m}{min}\left(v_{ij}^{R}\right)-v_{ij}^{L}\right|^{T}+\left|\underset{1\le i\le m}{min}\left(v_{ij}^{L}\right)-v_{ij}^{R}\right|^{T}}{2}}. \quad (14)$$

The weighed fuzzy clearness τ_{ij}^{∇} between the i-th scheme and the j-th maximum fuzzy indicator field $v_{\Delta j}$ about indicator *j* is:

$$\rho_{i}^{\nabla}=\sum_{j=1}^{n}\left(w_{j}*\tau_{ij}^{\nabla}\right). \quad (15)$$

The comprehensive fuzzy nearness for the i-th scheme is expressed as:

$$\rho_{i}=1/\left(1+\left(\left|\frac{\rho_{i}^{\Delta}}{\rho_{i}^{\nabla}}\right|\right)^{2}\right). \quad (16)$$

3.5 MULTILEVEL FUZZY COMPREHENSIVE EVALUATION

Based on the optimization principle of fuzzy clearness, if there is:

$$\rho_{max}=max\left(\rho_{1},\cdots,\rho_{i},\cdots,\rho_{m}\right)=\rho_{k}, \quad (17)$$

where *k* is the optimal scheme in scientific and research performance evaluation, indicating that k-th scheme has the best teaching ability and research ability among institutes of higher learning.

The specific steps of evaluation are as follows:

Step 1: Select evaluation indicators according to relevant principles and construct the evaluation system.

Step 2: Standardize different qualitative indicators as Table 2, and subject quantitative indicators to expression (7) to (9);

Step 3: Select the first-class indicator set and the second-class indicator set based on (1) to (3) and acquire the scheme set of scientific and research performance evaluation of institutes of higher learning;

Step 4: Allocate weight to indicators according to expression (4) to (6);

Step 5: Standardize indicators according to (10) to (11);

Step 6: Acquire the evaluation scheme and the fuzzy nearness of the minimum fuzzy indicator field about different indicators;

Step 7: Acquire weighed fuzzy nearness according to (13) to (15) and get the comprehensive fuzzy nearness through (16);

Step 8: Get the optimal scheme based on comprehensive fuzzy nearness.

4 Case Studies and Test

This paper takes scientific and research performance evaluation of key institutes of higher learning in a province as an example. Based on the indicator system, we can get the value of a quantity of indicators, as shown in Table 3.

TABLE 3 Value of a quantity of indicators of scientific and research performance evaluation

Criterion layer	Weight	Indicators	weight	Value of a quantity of indicators		
				Institute A	Institute B	Institute C
Teaching ability u_1	0.286	Ratio of full-time teachers to students u_{11}	0.25	12	12	16
		Proportion of professional teachers u_{12}	0.12	0.42	0.51	0.46
		Number of competitive classes u_{13}	0.15	53	62	46
		Professional knowledge degree of teaching content u_{14}	0.10	0.80-0.85	0.80-0.85	0.85-0.90
		Advanced level of teaching method u_{15}	0.18	0.80-0.85	0.85-0.90	0.80-0.85
		Software and hardware u_{16}	0.20	0.90-0.95	0.85-0.90	0.85-0.90
Research ability u_2	0.247	The number of senior number reserve u_{21}	0.16	135	182	167
		The number of papers issued u_{22}	0.13	2028	1986	2476
		The number of patent u_{23}	0.10	362	419	275
		Key laboratory above provincial level u_{24}	0.23	18	26	18
		The number of research project above provincial level u_{25}	0.15	235	271	184
		The number of award above provincial level u_{26}	0.23	13	10	8
Talent training ability u_3	0.163	Graduation rate u_{31}	0.20	0.925	0.937	0.956
		Unqualified rate u_{32}	0.23	0.03	0.05	0.02
		The number of award of students above provincial level u_{33}	0.25	22	19	15
		Student's innovation ability u_{34}	0.16	0.80-0.85	0.85-0.90	0.80-0.85
		Student's learning ability u_{35}	0.16	0.85-0.90	0.80-0.85	0.80-0.85
Comprehensive service u_4	0.208	Turning rate of scientific research u_{41}	0.30	0.26	0.38	0.32
		Social satisfaction on student u_{42}	0.30	0.85-0.90	0.85-0.90	0.80-0.85
		Scientific service ability u_{43}	0.20	0.80-0.85	0.85-0.90	0.85-0.90
		The number of identification of scientific and technological achievements u_{44}	0.20	127	201	162
Potential for development u_5	0.096	Teaching input u_{51}	0.30	0.80-0.85	0.85-0.90	0.85-0.90
		Scientific input u_{52}	0.30	0.85-0.90	0.85-0.90	0.80-0.85
		Overall management quality u_{53}	0.20	0.85-0.90	0.85-0.90	0.80-0.85
		Social awareness u_{54}	0.20	0.75-0.80	0.75-0.80	0.80-0.85

Given different weight of indicators, we can get the weighed fuzzy nearness ρ_i^{Δ} and ρ_i^{∇}, as is shown in Table 4.

TABLE 4 Fuzzy nearness of indicators in scientific and research performance evaluation of institutes of higher learning

Criterion layer	Institute A		Institute B		Institute C	
	ρ_i^{Δ}	ρ_i^{Δ}	ρ_i^{Δ}	ρ_i^{Δ}	ρ_i^{Δ}	ρ_i^{Δ}
Teaching ability u_1	0.019	0.026	0.007	0.038	0.041	0.004
Research ability u_2	0.042	0.035	0.003	0.057	0.063	0.013
Talent training ability u_3	0.034	0.015	0.009	0.009	0.017	0.002
Comprehensive service u_4	0.039	0.004	0	0.041	0.023	0.019
Potential for development u_5	0.004	0.003	0.002	0.004	0.004	0.004

Through comprehensive fuzzy nearness model, we can get the comprehensive fuzzy nearness sequence of three institutes of higher learning, namely,

$$\rho = (\rho_A, \rho_B, \rho_C) = (0.266, 0.833, 0.077) \cdot$$

Institute B has the optimal scientific and research level followed by A and then C. Therefore, Institute B is the priority of development.

4 Conclusions

This paper constructs a scientific and research performance evaluation system for institutes of higher learning and proposes an evaluation model based on multilevel fuzzy comprehensive decision analysis. After indicators are standardized, we can get the fuzzy nearness by fuzzy comprehensive decision analysis and Analytical Hierarchy Process for evaluating various schemes to find out the optimal scheme. This model is clear and easy to calculate for computer aid design and intelligence design. Case study has proved that the model and the algorithm are effective. The evaluation method proposed by this paper provides a solution to analyze scientific and research performance evaluation of institutes of higher learning as well as a support to computer program.

Acknowledgments

This research was supported by Education Office Foundation of Guangxi with project No. 201202ZD010. It is the Automatic monitoring system of optical mark method based on PON.

References

[1] Paskaleva K 2009 *International Journal of Innovation and Regional Development* **1(**4) 405-22
[2] Arora A 2012 *Traffic Engineering and Control* **53**(10)375-8
[3] Chen B Cheng H H 2010 *IEEE Trans. Intelligent Transportation Systems* **11(**2) 485–97
[4] Seng Dewen 2012 *Applied Mechanics and Materials* **189** 482-5
[5] Seng Dewen, LI Zhongxue 2008 *Journal of Liaoning Technical University* **27**(1) 9-12
[6] Dikaiakos M D 2012 *Minersoft: ACM Transactions on Internet Technology* **12(**1)
[7] Seng Dewen, Shu Yueqing 2013 *Advances in Intelligent Systems and Computing* 393-400
[8] Di Lecce, Amato V A 2011 *IET Intelligent Transport Systems* **5(**3), 149-58
[9] Prashant T, Kavita D, Manish S 2012 *International Journal of Soft Computing and Engineering* **2**
[10] Lin T Y 1974 Granular Computing: From Rough Sets and Neighborhood Systems to Information Granulation and Computing in Words. *European Congress on Intelligent Techniques and Soft Computing, Heidelberg, Germany*, 1602-06, September 8-12, 1974

Authors

Biao Song, born in January 1977, Suixi, Anhui

Current position, grades: lecturer of Nanjing University of Information Science & Technology
Scientific interests: Information processing, educational management

Social welfare, climate change and strategy selections for developing countries

Liao Yiqin, Chen Xun*

School of Economics and Business Administration, Chongqing University, 400044, Shapingba Chongqing China.

Abstract

This paper introduces external effect of carbon emission in social welfare function, constructs two-stage trade game among three countries, analyzes partially equilibrium output of the three countries and then discusses the influences of different carbon tax policies on social welfare in each country. The study shows it is more effective for developing countries to adopt strategies about founding free trade area and domestic carbon tax collection, up to a higher social total welfare lever, when they face carbon border tax adjustments(BTAs) from developed countries under Nash game conditions. By further studying, the efficiency of domestic carbon tax policy depends on the carbon intensity relation of each country; a higher relative intensity of carbon abroad decreases the negative external effect value caused by the carbon emissions.

Keywords: border tax adjustments, social welfare function, game theory

1 Introduction

In order to limit global warming, the Intergovernmental Panel on Climate Change (IPCC) noted that worldwide annual carbon emissions need to be cut approximately in half by 2050. As an essential part of post-Kyoto international climate negotiations, carbon-based border tax adjustments (BTAs) have been proposed to "level the playing field" by the US, EU and other OECD countries against countries without compatible emissions-reduction commitments, including China [1-2]. The US House of Representatives (2009) passed the American Clean Energy and Security Act of 2009 (HR2998) on June 26, 2009, in which a carbon-based border-adjustment provision was proposed to protect the competitive advantages of American produers against their competitors in countries without comparable emissions-reduction commitments. In the EU, the EC-commissioned High Level Group on Competitiveness, Energy and Environmental Policies proposed the BTA issue in its second report early in 2006. Moreover, BTAs have been recommended as useful policy tools to protect the competitiveness of domestic industries in the EU [3-5].

Such border tax adjustments by participating countries are driven by two related objectives. One is to provide competitiveness (of energy-intensive industries) offsets for domestic producers since the added costs for domestic producers involved with domestic carbon pricing impose a competitive disadvantage on them [6]. The other is carbon leakage, that the reductions of carbon emissions in participating countries such as the EU, US and other OECD countries [7-8]. Alexeeva-Talebi utilized CGE model to analyze the effects of carbon tariff on importing and exporting countries and considered carbon tariff could effectively protect domestic competitiveness of importing countries [9]. Weber and Peters drew the conclusion that carbon tariff neither violates WTO agreement nor influenced on international industrial competitions through analysis of America implementing carbon tariff policy [2]. Based on GTAP data, Hubler analysis drew the conclusion that carbon tariff policy does not only contribute to global emission reduction, but also worsened social welfare of developing countries [10]. Manders and Veenendaal found implementation of carbon tariff policy could effectively reduce carbon leakage, which would be beneficial to EU but damage the welfare of other countries, under EU emission system [11]. Siqueira applied a two-country model to analyze the effects of domestic politics on international externality [12]. He also investigated two kinds of countries which are cooperated with carbon emission or not.

A number of researchers have examined the impacts of BTAs and related policies. Most of the researchers have focused on the effectiveness of BTAs for protecting competitiveness and avoiding carbon leakage. But no general agreement has been found to date. Gros found that BTAs would increase global welfare [13]. Dissou and Eyland found that competitiveness would be hindered by BTAs in Canada [14]. Fischer and Fox suggested that border carbon adjustments would be beneficial for domestic production but not be effective to reduce global emissions [15]. Kuik and Hofkes focused on the carbon leakage-avoidance effects of the EU Emissions Trading System and suggested that BTAs might reduce the sector leakage rate of the iron and steel industry, but the overall leakage-reduction effect was modest [16].

Most of the existing studies focused on the effects of BTAs in developed countries, only a few studies examine the international trade impacts of a carbon border tax adjustments and none seems to look at the welfare implications from a developing country point of view, the country that BTAs mainly target, either implicitly or

* ***Corresponding author*** e-mail: dining16072@163.com

explicitly. This paper focus on how the developing countries international trade patterns could be used, for the fulfillment of climate policy and social welfare objectives. While this paper focus on the strategies selected by the developing countries to maximize the social welfare that are heavily export-oriented, the developing countries trading partners (notably the developing country and the developed country) are clearly also important subjects of discussion and are discussed where relevant.

It is argued that the social welfare of free trade area agreements and domestic carbon tax strategies are higher than the strategy of passive coping. By increasing focus on the efficiency of domestic carbon tax policy, a higher relative intensity of carbon abroad increases the desirability of high import tariff imposed by the home country because a border tax shifts production to the importing country, which in this case leads to lower environmental costs.

2 Model assumption and description

On the basis of analytic framework of traditional international trade theory, this paper blends in influencing factors of carbon tariff and constructs a partial equilibrium model considering climate changes to analyze the effects of importing countries collecting carbon tariff on welfare of importing and exporting countries and analyze the countermeasures of importing countries.

2.1 FUNDAMENTAL ASSUMPTION

Consider a reciprocal market model of intra-industry trade in homogeneous goods. There are three countries (G1, G2 and G3), and each country has a manufacturer which can be regarded as the aggregation of domestic enterprises to produce the same commodity which is both sold in domestic market and exported. The developed country imports goods from developing countries but does not export goods to other countries. The market of the goods in each country is segmented and three-oligarch by Home firm and two foreign firms. A static representation of a Cournot game is useful to map equilibrium options of various tax schemes. Without loss of generality, one unit of labor produces one unit of the goods, so that the wage rate is internationally fixed to unity. Production of the goods is nationally monopolized under a constant marginal cost c ≥ 0, and emits a proportional emission. Letting means the product quantity provided to Country j by the manufacturer of Country i. Suppose international trade transportation cost is 0; the market of each country is completely divided effectively; the information among governments, between governments and enterprises as well as among enterprises is complete. Utility maximization under the budget constraint yields linear inverse demand functions:

$$P_j = P_j(Q_j) = A - Q_j \text{ , where } Q_j = \sum_{i=1}^{3} q_j^i$$

Without loss of generality, we suppose that G1 is a developed country and the pure importing country; G2 and G3 are developing countries with both import and export trade. As the representative of developing countries, China is expressed as G2. Assume all countries carry out non-discriminatory special tariff for the imported products from the countries which do not sign trade agreement with the country. (i≠j) means special import tariff collected from the manufacturer in Country i by Country j. The importing countries collect extra carbon tariff (is used to express carbon tariff collected from Country i by G1) from the countries which do not take emission reduction obligation while collecting emission tax (expressed as) domestically so as to reach the purpose of limiting emission reduction of exporting countries and relieving climatic variations.

2.2 ENTERPRISE PROFIT FUNCTION AND CONSUMER SURPLUS

On each market, manufacturers' products face competetions against other manufacturers. All manufacturers make decisions simultaneously and confirm their own output in the condition of the other manufacturers is given to maximize the profit. Enterprise profit function of the importing country (G1) is:

$$\pi_1 = q_1^1(A - c - \sum_{i=1}^{3} q_1^i - \varepsilon) \quad . \tag{1}$$

Enterprise profit functions of the importing countries (G2 and G3) are:

$$\pi_2 = \sum_{j=1}^{3} q_j^2(A - c - \sum_{i=1}^{3} q_j^i - t_j^2) - \beta_1^2 q_1^2 \quad , \tag{2}$$

$$\pi_3 = \sum_{j=1}^{3} q_j^3(A - c - \sum_{i=1}^{3} q_j^i - t_j^3) - \beta_1^3 q_1^3 \quad . \tag{3}$$

To simplify the formula, make $B = A - c$.

$$CS_j = \frac{1}{2}(Q_j^d)^2 \quad . \tag{4}$$

Consumer surplus of Country j is the function of market product demand; product demand of Country j is $Q_j^d = \sum_{i=1}^{3} q_j^i$.

2.3 EXTERNAL EFFECT OF CARBON EMISSION

This paper adopts the processing method of Daniel Gros (2009) to set negative externality of CO_2 as invariant parameter σ. In order to account for potential differences in carbon technologies and intensities in each country, it is not assumed that the production of each unit leads to the same negative external effect. Instead production abroad takes place with a potentially different carbon intensity, which relative to country j is denoted by γ_j. Furthermore there is

α_j households at Country j. Global social welfare of carbon emission is thus given by:

$(\alpha_1+\alpha_2+\alpha_3)\sigma(\gamma_1 q_1^s+\gamma_2 q_2^s+\gamma_3 q_3^s)$.

External effects of carbon emission for country j should be the function of the total supply (q_j^s) of the country:

$$W_j^{carbon}=(\alpha_1+\alpha_2+\alpha_3)\sigma\gamma_j q_j^s=M\gamma_j q_j^s \quad , \tag{5}$$

$M=(\alpha_1+\alpha_2+\alpha_3)\sigma$ for simplicity.

2.4 SOCIAL TOTAL WELFARE FUNCTION

The gain of a country as the game player is social total welfare they care for. It is composed of four parts: consumer surplus (CS), domestic enterprise profit (π); government tax revenue (TR) and subtract the external effect of carbon emission (W^{carbon}):

$$W_j=CS_j+\pi_j+TR_j-W_j^{carbon}. \tag{6}$$

3 Carbon tax policies under non-cooperative and ccooperative game equilibrium

3.1 SITUATION OF EXPORTING COUNTRIES ARE CONFRONTED WITH BTAS OF IMPORTING COUNTRIES

G1 collects carbon emission tax at home and collects carbon border tax adjustments (BTAs) for imported products of the countries which do not adopt emission reduction measures. G2 and G3 adopt Nash equilibrium of negative coping. Manufacturers of each country are equal competitive relations and take actions simultaneously when seeing carbon emission tax and BTAs policies of each government. The game sequence is that each government first formulates their own tax policy and manufacturers of each country meanwhile maximize their won profits. To solve this game with backward induction, we have:

The optimal tax policy schemes of each government are:

$$\varepsilon=3M\gamma_1-B$$

$$t\beta_1^2=t\beta_1^3=\frac{1}{2}M\gamma_1, \tag{7}$$

$$t_3^2=\frac{1}{130}(50B+30M\gamma_1+24M\gamma_2-54M\gamma_3)$$

$$t_2^3=\frac{1}{130}(50B+30M\gamma_1-54M\gamma_2+24M\gamma_3)$$

Where, $B=A-c$. Since the collecting objects and coefficients of tariff t_j^i and carbon border tax adjustments β_j^i of G1 are the same, let $t\beta_1^2=t_1^2+\beta_1^2$ and $t\beta_1^3=t_1^3+\beta_1^3$ for simplicity. Substitute the optimal tax policies confirmed by the governments into the social welfare and gain the optimal outputs of manufactures of each country are:

$$q_1^1=B-2M\gamma_1 \quad q_1^2=q_1^3=\frac{1}{2}M\gamma_1$$

$$q_2^2=\frac{1}{3}(B+t_2^3) \quad q_2^3=\frac{1}{3}(B-2t_2^3) \tag{8}$$

$$q_3^2=\frac{1}{3}(B-2t_3^2) \quad q_3^3=\frac{1}{3}(B+t_3^2)$$

3.2 SITUATION OF EXPORTING COUNTRIES SET UP FTA TO COPE WITH BTAS IMPORTING COUNTRIES

When G2 and G3 form a free trade area (FTA), internal tariff among the member countries is 0, so $t_2^3=t_3^2=0$. Each member country selects external optimal tariff to make the social welfare maximize. The game sequence is that each government first formulates their own tax policy and meanwhile manufacturers of each country maximize their won profits after seeing the tax policies.

Enterprise profit target function of G1 is the same with Formula (1). Enterprise profit target functions of G2 and G3 changes.In the second stage, enterprises of the three countries maximize their own profit through selecting the optimal output. The solving process is the same with that in Section 3.1. The optimal outputs of manufacturers of each country are:

$$q_1^1=\frac{1}{4}(B-3\varepsilon+t_1^2+\beta_1^2+t_1^3+\beta_1^3)$$

$$q_1^2=\frac{1}{4}(B+\varepsilon-3t_1^2-3\beta_1^2+t_1^3+\beta_1^3) \tag{9}$$

$$q_1^3=\frac{1}{4}(B+\varepsilon+t_1^2+\beta_1^2-3t_1^3-3\beta_1^3)$$

$$q_2^2=q_2^3=q_3^2=q_3^3=\frac{1}{3}B$$

In the first stage, each government selects tax policies to maximize social welfare function (6). The tax policies of the governments are:

$$\varepsilon=3M\gamma_1-B \quad t\beta_1^2=t\beta_1^3=\frac{1}{2}M\gamma_1 . \tag{10}$$

From here we can see that domestic carbon tax and BTAs of G1 are not influenced by FTA set up by G2 and G3. The optimal social welfare function of G1 is the same with that in Section 3.1.

3.3 SITUATION WHERE G2 ADOPTS CARBON TAX

COLLECTION AT HOME TO COPE WITH BTAS

When G3 does not take any measure, but G2 collects carbon tariff at home to cope with carbon border tax adjustments of G1, what different from setting up free trade area is that the two exporting countries select the optimal tariff by themselves to maximize their own welfare. The same batch of commodities should pay both carbon tax collected by exporting countries and BTAs collected by importing countries. According to the principle of avoidance of double taxation, G2 collects carbon tax for domestically produced products, and G1 does not collect extra carbon tariff any more for the products imported from G2.

Enterprise profit target function of G1 is the same with Formula (1). Enterprise profit target functions of G2 and G3 change. Adopting backward induction method to solve tax policies of each country and the optimal outputs of manufacturers of each country are as follows:

$$q_1^1 = B - 2M\gamma_1 + \frac{1}{7}\varepsilon^* \quad q_1^2 = \frac{1}{2}M\gamma_1 - \frac{3}{7}\varepsilon^*$$

$$q_1^3 = \frac{1}{2}M\gamma_1 + \frac{1}{7}\varepsilon^* \quad q_2^2 = \frac{1}{3}(B - 2\varepsilon^* + t_2^3) \qquad (11)$$

$$q_2^3 = \frac{1}{3}(B + \varepsilon^* - 2t_2^3) \quad q_3^2 = \frac{1}{3}(B - 2\varepsilon^* - 2t_3^2)$$

$$q_3^3 = \frac{1}{3}(B + \varepsilon^* + t_3^2)$$

Domestic carbon tax of G1 $\varepsilon = 3M\gamma_1 - B$ keeps unchanged; the tariff collected from G2 by G1 is $t_1^2 = \frac{1}{2}M\gamma_1 - \frac{3}{7}\varepsilon^*$; tariff policy of G1 for G3 is the same with the situation in Section 3.1.

4 Contrastive analysis and discussion

4.1 COMPARISON OF SOCIAL TOTAL WELFARE OF G2 UNDER 3 GAME CONDITIONS

When developed country G1 adopts import carbon tariff collection policy, we compare the social welfare about developing country G2 under three different strategies. The first is the situation adopting negative coping, the second is set up free area with G3 and the third is collecting carbon tax at home.

$$W_2^2 - W_2^1 = \frac{1}{6}t_2^3(\frac{4}{3}B - \frac{1}{3}t_2^3) + (\frac{2}{3}t_3^2 - \frac{1}{3}t_2^3)(M\gamma_1 + \frac{4}{3}B - M\gamma_2 - \frac{2}{3}t_3^2 + \frac{1}{3}t_2^3) > 0$$

Conclusion 1: on the condition of Nash game, the strategy of G2 and G3 setting up FTA can achieve higher social total welfare level than adopting negative coping strategy.

$$W_2^3 - W_2^1 = \frac{2787}{882}(\varepsilon^*)^2 + (-\frac{324}{126}B - \frac{92}{126}t_2^3 + \frac{148}{63}t_3^2 - \frac{37}{21}M\gamma_1 + \frac{37}{21}M\gamma_2)\varepsilon^* + \frac{1}{3}t_2^3(B - 2t_2^3)$$

$W_2^3 - W_2^1$ can be regarded as a unitary quadratic function about domestic carbon tax ε^* of G2. Due to $\frac{2787}{882} > 0$ and $t_2^3(B - 2t_2^3) > 0$, the opening of this function is upward and the intercept term is higher than 0. The solving equation $W_2^3 - W_2^1 = 0$ has two solutions greater than zero, i.e. ε_1^* and ε_2^* ($0 < \varepsilon_1^* < \varepsilon_2^*$). When $\varepsilon_1^* < \varepsilon^* < \varepsilon_2^*$, we have $W_2^3 < W_2^1$. Thus, for developing country G2, social welfare resulted from domestic carbon tax collection is lower than that resulted from negative coping strategy. Thus, developing countries would rather show no interest in BTAs policy of developed countries. When $0 < \varepsilon^* < \varepsilon_1^*$ or $\varepsilon^* > \varepsilon_2^*$, we have $W_2^3 > W_2^1$. Developing country G2 can effectively improve social welfare level when collecting carbon tax at home.

Conclusion 2: on the condition of Nash game, G2 which adopts domestic carbon tax collection policy is not strictly superior to negative coping strategy. When carbon tax planted to be collected by developing countries is in $[\varepsilon_1^*, \varepsilon_2^*]$, negative coping strategy can realize higher social welfare level than carbon tax collection strategy.

$$W_2^3 - W_2^2 = \frac{2728}{882}(\varepsilon^*)^2 + (-\frac{324}{126}B - \frac{92}{126}t_2^3 + \frac{148}{63}t_3^2 - \frac{37}{21}M\gamma_1 + \frac{37}{21}M\gamma_2)\varepsilon^* + \frac{1}{3}t_2^3(B - 2t_2^3)$$
$$- \frac{1}{6}t_2^3(\frac{4}{3}B - \frac{1}{3}t_2^3) - (\frac{2}{3}t_3^2 - \frac{1}{3}t_2^3)(M\gamma_1 + \frac{4}{3}B - \frac{2}{3}t_3^2 + \frac{1}{3}t_2^3) - M\gamma_2(\frac{1}{3}t_2^3 - \frac{2}{3}t_3^2)$$

Solve $W_2^3 - W_2^2 = 0$. Since the coefficient of the quadratic term is greater than 0 and the intercept term is less than 0, it is known that this equation has one positive solution and one negative solution according to the function properties, i.e. $\hat{\varepsilon}_1^*$ and $\hat{\varepsilon}_2^*$ ($\hat{\varepsilon}_1^* < 0 < \hat{\varepsilon}_2^*$). When $\varepsilon^* \geq \varepsilon_2^*$, there is $W_2^3 - W_2^2 \geq 0$. For G2, social total welfare resulted from high-level carbon tax collection at home is higher than that brought by G3 setting up free trade area. The government of G2 will choose the coping strategy of carbon tax collection at home. If $\varepsilon^* \in [0, \hat{\varepsilon}_2^*]$, we have $W_2^3 < W_2^2$. For G2, social total welfare resulted from carbon tax collection at home is lower than that brought by G3 setting up free trade area. In this case, the government of G2 will choose to set up free trade area with G3.

Conclusion 3: under the objective of pursuing maximization of social total welfare, if the carbon tax G2 plants to collect is in $[0, \hat{\varepsilon}_2^*]$, social total welfare resulted from carbon tax collection at home is lower than that brought by G3 setting up free trade area. For G2, carbon tax collection at home is not the optimal selection.

4.2 COMPARISON OF EXTERNAL EFFECTS OF CARBON EMISSION UNDER 3 GAME CONDITIONS

Under the kth game situation, external effect of global carbon emission is the function of aggregate supply of each country:

$$WW_{carbon}^{k}=\sum_{j=1}^{3}M\gamma_j q_j^s \ . \tag{12}$$

The D-value between global carbon emission externality brought by G2 and G3 setting up free trade area and global carbon emission externality brought by negative coping strategy is:

$$WW_{carbon}^{2}-WW_{carbon}^{1}=\frac{1}{3}M\gamma_2(2t_3^2-t_2^3)+\frac{1}{3}M\gamma_3(2t_2^3-t_3^2) \ . \tag{13}$$

Since carbon intensity y_j of unit product of each country is constant greater than 0, the optimal tariff between G2 and G3 (see Section 3.1) has $2t_3^2>t_2^3$ and $2t_2^3>t_3^2$. In this case we have $WW_{carbon}^{2}-WW_{carbon}^{1}>0$.

Conclusion 4: the strategy of G2 and G3 setting up free trade area are results in greater negative external effect of carbon emission than negative coping strategy.

$$WW_{carbon}^{3}-WW_{carbon}^{1}=\frac{1}{21}M\varepsilon^{*}(3\gamma_1-16\gamma_2+17\gamma_3) \ . \tag{14}$$

The influence of carbon tariff strategy of G2 collecting domestic carbon tax depends on the relations of carbon intensity of each country. When $16\gamma_2\geq 3\gamma_1+17\gamma_3$, $WW_{carbon}^{3}\leq WW_{carbon}^{1}$. G2 collecting carbon tax at home brings less negative external effect of carbon emission than negative coping strategy. In other words, this strategy has promotion effect on reduction of global carbon emission externality. When $16\gamma_2<3\gamma_1+17\gamma_3$, we have $WW_{carbon}^{3}>WW_{carbon}^{1}$. G2 collecting carbon tax at home brings larger negative external effect of carbon emission than negative coping strategy. In other words, this strategy is not beneficial to reduction of global carbon emission externality.

Conclusion 5: the influences of G2 collecting carbon tax at home on environment depends on carbon intensity relations of each country. Higher carbon intensity of developing countries means carbon tax collection is more beneficial to reduction of external effect of carbon emission.

Conclusions

This paper analyzes how developing countries dynamically choose and confirm the optimal strategy in international trade competitions when faced with BTAs and considering negative external effect of carbon emission and strategy selection of other developing countries so as to realize maximization of social total welfare level. This paper depicts the process of carbon emission externality influencing government decision-making through introducing the influence of carbon emission on environment into social total welfare function, analyzes how manufacturers of various countries dynamically decide their outputs and corresponding strategies to gain the maximum profits.

The results show when developing countries are faced with developed countries collecting carbon tax, it is more superior for them to choose free trade area strategy than negative coping strategy, but domestic carbon tax collection strategy is not strictly superior to negative coping strategy. This conclusion is different from that Dieter et al. (2009) consider carbon policy is sub-game perfect Nash equilibrium solution. Social total welfare level caused by setting up free trade area and collecting carbon tax domestically is influenced by domestic carbon tax. When the carbon tax level developing countries plan to collect is in a certain interval, it is more superior for developing countries to choose free trade area strategy than the other strategies. In terms of the influences of carbon emission on external environment, the strategy of setting up free trade area is beneficial to enterprises in developing countries expanding market to gain larger profits and meanwhile causes more negative external effects of carbon emission. The effects of domestic carbon tax collection on environment depend on the carbon intensity relation of each country. Higher carbon intensity of developing countries means carbon tax collection is more beneficial to reduction of external effect of carbon emission.

Acknowledgement

In the course of writing the paper, I got many people's help, by which I am moved deeply.

Reference

[1] Dong Y and Whalley J 2011 Carbon motivated regional trade arrangeme-nts: Analytics and simulations *Economic Modeling.* **28** 2783-92
[2] Weber C L, Peters G P 2009 Climate change policy and international trade: Policy considerations in the US *Energy Policy*. **37** 432-40
[3] Van Asselt H, Biermann F 2007 European emissions trading and the international competitiveness of energy-intensive industries: a legal and political evaluation of possible supporting measures, *Energy Policy*. **38** 497-506
[4] Monjon S, Quirion P 2010 How to design a border adjustment for the European Union Emissions Trading System? *Energy Policy* **38** 5199-207
[5] Rivers N 2010 Impacts of climate policy on the competitiveness of Cana-dian industry: How big and how to mitigate? *Energy Economics*. **32** 1092-104
[6] Van Asselt H, Brewer T 2010 Addressing competitiveness and leakage concerns in climate policy. An analysis of border adjustment measures in the US and the EU *Energy Policy* **38** 42-51
[7] Antimiani A, Costantini V, Martini C, Salvatici L, Tommasino M C 2012 Assessing alternative solutions to carbon leakage *Energy Economics* **36** 299-311
[8] Springmann M 2012 A look inwards: Carbon tariffs versus internal impr-ovements in emissions-trading systems *Energy Economics* **34** 5228-39
[9] Alexeeva-Talebi V, Löschel A, Mennel T 2008 Climate Policy and the Problem of Competitiveness Border Tax Adjustments or Inte-

grated Emission Trading *Centre for European Economic Research (ZEW)* Discussing Paper No. 08-061
[10] Hubler M 2009 Can Carbon Based Import Tariffs Effectively Reduce Carbon Emissions? *Kiel Institute for the World Economy* Working Paper No. 1565.
[11] Mander T, Veenendaal P 2009 Border Tax Adjustments and the EU-ETS-A Quantitative Assessment *CPB Document* **171** 36
[12] Siqueira K 2003 International externalities, strategic interaction, and do-mestic politics *Journal of Environmental Economics and Management* **45** 674-91
[13] Gros D 2009 Global Welfare Implications of Carbon Border Taxes *Resource and Environment Economics* CESifo Working Paper Series 2790
[14] Dissou Y, Eyland T 2011 Carbon control policies, competitiveness and border tax adjustments *Energy Economics* **33** 556-64
[15] Fischer C, Fox A K 2012 Comparing policies to combat emissions le-akage: Border carbon adjustments versus rebates *Journal of Environmental Economics and Management* **64** 199-16
[16] Kuik O, Hofkes M 2010 Border adjustment for European emissions trading: Competitiveness and carbon leakage *Energy Policy* **38** 1741-48

Authors

Yiqin Liao, born in February, 1984, Wuhan, Hubei, China

Current position, grades: PhD student
University studies: Quantitative economics
Scientific interest: Environment Economy
Publications: 2 papers
Experience: PhD student in School of Economics and Business Administration, Chongqing University. Studied on the resources economics and policies in developing countries in the background of limit global warming and search the way of sustainable development

Xun Chen, born in September 1950, Gongyi, Henan, China

Current position, grades: Professor of economics
University studies: investment economy, industrial economy and regional economy
Scientific interest: investment economy, industrial economy and regional economy
Publications: 80 papers
Experience: Professor of economics, PhD supervisor in School of Economics and Business Administration, Chongqing University. Main research direction is investment economy, industrial economy and regional economy.

An evaluation model of sustainable development of sports tourism industry based on matter-element theory

Guishen Yu*

Henan Institute of Science and Technology, Xinxiang, Henan, China

Abstract

The sustainable development of sports tourism industry is concerned with many factors. Its evaluation is a complex system engineering. This paper studies the complexity and diversity of factors that influence the sustainable development of sports tourism industry and proposes an evaluation model of sustainable development based on matter-element theory. An indicator system is put in place. Evaluation indicators of classical field matter-elements model, section domain matter-element model and evaluation objects matter-element model are constructed based on matter-element theory. Different methods of calculating extension degree are adopted according to characteristics of the evaluation objects matter-element model to calculate the comprehensive extension degree between evaluation objects matter-element model and classical field matter-elements model. This extension degree refers to the layer of sustainable development capability of evaluation objects. It will provide strategic support for the development of sports tourism industry. Case study has proved that the model and the algorithm are effective.

Keywords: sports tourism industry; sustainable development; matter-element theory; extension degree; evaluation model

1 Introduction

With social and economic development, people have a higher requirement on life quality and their lifestyles become diversified. While enjoying a wealthy life, they are paying more attention to their health. Sports tourism industry is an emerging tourism industry. On one hand, it brings spiritual experience to tourism. On the other, it integrates sports into tourism to reach the purpose of health cultivation. Sports tourism industry is expected to have a bright future [1-4].

Limited by regions and populations, sports tourism industry needs long-term, sustainable and rapid development. And it is significant to evaluate the industry's ability to achieve sustainable development [5-8]. However, such evaluation is complicated given that many factors have to be taken into account. Some factors can be quantified and clear while others are fuzzy and require qualitative descriptions.

Thus, this paper draws merits from previous researches and proposes an evaluation model of the sustainable development of sports tourism industry based on matter-element theory [9-14]. Then the grade of the ability to achieve sustainable development is acquired. Case study proves that the model and the algorithm are feasible and effective.

2 Indicator system of Sustainable development of sports tourism evaluation

Generally speaking, the boom of the sports tourism industry has much to do with geographical locations. Though it is one of the shining points of this industry, destruction on environment may also bear its own consequences. Therefore, many factors should be taken into consideration to evaluate the ability to achieve sustainable development of sports tourism industry. Scientific, objective and effective indicators are selected tailored to real situation. Service-oriented purpose should also be emphasized to select comprehensive, key and feasible indicators.

Therefore, this paper analyzes common indicators from five perspectives, namely, social factors, economic factors, environmental factors, sports factors and tourism factors and constructs a scientific and effective indicator system of sustainable development of sports tourism evaluation, as is shown in Table 1.

TABLE 1 Indicator system of Sustainable development of sports tourism evaluation

Target layer	Criterion layer	Indicator layer
Indicator system of Sustainable development of sports tourism evaluation A	social factors A_1	upgrading of regional social force a_{11}
		government support a_{12}
		tourist satisfaction a_{13}
		public support a_{14}
		sustainable development of related industries a_{15}
	economic factors A_2	input cost a_{21}
		investment returns a_{22}
		market growth a_{23}
		sustainable industry scale a_{24}

* ***Corresponding author*** e-mail: guishen812@163.com

	environmental factors A_3	ecological environment protection a_{31}
		influence on residents a_{32}
		sustainable environment protection efforts a_{33}
	sports factors A_4	green life a_{41}
		sports event brand a_{42}
		sports event appeal a_{43}
		sports event competitiveness a_{44}
	tourism factors A_5	low-carbon tourism a_{51}
		tourism event brand a_{52}
		tourism event appeal a_{53}
		tourism event competitiveness a_{54}

3 Evaluation model of Sustainable development of sports tourism based on matter-element theory

3.1 MATTER-ELEMENT THEORY

As one of the pillars of Extenics, matter-element theory has a promising future. The analysis is based on matter-element design or object analysis and adopts extension qualitative and quantitative analysis with matter-element as the logic cell. A comprehensive use of extension mathematics, extension transformation and extension logic is adopted to address the problems as a part of the extension engineering.

Matter-element, as one of the logic cells of Extenics, describes the design objects by constructing a sequence groups with three elements $R=(N,C,V)$. N refers to the name of the design object. C refers to matter-element characteristics of the design object and V refers to the value of a quantity of the design object N about matter-element characteristic C. If the design object has only one characteristic, then $R=(N,C,V)$ is the matter-element extension model with one dimension, which is called the basic element; In particular, if the object has multiple characteristics, $R=(N,C,V)$ is the matter-element extension model with multiple dimensions. At this moment,

$$C=(c_1,\cdots,c_i,\cdots,c_n)^T,$$

$$V=(v_1,\cdots,v_i,\cdots,v_n)^T,$$

n refers to the number of dimensions of $R=(N,C,V)$. Extension distance can be used to measure the extension correlation degree between matter-element models or matter-element characteristics. Suppose the matter-element characteristics of the reference object is $X=[x_1,x_2]$, that of the target object is v, then the extension distance ρ between the two is:

$$\rho=\left|v-\frac{x_1+x_2}{2}\right|-\frac{x_2-x_1}{2}. \quad (1)$$

3.2 CONSTRUCTING THE MATTER-ELEMENT MODEL OF THE SUSTAINABLE DEVELOPMENT OF SPORTS TOURISM INDUSTRY

Definition 1 Matter-element in classic field for sustainable development evaluation the sustainable development of the sports tourism industry is in different states or has different evaluation grade. Every state or grade corresponds to a characteristic value of the matter-element model.

Thus, suppose there are n characteristics $c_1,c_2,\cdots c_n$ and their corresponding values of a quantity are $v_1,v_2,\cdots v_n$. Construct the evaluation matter-element model $R_n^C(i)$, which is the matter-element in classic field for sustainable development evaluation.

$$R_n^C=\begin{bmatrix} N_C & c_1 & v_1(i) \\ & c_2 & v_2(i) \\ & \vdots & \vdots \\ & c_n & v_n(i) \end{bmatrix}=\begin{bmatrix} N_C & c_1 & \left[v_1^L(i),v_1^R(i)\right] \\ & c_2 & \left[v_2^L(i),v_2^R(i)\right] \\ & \vdots & \vdots \\ & c_n & \left[v_n^L(i),v_n^R(i)\right] \end{bmatrix}. \quad (2)$$

Definition 2 Matter-element in section field for sustainable development evaluation Section field is defined as the collection of the development state or evaluation grade of sustainable development of the sports tourism industry.

The extreme values of n characteristics $c_1,c_2,\cdots c_n$ and their values of a quantity are part of the section field matter-element model R_n^O with n dimensions:

$$R_n^O = \begin{bmatrix} N_O & c_1 & v_1 \\ & c_2 & v_2 \\ & \vdots & \vdots \\ & c_n & v_n \end{bmatrix} = \begin{bmatrix} N_O & c_1 & \left[v_{O1}^L, v_{O1}^R\right] \\ & c_2 & \left[v_{O2}^L, v_{O2}^R\right] \\ & \vdots & \vdots \\ & c_n & \left[v_{On}^L, v_{On}^R\right] \end{bmatrix}. \quad (3)$$

Where, $v_{Oj}^L = \min_{1\le i\le m}\left(v_j^L(i)\right), v_{Oj}^R = \max_{1\le i\le m}\left(v_j^R(i)\right)$. m refers to the number of development state or evaluation grade of sustainable development of sports tourism industry.

3.2.1 EXTENSION degree evaluation model of sustainable development of sports tourism industry

Suppose the matter-element model for evaluation is R_d:

$$R_d = \begin{bmatrix} N_d & c_{d1} & v_{d1} \\ & c_{d2} & v_{d2} \\ & \vdots & \vdots \\ & c_{dn} & v_{dn} \end{bmatrix} = \begin{bmatrix} N_d & c_{d1} & \left[v_{d1}^L, v_{d1}^L\right] \\ & c_{d2} & \left[v_{d2}^L, v_{d2}^L\right] \\ & \vdots & \vdots \\ & c_{dn} & \left[v_{dn}^L, v_{dn}^L\right] \end{bmatrix}. \quad (4)$$

To effectively calculate the extension degree between the matter-element model R_d and matter-element in classic field of different state or evaluation degree, this paper discusses the followings:

(1) If the characteristic value of a quantity of the matter-element model for evaluating R_d is a point value for qualitative description, then the extension distance $\rho\left(R_d^j \mid R_n^C(i)\right)$ between R_d and the matter-element model $R_n^C(i)$ about characteristics j is:

$$\rho\left(R_d^j \mid R_n^C(i)\right) = \left|v_{dj} - \frac{v_j^L(i)+v_j^R(i)}{2}\right| - \frac{v_j^R(i)-v_j^L(i)}{2}. \quad (5)$$

Similarly, the extension distance $\rho\left(R_d^j \mid R_n^O(i)\right)$ between R_d and the matter-element model $R_n^O(i)$ about characteristics j is:

$$\rho\left(R_d^j \mid R_n^O(i)\right) = \left|v_{dj} - \frac{v_{Oj}^L(i)+v_{Oj}^R(i)}{2}\right| - \frac{v_{Oj}^R(i)-v_{Oj}^L(i)}{2}. \quad (6)$$

(2) If the characteristic value of a quantity of the matter-element model for evaluating R_d is a fuzzy interval value of quantity and if the optimal value of a quantity of the matter-element in classic field is at the middle of the interval, then the extension distance $\rho\left(R_d^j \mid R_n^C(i)\right)$ between R_d and the matter-element model $R_n^C(i)$ about characteristic j is:

$$\rho\left(R_d^j \mid R_n^C(i)\right) = \frac{1}{2}\left(\rho\left(v_{dj}^L \mid R_n^C(i)\right) + \rho\left(v_{dj}^R \mid R_n^C(i)\right)\right). \quad (7)$$

Substitute (1) to (7) and get:

$$\rho\left(R_d^j \mid R_n^C(i)\right) = \frac{1}{2}\left(\left|v_{dj}^L - \frac{v_j^L(i)+v_j^R(i)}{2}\right| - v_j^R(i) + v_j^L(i) + \left|v_{dj}^R - \frac{v_j^L(i)+v_j^R(i)}{2}\right|\right). \quad (8)$$

Similarly, the extension distance $\rho\left(R_d^j \mid R_n^O(i)\right)$ between R_d and the matter-element model $R_n^O(i)$ about characteristic j is:

$$\rho\left(R_d^j \mid R_n^O(i)\right) = \frac{1}{2}\left(\rho\left(v_{dj}^L \mid R_n^O(i)\right) + \rho\left(v_{dj}^R \mid R_n^O(i)\right)\right). \quad (9)$$

Substitute (1) to (7) and get:

$$\rho\left(R_d^j \mid R_n^O(i)\right) = \frac{1}{2}\left(\left|v_{dj}^L - \frac{v_{Oj}^L(i)+v_{Oj}^R(i)}{2}\right| - v_{Oj}^R(i) + v_{Oj}^L(i) + \left|v_{dj}^R - \frac{v_{Oj}^L(i)+v_{Oj}^R(i)}{2}\right|\right). \quad (10)$$

(3) If the characteristic value of a quantity of the matter-element model for evaluation R_d is a fuzzy interval value of quantity and if the optimal value of a quantity of the matter-element in classic field is $v_j^0(i) \in \left[v_j^L(i), \frac{v_j^L(i)+v_j^R(i)}{2} \right]$, the extension distance $\rho\left(R_d^j \mid R_n^C(i)\right)$ between R_d and the matter-element model $R_n^C(i)$ about characteristic j is:

$$\begin{cases} \rho\left(R_d^j \mid R_n^C(i)\right) = \frac{1}{2}\left(\rho\left(v_{dj}^L \mid R_n^C(i)\right) + \rho\left(v_{dj}^R \mid R_n^C(i)\right)\right) \\ \rho\left(v_{dj}^x \mid R_n^C(i)\right) = v_j^L(i) - v_{dj}^x & v_{dj}^x \le v_j^L(i) \\ \rho\left(v_{dj}^x \mid R_n^C(i)\right) = v_{dj}^x - v_j^R(i) & v_{dj}^x \ge v_j^0(i) \\ \rho\left(v_{dj}^x \mid R_n^C(i)\right) = (v_j^R(i) - v_j^0(i))(v_{dj}^x - v_j^L(i))/(v_j^L(i) - v_j^0(i)) & v_{dj}^x \in (v_j^L(i), v_j^0(i)) \end{cases} . \quad (11)$$

If the optimal value of a quantity of matter-element in classic field is $v_j^0(i) \in \left[\frac{v_j^L(i)+v_j^R(i)}{2}, v_j^R(i) \right]$, the extension distance $\rho\left(R_d^j \mid R_n^C(i)\right)$ between R_d and the matter-element model $R_n^C(i)$ about characteristic j is:

$$\begin{cases} \rho\left(R_d^j \mid R_n^C(i)\right) = \frac{1}{2}\left(\rho\left(v_{dj}^L \mid R_n^C(i)\right) + \rho\left(v_{dj}^R \mid R_n^C(i)\right)\right) \\ \rho\left(v_{dj}^x \mid R_n^C(i)\right) = v_j^L(i) - v_{dj}^x & v_{dj}^x \le v_j^0(i) \\ \rho\left(v_{dj}^x \mid R_n^C(i)\right) = v_{dj}^x - v_j^R(i) & v_{dj}^x \ge v_j^R(i) \\ \rho\left(v_{dj}^x \mid R_n^C(i)\right) = (v_j^L(i) - v_j^0(i))(v_j^R(i) - v_{dj}^x)/(v_j^R(i) - v_j^0(i)) & v_{dj}^x \in (v_j^0(i), v_j^R(i)) \end{cases} . \quad (12)$$

Similarly, the extension distance $\rho\left(R_d^j \mid R_n^O(i)\right)$ between R_d and the matter-element model $R_n^O(i)$ about characteristics j has two forms. If there is $v_j^0(i) \in \left[v_{Oj}^L(i), \frac{v_{Oj}^L(i)+v_{Oj}^R(i)}{2} \right]$, then:

$$\begin{cases} \rho\left(R_d^j \mid R_n^O(i)\right) = \frac{1}{2}\left(\rho\left(v_{dj}^L \mid R_n^O(i)\right) + \rho\left(v_{dj}^R \mid R_n^O(i)\right)\right) \\ \rho\left(v_{dj}^x \mid R_n^O(i)\right) = v_{Oj}^L(i) - v_{dj}^x & v_{dj}^x \le v_{Oj}^L(i) \\ \rho\left(v_{dj}^x \mid R_n^O(i)\right) = v_{dj}^x - v_{Oj}^R(i) & v_{dj}^x \ge v_j^0(i) \\ \rho\left(v_{dj}^x \mid R_n^O(i)\right) = (v_{Oj}^R(i) - v_j^0(i))(v_{dj}^x - v_{Oj}^L(i))/(v_{Oj}^L(i) - v_j^0(i)) & v_{dj}^x \in (v_{Oj}^L(i), v_j^0(i)) \end{cases} . \quad (13)$$

If there is $v_j^0(i) \in \left[\frac{v_{Oj}^L(i)+v_{Oj}^R(i)}{2}, v_{Oj}^R(i) \right]$, then:

$$\begin{cases} \rho\left(R_d^j \mid R_n^O(i)\right) = \frac{1}{2}\left(\rho\left(v_{dj}^L \mid R_n^O(i)\right) + \rho\left(v_{dj}^R \mid R_n^O(i)\right)\right) \\ \rho\left(v_{dj}^x \mid R_n^O(i)\right) = v_{Oj}^L(i) - v_{dj}^x & v_{dj}^x \le v_j^0(i) \\ \rho\left(v_{dj}^x \mid R_n^O(i)\right) = v_{dj}^x - v_{Oj}^R(i) & v_{dj}^x \ge v_{Oj}^R(i) \\ \rho\left(v_{dj}^x \mid R_n^O(i)\right) = (v_{Oj}^L(i) - v_j^0(i))(v_{Oj}^R(i) - v_{dj}^x)/(v_{Oj}^R(i) - v_j^0(i)) & v_{dj}^x \in (v_j^0(i), v_{Oj}^R(i)) \end{cases} . \quad (14)$$

Thus, we can get the extension degree $K\left(R_d^j \mid R_n^C(i)\right)$ between R_d and the matter-element model $R_n^C(i)$ about characteristics j:

$$K\left(R_d^j \mid R_n^C(i)\right)=\begin{cases}-\dfrac{\rho\left(R_d^j \mid R_n^C(i)\right)}{\left|v_j(i)\right|} & \left[v_{dj}^L, v_{dj}^R\right]\in\left[v_j^L(i), v_j^R(i)\right] \\ \dfrac{\rho\left(R_d^j \mid R_n^C(i)\right)}{\rho\left(R_d^j \mid R_n^O(i)\right)-\rho\left(R_d^j \mid R_n^C(i)\right)} & \left[v_{dj}^L, v_{dj}^R\right]\notin\left[v_j^L(i), v_j^R(i)\right]\end{cases}. \tag{15}$$

If weight of the matter-element characteristic is considered, the weighted extension degree $\phi\left(R_d \mid R_n^C(i)\right)$ is:

$$\phi\left(R_d \mid R_n^C(i)\right)=\sum_{j=1}^{n}\left(w_j * K\left(R_d^j \mid R_n^C(i)\right)\right). \tag{16}$$

Normalize the weighted extension degree $\phi\left(R_d \mid R_n^C(i)\right)$ and get:

$$\bar{\phi}\left(R_d \mid R_n^C(i)\right)=\frac{\phi\left(R_d \mid R_n^C(i)\right)-\min\limits_{1\le i\le m}\phi\left(R_d \mid R_n^C(i)\right)}{\max\limits_{1\le i\le m}\phi\left(R_d \mid R_n^C(i)\right)-\min\limits_{1\le i\le m}\phi\left(R_d \mid R_n^C(i)\right)}. \tag{17}$$

If there is

$$\bar{\phi}\left(R_d \mid R_n^C(k)\right)=\max\left(\bar{\phi}\left(R_d \mid R_n^C(1)\right),\bar{\phi}\left(R_d \mid R_n^C(2)\right),\cdots,\bar{\phi}\left(R_d \mid R_n^C(m)\right)\right). \tag{18}$$

Then it indicates that the evaluation grade of sustainable development of sports tourism industry is of grade k.

4 Case study and test

This paper intends to combine forest areas' geographical features and regional advantage to do the analysis of the sustainable development assessment of sports tourism industry. And based on the above to do the verification and instructions for models and algorithms. The forest yard, their geographic environment is superior, the river in the forest area of diverse, both gentle rapids of the river there are rapids, relatively abundant rainfall in summer, while winter snow sources are abundant, therefore, based on the existing strengths suited to carry out forest skiing and rafting sports tourism projects. To this end, by seeking expert advice on the basis of the forest to carry out skiing, rafting and other sports tourism industry sustainability assessment into force excellent, good, fair, poor four levels, namely the force corresponding to the four sustainability assessment Classic domain matter element, the specific results is as shown in Table 2

TABLE 2 Analysis on evaluation of sustainable development of forest sports tourism industry

indicator layer	Characteristics value of quantity of matter-element				
	Classic field I	Classic field II	Classic field III	Classic field IV	Indicator value
upgrading of regional social force a_{11}	0-0.20	0.20-0.40	0.40-0.70	0.70-1.0	0.50
government support a_{12}	0-0.20	0.20-0.50	0.50-0.80	0.80-1.0	0.60
tourist satisfaction a_{13}	0-0.20	0.20-0.50	0.50-0.80	0.80-1.0	0.70-0.80
public support a_{14}	0-0.20	0.20-0.50	0.50-0.80	0.80-1.0	0.60-0.70
sustainable development of related industries a_{15}	0-0.20	0.20-0.40	0.40-0.60	0.60-1.0	0.30
input cost a_{21}	80-100	60-80	40-60	0-40	80
investment returns a_{22}	0-0.10	0.10-0.30	0.30-0.60	0.60-1.0	0.30
market growth a_{23}	0-0.10	0.10-0.30	0.30-0.60	0.60-1.0	0.20
sustainable industry scale a_{24}	0-20	20-40	40-60	60-100	40
ecological environment protection a_{31}	0-0.60	0.60-0.80	0.80-0.90	0.90-1.0	0.80

influence on residents a_{32}	0-0.10	0.10-0.30	0.30-0.60	0.60-1.0	0.40-0.50
sustainable environment protection efforts a_{33}	0-0.40	0.40-0.60	0.60-0.80	0.80-1.0	0.80
green life a_{41}	0-0.20	0.20-0.40	0.40-0.60	0.60-1.0	0.60-0.80
sports event brand a_{42}	0-0.20	0.20-0.40	0.40-0.60	0.60-1.0	0.80
sports event appeal a_{43}	0-0.20	0.20-0.40	0.40-0.60	0.60-1.0	0.60
sports event competitiveness a_{44}	0-0.20	0.20-0.40	0.40-0.60	0.60-1.0	0.60
low-carbon tourism a_{51}	0-0.10	0.10-0.30	0.30-0.60	0.60-1.0	0.80
tourism event brand a_{52}	0-0.20	0.20-0.40	0.40-0.60	0.60-1.0	0.40
tourism event appeal a_{53}	0-0.20	0.20-0.40	0.40-0.60	0.60-1.0	0.60
tourism event competitiveness a_{54}	0-0.20	0.20-0.40	0.40-0.60	0.60-1.0	0.60

According to the proposed model and algorithm, we can get the extension distance and extension degree of forest sports tourism, such as skiing, rafting, as are shown in Table 3 and Table 4.

TABLE 3 Extension distance of sustainable development of forest sports tourism industry

Indicator layer	**Classic field I**	**Classic field II**	**Classic field III**	**Classic field IV**	**Section field**
upgrading of regional social force a_{11}	0.30	0.10	-0.10	0.20	-0.50
government support a_{12}	0.40	0.10	-0.10	0.20	-0.40
tourist satisfaction a_{13}	0.55	0.25	-0.015	0.05	-0.25
public support a_{14}	0.45	0.15	-0.10	0.15	-0.35
sustainable development of related industries a_{15}	0.10	-0.10	0.10	0.30	-0.30
input cost a_{21}	0	0	20	40	-20
investment returns a_{22}	0.20	0	0	0.30	-0.30
market growth a_{23}	0.10	-0.10	0.10	0.40	-0.20
sustainable industry scale a_{24}	20	0	0	20	-40
ecological environment protection a_{31}	0.20	0	0	0.10	-0.20
influence on residents a_{32}	0.35	0.15	-0.10	0.15	-0.45
sustainable environment protection efforts a_{33}	0.40	0.20	0	0	-0.20
green life a_{41}	0.50	0.25	0.10	-0.10	-0.30
sports event brand a_{42}	0.60	0.40	0.20	-0.20	-0.20
sports event appeal a_{43}	0.40	0.20	0	0	-0.40
sports event competitiveness a_{44}	0.40	0.20	0	0	-0.40
low-carbon tourism a_{51}	0.70	0.50	0.20	-0.20	-0.20
tourism event brand a_{52}	0.20	0	0	0.20	-0.40
tourism event appeal a_{53}	0.40	0.20	0	0	-0.40
tourism event competitiveness a_{54}	0.40	0.20	0	0	-0.40

TABLE 4 Extension degree of sustainable development of forest sports tourism industry

indicator layer	Classic field I	Classic field II	Classic field III	Classic field IV
upgrading of regional social force a_{11}	-0.375	-0.167	0.333	-0.286
government support a_{12}	-0.50	-0.20	-0.10	-0.333
tourist satisfaction a_{13}	-0.688	-0.50	0.168	-0.20
public support a_{14}	-0.563	-0.300	0.333	-0.300
sustainable development of related industries a_{15}	-0.25	0.50	-0.25	-0.50
input cost a_{21}	0	0	-0.50	-0.667
investment returns a_{22}	-0.40	0	0	-0.50
market growth a_{23}	-0.333	0.50	-0.333	-0.667
sustainable industry scale a_{24}	-0.333	0	0	-0.333
ecological environment protection a_{31}	-0.50	0	0	-0.334
influence on residents a_{32}	-0.438	0.25	0.337	0.25
sustainable environment protection efforts a_{33}	-0.667	-0.50	0	0
green life a_{41}	-0.625	-0.455	-0.25	0.25
sports event brand a_{42}	-0.75	-0.667	-0.50	0.50
sports event appeal a_{43}	-0.50	-0.333	0	0
sports event competitiveness a_{44}	-0.50	-0.333	0	0
low-carbon tourism a_{51}	-0.778	-0.714	-0.50	0.50
tourism event brand a_{52}	-0.333	0	0	-0.333
tourism event appeal a_{53}	-0.50	-0.333	0	0
tourism event competitiveness a_{54}	-0.50	-0.333	0	0

Weigh the extension degree of each indicator and get the membership degree sequence between ability to achieve sustainable development of forest sports tourism, such as skiing, rafting, and every classic field. The conclusion reached is that the forest region is in a good state of sustainable development.

5 Conclusions

This paper proposes an evaluation model of sustainable development based on matter-element theory. An indicator system is put in place. Evaluation indicators of classical field matter-element model, section domain matter-element model are constructed to calculate the extension distance and the extension degree between different evaluation indicators and matter-element in classic field. The grade of the sports tourism industry is known. This paper provides a solution to optimize the sports tourism industry. Case study has proved that the model and the algorithm are effective.

References

[1] Xu Hong, Fan Qing 2008 A Studying on the Obstacles and Competitiveness Promotion Strategy of Chinese Tourism Industrial Convergence *Tourism Science* **22**(4) 1-5
[2] Chen Yang, Zuo Shan 2014 Based on Rough Set Theory of Community Sports Service Public Satisfaction Evaluation of Empirical Study *Journal of Xiangtan University (Philosophy and Social Sciences)* **38**(1) 95-101
[3] Chang Huajun, Han Xiaoyan 2003 Study on China's Sports Tourism Status Quo and Prospects *Sports Culture Guide* (1) 33-4
[4] Gibson H 2003 Sport tourism: an introduction to the special issue *Journal of Sport Management* (17) 205-13
[5] Yu Feng, Wu Yi, Shao Xianming 2013 Matter-Element Evalucation Of The Sustainable Development Of Skiing Tourism In Grey Area *Tourism economy* (3) 121-2

[6] Yu Feng, Wu Yi 2013 Fuzzy Comprehensive Evaluation of College Sunshine Sports Based on Grey Correlation Coefficient *Frontier* (8) 109-10
[7] Li Ping 2014 Competitiveness of Urban Sports Tourism Industry in China and Abroad Based on PCA and AHP *Journal of Shenyang Sport University* **33**(1) 28-31+36
[8] Luo Zhibo, Xiong Maoxiang, Wen Tingxiao 2013 Study and Assessment on the Competitiveness of National Competitive Sports *Journal of Beijing Sport University* **36**(2) 21-6
[9] Meng Lisha, Wang Zhenhua, Shen Zhonghua 2013 Establishment and Application of Risk Evaluation Matter-element Model for Venture Capital Project *Science and Technology Management Research* (10) 933-196+201
[10] Cai Wen, Yang Chunyan 2013 Basics and Methodologies of Extenics *Chinese Science Bulletin* **58**(13) 1190-9
[11] Wang Ti-chun, Yang Ai-jun, Liang-feng B U 2013 Mechanism scheme design based on multi-attribute extension gray relevant optimized decision-making model *Systems Engineering – Theory & Practice* **33**(9) 2321-9
[12] Ti-chun Wang, Ai-jun Yang, Shi-sheng Zhong 2014 Multi-Attribute Extension Fuzzy Optimized Decision-Making Model Of Scheme Design *Tehnički vjesnik/Technical Gazette* **21**(2) 239-47
[13] Cai Wen, Yang Chunyan 2010 The Application Research, Popularization and Generalization of Extenics *Mathematics In Practice And Theory* **40**(7) 214-20
[14] Zhao Y W, Zhang G X 2012 A New Integrated Design Method Based On Fuzzy Matter-Element Optimization *Journal of Materials Processing Technology* **129**(1-3) 612-18

Authors

Guishen Yu, born on August 12, 1978, Henan

Current position, grades: Associate professor, School of Physical Education, Henan Institute of Science and Technology
University studies: Henan Normal University, in the Institute of Physical Education (1998-2002)
Scientific interest: teaching theories and training of sports
Publications: 4 papers

A method of relative Grey relation degree combined with combination weights for materials selection

Xiqin Wen[1,2*], Kaibo Wu[2], Anhua Peng[1,2]

[1]*Engineering Training Center, Huaihai Institute of Technology, 59 Cangwu Road, Lianyungang, Jiangsu, China*

[2]*School of Mechatronic Engineering, China University of Mining and Technology, 1 University Road, Xuzhou, China*

Received 23 May 2014, www.cmnt.lv

Abstract

The quality and cost of a product rely heavily on suitable material selection, and therefore the ability to select the most appropriate material for a given application is the fundamental challenges faced by the design engineer. The general grey relational analysis (GRA) has three weaknesses, (i) the weight determination depends only on expert judgments, (ii) the qualitative indexes are simply quantified with exact numbers, and (iii) the general GRA only takes into account the relationship between the imaginarily best material and the candidate materials. Weights were determined by combining subjective and objective weights based on maximum deviation, the qualitative indexes were fuzzily quantified through trapezoidal fuzzy numbers (TFNs), and then ranked alternatives according to relative grey relation grade. The illustrative example showed that the results matched well with that using WAA and TOPSIS, proved the proposed method reasonable and trustworthy. And therefore the proposed method possesses important application values.

Keywords: material selection, combination weights (CW), relative Grey relation analysis (RGRA), fuzzy numbers, maximum deviation

1 Introduction

It is well known that materials play an important role in engineering designs. After the conceptual design stage, designers always need to select materials with specific properties which can guarantee optimum system performance by satisfying all existing constraints [1]. And it is an important step in engineering designs, since an inappropriate choice of material(s) can adversely affect the productivity, profitability, and reputation of a manufacturing organization as well [2]. When selecting materials for engineering designs, a clear understanding of the functional requirements for each individual component is required and various important criteria or attributes need to be simultaneously considered. These attributes include not only the traditional ones such as usability, machinability, and cost, but also material impact on environment, recycling, and even cultural aspects. They contradict and even conflict with each other, and furthermore Deng and Edwards [3] emphasized that the process of materials selection should be combined with structural optimazation. And therefore the ability to select the most appropriate material for a given application is the fundamental challenges faced by the design engineers. There have been much literature dealing with the material selection, and so great progress has been made in this field. Zhou et al. [4] proposed an integration of artificial neural networks (ANN) with genetic algorithms (GA) to optimize the multi-objectives of material selection, and applied it to selecting proper materials for drink containers. [5] proposed fuzzy inference method and applied it to material selection for a liquid nitrogen storage tank and spar of an aircraft wing. [6] also proposed fuzzy inference method for material substitution selection in electric industry, while combined with fuzzy weight average to extend fuzzy inference to uncertain environment. The advantage of fuzzy inference is that it does not require normalizations to ratings, but identifying membership function for each attribute is strongly subjective, depending entirely on expert's experiences. [7] presented an intelligent method to deal with the materials selection problems where the design configurations, working conditions, as well as the design-relevant information are not precisely known, and applied it to selecting optimal materials for robotic components at early stage of design.. [8] presented digital tools for material selection in product design, where about three hundred software, database and website references were collected, and 87 were selected to try to answer a few important questions to help designers, engineer students, and all kinds of professionals perform materials selection for product design.

The grey system theory proposed by Deng in 1982 [9] has been proven to be useful for dealing with problems with poor, insufficient, and uncertain information. According to it, systems can be divided into three classes: white systems, which have completely clear information, black systems, which have completely unknown information, and grey systems, which lie between white systems and black systems [10]. The grey relational analysis based on this theory can further be effectively adopted for solving the complicated interrelationships among the designated performance characteristics. Through this analysis, a grey

[*] ***Corresponding author's*** e-mail: xqw999_1@163.com

relational grade (GRD) is favourably defined as an indicator of multiple performance characteristics for evaluation. In recent years, grey relational analysis has become a powerful tool to multi-attribute decision making (MADM). In [11] used GRA to compute the weights of each decision maker: a larger weight is assignned to the expert whose preferences are more similar to the others' preferences, while a less weight is assigned to the expert whose preferences are less similar to the others' preferences. [12] adopted grey relation entropy analysis to evaluate information technology impact on business performance of biotechnology industry. There is a great deal of literature on the applications of GRA to material selection. In [13] presented an integrated methodology of performing an order pair of materials and end-of-life strategy for the purpose of material selection, where the GRA is employed to calculate the grey relational grade between the imaginarily best material and candidate materials. The material with the greatest grey relational grade is the best choice. [14] also presented GRA to select handbag materials for a leading handbag manufacturer in Guangdong Province, while combined with binary dominance matrix to specify the weights. However, the general GRA has three weaknesses, (i) the weight determination depends only on expert judgments, (ii) the qualitative indexes are simply quantified with exact number, and (iii) the general GRA only takes the relationship between the imaginarily best material and the candidate materials into account, but takes into no account the relationship between the imaginarily worst material and the candidate materials. Therefore the general GRA in a sense has limitations. The methodology of relative GRA (RGRA) is proposed, not only taking into the account the relationship between the imaginarily best material and the candidate materials but also into the account the relationship between the imaginarily worst material and the candidate materials. It is predicted that the results of RGRA are more reliable than that of general GRA. Weights are determined by combining subjective and objective weights based on maximum deviation, and the qualitative indexes are fuzzily quantified through trapezoidal fuzzy number (TFN).

The paper is structured as follows. Section 2 enunciates the mechanism of maximum deviation and gives the corresponding formulations to calculate combination weights. Section 3 explains the concept of trapezoidal fuzzy numbers and their operation rules. The calculation procedure for relative grey relation grade is enunciated in Section 4. Section 5, taking the bearing material selection for example, details the decision making process, and each ways to determine weights. Section 6 closes the paper with a short discussion of the issues raised and pointing the way to future research direction.

2 Combination weights based on maximum deviation

The way of identifying weights largely includes subjective weighting and objective weighting. The former identifies weights depending only on the subjective preferences or experiences of an expert, such as Delphi method, analytic hierarchy process (AHP) etc., while the latter does depending only on the information of a matrix of decision making and the mathematical model based on it, like entropy weight, principal component analysis, multi-objective optimization, etc [15]. Subjective weights fully reflect decision makers' empirical judgments, not violating common sense in identifying the relative importance of attributes, but with much more arbitrariness and poorer accuracy and reliability. However, objective weights enjoy objective criteria, but neglect subjective preferences or experiences of an expert, sometimes resulting in irrational practice. To make the selection more scientific, the subjective weights and objective ones are combined to obtain the better ones, combination weights (CW), which reflect not only subjective information but also objective one. Supposing a MADM problem, there are m alternatives, expressed as $\mathbf{s}=\{s_1, s_2, \cdots, s_m\}$, and n attributes, expressed as $\mathbf{p}=\{p_1, p_2, \cdots, p_n\}$. Letting a_{ij} be the ratings of s_i with respect to p_j, $i=1,2,\cdots,m$, $j=1,2,\cdots,n$ and $A=(a_{ij})_{m\times n}$ be decision making matrix, ratings may be in different units (e.g. material cost expressed in dollars, yield strength expressed in MPa), resulting in incommensurability, which hence entails normalizations. Suppose the normalized decision making matrix is $\mathbf{b}=(b_{ij})_{m\times n}$ with any element within [0, 1], and the larger the b_{ij}, the better the performance.

If the ratings of the *j*-th attribute are nearly same to all alternatives, then it contribute little or nothing to the alternative rank, and zero should be assigned to the weight of the attribute. Otherwise, if the ratings of the *j*-th attribute vary greatly to all alternatives, then it contributes greater to the alternative rank, and greater weight should be assigned to it. Deviation is the index in statistics reflecting the differences to all alternatives under an attribute, and as such, the weighting vector $\boldsymbol{\omega}_c$ should be selected, which enables the total deviation sum of all n attributes to reach the maximum, called maximum deviation principle.

Suppose there are l methods of weighting and the weight vector according to *k*-th method is:

$$\boldsymbol{\omega}_k=(\omega_{1k}, \omega_{2k}, \cdots, \omega_{nk})^T \quad k=1,2,\cdots,l, \tag{1}$$

where $\omega_{jk}\geq 0$ and $\sum_{j=1}^{n}\omega_{jk}=1$. Combination weight vector can be taken as $\boldsymbol{\omega}_c=(\omega_{c1}, \omega_{c2}, \cdots, \omega_{cn})$, where $\boldsymbol{\omega}_c=\theta_1\boldsymbol{\omega}_1+\theta_2\boldsymbol{\omega}_2+\cdots+\theta_l\boldsymbol{\omega}_l$ and $\theta_1, \theta_2, \cdots, \theta_l$ represent linear coefficient respectively of $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_l$ satisfying $\theta_k\geq 0, k=1,2,\cdots,l$, and $\sum_{k=1}^{l}\theta_k^2=1$. According to [16],

the value of θ_k based on maximum deviation principle can be taken as:

$$\theta_k = \mathbf{C}\boldsymbol{\omega}_\mathbf{k} / \sqrt{\sum_{k=1}^{l} (\mathbf{C}\boldsymbol{\omega}_\mathbf{k})^2}\,, \tag{2}$$

where $\mathbf{C}$ is a row vector with n dimensions and can be calculated as:

$$\mathbf{C} = (c_1, c_2, \cdots, c_j, \cdots, c_n) = \left(\sum_{i=1}^{m}\sum_{i_1=1}^{m} \left| b_{i1} - b_{i_1 1} \right|, \sum_{i=1}^{m}\sum_{i_1=1}^{m} \left| b_{i2} - b_{i_1 2} \right|, \cdots, \sum_{i=1}^{m}\sum_{i_1=1}^{m} \left| b_{in} - b_{i_1 n} \right| \right). \tag{3}$$

3 Fuzzy quantification for linguistic information

3.1 TRAPEZOIDAL FUZZY NUMBERS

Suppose $\tilde{A}$ is a bounded and convex fuzzy subset in the domain of R $(R \in [0,1])$, and possesses the continuous membership function $u_{\tilde{A}}(x)$ as:

$$u_{\tilde{A}}(x) = \begin{cases} u_{\tilde{A}}^{L}(x) & a < x < m \\ 1 & m \le x \le n \\ u_{\tilde{A}}^{R}(x) & n < x < \beta \\ 0 & others \end{cases}, \tag{4}$$

where $u_{\tilde{A}}^{L}(x)$ is strictly increasing function in the domain of $[a,m]$, called the left membership function, while $u_{\tilde{A}}^{R}(x)$, strictly decreasing function in the domain of $[n,\beta]$, called the right membership function. Trapezoidal fuzzy numbers (TFNs) are the widely used fuzzy numbers, whose membership function is shown in Figure 1, and can be expressed as (a,m,n,β), also as $(m,n;\gamma,\delta)$ where $\gamma = m-a, \delta = \beta - n$. If $m = n$, then trapezoidal fuzzy numbers reduce to triangle fuzzy numbers, and it can be said that triangle fuzzy numbers are the special case of trapezoidal fuzzy numbers.

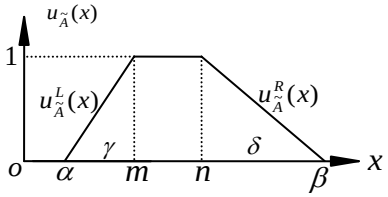


FIGURE 1 Membership function of TFN

For qualitative attributes, they are usually expressed in linguistic information. Table 1 shows their corresponding TFNs, while Figure 2 shows their membership function.

TABLE 1 Representations of linguistic information in trapezoidal fuzzy numbers

Linguistic information	**In the form of** (α,m,n,β)	**In the form of** $(m,n;\gamma,\delta)$
Very poor (VP)	(0.0, 0.0, 0.0, 0.2)	(0.0, 0.0; 0.0, 0.2)
Poor (P)	(0.0, 0.0, 0.1, 0.3)	(0.0, 0.1; 0.0, 0.2)
Medium poor (MP)	(0.0, 0.2, 0.2, 0.4)	(0.2, 0.2; 0.2, 0.2)
Fair (F)	(0.3, 0.5, 0.5, 0.7)	(0.5, 0.5; 0.2, 0.2)
Medium good (MG)	(0.6, 0.8, 0.8, 1.0)	(0.8, 0.8; 0.2,0.2)
Good (G)	(0.7, 0.9, 1.0, 1.0)	(0.9, 1.0; 0.2, 0.0)
Very good (VG)	(0.8,1.0,1.0,1.0)	(1.0, 1.0; 0.2, 0.0)

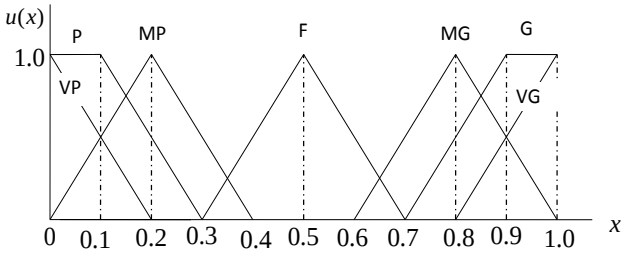


FIGURE 2 Representations of linguistic information in TFNs

3.2 OPERATIONS FOR TFNS

Suppose the two trapezoidal fuzzy numbers $M_1 = (m_1,n_1;\gamma_1,\delta_1)$ and $M_2 = (m_2,n_2;\gamma_2,\delta_2)$, then according to [17] the operators of multiplication and division can be respectively defined as:

$$M_1 \cdot M_2 \approx (m_1 m_2, n_1 n_2; m_1 \cdot \gamma_2 + m_2 \cdot \gamma_1 - \gamma_1 \cdot \gamma_2, n_1 \cdot \delta_2 + n_2 \cdot \delta_1 - \delta_1 \cdot \delta_2), \tag{5}$$

$$M_1 / M_2 \approx [m_1 / n_2, n_1 / m_2; (m_1 \cdot \delta_2 + n_2 \cdot \gamma_1) / n_2(n_2 + \delta_2), (n_1 \cdot \gamma_2 + m_2 \cdot \delta_1) / m_2(m_2 + \gamma_2)]. \tag{6}$$

By means of the approximate calculation, the operation of trapezoidal fuzzy number can be dramatically simplified, and the calculation precision in most cases meets the project requirements, but it cannot be repeatedly used, otherwise, the cumulative error would lead to erroneous results.

3.3 TOTAL EXPECTATIONS OF TFNS

Supposing $\tilde{A}$ is a fuzzy number, and $g_{\tilde{A}}^{L}(y)$, $g_{\tilde{A}}^{R}(y)$ respectively the inverse function of $u_{\tilde{A}}^{L}(x)$, $u_{\tilde{A}}^{R}(x)$. Letting $I_L(\tilde{A}) = \int_0^1 g_{\tilde{A}}^{L}(y)dy$ and $I_R(\tilde{A}) = \int_0^1 g_{\tilde{A}}^{R}(y)dy$, then $I(\tilde{A}) = [I_L(\tilde{A}) + I_R(\tilde{A})]/2$ is called as the total expectation of fuzzy number $\tilde{A}$. As for the TFN $\tilde{A} = (m,n;\gamma,\delta)$, the total expectation can be calculated as:

$$I(\tilde{A}) = (2m + 2n + \delta - \gamma)/4. \tag{7}$$

4 Relative Grey relation analysis (RGRA)

4.1 NORMALIZATION

The purpose of normalization is to obtain dimensionless values of the different criteria so that all of them can be compared with each other, which, in grey system theory, is also called grey relation generation.

1) For benefit type attributes

$$b_{ij} = \frac{a_{ij}}{\max\limits_i a_{ij}}. \tag{8}$$

2) For cost type attributes

$$b_{ij} = \frac{\min\limits_i a_{ij}}{a_{ij}}. \tag{9}$$

3) For fixation type attributes

$$b_{ij} = \begin{cases} \dfrac{\min\limits_i \left|a_{ij} - g^j\right|}{\left|a_{ij} - g^j\right|} & a_{ij} \neq g^j \\ 1 & a_{ij} = g^j \end{cases}, \tag{10}$$

where, g^j is the optimum value to fixation attribute *j*.

4.2 GREY RELATION COEFFICIENT

Suppose comparison sequences are $\mathbf{b_i}(i = 1,2,\cdots,m)$, $\mathbf{b_i} = (b_{i1}, b_{i2}, \cdots, b_{ij}, \cdots, b_{in})$, namely, the ratings of the each alternative material. Suppose the positive reference sequence is $\mathbf{b_o^+}$, $\mathbf{b_o^+} = (b_{o1}^+, b_{o2}^+, \cdots, b_{oj}^+, \cdots b_{on}^+)$, $b_{oj}^+ = \max\{b_{1j}, b_{2j}, \cdots, b_{mj}\}$, namely, the ratings of the hypothetically optimal material, and the negative reference sequence is $\mathbf{b_o^-}$, $\mathbf{b_o^-} = (b_{o1}^-, b_{o2}^-, \cdots, b_{oj}^-, \cdots b_{on}^-)$, $b_{oj}^- = \min\{b_{1j}, b_{2j}, \cdots, b_{mj}\}$, namely, the ratings of the hypothetically worst material. Then grey relational coefficient between the hypothetically optimal material $\mathbf{b_o^+}$ and the alternative material $\mathbf{b_i}(i = 1,2,\cdots,m)$ in the *j*-th attribute is defined as:

$$r_{io}^+(j) = \frac{\min\limits_i \min\limits_j \left|b_{oj}^+ - b_{ij}\right| + \rho \max\limits_i \max\limits_j \left|b_{oj}^+ - b_{ij}\right|}{\left|b_{oj}^+ - b_{ij}\right| + \rho \max\limits_i \max\limits_j \left|b_{oj}^+ - b_{ij}\right|}. \tag{11}$$

The grey relational coefficient between the hypothetically worst material $\mathbf{b_o^-}$ and the alternatives $\mathbf{b_i}(i = 1,2,\cdots,m)$ in the *j*-th attribute is defined as follows:

$$r_{io}^-(j) = \frac{\min\limits_i \min\limits_j \left|b_{oj}^- - b_{ij}\right| + \rho \max\limits_i \max\limits_j \left|b_{oj}^- - b_{ij}\right|}{\left|b_{oj}^- - b_{ij}\right| + \rho \max\limits_i \max\limits_j \left|b_{oj}^- - b_{ij}\right|}, \tag{12}$$

where ρ is the distinguishing coefficient with $\rho \in [0,1]$, usually set as 0.5 in this study. Suppose ω_{cj} is the CW of the *j*-th attribute, and then grey relational grade (GRD), the weighted sum of the grey relational coefficient, can be defined as:

$$\gamma_{io}^+ = \sum_{j=1}^n \omega_{cj} \gamma_{io}^+(j), \tag{13}$$

$$\gamma_{io}^- = \sum_{j=1}^n \omega_{cj} \gamma_{io}^-(j). \tag{14}$$

4.3 RELATIVE GREY RELATION GRADE

The relative grey relation grade can be calculated as:

$$\gamma_{io} = \frac{\gamma_{io}^+}{\gamma_{io}^+ + \gamma_{io}^-} \quad (i = 1,2,\cdots,m) \tag{15}$$

The larger the value γ_{io}, the better the performance of the alternative. It should be noted that the greater γ_{io}^+ does not always lead to the less γ_{io}^-, and therefore selecting materials only depending on γ_{io}^+ can be biased. It can be inferred from Equation (14) that the greater γ_{io}^+ and the less γ_{io}^- must result in the greater γ_{io}.

5 Case Study

Suppose a bearing works under the conditions of higher speed with stable, light load, and the task is to select the best material for the bearing. According to the rigid attributes, which, if a material to be accepted, must be fully satisfied, the materials not satisfying any of the requirements of rigid attributes are firstly eliminated. The materials initially screened out and their ratings are shown in Table 2. The procedure for the material selection is as follows.

5.1 NORMALIZATIONS

For the qualitative attributes, such as fatigue durability, corrosion durability, abrasion durability, and anti-seizing, they were expressed in TFNs. Benefit type, like fatigue durability, corrosion durability, abrasion durability, and anti-seizing, was normalized by Equations (6) and (8) and then converted into total expectations by Equation (7), fixation type, like hardness with optimum value of hardness specified as 45HBS normalized using Equation (10), and cost type by Equation (9). The normalized ratings are shown in Table 3.

TABLE 2 Raw ratings of the candidate materials

No.	Materials	Cost	Hardness (HBS)	Fatigue durability	Corrosion durability	Abrasion durability	Anti-seizing
1	*Zchsn3*	*15*	*24*	MP	VG	MP	VG
2	*Zchpb1*	*100*	*30*	MP	G	MG	VG
3	*ZQpb30*	*70*	*25*	VG	MG	MG	MP
4	*ZznM7-5*	*80*	*100*	G	VG	VG	MP

TABLE 3 Normalized ratings

No.	Materials	Cost	Hardness (HBS)	Fatigue durability	Corrosion durability	Abrasion durability	Anti-seizing
1	*Zchsn3*	1.0000	*0.7143*	0.1933	1.0000	0.1933	1.0000
2	*Zchpb1*	*0.1500*	1.0000	0.1933	0.9496	0.8319	1.0000
3	*ZQpb30*	*0.2143*	*0.7500*	1.0000	0.8319	0.8319	0.1933
4	*ZznM7-5*	*0.1875*	*0.2727*	0.9496	1.0000	1.0000	0.1933

5.2 COMBINATION WEIGHTS

5.2.1 Objective weights

1) Entropy method: according to Shannon's entropy method [18], if the ratings of each alternative under an attribute have more obvious differences, such an attribute plays a more important role in choosing the best alternative, and a greater weight should be assigned to it. The entropy value for the attribute j can be defined as:

$$H_j = -k\sum_{i=1}^{m} f_{ij} \ln f_{ij}\,, \tag{16}$$

where $k = 1/\ln m$ (m denoting the number of alternatives), $f_{ij} = b_{ij} / \sum_{i=1}^{m} b_{ij}$. The weight of attribute j can be defined as:

$$\omega_j = \frac{1 - H_j}{\sum_{j=1}^{n}(1 - H_j)}. \tag{17}$$

In the special case where under an attribute, say attribute *j*, the ratings of all alternatives are the same, it can be calculated $H_j = 1$ and $\omega_j = 0$. According to Table 3, the entropy weight vector $\boldsymbol{\omega}_1$ can be calculated as

$$\boldsymbol{\omega}_1 = (0.3332, 0.0795, 0.2334, 0.0027, 0.1127, 0.2386)^T.$$

2) Principal component analysis (PCA). Its main advantage is significantly alleviating loading and complexion of information by simplifying several correlated variables into fewer uncorrelated and independent principal components, at the same time preserving as much original information as possible using linear combination. In recent time, PCA has gradually become an analytical tool for the optimization of a system with multiple performance characteristics. Firstly calculate the correlation coefficient matrix to the normalized decision making matrix $\mathbf{b} = (b_{ij})_{m\times n}$ using the function $\mathbf{S} = corrcoef\,(\mathbf{b})$ in Matlab, then calculate eigenvalues and corresponding eigenvectors, and finally select the eigenvector corresponding to the maximum eigenvalue, after normalization, as the weight vector $\boldsymbol{\omega}_2$, and the resultant weight vector is

$$\boldsymbol{\omega}_2 = (0.1747, 0.1100, 0.2080, 0.1067, 0.1923, 0.2084)^T$$

5.2.2 Subjective weights

1) Analytic Hierarchy Process (AHP) [19-20], developed first by Satty as a popular tool for MADM, has been increasingly widely used in more and more domains. Its main steps include: (i) Against total goal, the six criteria are compared in pair, consequently constituting a judgment matrix; (ii) Calculating eigenvalues and corresponding eigenvectors to the judgment matrix; and finally (iii) selecting the eigenvector corresponding to the maximum eigenvalue, after normalization, as the weight vector $\boldsymbol{\omega}_3$. The resultant weight vector is

$$\boldsymbol{\omega}_3 = (0.2856, 0.2380, 0.1904, 0.0476, 0.0952, 0.1428)^T \quad .$$

2) Delphi method: suppose the weight vector is

$$\boldsymbol{\omega}_4 = (0.1700,\ 0.1600,\ 0.1900,\ 0.1000,\ 0.1800,\ 0.2000)^T.$$

5.2.3 Combination weights

The row vector $\mathbf{C}$ according to Equation (3) can be calculated as $\mathbf{C} = (5.1400, 4.4352, 6.300, 1.1002, 4.8000\ 6.4000)$, then the linear coefficient according to Equation (2) as $\theta_1 = 0.5371, \theta_2 = 0.4857, \theta_3 = 0.4929, \theta_4 = 0.4823$, and finally the combination weight as

$$\boldsymbol{\omega}_c = (0.2436,\ 0.1455,\ 0.2062,\ 0.0623, 0.1440,\ 0.1984)^T.$$

[21] provides a method to determine θ_l according to the consistency degree between any two ranking vectors generated by the two weight ones. Supposing the two ranking vectors as $p^{(t)} = (p_1^{(t)}, p_2^{(t)}, \cdots, p_1^{(t)})$, $p^{(k)} = (p_1^{(k)}, p_2^{(k)}, \cdots, p_1^{(k)})$ according respectively to the *t*-th and *k*-th weighting methods, the consistency degree in the light of Speaman rank correlation coefficient can be defined as:

$$\rho_{kt} = 1 - \frac{6}{n(n^2-1)}\sum_{j=1}^{n}(p_j^{(k)} - p_j^{(t)})^2\,, \tag{18}$$

where *n* is the number of attributes.

And then the consistency degree of the *k*-th weighting method can be averaged as:

$$\rho_k = \frac{1}{l-1}\sum_{t=1,t\neq k}^{l}\rho_{kt}, \tag{19}$$

θ_k can be calculated as:

$$\theta_k = \rho_k / \sum_{k=1}^{l}\rho_k . \tag{20}$$

With the method in [21], the combination weight

$\hat{\boldsymbol{\omega}}_c = (0.2364, 0.1307, 0.2086, 0.0650, 0.1520, 0.2072)^T$

can be obtained. The Speaman rank correlation coefficient between $\hat{\boldsymbol{\omega}}_c$ and $\boldsymbol{\omega}_c$ is 0.9429, implying strong correlation and consistency degree between the two methods.

According to Table 3, the positive reference sequence is $\mathbf{b}_o^+$, $\mathbf{b}_o^+ = (1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000)$, and the negative reference sequence, $\mathbf{b}_o^-$, $\mathbf{b}_o^- = (0.1500, 0.2727, 0.1933, 0.8319, 0.1933, 0.1933)$.

According to Equations (11)-(15), the $\gamma_{oi}^+, \gamma_{oi}^-, and\ \gamma_{oi}$ are respectively calculated shown in Table 4. It can be concluded from Table 4 that in the light of $\gamma_{oi}^+, or\ \gamma_{oi}$ the material 1 is the best choice, followed by material 2, material 3, and material 4. The methods, weighted arithmetic averaging (WAA) and technique for order preference by similarity to ideal solution (TOPSIS), are widely used two ones in multi-attributes decision making, the results of which are also tabulated in Table 4. It can be seen that whether using the WAA or the TOPSIS the ranking results is the same, which proves selecting the *Zchsn3* for the bearing is trustworthy and reliable.

5.3 RELATIVE GREY RELATION GRADE

TABLE 4 Ranking results

No.	Materials	γ_{io}^+	Rank	γ_{io}^-	Rank	γ_{io}	Rank	WAA	Rank	TOPSIS	Rank
1	Zchsn3	0.7121	1	0.6159	1	0.5363	1	0.6759	1	0.5179	1
2	Zchpb1	0.6511	2	0.6783	3	0.4913	2	0.5993	2	0.5164	2
3	ZQpb30	0.5996	3	0.6695	2	0.4725	3	0.5775	3	0.4910	3
4	ZznM7-5	0.5964	4	0.7363	4	0.4475	4	0.5258	4	0.4519	4

6 Conclusions

From the performed research work using relative grey relation degree combined with combination weights, the main conclusions and directions for future research can be summarized as follows:

1) The proposed method expresses qualitative attributes with trapezoidal fuzzy number, more in conformity with actual situations than the traditional method with exact numbers.
2) Weights play a very significant role in the ranking results of the materials, Combination weights based on maximum deviation both considering subjective and objective weights are more appropriate for determining criteria weights than the subjective or objective weights alone. In comparison with the results by the method in literature [21], the speaman rank correlation coefficient is 0.9429, further implying the proposed method of computing combination weights is feasible.
3) The result using the proposed method in this paper matches well with the other methods such as WAA, TOPSIS, further indicating its feasibility and validity.

References

[1] Lennart Y L 2007 Materials selection and design for development of sustainable products *Materials and Design* **28** 466-79
[2] Ashby M F 2000 Multi-objective optimization in material design and selection *Acta Materialia* **48** 359-69
[3] Deng Y M, Edwards K L 2007 The role of materials identification and selection in engineering design *Materials and Design* **28** 131-9
[4] Zhou C C, Yin G F, Hu X B 2009 Multi-objective optimization of material selection for sustainable products: Artificial neural networks and genetic algorithm approach *Materials and Design* **30** 1209-15
[5] Khabbaz R S, Manshadi B D, Abedian A, Mahmudi R 2009 A simplified fuzzy *Materials and Design* for materials selection in mechanical engineering design *Mater Des* **30** 681-9
[6] Lin K P, Ho H P 2012 Combining fuzzy weight average with fuzzy inference system for material substitution selection in electric industry *Computers & Industrial Engineering* **62** 1034-45
[7] Sharif Ullah A M M, Harib H K 2008 An intelligent method for selecting optimal materials and its application *Advanced Engineering Informatics* **22** 473-83
[8] Ramalhete P S, Senos A M R, Guiar C A 2010 Digital tools for material selection in product design *Materials and Design* **31** 2275-85
[9] Deng J L 1989 Introduction of grey system *Journal of Grey Systems* **1** 1-24
[10] Lu H S, Chang C K, Hwang N C 2009 Grey relational analysis coupled with principal component analysis for optimization design of the cutting parameters in high-speed end milling *Journal of Materials Processing Technology* **209** 3808-17
[11] Liu X Y, Ju Y B, Wang A H 2012 A multiple attribute decision making method with its application to emergency alternative assessment. *Journal of Convergence Information Technology* **7** 75-82
[12] Tseng S F, Lin J H 2012 Evaluating information technology impact on business performance of biotechnology industry using grey relation entropy analysis *Journal of Convergence Information Technology* **14** 297-304
[13] Chan J W K, Tong T K L 2007 Multi-criteria material selection and end-of-life product strategy: Grey relational analysis approach. *Materials and Design* **28** 1539-46
[14] Zhao R, Neighbour G, Deutz P, McGuire M 2012 Materials selection for cleaner production: An environmental evaluation approach *Materials and Design* **37** 429-34

[15] Jahan A, Mustapha F Z, Sapuan S M, Ismail Y, Bahraminasab M 2012 A framework for weighting of criteria in ranking stage of material selection process *International Journal of Advanced Manufacturing Technology* **58** 411-20
[16] Chen H Y 2004 Combination weights based on maximum deviation in multiple attribute decision making *Systems Engineering and Electronics* **26** 194-7
[17] Wei G W 2012 Hesitant fuzzy prioritized operators and their application to multiple attribute decision making. *Knowledge-Based Systems* **31** 176-82
[18] Ye J 2010 Multi-criteria fuzzy decision-making method using entropy weights-based correlation coefficients of interval-valued intuitionistic fuzzy sets *Applied Mathematical Modelling* **34** 3864-70
[19] Bevilacqua M, Braglia M 2000 The analytic hierarchy process applied to maintenance strategy selection *Reliability Engineering and System Safety* **70** 71-83
[20] Yue R, Wang Z B 2012 Multi-attribute group decision making based on set pair analysis. *International Journal of Advancements in Computing Technology* **4** 205-13
[21] Xu J X, Liu Y J, Cai H P 2005 Methods of the linear combination weighting for multiple attribute decision making Journal of National University of Defense Technology 27 121-4 (*in Chinese*)

Authors

Xiqin Wen, born in September 1962, Hanchuan County, Hubei Province, P.R. China.

Current position, grades: Professor of Engineering Training Center, Huaihai Institute of Technology, China.
University studies: MSc at China University of Mining and Technology in 1997, PhD in 2006.
Scientific interests: monitor for dynamic characteristics and fault diagnosis in mechanical engineering.
Publications: More than 30 papers.
Experience: teaching and scientific research experience of more than 30 years, more than 10 scientific research projects.

Kaibo Wu, born in September 1989, Luoyang County, Henan Province, P.R. China.

Current position: Master Degree Candidate in China University of Mining and Technology, China.
University studies: BSc in Mechanical Design, Manufacturing and Automation at Zhongyuan University of Technology in 2013.
Scientific interests: digital and intelligent design and manufacturing.

Anhua Peng, born in January 1973, Jian County, Jiangxi Province, P.R. China.

Current position, grades: associate professor of Engineering Training Center, Huaihai Institute of Technology, China.
University studies: MSc at Yangzhou University in 2008.
Scientific interests: Multi-criteria decision making and its applications to advanced manufacturing technology.
Publications: more than 50 papers.
Experience: teaching and scientific research experience of about 20 years.

An analysis of landscape characteristics of urban rivers – with the case of Chengdu

Jian Zhang, Zhongwei Shen, Hao Shen*

School of Architecture, Southwest Jiaotong University, Chengdu, Sichuan, China

Received 1 September 2014, www.cmnt.lv

Abstract

Urban river landscape is a key ecological regulating system in cities, attracting increasing attention from people. Our research conducted a systematic and comprehensive research of landscape characteristics of main rivers in Chengdu from some new perspectives, including the city in the large, city location and Single River. This research is intended to better understand the landscape structure along main rivers in Chengdu, to facilitate the evaluation of macro-regional ecological environment, and to provide references for the study of regional ecological environment.

Keywords: urban rivers, landscape characteristic, city location, tree crown sizes, canopy contribution rate, landscape area

1 Background research

Since the birth of city, it has formed irreconcilable opposite to the natural ecosystem [1], which is rooted in the absolute destruction and demand of ecosystem during the urban development. The urban ecologicalization is an inexorable trend in urban development, as well as a new trend of world urban development [2]. Urban rivers are key constituent parts in the urban ecosystem. These urban rivers provide important water sources and material transport channels for cities, which also enhance the diversity of urban landscape and enrich the lives of residents. Urban rivers prepare a vital material foundation for the comfort, stability and sustainability of cities.

Urban rivers indicate rivers or river sections originated from urban areas or running through urban areas, including some historical canals excavated by human but now showing characteristics of natural rivers after years of evolution [3]. Human activities intensely influence the hydrological characteristics, physical structure and ecological environment of urban rivers. Besides, the socio-economic system and daily life of residents in cities depend on the benefits of urban rivers. Compared with natural rivers, the interaction between human and urban rivers is very significant [4-10]. Gallery of urban rivers are vital multi-functional service body in urban landscape. It can be defined as line and area combination zone connecting the aquatic-terrestrial ecotone at both banks in the urban area and forest cover with forest structure and functions. It includes the vegetation on the flood plain in streamside space of cities, in the protection forest of riparian zone, and in a certain area from the bunding. According to the widely accepted landscape type rating, waterfront space with natural banks, meandering shoreline, open sight, diverse space activities, and pseudo-classic architecture is relatively ideal. (1) According to the comprehensive rating of "Eight Scenes of Shahe River", from the high to low rating, Xinlv Grinding Grain (Z2)> North Lake(Z1)> Sandong Ancient Bridge (Z3)>Mashi Stones (Z5)> the Hakkas in Shahe River (Z6)>Spring in Tashan Mountain (Z7)>Cuihu Lake (Z8)>Technology Garden (Z4).

2 Research method and range

In recent years, the research of landscape characteristics of urban rivers has become a focus in the research area of landscape ecology and global change. It is hard to obtain high-resolution aerial photos so that Google Earth map was used in our research to replace aerial photos as the analysis data. From 500m at the overhead view (a comparatively optimal position in terms of the overhead view contour sharpness of canopy and river landscape patch), a total of 700 screenshots were used. These photos were processed in Photoshop, and generated the remote-sensing image of rivers in Chengdu. Next, the boundary of canopy and river landscape patch was drawn. A new file folder was created in ArcCatolog to establish personal data. In Arcmap, diagrams in Google Earth were used as base maps so as to establish canopy cover shape pictures with meter as the unit. These files were saved in the form of tables for further calculation and analysis. This research focused on the analysis of similarities and differences of landscape characteristics in the above mentioned areas. The landscape index and canopy cover rate of vegetables along rivers were used as major reference indexes. Indexes including landscape characteristics to the highest degree were selected to demonstrate the landscape condition in the research area so that the landscape structure along main rivers in Chengdu could be under-

* ***Corresponding author's*** e-mail:737006315@qq.com

stood. This could facilitate the evaluation of the macroscopic regional ecological environment, and could provide references for the ecological environment and management of the research area. Statistics of the index rating rsi are shown in Table 1:

TABLE 1 Statistics of index rating rsi

Index	Very Satisfactory (V1)	Satisfactory (V2)	Average (V3)	Unsatisfactory (V4)	Extremely Unsatisfactory (V5)
A	0.1080	0.3991	0.4554	0.0376	0.0000
B	0.1174	0.5258	0.2817	0.0751	0.0000
C	0.0845	0.3052	0.4272	0.1080	0.0751
D	0.1737	0.5070	0.2535	0.0376	0.0282
E	0.2911	0.5164	0.1268	0.0469	0.0188
F	0.1362	0.4178	0.3474	0.0986	0.0000
G	0.1502	0.4836	0.2817	0.0845	0.0000
H	0.1174	0.2441	0.4554	0.1643	0.0188
I	0.0563	0.2160	0.4554	0.2441	0.0282
J	0.0751	0.3709	0.4272	0.1174	0.0094
K	0.0798	0.4085	0.4554	0.0376	0.0188
A	0.1080	0.3991	0.4554	0.0376	0.0000

Waterfront vegetation rating was satisfactory, which was larger than 4.0. Ratings of naturalness of river banks, the breadth of vision, the degree of bending, river bank hydrophilia and river flow were also satisfactorily larger than 3.5. The ratings of other indexes were between 3.0 and 3.5.

The overall evaluation results of waterfront landscape reached a rating larger than 3.5. This was satisfactory.

According to the index of criterion layer, the ratings of indexes can be prioritized from high to low as: Natural Ecology>Entertainment and Ornamental Function>Social Environment. Among them, ratings of natural ecology and entertainment and ornamental function exceeded 3.5, which were satisfactory. The rating of social environment was average.

In this research, according to landscape characterristics, the entire research area were further divided into: inside 1st Ring Rd, 1st Ring Rd to 2nd Ring Rd, 2nd Ring Rd to 3rd Ring Rd, outside the 3rd Ring Rd. Besides, according to the regional relation of Chengdu, Chengdu was divided with Tianfu Square as the center into four research areas, including northeast area, northwest area, southwest and southeast areas.

The range of research areas: the east, north and west boundaries were ring roads of Chengdu. Due to the migration of city center towards the south, the south boundary was the intersection between Fuhe River and JiangAnhe River in Huayang Town, including the urban area of Chengdu. Rivers in our study were 4 main rivers with the greatest influence on the urban area of Chengdu. The average monthly designed flow in different water periods is shown in Table 2:

TABLE 2 Average monthly designed flow in different periods, Units: (m^3/s)

`	Dry Season	Prior Normal Water Period	Wet Season	Post Normal Water Period
Qingshuihe River (Supo Bridge)	1.1	12.4	35.3	19.9
Nanhe River (Baihuatan Bridge)	1.5	12.1	38.4	22.3
Shahe River (Gaoqiao Bridge)	19.6	23.5	52.3	37.2
Shahe River (Gate of Dongzikou)	15.0	14.8	17.4	16.6
Fuhe River (Branch of Dongzikou)	4.6	8.6	34.9	20.6

The river landscape model is shown in Equation (1):

$$NMSE = \frac{1}{N\sigma^2}\sum_{K=1}^{N}\left[x(k)-\hat{x}(k)\right]^2 . \quad (1)$$

3 Specific Research

This research first explored landscape characteristics of the entire research area. On this basis, the research area was divided into: inside 1st Ring Rd, 1st Ring Rd to 2nd Ring Rd, 2nd Ring Rd to 3rd Ring Rd and outside the 3rd Ring Rd. Next, with Tianfu Square as the center, Chengdu was divided into northeast area, northwest area, southwest and southeast areas. Finally, each river was systematically researched.

3.1 ANALYSIS OF THE ENTIRE RESEARCH AREA

Comparison of the canopy cover rate of each river in the research area is shown in Table 3. According to Table 3, it is known that the total canopy cover area of trees of urban river landscape in the research area was 828 hm2, and the canopy cover rate was 45.2%.Within the research area, the difference of canopy cover rates of each river in the research area was significant: Shahe River showed the highest canopy cover rate of 63.52%, which was mainly attributed to the role of Shahe River as the protected water source in Chengdu. Therefore, the ecological environment and the surrounding green space system of Shahe River are the best in Chengdu, thus generating the highest tree coverage rate of landscape around the river. It was followed by 46.8% (Fuhe River), 45.17% (Qing-

shuihe River and Modihe River), 44.36% (Dongfeng Canal) and 31.8% (JiangAnhe River). In terms of the comparison of canopy cover area, Fuhe River has the largest canopy cover area, followed by Dongfeng Canal > Shahe River > JiangAnhe River > Qingshuihe River and Modihe River.

TABLE 3 Comparison of canopy cover rates in each river of the research area

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area of All Rivers (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in Landscape Area of All Rivers (%)
Fuhe River	635.1599	34.66	297.1229	46.8	35.87
Qingshuihe River	148.3505	8.09	67.012	45.17	8.09
Shahe River	223.4577	12.19	141.9364	63.52	17.14
JiangAnhe River	350.7365	19.14	111.5229	31.8	13.46
Dongfeng Canal	475.0198	25.92	210.7339	44.36	25.44
Total	1832.7244	100	828.3281	45.2	100

The data model is shown in Equation (2):

$$\hat{x}_{(k+1)}^{(1)} = \left(x_{(1)}^{(0)} - \frac{u}{a} \right) e^{-ak} + \frac{u}{a}. \tag{2}$$

Subsequently, the contribution level of canopy cover of each river was compared. Shahe River contributes a canopy cover rate of 17.14% with 12.19% land area, followed by the canopy cover rate of Fuhe River, Qingshuihe River and Modihe River, Dongfeng Canal (35.87%, 8.09% and 25.44%) with a land area of 34.66%, 8.09% and 25.92% respectively. Lastly, JiangAnhe River contributes a canopy cover rate of 13.46% with a land area of 19.14%. The control rate of fracture surface is shown in Table 4:

TABLE 4 Control rate of fracture surface at different periods

Fracture Surface	Dry Season	Prior Normal Water Period	Wet Season	Post Normal Water Period
Fuhe River Entry	√	√	√	√
Xibei Bridge	√	√	√	√
DaAn Bridge	√	√	√	√
Shahe River Entry	√	√	√	√
Gantachang	√	√	√	√
Shanban Bridge	√	√	√	√
Chengren Bridge	√	√	√	√
Baihua Bridge	×	√	√	√
Anshun Bridge	×	√	√	√
Hejiangting	×	√	√	√
YongAn Bridge	×	√	√	√
Zhonghe	×	×	√	√
Yuzui	×	×	×	×
Huanglongxi Brook	×	×	×	×
Control Rate of Fracture Surface	50%	79%	85.7%	85.7%

3.2 ANALYSIS OF FOUR MAIN RESEARCH AREAS

The comparison of the canopy cover rate of river landscape in each ring road of the research area is shown in Table 5. According to Table 5, the areas with the canopy cover rate ranking from high to low are respectively: 2nd Ring Rd to 3rd Ring Rd, 1st Ring Rd to 2nd Ring Rd, inside 1st Ring Rd and outside the 3rd Ring Rd. Among them, the canopy cover rates of 2nd Ring Rd to 3rd Ring Rd and that of 1st Ring Rd to 2nd Ring Rd were similarly highest. This could be attributed to the distribution of dense resident complexes near rivers from 2nd Ring Rd to 3rd Ring Rd and from 1st Ring Rd to 2nd Ring Rd. Therefore, the mode of landscape was relatively stable. Besides, due to the requirements of residents on river landscape, the landscape was diverse. However, the canopy cover rate inside 1st Ring Rd was lower than the above two areas because the area inside 1st Ring Rd was the urban center. The rivers are mainly surrounded by commercial lands or lands for other purposes, with huge visitors' flow rate. Since the human activity imposes huge demand on river landscape, the area of green space was narrowed, and the canopy cover rate became smaller. Moreover, the area outside the 3rd Ring Rd was close to the nature and the density of vegetation was smaller than that of the urban area, and the cover rate of trees was the lowest.

TABLE 5 Comparison of canopy cover rate of river landscape in each ring road of our research area

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area of All Rivers (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas (%)
Inside 1st Ring Rd	56.9955	3.11	25.0592	43.97	3.02
1st Ring Rd to 2nd Ring Rd	84.9036	4.33	45.0156	53.02	5.44
2nd Ring Rd to 3rd Ring Rd	432.8825	23.62	230.3341	53.21	27.81
Outside 3rd Ring Rd	1258.0248	68.64	527.9192	41.96	63.73
Total	1832.7244	100	828.3281	45.2	100

In terms of contribution rate: the area from 1st Ring Rd to 2nd Ring Rd had the highest contribution rate, which contributes a canopy cover rate of 5.44% with a land area of 4.33%. This was mainly caused by the contribution of large ecological gardens like Huahuaxi Park. This was followed by that from 2nd Ring Rd to 3rd Ring Rd, which contributed a canopy cover rate of 27.81% with a land area of 23.62%. This was mainly caused by the considerable number of ecological wetland parks in the flow area of Shahe River from 2nd Ring Rd to 3rd Ring Rd. The ecological wetland parks remarkably improved the tree coverage in this area.

It can be seen from the results that the ratings of evaluation indexes of landscape types prioritized from high to low are as follows: (1) river flow (A): A3 gentle slope waterfront for people to enjoy > A1 water level lower than the river bank by 0-2m > A2 water level lower than the river bank by 2 – 5m; the rating model of river landscape evaluation indexes is shown in Equation (3):

$$\hat{x}^{(0)}(k) = (x^{(0)}_{(1)} + \frac{u}{a})e^{ak} - \frac{u}{a}. \quad (3)$$

Degree of river bending (B): B3 relatively meandering > B2 slightly meandering > B1 straight;

Environmental quality of waterfront (C): C1 Waterfront vegetation width larger than 50m> C2 waterside vegetation width between 20 and 50m > C3 waterside vegetation width less than 20m;

Naturalness of River Banks (D): D3 natural and Nature-Imitated banks > D2 stacking of natural stones > D1 grass planted on artificial banks;

Waterfront vegetation (E): E3 mainly grass + shrub + tree > El mainly grass + tree > E2 mainly grass + shrub;

Hydrophily of river banks (F): F2 with fishing available along rivers > F3 with viewing platform > F1 with walking along banks allowed;

Next, the area inside 1st Ring Rd and outside the 3rd Ring Rd contributed a canopy cover rate of 3.02% and 63.73% respectively with a land area of 3.11% and 68.64%. The data model of waterside vegetation and water flow is shown in Equation (4):

$$X^{(0)}_{(t)} = \left\{x^{(0)}_{(1944)}, x^{(0)}_{(1945)}, ..., x^{(0)}_{(1990)}\right\} = \{261, 465, ..., 432\}. \quad (4)$$

3.2.1 Comparison of canopy cover rates of different rivers inside 1st Ring Rd

The comparison of canopy cover rates of different rivers inside 1st Ring Rd is shown in Table 6. According to Table 6, it can be found that the canopy cover rate of Fuhe River, Qingshuihe River and Modihe River inside 1st Ring Rd was almost identical, so was their contribution to the canopy cover rate. Fuhe River, Qingshuihe River and Modihe River inside 1st Ring Rd contributed a canopy cover rate of 56.74% and 43.26% with a land area of 55.83% and 44.17% respectively. This indicates that the establishment and operation modes of these river landscapes inside 1st Ring Rd were almost identical. Meanwhile, a limited canopy cover rate also reflects the large demand from human activity, which is closely related with the land nature of commercial space around rivers.

TABLE 6 Comparison of canopy cover rate of different rivers inside 1st ring rd of the research area

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area Inside 1st Ring Rd (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas Inside 1st Ring Rd (%)
Fuhe River	31.8389	55.83	14.2292	44.62	56.74
Qingshuihe River and Modihe River	25.1566	44.17	10.83	43.07	43.26
Total	56.9955	100	25.0592	43.92	100

3.2.2 Comparison of canopy cover rates of different rivers inside 1st Ring Rd to 2nd Ring Rd

The comparison of canopy cover rates of different rivers from 1st Ring Rd to 2nd Ring Rd is shown in Table 7. According to Table 7, within the area from 1st Ring Rd to 2nd Ring Rd, the canopy cover rate of three rivers showed remarkable significance. Fuhe River had the highest canopy cover rate of 60.91% mainly because Fuhe River runs through some urban parks with flourishing vegetations like Wangjiang Park. These parks substantially increased the tree coverage rate of river landscape in this area. This was followed by the tree coverage rate of 50.93% of Shahe River. Compared the

overall tree coverage rate of Shahe River, the tree coverage rate of this section was reduced because it mainly flowed through the resident area of old cities. The high participation rate of citizens influenced the river landscape area of Shahe River to a certain degree. Next, the tree coverage rate of Qingshuihe River and Modihe River reached 47.84. In terms of tree coverage contribution rate, Fuhe River, Shahe River, Qingshuihe River and Modihe River contributed a tree coverage rate of 04%, 33.34% and 30.62% respectively with a land area of 31.32%, 34.74% and 33.94%.

TABLE 7 Comparison of canopy cover rate of different rivers from 1st ring rd to 2nd ring rd of the research area

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area from 1st Ring Rd to 2nd Ring Rd (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas from 1st Ring Rd to 2nd Ring Rd (%)
Fuhe River	26.6437	31.32	16.228	60.91	36.04
Qingshuihe River and Modihe River	28.8636	33.94	13.797	47.84	30.62
Shahe	29.3963	34.74	14.9906	50.93	33.34
Total	84.9036	100	45.0156	53.04	100

3.2.3 Comparison of canopy cover rates of different rivers inside 2nd Ring Rd to 3rdc Ring Rd

The comparison of the canopy cover rate of different rivers from 2nd Ring Rd to 3rd Ring Rd is shown in Table 8. According to Table 8, rivers with the tree coverage rate ranking from high to low are Shahe River (65.42%), Dongfeng Canal (45.47%), Qingshuihe River and Modihe River (42.07%), and Fuhe River (41.91%). The tree coverage rate of Shahe River was the highest in this section, which was mainly attributed to the large number of ecological wetland parks and urban parks at both sides of it, as well as the extensive green belts. In terms of tree coverage contribution rate, Shahe River topped by contributing a tree coverage rate of 55.11% with a land area of 44.84%, followed by that of Fuhe River, Qingshuihe River and Modihe River as well as Dongfeng Canal. They contributed a tree coverage rate of 19.84%, 6.93% and 18.12% with a land area of 25.2%, 8.76% and 21.2%.

TABLE 8 Comparison of canopy cover rate of different rivers from 2nd ring rd to 3rd ring rd of the research area

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area from 2nd Ring Rd to 3rd Ring Rd (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas from 2nd Ring Rd to 3rd Ring Rd (%)
Fuhe River	109.0456	25.2	45.7025	41.91	19.84
Qingshuihe River and Modihe River	37.9232	8.76	15.9572	42.07	6.93
Shahe	194.0614	44.84	126.9458	65.42	55.11
Dongfengqu Canal	91.7703	21.2	41.7286	45.47	18.12
Total	432.8005	100	230.3341	53.21	100

3.2.4 Comparison of canopy cover rates of different rivers outside 3rdc Ring Rd

The comparison of canopy cover rates of different rivers outside the 3rd Ring Rd is shown in Table 9. According to Table 9, the tree coverage rate was realized from high to low by Fuhe River, Qingshuihe River and Modihe River, Dongfeng Canal, as well as JiangAnhe River. The tree coverage contribution rate was realized from high to low by Qingshuihe River and Modihe River, Fuhe River, Dongfeng Canal and JiangAnhe River. They contributed a tree coverage rate of 5.01%, 41.86%, 32.01% and 21.12% with a land area of 4.48%, 37.17%, 30.46% and 27.89% respectively. This indicated that the urbanization level of areas near Fuhe River, Qingshuihe River and Dongfeng Canal was comparatively consistent. Meanwhile, the overall urbanization level of areas near JiangAnhe River was lower than that of the other three rivers.

TABLE 9 Comparison of canopy cover rate of different rivers outside the 3rd ring rd of the research area

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area outside 3rd Ring Rd (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas outside 3rd Ring Rd (%)
Fuhe River	467.6317	37.17	220.9632	47.25	41.86
Qingshuihe River and Modihe River	56.4071	4.48	26.4278	46.85	5.01
Shahe	0	0	0	0	0
Dongfengqu Canal	383.2495	30.46	169.0053	44.1	32.01
JiangAnhe River	350.7365	27.89	111.5229	31.8	21.12
Total	1258.0248	100	527.9192	41.96	100

3.3 ANALYSIS OF REGIONAL LOCATION OF CHENGDU

In this research, Chengdu was divided with Tianfu Square as the center into four research areas, including northeast area, northwest area, southwest and southeast areas. Besides, this division was integrated with 1st Ring Rd, 2nd Ring Rd, 3rd Ring Rd, outside the 3rd Ring Rd to analyze the river landscape of Chengdu. The comparison of the canopy cover rate of river landscapes at different locations is shown in Table 10. According to Table 10, the changing patterns of canopy cover rates in northwest area, northeast area and southeast area are: the canopy cover rate was the highest from 2nd Ring Rd to 3rd Ring Rd, followed by that in the area from 1st Ring Rd to 2nd Ring Rd, and that inside 1st Ring Rd and outside the 3rd Ring Rd. This was consistent with the previous conclusion without the division of directions. The southwest area, however, was an exception. In this area, the canopy cover rate was highest in the area from 1st Ring Rd to 2nd Ring Rd, followed by that inside 1st Ring Rd, the area from 2nd Ring Rd to 3rd Ring Rd and finally the area outside the 3rd Ring Rd. This was mainly attributed to the Huanhuaxi park in the section of Qingshuihe River and Modihe River. Huanhuaxi Park is located in the area between 1st Ring Rd and 2nd Ring Rd in the southwest, thus substantially increasing the canopy cover rate of the area. Around the rivers from 2nd Ring Rd to 3rd Ring Rd in this area, there were the main resident complexes of Chengdu, with large density of populations. Residents present strong demand on the activities in the river landscape. Hence, the river landscape area was comparatively narrow, and the canopy cover rate was lower than other areas. Variance analysis of location difference rating is shown in Table 11.

TABLE 10 Comparison of canopy cover rate of river landscape in different regions

Type	Within 1st Ring Rd		1st Ring Rd to 2nd Ring Rd		2nd Ring Rd to 3rd Ring Rd		Outside 3rd Ring Rd	
	Area (hm2)	Canopy Cover Rate (%)	Area (hm2)	Canopy Cover Rate (%)	Area (hm2)	Canopy Cover Rate (%)	Area (hm2)	Canopy Cover Rate (%)
Northwest Area	4.4298	43.91	18.6901	47.75	94.3988	49.84	362.419	42.83
Northeast Area	18.3359	44.01	29.3963	50.99	165.0332	55.46	150.9815	43.52
Southwest Area	14.9724	43.52	23.4515	62.55	18.0612	41.52	383.791	39.95
Southeast Area	19.2574	44.28	13.3657	48.13	155.3073	54.26	360.8333	42.58
Total	56.9965		84.9036		432.8825		1258.0248	

TABLE 11 Variance analysis of location difference rating

Evaluation Index	Near Shahe River	Urban Area of Chengdu	Other Cities	F Value	Significance Probability	Multiple Comparison Appraisal
River Flow	3.56	3.57	3.63	0.332	0.717	(n,s)
Bending Degree of River	3.71	3.68	3.66	0.128	0.880	(n,s)
Waterside Environment Quality	2.95	3.35	3.41	10.017	0.000 ¨	(1,2)(1,3)
Hydrophily of Banks	3.80	3.73	3.73	0.330	0.719	(n,s)
Breath of Vision	3.92	4.08	4.08	1.701	0.184	(n,s)
Waterside Architectural Style	3.49	3.66	3.67	2.329	0.099	(n,s)
Waterside Land Utilization	3.63	3.76	3.72	1.020	0.362	(n,s)
Traffic	3.15	3.38	3.33	2.477	0.085	(1,2)
Safety	2.85	3.09	3.19	6.005	0.003 ¨	(1,2)(1,3)
Total Average	3.13	3.52	3.59	14.971	0.000 ¨	(1,2)(1,3)

As can be drawn from Table 11, with different distances from the river, significant differences were observed among the waterside environment quality, waterside land utilization and traffic indexes, showing a small degree of influence (3/12 = 25%). Our assumption that the difference of resident locations would influence the overall evaluation of waterside landscape was partially established. We conducted multiple comparisons through the L.S.D least significant difference method, and obtained that differences of the three indexes were mainly reflected between areas near river, urban area of Chengdu and other cities. According to the above indexes, the rating of population living near the river was the lowest. It was because overall the waterside landscape had a width smaller than 50m, except the large width of vegetation buffer area. A large part of the river sections between 2nd Ring Rd and 1st Ring Rd was less than 20m wide. Its inner relation can be expressed by Equation (5):

$$\bar{A} = \sum_{i=1}^{M} \left| q^{(0)}(j) \right| M , \tag{5}$$

In terms of changes in the river patch area, changes at four directions were different. In the northwest, the river patch area increased successively from inside the city to outside the city, which was consistent with our previous conclusion without the division of directions. In the northeast, the river patch area increased successively from 1st Ring Rd to 3rd Ring Rd. However, compared with the patch area from 2nd Ring Rd to 3rd Ring Rd, the patch area outside the 3rd Ring Rd decreased again. This could be attributed to the main drainage basin of Shahe River that is located in the area from 2nd Ring Rd to 3rd Ring Rd. This substantially increased the river landscape area of the

area. In the southwest, the overall trend is consistent with our previous findings. This was because the river flew through the main resident complexes of Chengdu in the area from 2nd Ring Rd to 3rd Ring Rd. Residents' demand on activities dominated so that the river landscape area was limited and the changes in the trend were produced. In the southeast, in addition to a decreasing trend in the area from 1st Ring Rd to 2nd Ring Rd, trends in other sections were increasing. The declining trend from 1st Ring Rd to 2nd Ring Rd was attributed to the geological location of the river section, which is close to key roads of Chengdu and the campus of Sichuan University. In order to ensure the smooth traffic and sufficient campus area, the river landscape area was greatly reduced.

The comparison of river landscape patch number in different directions and the average patch area is shown in Table 12. According to Table 12, in northwest, the patch number decreased from the area outside the 3rd Ring Rd, from 2nd Ring Rd to 3rd Ring Rd, from 1st Ring Rd to 2nd Ring Rd and the area inside 1st Ring Rd. The average area of patch decreased successively from the area outside the 3rd Ring Rd, the area inside 1st Ring Rd, from 2nd Ring Rd to 3rd Ring Rd, and from 1st Ring Rd to 2nd Ring Rd. In northwest, the patch number and patch area fell from the area outside the 3rd Ring Rd, from 2nd Ring Rd to 3rd Ring Rd, inside 1st Ring Rd, and from 1st Ring Rd to 2nd Ring Rd; from 2nd Ring Rd to 3rd Ring Rd, outside the 3rd Ring Rd, 1st Ring Rd to 2nd Ring Rd, inside 1st Ring Rd; In southwest, the patch number and patch area fell from the area outside the 3rd Ring Rd, 1st Ring Rd to 2nd Ring Rd, inside 1st Ring Rd, 2nd Ring Rd to 3rd Ring Rd; outside the 3rd Ring Rd, 2nd Ring Rd to 3rd Ring Rd, 1st Ring Rd to 2nd Ring Rd, inside 1st Ring Rd; In southeast, the patch number and patch area fell from the area outside the 3rd Ring Rd, 2nd Ring Rd to 3rd Ring Rd, inside 1st Ring Rd, 1st Ring Rd to 2nd Ring Rd; outside the 3rd Ring Rd, 2nd Ring Rd to 3rd Ring Rd, 1st Ring Rd to 2nd Ring Rd and the area inside 1st Ring Rd. Thus, it can be concluded that the patch number in northwest progressively increased from the urban area to external areas. In northwest, the patch area was smallest in the urban area and larger in external areas. In northeast, the patch number progressively increased from 1st Ring Rd and the urban area to external areas. The patch area progressively increased from the area inside 1st Ring Rd to 3rd Ring Rd. Area outside the 3rd Ring Rd had smaller patch area than that from 2nd Ring Rd to 3rd Ring Rd. Compared with the situation in southeast, these trends were similar in southwest. They both presented a smallest number of patches in urban area, which was smaller than external areas. The changing trend of patch areas was the progresssive increase from the urban center to the external.

TABLE 12 Comparison of patch number and average patch area in different regions of the research area

Type	Within 1st Ring Rd		1st Ring Rd to 2nd Ring Rd		2nd Ring Rd to 3rd Ring Rd		Outside 3rd Ring Rd	
	Average Patch Area (hm2)	Patch Number	Average Patch Area (hm2)	Patch Number	Average Patch Area (hm2)	Patch Number	Average Patch Area (hm2)	Patch Number
Northwest Area	1.4766	3	0.49	38	1.15	82	2.88	126
Northeast Area	0.87	21	1.55	19	3.93	42	2.8	54
Southwest Area	1.07	14	1.15	21	3.01	6	4.41	87
Southeast Area	0.71	27	1.34	10	4.85	32	10.31	35

Comparison of patch density in different regions of the research area is shown in Table 13. According to Table 13, through the calculation of the river landscape area in per square kilometer, we obtained the size of river landscape in different locations of Chengdu and their trend of variation. In northwest, the patch density decreased from the area from 2nd Ring Rd to 3rd Ring Rd, outside the 3rd Ring Rd, 1st Ring Rd to 2nd Ring Rd and the area inside 1st Ring Rd. In northeast, the patch density decreased from the area from 1st Ring Rd to 2nd Ring Rd, 2nd Ring Rd to 3rd Ring Rd, inside 1st Ring Rd and outside the 3rd Ring Rd. In southwest, the patch density decreased from the area from 1st Ring Rd to 2nd Ring Rd, outside the 3rd Ring Rd, inside 1st Ring Rd and from 2nd Ring Rd to 3rd Ring Rd. In southeast, the patch density decreased successively from the area outside the 3rd Ring Rd, from 2nd Ring Rd to 3rd Ring Rd, inside 1st Ring Rd, and from 1st Ring Rd to 2nd Ring Rd. It can be concluded that only the patch density in the area outside the 3rd Ring Rd in northwest declined. In other areas, the patch density generally increased from the urban area to the external areas. In northeast, the patch density was high in the middle and lower at both ends. The patch density of the inner city area was higher than that of the external areas. In southwest, the patch density was lower in the area from 2nd Ring Rd to 3rd Ring Rd, and the patch density of other sections remained similar. In southeast, the patch density in the area from 1st Ring Rd to 2nd Ring Rd was slightly lower but presented an increasing trend from urban area to external areas.

TABLE 13 Comparison of patch density in different regions of the research area

Type	Within 1st Ring Rd		1st Ring Rd to 2nd Ring Rd		2nd Ring Rd to 3rd Ring Rd		Outside 3rd Ring Rd	
	Average Patch Area (hm2)	Patch Density (hm2/km2)	Average Patch Area (hm2)	Patch Density (hm2/km2)	Average Patch Area (hm2)	Patch Density (hm2/km2)	Average Patch Area (hm2)	Patch Density (hm2/km2)
Northwest Area	4.4298	0.65	18.6901	2.46	94.3988	3.46	362.419	3.01
Northeast Area	18.3359	2.19	29.3963	4.39	165.033	3.81	150.982	1.5
Southwest Area	14.9724	2.52	23.4515	3.07	18.0612	0.74	383.791	3.01
Southeast Area	19.2574	2.91	13.3657	1.43	155.307	4.28	360.833	4.43

3.4 ANALYSIS OF THE TEGIONAL LOCATION OF UNDIVIDUAL RIVER

The comparison of canopy cover rate in each section of Fuhe River landscape is shown in Table 14. In Table 14, the variation patterns of river patch area and canopy cover area were identically increasing from center area to the external areas. The area outside 3rd Ring Rd in the direction of southwest had the highest canopy cover rate, which was mainly caused by the Fuhe River that flew through the new city area of Chengdu. The river banks of the new city area were larger than those of old city areas. Besides, the vegetable types were greatly diverse. Therefore, this area had the highest canopy cover rate, followed by that from 1st Ring Rd to 2nd Ring Rd in southeast, from 3rd Ring Rd to 2nd Ring Rd in northwest, from 2nd Ring Rd to 1st Ring Rd in northwest, inside 1st Ring Rd in southeast, inside 1st Ring Rd in northwest, expressway to 3rd Ring Rd in northwest, outside the 3rd Ring Rd in southeast and the area inside 1st Ring Rd in northeast. We could come up with a conclusion that the area of Fuhe River in southeast had the highest canopy cover rate, followed by that in northwest. Their variation pattern was constant increase of canopy cover rate from the urban center to areas outside the city. The variation pattern of canopy cover rate in another two directions was larger canopy cover rate in external areas than urban areas.

TABLE 14 Comparison of canopy cover rate in the landscape of each section of Fuhe River

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area of Fuhe River (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas of Fuhe River (%)
Expressway to 3rd Ring Rd (Northwest)	221.4725	34.87	93.9964	42.44	31.63
3rd Ring Rd to 2nd Ring Rd (Northwest)	54.9638	8.65	28.1191	51.16	9.46
2nd Ring Rd to 1st Ring Rd (Northwest)	3.1922	0.5	1.5598	48.86	0.52
Inside 1st Ring Rd (Northwest)	4.4298	0.6	1.9452	43.91	0.65
Inside 1st Ring Rd (Northeast)	18.3359	2.89	8.0705	40.15	2.71
Inside 1st Ring Rd (Southeast)	9.0732	1.43	4.2135	46.44	1.41
1st Ring Rd to 2nd Ring Rd (Southeast)	23.4515	3.69	14.6682	62.55	4.93
2nd Ring Rd to 3rd Ring Rd (Southeast)	54.0818	8.52	17.5834	32.51	5.91
Outside 3rd Ring Rd (Southeast)	169.7264	26.72	69.0292	40.67	23.23
Outside 3rd Ring Rd (Southwest)	76.4328	12.03	57.9376	75.8	19.5
Total	635.1599	100	297.1229	46.78	100

Comparison of canopy cover rate in the landscape of each section of Shahe River is shown in Table 15. It can be drawn from Table 15 that river patch area and canopy cover area presented a pattern of increasing from urban center to the area outside the city. The highest canopy cover rate and canopy contribution rate occurred in areas from 3rd Ring Rd to 2nd Ring Rd in northeast and southeast. Both presented a pattern of larger canopy cover rate and canopy contribution rate in external areas than those in urban center.

TABLE 15 Comparison of canopy cover rate in the landscape of each section of Shahe River

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area of Shahe River (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas of Shahe River (%)
3rd Ring Rd to 2nd Ring Rd (Northwest)	19.573	8.76	10.471	53.5	7.34
3rd Ring Rd to 2nd Ring Rd (Northeast)	55.7578	24.95	40.6303	72.87	28.63
2nd Ring Rd to 1st Ring Rd (Northeast)	29.3963	13.16	14.9906	50.99	10.56
2nd Ring Rd to 3rd Ring Rd (Northeast)	31.0193	13.88	14.3813	46.36	10.13
2nd Ring Rd to 3rd Ring Rd (Southeast)	87.7113	39.25	61.4632	70.07	43.34
Total	223.4577	100	141.9364	63.52	100

Comparison of canopy cover rate in the landscape of each section of Dongfeng Canal is shown in Table 16. In Table 16, river patch area and canopy cover area of Dongfeng Canal presented a pattern of successively increasing from urban center to the area outside the city. The canopy cover rate from 3rd Ring Rd to 2nd Ring Rd in northeast was highest, and that from 2nd Ring Rd to 3rd Ring Rd in southeast was lowest. Generally, the canopy cover rate of each section was similar, which could mainly be attributed to the artificial excavation of Dongfeng Canal with basically consistent channel width and river landscape width. Therefore, the canopy cover rate was basically identical.

TABLE 16 Comparison of canopy cover rate in the landscape of each section of Dongfeng Canal

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area of Dongfengqu Canal (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas of Dongfengqu Canal (%)
Expressway to 3rd Ring Rd (Northwest)	41.1611	8.67	18.6568	45.33	8.85
Expressway to 3rd Ring Rd (Northeast)	150.9815	31.78	65.7182	43.53	31.16
3rd Ring Rd to 2nd Ring Rd (Northeast)	78.2561	16.47	36.5122	46.66	17.33
2nd Ring Rd to 3rd Ring Rd (Southeast)	13.5142	2.84	5.2164	38.6	2.48
3rd Ring Rd to Expressway (Southeast)	191.1069	40.22	84.6303	44.28	40.18
Total	475.0198	100	210.7339	44.36	100

Comparison of canopy cover rate in the landscape of each section of Jianganhe River is shown in Table 17. According to Table 17, although the main drainage basin of JiangAnhe River was outside 3rd Ring Rd, its canopy cover rate showed significant difference. This was mainly because Jianganhe River flew through farmlands outside the 3rd Ring Rd, then the towns. Since the tree density of towns is much larger than that in farmlands, this difference has been formed.

TABLE 17 Comparison of canopy cover rate in the landscape of each section of Jianganhe River

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area of JiangAnhe River (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas of JiangAnhe River (%)
Outside 3rd Ring Rd (Northwest)	43.3783	12.38	16. 138	37.2	14.17
Outside 3rd Ring Rd (Southwest)	307.3582	87.62	95.3849	31.03	85.53
Total	350.7365	100	111.5229	31.8	100

Comparison of canopy cover rate in the landscape of each section of Qingshuihe River and Modihe River is shown in Table 18. According to Table 18, river patch area and canopy cover area of Qingshuihe and Modihe River presented a pattern of successively increasing from urban center to the area outside the city. The canopy cover rate was highest in the area from 2nd Ring Rd to 1st Ring Rd in southwest, generally performing the feature of high in the middle but low at both ends.

TABLE 18 Comparison of canopy cover rate in the landscape of each section of Qingshuihe and Modihe Rivers

Type	Floor Area and Ratio		Canopy Cover Characteristics		
	Area (hm2)	Percentage in Landscape Area of Qingshuihe and Modihe River (%)	Canopy Cover Area (hm2)	Canopy Cover Rate (%)	Percentage in All the Canopy Areas of Qingshuihe and Modihe River (%)
Expressway to 3rd Ring Rd (Northwest)	56.4071	38.02	26.4278	46.85	39.44
3rd Ring Rd to 2nd Ring Rd (Northwest)	19.862	13.39	8.4574	42.58	12.62
3rd Ring Rd to 2nd Ring Rd (Southwest)	18.0612	12.18	7.4998	41.52	11.19
2ndRing Rd to 1st Ring Rd (Northwest)	15.4979	10.45	7.3647	47.52	10.99
2nd Ring Rd to 1st Ring Rd (Southwest)	13.3657	9.01	6.4323	48.12	9.6
Inside 1st Ring Rd (Southwest)	14.9724	10.09	6.5162	43.52	9.72
Inside 1st Ring Rd (Southeast)	10.1842	6.86	4.3138	42.36	6.44
Total	148.3505	100	67.012	45.17	100

To sum up, the patch number, average patch area and the total patch area of urban river landscape generally presented a trend of successively increasing from urban center to areas outside the city. This can be attributed to the fact that most river landscapes were artificial landscapes in the city, showing substantial interference from human. Besides, the degree of landscape fragmentation was high. In cities, when rivers flow through old resident areas, the patch area and canopy cover rate of river landscape become minimal. However, when rivers flow through new resident areas, the patch area and canopy cover rate of river landscape become maximum. The patch area of river landscape in different regions should be balanced to the greatest extent so as to realize the maximum balance of river ecosystems in different regions of a city. In this way, the urban ecosystem can be regulated to the largest extent.

4 Conclusions

Urban river landscape is a vital ecological regulating system for cities. In our research, from perspectives of the overall urban, city location and single river, a systematic and comprehensive research of landscape characteristics of main rivers in Chengdu was conducted so that the landscape characteristics of major rivers in Chengdu could be perceived. This research is intended to provide references for evaluating the macro-regional ecological environment of Chengdu, the ecological environment and management.

Acknowledgements

The heartfelt gratitude goes to the National Natural Science Foundation of China for its sponsor of this research (No.51208428).

References

[1] Cheng X, Hu Y 2006 Theory and practice on ecological garden Peking *China Forestry Press (in Chinese)*
[2] Peng Zhenhua 2003 *Modern urban forest development in Shanghai* Peking, China Forestry Press *(in Chinese)*
[3] Song Q, Yang Z 2005 Discussion of issues on urban river manage *Peking Planning and design research*
[4] Song Q, Yang Z 2002 Reflection on river integrated management in China *Water science development (in Chinese)*
[5] Wu J 2007 Landscape Ecology Peking *Higher Education Press (in Chinese)*
[6] Fu F, Dong L 2012 Analysis of research status on urban river landscape planning and design *Peking, Research on Urban Development* (12) 8-11 *(in Chinese)*
[7] Newbold J D, Erman D C, Roby K B 1980 Effects of logging on macro invertebrates in streams with and without buffer strips *Canadian Journal of Fisheries and Aquatic Science* **37**(7) 1076-85
[8] Large A R G, Petts G E 1996 Rehabilitation of river margins *River Restoration* **71** 106-23
[9] Rohling J 1998 Corridors of Green *Wildlife in North Carolina* **5** 22-7
[10] Fu F 2011 Spatial analysis of Urban river landscape space by ecological corridor principle *Peking, Chinese Garden (in Chinese)*

Authors

Jian Zhang, born in October 1987, Chengdu, Sichuan Province, P.R. China.

Current position, grades: Doctor at the School of Architecture, Southwest Jiaotong University, China.
University studies: master's degree in architecture at Southwest Jiaotong University in China.
Scientific interests: human settlement and river landscape.
Publications: 3 papers.
Experience: participated 4 scientific research projects.

Zhongwei Shen, born in October 1965, Chengdu, Sichuan Province, P.R. China.

Current position, grades: professor at the School of Architecture, Southwest Jiaotong University, China.
University studies: master's degree in Logistical Engineering University of PLA in China.
Scientific interests: human settlement and traffic building.
Publications: more than 20 papers.
Experience: teaching experience of more than 20 years, completed 6 scientific research projects.

Hao Shen, born in June 1988, Chengdu, Sichuan Province, P.R. China.

Current position, grades: master at the School of Architecture, Southwest Jiaotong University, China.
University studies: bachelor's degree in architecture from Southwest Jiaotong University in China.
Scientific interests: human settlement and river landscape.
Publications: 1 paper.
Experience: 1 scientific research project.

Analysis of flood disasters from 206 BC to 1949 in China

Zhenhua Liu*

Department of Municipal Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou, China, 310018

Received 3 September 2014, www.cmnt.lv

Abstract

Flood disasters produces not only direct economic losses but also serious environmental problems, the government attaches great importance to flood disasters prevention from ancient times to the present. A variety of analytic methods are adopted, for example mathematical model analysis, chart analysis, qualitative and quantitative analysis. Some mathematical models are constructed by means of statistical analysis of historical data for flood disasters, such as mathematical model for flood cycle each dynasty. There are the following results. First, there are 1037 times flood disasters from 206 BC to 1936, once every 2.07 years. According to analysis of mathematical model for flood cycle each dynasty from 206 BC to 1936, the flood cycle is the declining concave curve, except for the Five Dynasties. Second, in modern history, the most number of catastrophic floods is the Yellow River reaching 8 times, the Yangtze River 5 times. Finally, water pollution caused by floods can not be ignored. In 2012 92 percent of untreated rural sewage was discharged dispersedly.The vast majority of pollution load on the ground runs into directly riverway through stormwater runoff. Water wells are easily contaminated for germs and parasite. We should pay attention to water pollution caused by floods and combine organically flood control with water environment protection.The government coordinates closely ideas, rule of law, technology and investment with rural residents for governing rural sewage to reduce the risk of water pollution caused by floods.

Keywords: flood disasters, flood cycle model, qualitative and quantitative analysis, rural sewage, water pollution

1 Introduction

The flood is one of the major natural disasters not only in China but also in the rest of the world, governments around the world are very concerned about flood control issues. The Chinese government has pay high attention to flood and drought disasters involving livelihood issues [1,2]. The number of serious flood disasters from 1950 to 2004 reached 2606 times all over the world, involving 172 countries and regions, the total affected population of 2.75 billion, the economic loss of 347.235 billion dollars [3]. According to historical records and incomplete statistics, there is about once every two years [4]. The Yangtze River is vulnerable to flood disasters, during the 20th century, there are basin-wide floods in 1998, 1954 and 1931 [5]. The impact of floods on China's national economy is much larger than the impact on the United States [6]. It is necessary to research on flood disaster, which can provide a reference for government policy makers.

Floods has also produced a series of environmental issues, for example, the destruction of the ecological environment, the destruction of arable land, the destruction of the river system, the pollution of the water environment,especially water pollution.In 2012, the discharge of rural waste water across the country was 32.2 million t every day, most of sewage is discharged directly. Rural sewage, acid rain, agricultural pollutants, and other rural pollutants will enter into natural water environment by surface runoff. Therefore, we need to focus on the impact of floods on the water environment.

There are many effects of flood hazards.Analysis of flood disasters should choose a variety of analytic methods combined with qualitative analysis and quantitative analysis, such as mathematical models, mapping analysis, spreadsheet analysis, comparative analysis.

2 The national flood disasters in ancient times

Chinese history is usually divided into ancient history, modern history and New China, ancient history means the period before 1840, modern history means the period from 1840 to 1949, New China means the period since 1949. In ancient times in China, there are many legends about disasters of flood and waterloging, due to lack of historical data records and physical proof, it is difficult to analyze the conditions of disasters of flood and waterloging in detail. Until the Qin and Han dynasties, there are more historical data, historical records are quite specific. According to the history of the Chinese famine writed by Deng Tuo [7], there are 2142 years during the period from 206 BC to 1936, the flood disasters statistics is shown in Table 1, the flood frequency is shown in Figure 1. The mathematical model of the flood cycle is constructed by means of statistical analysis of historical data for flood disasters. 1 corresponds to the Han dynasty. 2 corresponds to the Weijin dynasty. 3 corresponds to the Northern and Southern Dynasties. 4 corresponds to the Sui and Tang Dynasties. 5 corresponds to the Five Dynasties. 6 corresponds to the Song dynasty. 7 corresponds to the Yuan dynasty. 8 corresponds to the Ming dynasty. 9

* ***Corresponding author's*** e-mail: liuzhh@zjweu.edu.cn

corresponds to the Qing dynasty. 10 corresponds to the Republic of China for the mathematical model of the flood cycle as follows:

$$y = 5.9307x^{-0.6763}, \quad (1)$$

where y is the flood cycle of dynasty, x is 1, 2, …, 10.

The flood cycle is the declining concave curve from 206 BC to 1936 in China, except for 5 representing for the Five Dynasties.

TABLE 1 The flood disasters statistics from 206 BC to 1936

Dynasty	Number of flood disasters
The Han dynasty	76
The Weijin dynasty	56
The Northern and Southern Dynasties	77
The Sui and Tang Dynasties	120
The Five Dynasties	11
The Song dynasty	193
The Yuan dynasty	92

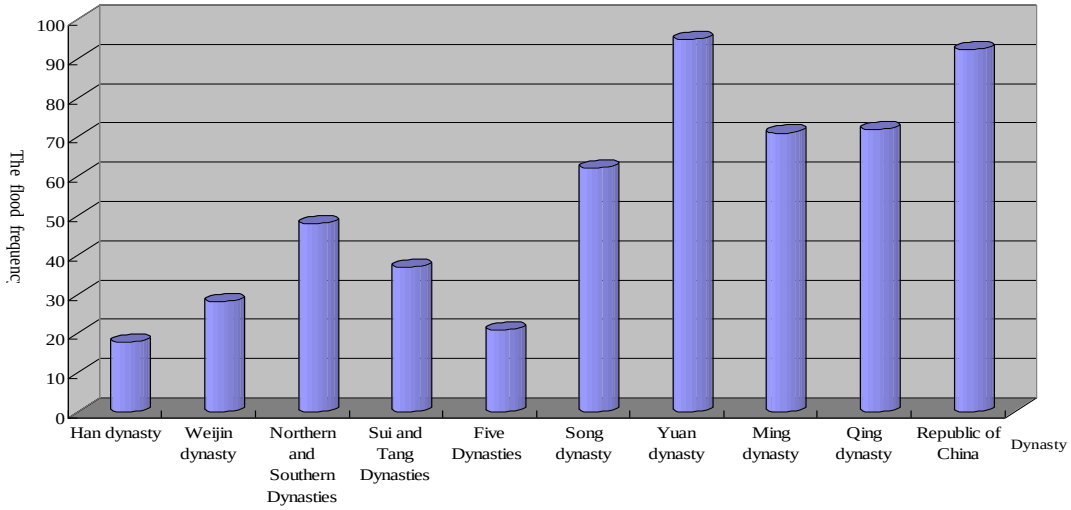


FIGURE 1 The flood frequency (%) of flood disasters from 206 BC to 1936

Table 1 shows that there are 1037 times of flood disasters from 206 BC to 1936 in China.At the same time, changing characteristics of flood frequency is depicted by graphing method. Figure 1 shows that the trends in flood frequency is divided into two phases in ancient times in China. The first stage from the Han dynasty to the Five Dynasties, flood frequency is relatively lower. The second stage from the Song Dynasty to the Qing Dynasty, flood frequency is relatively higher. There are 1165 years from the Han dynasty to the Five Dynasties,and 340 times of flood disasters, the flood frequency of 29.2 percent on average; 951 years from the Song Dynasty to the Qing Dynasty, times years of flood disasters,the average flood frequency 70.8 percent. However, the average flood frequency in Yuan Dynasty reached 94.8 percent, which is the highest flood frequency in the ancient Chinese dynasties, and is the dynasty of the most frequent occurrence of floods. Meanwhile, the flood frequency reached 92.3 percent from 1911 to 1936 in Republic of China. The average flood frequency from the Song Dynasty to the Qing Dynasty is 2.4 times from the Han dynasty to the Five Dynasties. Therefore, whether it is the first stage or second stage, basically first increase and then decrease, but generally speaking, the flood frequency is the rising trend from the Han dynasty to the Qing Dynasty. In ancient flood statistics, due to various conditions limits,it is difficult for precise statistics,but there is a very obvious trend that the flood frequency is higher and higher.

3 The flood disasters in ancient times in the yellow river valley

The Yellow River valley is the China's political, economic and cultural center in ancient times, in the Spring and Autumn and Warring States Period, there are historical records of flooding of the Yellow River. According to the records statistics of history books and local chronicles [7], the flood disasters statistics in the Yellow River valley is shown in Figure 2 in ancient times. Here, the flood disasters indicates the scope of flood disasters including most parts of middle reaches or the middle and lower reaches of the Yellow River,the flood situation is very serious disaster.

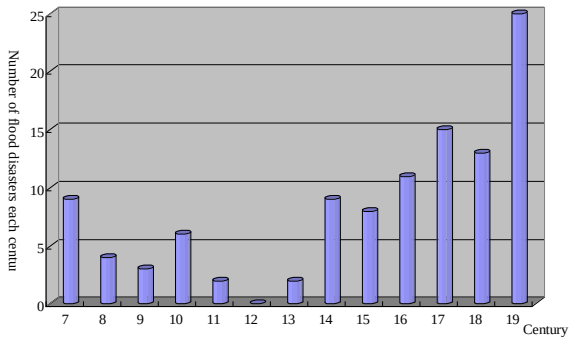


FIGURE 2 The flood disasters situation in the Yellow River valley from the 7th century to 19th century

The mathematical model of flood disasters is constructed in the Yellow River valley from the 12th century to 19th century. If 1 means the 13th century, 2 means the 15th century, 3 means the 17th century, 4 means the 19th century,as follows for the mathematical model of flood disasters:

$$y = 7.6x - 6.5, \quad (2)$$

where y is the number of flood disasters each century, x is 1,2,3,4.

1 corresponds to the 12th century, 2 corresponds to the 14th century, 3 corresponds to the 16th century, 4 corresponds to the 18th century for the mathematical model of flood disasters as follows:

$$y = -1.75x^2 + 12.85x - 10.75, \quad (3)$$

where y is the number of flood disasters each century, x is 1,2,3,4.

At the same time, changing characteristics of flood disasters is depicted by graphing method. Figure 2 shows that on the whole number of flood disasters first decrease and then increase in the Yellow River valley from the 7th century to 19th century. Number of flood disasters reached 26 from the 7th century to 13th century, average 3.7 times per century, in other words, floods occur once every 27 years. At the same time, the flood did not happen in the 12th century. Number of flood disasters reached 81 from the 14th century to 19th century, average 13.5 times per century, once every 7.4 years. On the whole, number of flood disasters reached 107 from the 7th century to 19th century, average 8.2 times per century, once every 12.2 years, but once every 4 years in the 19th century. It can be seen that the flood disasters is relatively severe from the 16th century to 19th century, particularly in the 19th century.

4 The flood disasters in ancient times in other river basins and region

In other river basins in China, floods also often occur, and caused huge losses of people's lives and property. According to the historical data of the main river and region in china, through further statistical analysis, flood disasters statistics of the main river and part of region in china is shown in Table 2 in ancient times [7-19], the flood cycle is shown in Figure 3.

TABLE 2 The flood disasters statistics of the main river and part of region

River and region	Time	The flood frequency (%)
The Yellow River	From 602 BC to 1938	62.60
The Huaihe River	From 1279 to 1840	12.46
The Haihe River	From 1368 to 1840	65.54
The Yangtze River	From 618 to 1911	17.23
The Yellow Sea coast	From 798 to 1949	125.00
The Southeast coast	From 66 to 1911	91.39
The Jiangsu and Zhejiang coast	From 1368 to 1911	3.13
Hunan province	From 221 BC to 1644	3.16
The Jianghan Plain	From 1276 to 1911	55.56
The Guanzhong Reach of Weihe River	From 1600 to 1859	41.15
Kaifeng Area	From 1644 to 1911	29.10
The middle reaches of river Fenhe	From 1644 to 1911	38.81
The Sushui River	From 1368 to 1911	16.18
Jinan city	From 1644 to 1911	41.04
Wuzhong region of Ningxia	From 1644 to 1911	34.33
Yulin Area	From 1369 to 1644	42.75
Jinghe River valley	From 618 to 907	28.28
Luohe River Basin	From 1368 to 1644	20.22
Liaoning Province	From 1791 to 1911	64.46

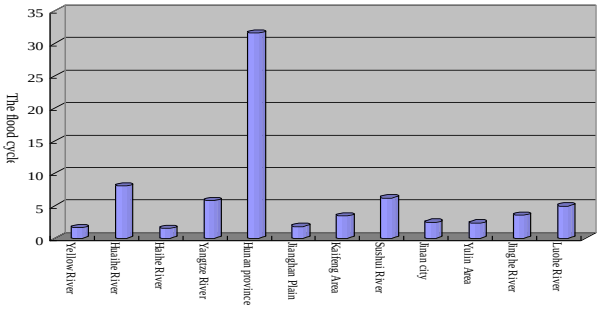


FIGURE 3 The flood cycle in the main river and part of region

According to records of governance and development of the Yellow River, the Yellow River flooding reached 1590 from 602 BC to 1938, there are a total of 543 years of flooding, the larger watershed change course 26 times. On the basis of the rough estimate based on historical data, the Yangtze river basin flooding reached 223 from 618 to 1911, once every 5.8 years on average. Among them, 16 times in Tang dynasty, once every 18 years; 79 times in Song and Yuan dynasty, once every 5.2 years;128 times in Ming and Qing dynasty, once every 4.2 years. Figure 3 shows that Hunan's flood cycle is longest, once every 31.6 years; The shortest cycle in Haihe River, once every 1.5 years, which shows a very high flood frequency of Haihe River; The remaining flood cycle, once every 3.8 years, in addition to Hunan Province.

Because historical records of floods is relatively simple in ancient times in China, it is difficult for accurate statistics of flood conditions. But some historical documents record the floods situation of part of the river and the areas, it can be used as one of research information. Currently, there are some other scholars researching on flooding of part of the rivers and regions, some research results are as follows. The floods disasters in the Tang Dynasty in Jinghe River valley can be divided into three periods, the medium-term flood-prone period, early and late period of less frequent floods [18]. Annual total flood duration day's and annual average flood duration days increase in the period from 1736 to 1948 in china [19]. The dominant periods are different in different times, which are the centennial scale oscillation during from 1644 to 1833 [20].130 flood disasters occurred in the Guanzhong region of the Weihe River Basin in the Qing Dynasty, once every 2.06 years on average, the occurring frequency of flood disasters increased obvio-

usly in the middle and late stages of the Qing Dynasty [21].

The major kinds of floods were mountain torrents and river torrents in Ancient Hunan, which brokes out more and more frequently from distribution of time [8].The earliest record of flood disasters in Liaoning Province is the first year of Ming King Emperor in 237. According to the historical records of Wei Kingdom Survey in History of the Three kingdoms, in the fall of the July, there were ten days of continuous rainfall, the water level of the Liaohe River rised. According to the statistical analysis of recorded data, there were a total of 78 floods during the period from 1791 to 1911,of which: catastrophic floods 4 times, serious floods 24 times, common floods 50 times; consecutive 3-year floods 3 times, consecutive 4-year floods once, consecutive 5-year floods 5 times, consecutive 6-year floods 2 times, consecutive 7-year floods once, consecutive 9-year floods once; in flood years, the interval of 1 year 11 times, the interval of 2 year 4 times, the interval of 3 year 4 times, the interval of 5 year once. Flood period in Liaoning Province is 1.6 years on average once, the average flood frequency is 64.5 percent.

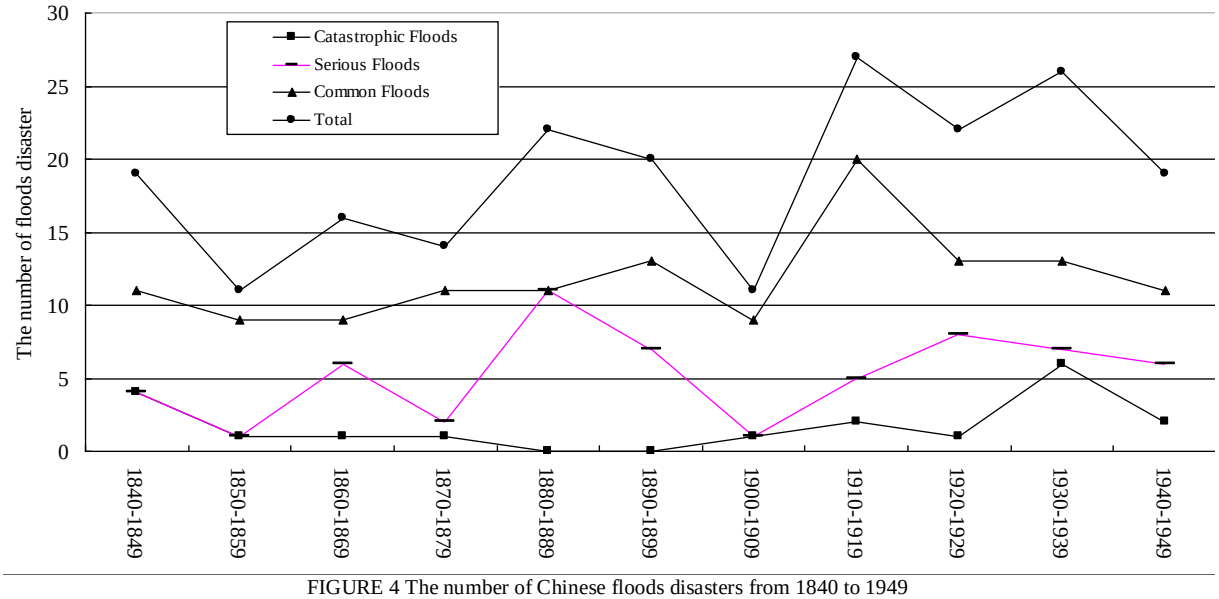


FIGURE 4 The number of Chinese floods disasters from 1840 to 1949

5 The national flood disasters in modern times

Although there was still in semifeudal and semicolonial society, the historical records of flood disasters in modern history is relatively better than in ancient times in China.The national and local flood situations were recorded in accordance with flood time, flood danger and the different basins. In order to better record the actual situation of the flood, the flood is divided into catastrophic floods, serious floods and common floods. Catastrophic floods means flood frequency below 5 percent, serious floods from 5 percent to 10 percent, common floods from 10 percent to 20 percent. Meanwhile, according to the distribution of China's major rivers, the flood situation was recorded in different rivers. According to the Chinese flood disasters chronology from 1840 to 1949 [4,7], 10 years as a unit of measurement, the number of Chinese floods disasters is shown in Figure 4 from 1840 to 1949, the number of affected counties is shown in Figure 5 every 10 years.

1 corresponds to the period from 1840 to 1849, 2 corresponds to the period from 1850 to 1859, 3 corresponds to the period from 1860 to 1869, 4 corresponds to the period from 1870 to 1879, 5 corresponds to the period from 1880 to 1889, 6 corresponds to the period from 1890 to 1899, 7 corresponds to the period from 1900 to 1909, 8 corresponds to the period from 1910 to 1919, 9 corresponds to the period from 1920 to 1929, 10 corresponds to the period from 1930 to 1939, 11 corresponds to the period from 1940 to 1949 for the mathematical model of the flood cycle as follows:

$$y = 0.12x^2 - 1.3224x + 4.1394, \quad (4)$$

where y is the number of catastrophic floods, x is 1, 2,…, 11.

Figure 4 shows the number of floods disasters in different time periods, 10-year as a time period, including the number of catastrophic floods, serious floods, common floods and total floods. The most number of catastrophic floods is the period from 1930 to 1939 reaching 6 times, secondly from 1940 to 1949 4 times, then from 1910 to 1919 and from 1940 to 1949 2 times, but from 1880 to 1889 and from 1890 to 1899 zero time. The most number of serious floods is the period from 1880 to 1889 reaching 11 times, secondly from 1920 to 1929 8 times, then from 1890 to 1899 and from 1930 to 1939 7 times. The number of catastrophic floods and serious floods is the most period from 1930 to 1939 totaling 13 times, meanwhile,the affected counties most totaling 2339.

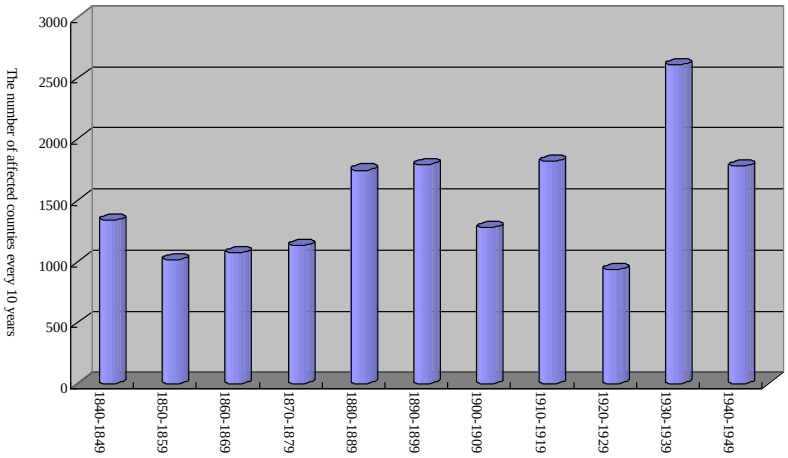


FIGURE 5 The number of the covered counties every 10 years from 1840 to 1949

Figure 5 shows that the number from 1840 to 1949 is the most affected counties in 1931 reaching 592, secondly in 1935 368. The catastrophic floods frequency or serious floods frequency is 130 percent during the period from 1930 to 1939; from 1880 to 1889 110 percent, but no catastrophic floods; from 1920 to 1929 90 percent; from 1840 to 1849 and from 1940 to 1949 80 percent; only the lowest frequency in the period from 1850 to 1859 and from 1900 to 1909 20 percent. It can clearly be seen that there is a increasing trend in floods disasters in the late modern history, especially during the period from 1930 to 1939.

6 The flood disasters of the major rivers in modern history

There are seven major rivers in China, including the Yangtze River, the Yellow River, the Pearl River, the Huaihe River, the Hailuanhe River, the Liaohe River, the Songhua River. In modern history, the major rivers in china occurred the serious floods, according to the Chinese flood disasters chronology from 1840 to 1949[4], by further analysing, the number of floods disasters of the major rivers is shown in Figure 6, the flood cycle is shown in Figure 7, the size distribution of flood is shown in Figure 8.

Figure 6 shows the number of floods disasters in different river. Figure 8 shows the size distribution of flood in modern history. China's seven major river floods is mainly the common floods, secondly the serious floods, and then the catastrophic floods. The most number of catastrophic floods is the Yellow River reaching 8 times, secondly the Yangtze River 5 times, then the Pearl River 3times, the Huaihe River 2 times, the Hailuanhe River and the Songhua River once, but the Liaohe River no catastrophic floods. The most number of serious floods is the Yangtze River reaching 21 times, secondly the Yellow River 11 times, and then the Liaohe River 7 times, the least the Songhua River only 2 times. Basically, the trend of the common floods and the total number of floods is same as the serious floods, the most number the Yangtze River, secondly the Yellow River, and then the Liaohe River, the least the Songhua River. This shows that china's seven major river floods are basically the same trends with the exception of catastrophic floods, but catastrophic floods mainly focusing on the Yellow River, the Yangtze River and the Pearl River, especially the Yellow River and the Yangtze River.

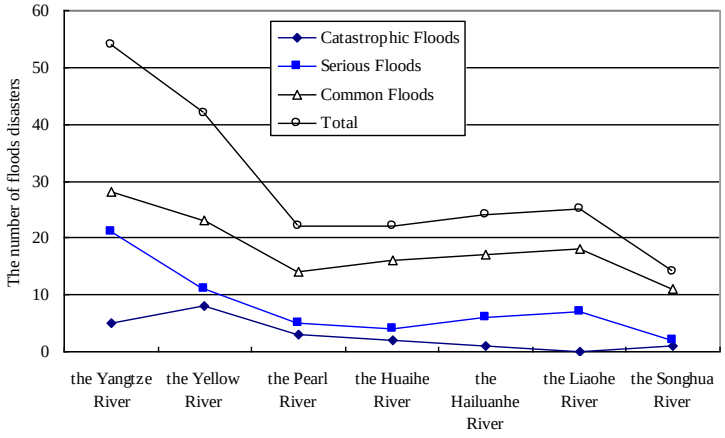


FIGURE 6 The number of floods disasters of the major rivers from 1840 to 1949

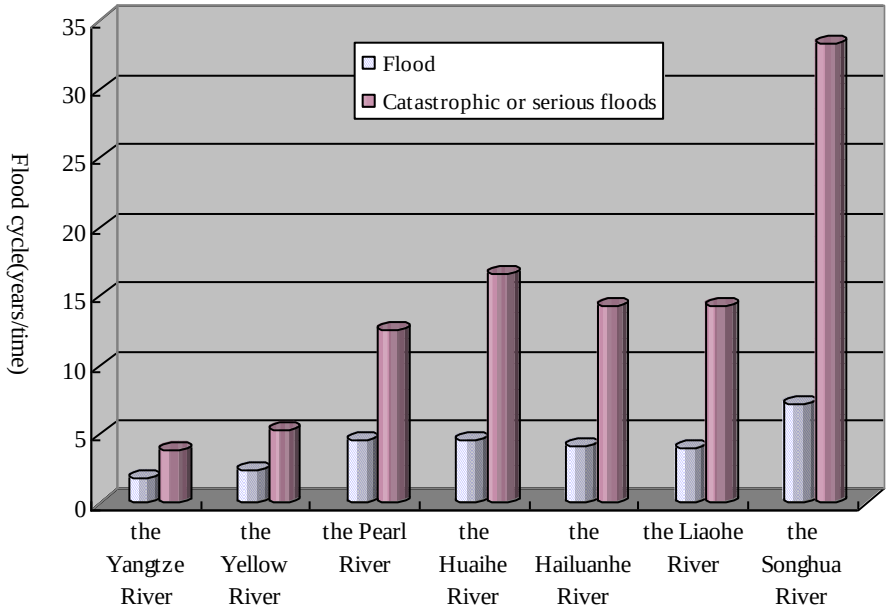


FIGURE 7 The flood cycle of the major rivers from 1840 to 1949

Figure 7 shows the flood cycle of the major rivers. The shortest cycle of the major rivers in China is the Yangtze River, the flood cycle once every 1.85 years, the catastrophic floods or serious floods once every 3.85 years; Secondly, the flood cycle of the Yellow River once every 2.4 years, the catastrophic floods or serious floods once every 5.3 years; This shows the flood happened in very high frequency in the Yangtze River and the Yellow River, on the whole the flood frequency of the Yangtze River higher than the Yellow River, But the number of catastrophic floods in the Yellow River is more than 60 percent of the Yangtze River. However, the longest cycle is the Songhua River, the flood cycle once every 7.1 years, the catastrophic floods or serious floods once every 33.3 years. Generally speaking, the variation trend of flood cycle is same as the catastrophic or serious floods with high similarity.

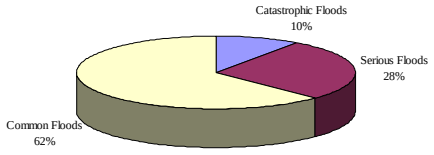


FIGURE 8 The size distribution of flood from 1840 to 1949

7 The water pollution caused by floods

Water pollution caused by floods can not be ignored. Rural sewage, garbage, dead animals, agricultural pollutants, toxic substance diffuse with the flow during the flood. Rivers, ponds and wells are easily contaminated for germs and parasite, which results in a variety of disease outbreaks and endanger seriously people's health. For example, in 1991, in the hardest-hit areas caused by floods in Anhui, the total number of bacteria was 10 times higher than standards for drinking water quality, escherichia coli was 700 times higher than standards for drinking water quality. In 2012, the discharge of rural waste water across the country was 11.7 billion t. Because operation and management for rural sewage treatment is relatively backward, sewage treatment rate is very low [22]. 92% of rural sewage untreated was discharged dispersedly on the ground. Agricultural pollution has become the dominant source of water pollution in China. In 2009, the total COD discharge of agricultural pollution was 12.8 million t, the total TN discharge was 3.46 million t, the total TP discharge was 460 thousand t [23]. Excessive amounts of pesticides and chemical fertilizer is on the ground. The vast majority of pollution load on the ground for rural sewage and agricultural pollutants runs into directly riverway through stormwater runoff. According to Report on the State of the Environment in China 2013, in 2013, among the 473 cities (counties) under monitoring, 210 cities (counties) had acid rain, accountting for 44.4%; 130 cities had acid rain frequency over 25%, taking up 27.5%; 43 had acid rain frequency over 75%, taking up 9.1%. In 2013, 29.6% of the cities (counties) had the annual average of precipitation pH value less than 5.6 (acid rain), 15.4% of the cities (counties) had the annual average of precipitation pH value less than 5.0 (relatively heavy acid rain) and 2.5% of the cities (counties) had the annual average of precipitation pH value less than 4.5 (heavy acid rain). For example, pH value for some reservoir in Zhejiang is low, which meet quality standards of water sources. In short, we should pay attention to the impact of flood on water environment.

8 Conclusion and suggestion

There are the characteristics of flood disasters in china including a wide range of floods, the frequent occurrence, the strong sudden and the large losses. In this paper, the flood disasters are analysed in ancient history, modern history and the impact of flood on water environment, conclusion and suggestion are the following:

First, before 206 BC, the relative lack of historical records of floods, until the Qin and Han dynasties, there are more specific historical records of flood. Therefore, the floods from 206 BC to 1840 are mainly analysed. There are 1037 times of flood disasters from 206 BC to 1936 in China. The flood cycle is the declining concave curve from 206 BC to 1936, except for the Five Dynasties. The trends in flood frequency is divided into two phases in ancient times in China, the first stage from the Han dynasty to the Five Dynasties, flood frequency is relatively lower; The second stage from the Song Dynasty to the Qing Dynasty, flood frequency is relatively higher. The flood disasters is relatively severe from the 16th century to 19th century in the Yellow River valley, particularly in the 19th century.

Second, in order to better record the actual situation of the flood, the flood is divided into catastrophic floods, serious floods and common floods. The most number of catastrophic floods is the period from 1930 to 1939 reaching 6 times, secondly from 1940 to 1949 4 times, then from 1910 to 1919 and from 1940 to 1949 2times,but from 1880 to 1889 and from 1890 to 1899 zero time. China's seven major river floods is mainly the common floods, secondly the serious floods, and then the catastrophic floods. The variation trend of flood cycle is same as the catastrophic or serious floods with high similarity.

Finally, we should pay attention to Water pollution caused by floods and combine organically flood control with water environment protection. By the rule of law and the media, the government strengthens publicity to raise public awareness of disaster prevention and water environmental protection, and organizes regularly the exercises of water pollution and flood disaster prevention. The government coordinates closely ideas, rule of law, technology and investment with rural residents for rural sewage governance. At present, there are no laws and regulations for rural sewage treatment in China. It is very important for urban and rural residents to change ideas of rural sewage treatment. It is good for rural sewage treatment, which can reduce the impact of floods on water pollution, especially the spread of germs and disease. Therefore, the government and residents should attach great importance to not only the direct loss of flooding but also the impact of the floods on water pollution.

Acknowledgements

The research results is supported by fund of Zhejiang planning project of philosophy and social sciences (No.14SWH17YB), the young and middle-aged teacher training. At the same time, we are also very grateful to the editorial board on Flood and Drought Disasters in China.

References

[1] Liu Z 2012 Research on Prevention Measures and Impact on Rural Safe Drinking Water Caused by Drought *Disaster Advances* **5**(4) 536-41
[2] Liu Z 2012 Comprehensive Analysis of Drought Disasters in China from 1483 to 2010 *Disaster Advances* **5**(4) 1053-9
[3] Jiang W, Li J, Wang L 2006 Compositive Analysis of Global Flood Disaster from 1950 to 2004 *Journal of Beijing Normal University (Natural Science)* **42**(5) 530-3 *(in Chinese)*
[4] Luo C, Le J 1997 Serious Floods in China - Summary of Flood Disaster China Bookstore Publishing House: Beijing *(in Chinese)*
[5] Zhang X, Tao S, Wei J 2006 An Analysis on the Basin-wide Catastrophic Floods in the Yangtze River during the 20th Century *Climatic and Environmental Research* **11**(6) 669-82 *(in Chinese)*
[6] Jiang F, Cheng X, Xiang L, Wu Y 2003 Review on the flood losses of the 20th century in the United States and comparative study on flood damages between China and USA in 1990s *Advances in Water Science* **14**(3) 384-8 *(in Chinese)*
[7] The National Flood Control and Drought Relief Headquarters Office, Nanjing Research Institution of Hydrology and Water Resources of Ministry of Water Resources 1997 Flood and drought disasters in China *China Bookstore Publishing House:* Beijing *(in Chinese)*
[8] Yang P 2003 Floods of Ancient Hunan *Xiangtan University Journal of Philosophy & Social Sciences Edition* **27**(1) 109-13*(in Chinese)*
[9] Xu X, Liu C 2011 The characteristics and its genetic analysis of flood disaster in the Jianghan Plain from during the Ming and Qing Dynasties *Journal of Xianning University* **31**(4) 1-4 *(in Chinese)*
[10] Zhao J, Long T, Chen Y 2010 Scale and Temporal Properties of Flood Sequence of Guanzhong Reach of Weihe River During the Last 400 Years *Bulletin of Soil and Water Conservation* **30**(2) 5-8 *(in Chinese)*
[11] Li Y, Zhao J 2010 Flood disasters and types in Kaifeng Area in Qing Dynasty *Journal of Arid Land Resources and Environment* **24**(3) 64-70 *(in Chinese)*
[12] Su H, Zha X 2010 Study on flood disasters in the middle reaches of river Fenhe during Qing Dynasty *Journal of Arid Land Resources and Environment* **24**(1) 90-4 *(in Chinese)*
[13] Wu P 2009 Study on flood disasters of the Sushui River Valley in the 1368-1911. *Journal of Arid Land Resources and Environment* **23**(12) 123-6 *(in Chinese)*
[14] Zhang X, Zhao J 2009 Study on flood disasters in the Qing dynasty in Jinan city *Journal of Shaanxi Normal University (Natural Science Edition)* **37**(4) 95-100 *(in Chinese)*
[15] Li Y, Zhao J 2009 Research on the flood disaster in Wuzhong region of Ningxia in Qing Dynasty *Journal of Arid Land Resources and Environment* **23**(4) 120-4 *(in Chinese)*
[16] Shao T, Zhao J 2009 The Characteristic Analysis on the Flood Disaster of Yulin Area in the Ming Dynasty *Journal of Arid Land Resources and Environment* **23**(1) 93-7 *(in Chinese)*
[17] Zhao J, Wang N, Long T 2010 Research on Flood Disasters of Jinghe River Valley in the Tang Dynasty *Journal of Arid Land Resources and Environment* **28**(3) 109-13 *(in Chinese)*

[18] Long T, Zhao J 2008 A Study on Fractal Structure of Flood Sequence in Luohe River Basin in Ming Dynasty *Marine Geology & Quaternary Geology* **28**(3) 115-9 *(in Chinese)*
[19] Zhou J, Shi P, Fang W 2001 Basic Analysis on Flooding Duration Days in China from 1736 to 1998 *Journal of Beijing Normal University (Natural Science)* **37**(3) 409-14 *(in Chinese)*
[20] Fang X, Chen L, Li S 2007 Changes of dominant periods of flood disaster in China during 1644-2004 *Advances in Water Science* **18**(5) 656-61 *(in Chinese)*
[21] Du J, Zhao J 2007 Study on Flood Disasters in the Guanzhong Region of the Weihe River Basin in the Qing Dynasty *Arid Zone Research* **24**(5) 598-603 *(in Chinese)*
[22] Yu N 2014 Research Progress of Rural Sewage Treatment in China *Journal of Anhui Agricultural Sciences* **42**(11) 323-5 *(in Chinese)*
[23] Li J, Li H, Xie L 2013 Potential and Efficiency of Agricultural Pollution Control in China and Its Influential Factors *Asian Agricultural Research* **5**(10) 56-60 *(in Chinese)*

Author

Zhenhua Liu, born in October 1977, Daxian County, Sichuan Province, P.R. China.

Current position, grades: associate professor of Department of Municipal Engineering, Zhejiang University of Water Resources and Electric Power, China.
University studies: Master Degree in Environmental Engineering at Qingdao Technological University in China.
Scientific interests: rural drinking water, rural sewage treatment, water pollution control, sewage and rainwater resources.
Publications: more than 10 papers.
Experience: teaching experience of 10 years, 3 scientific research projects.

Application of fuzzy mathematics models in hospital management evaluation

Mei Sun[1,2], Hui Feng[2], Siyuan Tang[2*], Ziqiang Luo[3]

[1]*Postdoctoral research station of school of basic medicine, Central south university, Changsha City, Hunan Province, China, 410013*

[2]*Nursing School of central south university, Changsha City, Hunan Province, China, 410013*

[3]*School of basic medicine, central south university, Changsha City, Hunan Province, China, 410013*

Received 15 September 2014, www.cmnt.lv

Abstract

To formulate a hospital management evaluation system and conduct empirical research, Delphi, analytic hierarchy process, and fuzzy comprehensive evaluation are used to build a hospital management evaluation system and conduct hospital management empirical research. And then a hospital management evaluation system is built. This system contains five first-level indices, namely, administrative management, human resource management, medical management, financial management and logistical support, and 23 second-level indices. Empirical research shows that the comprehensive evaluation of hospital management is above average and that service quality and infrastructure construction have the highest and lowest evaluation scores, respectively. This evaluation system is an effective tool for studying hospital management. Chinese hospital management requires further improvement, especially in terms of infrastructure construction.

Keywords: fuzzy mathematics, hospital management, evaluation system, analytic hierarchy process

1 Introduction

Hospital is a health and medical institution that treats and prevents diseases and protects people's health. Thus, hospital development and management are important concerns of all social sectors [1]. At present, hospital functions change from simple diagnosis and treatment into prevention, heath care, diagnosis and treatment, and rehabilitation, making services within hospitals extend forward and backward hospitals [2]. Hospital management shall adapt to such changes, adjust service structure and mode, and optimize and reorganize hospital resources. The increasingly fierce competitions in the medical market and the increasing medical service demands have brought new problems, changes, and challenges to hospital management. Hospital management innovation is an inevitable requirement of constantly changing objective hospital conditions and is an eternal theme of sustainable development of health cause. Hospitals should continuously innovate and improve management concept to remain invincible in the fierce market competition [3]. Therefore, hospital administrators need to actively innovate hospital management concept.

WHO introduces a new framework to analyze the performance of health systems in different countries and believes that the health system shall have three objectives [4], namely, promote good health, strengthen reaction capacity that people expect, and ensure financing fairness. It also proposes that a health system should possess four main functions, namely, management, financing, service provision, and raising resources. Hospital management quality can reflect the comprehensive strength and level of a hospital. The actual situation of Chinese hospital management reflects the urgent need for the guidance of hospital management innovation theory. Currently, all levels and categories of Chinese hospitals have inaccurate market and functional positioning, outdated operation and management concepts, low operating efficiency, and weak driving force and capital support for sustainable development [5]. Practical exploration and theoretical research of hospital management innovation should be vigorously conducted to make modern Chinese hospitals truly become subjects of market economy with independent management, self-discipline, and self-development. Changes in social medical and health demands urgently require hospitals to study management system innovation, management operation innovation, and management technological innovation, among others [6].

A few domestic and international studies have evaluated hospital management quality. Hospital management quality has its own features, and the particularity of comprehensive evaluation of hospital management quality is worthy of exploration and research. This study builds a hospital management evaluation index system by analytic hierarchy process (AHP). This process can scien-

[*]***Corresponding author's*** e-mail: sunmei1111@yeah.net

tifically, objectively, and accurately evaluate hospital management performance and promote hospitals to strengthen connotation construction. Then, fuzzy comprehensive evaluation is used to assess the management quality of six hospitals in a city. The evaluation results on hospital management quality are discussed for administrators to compare work quality and work development among hospitals in different times. This study may serve as a basis for evaluating hospital management quality.

2 Methods

2.1 EXPERT INTERVIEW

On the basis of literature reading, experience reference, and individual visit, this study begins with important and key contents that reflect hospital management, preliminarily selects hospital management evaluation indices, and initially builds a hospital management evaluation index system that contains 5 first-level indices and 37 second-level indices.

2.2 DELPHI

A qualified expert group and screen indices are selected from studies on social indices through Delphi, a scientific, relatively objective, and strongly operational method [7]. We seek advices of experts on selected hospital management evaluation index systems by distributing questionnaires and finally determine an evaluation system that contains 5 first-level indices and 23 second-level indices according to the advices and suggestions of experts.

2.3 AHP

Weight coefficients of all levels of indices can be calculated by AHP [8].

2.4 FUZZY COMPREHENSIVE EVALUATION

Hospital management involves a wide scope. Hospital management evaluation is a multifactor and multilevel process with highly complex influencing factors. Evaluation indices are both qualitative and quantitative. In addition, medical staff evaluates hospital management on the basis of subjective factors, fuzzy index factors, and asymmetric information. The introduction of the hybrid multi-attribute evaluation model can efficiently solve this problem. Triangular fuzzy numbers in fuzzy multi-attribute decision making can effectively handle index quantification and evaluators' subjective preference values [9,10]. Fuzzy comprehensive evaluation is realized by MATLAB7.0 program package.

3 Building of hospital management evaluation system

3.1 BUILDING OF INDEX SYSTEM

In accordance with the basic principles of SERVQUAL [11], a large number of hospital management indices are collected through literature and semi-structured interviews, and then Delphi is used to preliminarily select 37 indices and determine hospital management evaluation indices, including 5 first-level indices (i.e., administrative management, human resource management, medical management, financial management, and logistical support) and 23 second-level indices. An evaluation system is established through experts' scores of indices by pair-wise comparison combined with AHP. A questionnaire is designed according to the established system, and the questionnaire is scored by the Likert five-level scale. Finally, a model is established by fuzzy mathematical modeling combined with exploratory and confirmatory analyses.

3.2 DETERMINE INDEX WEIGHT

The weight of hospital management evaluation indices plays an important role in the evaluation system. It is related to the importance degree of hospitals' management of influencing factors. In terms of social evaluation, the weights of evaluation indices are currently directly determined by a few experts according to experience, which is an insufficient basis for data analysis [12]. In addition, their experience may significantly deviate from the actual situation, thereby affecting the accuracy of evaluation results. Thus, this study uses AHP to determine index weights. The index weight of the hospital management evaluation system is calculated by AHP according to the following steps:

Step 1: establish a hierarchical index system structure, i.e., build index hierarchy according to the basic relations of the evaluation index system.

Step 2: build a pair-wise comparison matrix.

Step 3: check consistency.

$$C.I. = \frac{\lambda_{\max} - n}{n-1}, \qquad C.R. = \frac{C.I.}{R.I.}. \tag{1}$$

In this study, CR = 0.017 < 0.10, indicating that weight coefficients are available.

The index weight coefficients of the evaluation index system can be determined according to the above three steps. The weight of all levels of indices can be finally determined as shown in Table 1 on the basis of the scores provided by 10 experts on 5 first-level indices and 23 second-level indices.

TABLE 1 Evaluation index system and weight

First-level index (U_i)	Weight (W_i)	Second-level index (U_{ii})	Weight (W_{ii})	Index connotation
Administrative management U1	0.18	Organizational structure U11	0.22	Rational hospital management organizational structure
		Leaders' management responsibilities U12	0.31	Leaders devote themselves to hospital management
		Management accountability U13	0.21	Management accountability, to implement reward and punishment system
		Development planning U14	0.26	Formulate and implement development planning and annual working plan
Human resource management U2	0.21	Human resource allocation U21	0.23	Rational allocation of department human resources
		Talent echelon construction U22	0.22	Talent echelon construction system, continuing education system
		Physician structure U23	0.18	Rational three-level physician employment structure
		Nursing staff U24	0.19	Rational amount and echelon structure of nursing staff
		Educational background and professional knowledge U25	0.18	Rational educational background and professional knowledge structure of medical staff
Medical management U3	0.32	Medical quality U31	0.21	Medical treatment, medical matters and other management organizations, and working system
		Department supervision U32	0.19	Manage and supervise the quality of clinical and medicine departments
		Medical risk warning U33	0.09	Medical risk warning system, response and handling capacities
		Doctor–patient relationship U34	0.19	Deal with medical disputes timely and properly, and coordinate doctor–patient relationship
		Emergency management U35	0.15	Deal with emergencies and public disasters
		Medical attitude U36	0.17	Attitude of medical staff toward patients
Financial management U4	0.13	Accounting posts U41	0.24	Set accounting posts scientifically according to needs
		Accounting U42	0.27	Set accounting items for financial accounting according to provisions
		Budgeting U43	0.22	Formulate and implement budget for revenues and expenditures scientifically and rationally
		Financial control U44	0.27	Strengthen cost accounting and control, and reduce operating costs
Logistical support U5	0.16	Basic facility construction U51	0.28	Carry out basic construction items according to national laws and regulations
		Equipment management U52	0.26	Manage equipment scientifically
		Equipment renewal U53	0.19	Perfect medical equipment purchasing, maintenance and renewal system
		Logistical work U54	0.27	Meet needs of clinical work and patients' treatment

3.3 EVALUATION METHODS

Fuzzy comprehensive evaluation mathematical models can be divided into one-level models and multilevel models. Analysis of evaluation factors shows that some factors are coordinating while some are causal. In other words, evaluation factors have different levels, causing a practical issue that objectively exists.

Multilevel fuzzy comprehensive evaluation model: low-level factors are initially comprehensively evaluated, and then results are subjected to high-level comprehensive evaluation by the following specific steps:

1) Confirm the factor set of evaluation object $F=\{f_1, f_2, \ldots, f_n\}$. For hospital management evaluation indices, first-level evaluation factors include f_1 (administrative management), f_2 (human resources), f_3 (medical management), f_4 (financial management), and f_5 (logistical support) while second-level evaluation sub-factors include those of the indices in Table 1.

2) Confirm evaluation category set $E=\{e_1, e_2, \ldots, e_n\}$. Hospital management evaluation grades are determined as E= {very high, quite high, general, quite low, very low} grades, which correspond to scores of 90, 80, 70, 60, and 50, respectively.

3) Determine weight set W_f.

4) Perform single-factor evaluation. Establish fuzzy mapping f from factor set F to comment set E, derive a fuzzy relation R_f from f, and its matrix representation can be denoted as $R_i = R_f = (\gamma_{ikj}\ m\times n)$.

5) Comprehensive evaluation. Calculate comprehendsive evaluation vector S and comprehensive evaluation value μ according to the following formulas. $S = W_f R$, $\mu = W_e S^T$.

4 Empirical research of hospital management based on fuzzy mathematics

A questionnaire is formulated in accordance with a hospital management evaluation system to investigate hospital management quality, and the questionnaire is scored by the Likert five-level scale with "1" representing "dissatisfied" and "5" representing "very satisfied." Medical staff members select according to perceived actual situation and thus complete the test of hospital management quality. A total of 600 questionnaires are sent out, and 586 valid ones are recovered with an effective recovery rate of 97.67%.

With different emphasis on factors in U, each factor shall be given different weights, which can be expressed as a fuzzy subset $\underset{\sim}{A}(a_1, a_2, \cdots, a_n)$ of U, and stipulate $\sum_{i=1}^{n} a_i = 1$.

With $\underset{\sim}{R}$ and $\underset{\sim}{A}$, comprehensive evaluation is $\underset{\sim}{B} = \underset{\sim}{A} \cdot \underset{\sim}{R}$, $\underset{\sim}{B} = (b_1, b_2, \cdots, b_m)$ is a fuzzy subset of V, where

$$b_j = \overset{n}{\underset{i=1}{V}}(a_i \wedge r_{ij})(j = 1, 2, \cdots, m). \quad (2)$$

The statistical data from the sample survey are substituted into the established model, and the vectors of all levels of fuzzy comprehensive evaluation are calculated.

4.1 ESTABLISH AN EVALUATION SET

An evaluation set consists the evaluation of all indices and all possible results of total evaluation. Hospital management evaluation may have five possible results. The evaluation set can be denoted as $R = \{r_1, r_2, r_3, r_4, r_5\}$ = {very low, quite low, general, quite high, very high}.

The index "organizational structure U11" is "very high" for 48 people, "quite high" for 169 people, "general" for 228 people, "quite low" for 37 people, and "very low" for 104 people. On the basis of normalization calculation, the evaluation set of organizational structure U11 is r1 = (0.178, 0.063, 0.389, 0.289, 0.081).

Similarly, the single-factor evaluation matrix of administrative management obtained by combining the evaluation grades of the sub-indices of administrative management is as follow:

$$R_1 = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix} =$$

$$\begin{pmatrix} 0.17848 & 0.063194 & 0.388557 & 0.288642 & 0.081127 \\ 0.099915 & 0.098207 & 0.447481 & 0.268147 & 0.086251 \\ 0.24509 & 0.020495 & 0.378309 & 0.288642 & 0.067464 \\ 0.207515 & 0.076003 & 0.41076 & 0.293766 & 0.011956 \end{pmatrix}$$

TABLE 2 Index evaluation scores

First-level index	Second-level index	Evaluation scores
Administrative management U1	Organizational structure U11	70.30743
	Leaders' management responsibilities U12	71.42613
	Management accountability U13	69.12895
	Development planning and management U14	68.26644
Human resource management U2	Human resource allocation U21	77.42101
	Talent echelon construction U22	76.13151
	Physician structure U23	78.94962
	Nursing staffs U24	78.65073
	Educational background and professional knowledge U25	69.12041
Medical management U3	Medical quality U31	76.65243
	Department supervision U32	78.924
	Medical risk warning U33	68.93254
	Doctor–patient relationship U34	68.94108
	Emergency management U35	68.7105
	Medical attitude U36	62.50213
Financial management U4	Accounting posts U41	68.77882
	Accounting U42	68.33476
	Budgeting U43	70.40137
	Financial control U44	63.57814
Logistical support U5	Basic facility construction U51	77.30999
	Equipment management U52	71.04184
	Equipment renewal U53	68.51409
	Logistical work U54	66.191289

According to fuzzy membership degree, the evaluation matrix shows that the organizational structure and leader management responsibilities are general while the management accountability and development planning are low. In terms of comprehensive evaluation scores, physician structure evaluation has the highest scores; whereas medical attitude evaluation has the lowest scores (Table 2).

Similarly, the single-factor evaluation matrix of human resource management obtained by combining the evaluation grades of sub-indices of human resource management is shown as follow:

r1	0.17848	0.043553	0.050384	0.312553	0.41503
r2	0.172502	0.072588	0.081981	0.315115	0.357814
r3	0.013664	0.003416	0.326217	0.387703	0.269001
r4	0.032451	0.051238	0.228864	0.393681	0.293766
r5	0.187874	0.097353	0.389411	0.265585	0.059778

In terms of fuzzy membership degree, the evaluation of human resource allocation, talent echelon construction, physician structure, and nursing staff is general, whereas the evaluation of educational background and professional knowledge is low. In terms of evaluation scores, human resource allocation has the highest scores, whereas educational background and professional knowledge have the lowest scores.

Similarly, the single-factor evaluation matrix of medical management obtained by combining the evaluation grades of sub-indices of medical management is shown as follow:

r1	0.184458	0.071734	0.052946	0.275833	0.415030
r2	0.028181	0.071734	0.169086	0.441503	0.289496
r3	0.220325	0.000854	0.482494	0.257899	0.038429
r4	0.189582	0.057216	0.458582	0.258753	0.035867
r5	0.247652	0.008540	0.429547	0.253629	0.060632
r6	0.232280	0.436379	0.252775	0.005978	0.072588

In terms of fuzzy membership degree, medical quality and department supervision have general evaluation, whereas medical risk warning, doctor–patient relationship, emergency management, and medical attitude have low evaluation. In terms of evaluation scores, medical quality has the highest scores, whereas medical attitude has the lowest scores.

Similarly, the single-factor evaluation matrix of financial management obtained by combining the evaluation grades of the sub-indices of financial management is shown as follow:

r1	0.222886	0.042699	0.40222	0.298036	0.034159
r2	0.245090	0.047822	0.360376	0.321947	0.024765
r3	0.168232	0.035013	0.455167	0.271563	0.070026
r4	0.203245	0.400512	0.273271	0.081127	0.041845

In terms of fuzzy membership degree, accounting posts, accounting, and financial control have low evaluation, whereas budgeting has general evaluation. In terms of evaluation scores, budgeting has general scores, the other aspects of financial management have low scores, and financial control has the lowest scores.

Similarly, the single-factor evaluation matrix of logistical support obtained by combining the evaluation grades of sub-indices of logistical support is shown as follow:

r1	0.222886	0.042699	0.402220	0.298036	0.034159
r2	0.245090	0.047822	0.360376	0.321947	0.024765
r3	0.168232	0.035013	0.455167	0.271563	0.070026
r4	0.203245	0.400512	0.273271	0.081127	0.041845

In terms of fuzzy membership degree, basic facility construction and equipment management have general eva-

luation, and equipment renewal and logistical work have low evaluation. Basic facility construction has the highest scores, whereas logistical work has the lowest scores.

4.2 ESTABLISH A WEIGHT SET

Weight set of second-level index layer: The corresponding weight of each index is provided according to the above expert scoring and AHP analysis results.

Weights of administrative management indices:
$A_1 = c$ **(0.22, 0.31, 0.21, 0.26)**
Weights of human resource management indices:
$A_2 = c$ **(0.23, 0.22, 0.18, 0.19, 0.18)**
Weights of medical management indices:
$A_3 = c$ **(0.21, 0.19, 0.09, 0.19, 0.15, 0.17)**
Weights of financial management indices:
$A_4 = c$ **(0.24, 0.27, 0.22, 0.27)**
Weights of logistical support indices:
$A_5 = c$ **(0.28, 0.26, 0.19, 0.27)**

Weight set of first-level index layer: The corresponding weight of each index is provided according to the AHP analysis results: $W = c$ (0.18, 0.21, 0.32, 0.13, 0.16).

4.3 COMPREHENSIVE EVALUATION

With the above weight sets and single factor evaluation sets, we can calculate the following:

The comprehensive evaluation vector of administrative management is:

$$B_1 = A_1 \times R_1 = (0.22 \quad 0.31 \quad 0.21 \quad 0.26)$$
$$\begin{pmatrix} 0.17848 & 0.063194 & 0.388557 & 0.288642 & 0.081127 \\ 0.099915 & 0.098207 & 0.447481 & 0.268147 & 0.086251 \\ 0.24509 & 0.020495 & 0.378309 & 0.288642 & 0.067464 \\ 0.207515 & 0.076003 & 0.41076 & 0.293766 & 0.011956 \end{pmatrix} =$$
$$(0.175662 \quad 0.068412 \quad 0.410444 \quad 0.283621 \quad 0.061862).$$

In terms of fuzzy membership degree, at administrative management has low evaluation, and its comprehensive evaluation scores are:

$$0.1756618\times 50 + 0.06841161\times 60 + 0.4104441\times 70 + 0.2836208\times 80 + 0.06186166\times 90 = 69.87609.$$

Similarly, the comprehensive evaluation vector of human resource management is:

$$B_2 = (0.121443 \quad 0.05386 \quad 0.201921 \quad 0.333604 \quad 0.289172).$$

In terms of fuzzy membership degree, human resource management has general evaluation, and its comprehendsive evaluation scores are:

$$0.1214432\times 50 + 0.05385995\times 60 + 0.2019214\times 70 + 0.3336038\times 80 + 0.2891716\times 90 = 76.15201.$$

Similarly, the comprehensive evaluation vector of medical management is:

$$B_3 = (0.176576 \quad 0.115107 \quad 0.281204 \quad 0.253245 \quad 0.173869).$$

In terms of fuzzy membership degree, medical management has general comprehensive evaluation, and its evaluation scores are:

$$0.1765756\times 50 + 0.1151067\times 60 + 0.2812041\times 70 + 0.2532451\times 80 + 0.1738685\times 90 = 71.32724.$$

Similarly, the comprehensive evaluation sector of financial management is:

$$B_4 = (0.211554 \quad 0.139001 \quad 0.367754 \quad 0.240103 \quad 0.041588).$$

In terms of fuzzy membership degree, financial management has quite low comprehensive evaluation, and its comprehensive evaluation scores are:

$$0.2115542\times 50 + 0.1390009\times 60 + 0.3677541\times 70 + 0.2401025\times 80 + 0.04158839\times 90 = 67.6117.$$

Similarly, the comprehensive evaluation sector of logistical support is:

$$B_5 = (0.197874 \quad 0.050325 \quad 0.324355 \quad 0.308121 \quad 0.119325).$$

In terms of fuzzy membership degree, logistical support has general comprehensive evaluation, and its comprehensive evaluation scores are:

$$0.1978736\times 50 + 0.05032451\times 60 + 0.3243553\times 70 + 0.3081213\times 80 + 0.1193254\times 90 = 71.007.$$

With the above comprehensive evaluation and weights of three first-level indices, the comprehensive evaluation scores of hospital management can be calculated as follow:

$$C = W \times \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \end{pmatrix} (0.18 \quad 0.21 \quad 0.32 \quad 0.13 \quad 0.16)$$
$$\begin{pmatrix} 0.176 & 0.068 & 0.410 & 0.284 & 0.062 \\ 0.121 & 0.054 & 0.202 & 0.334 & 0.289 \\ 0.177 & 0.115 & 0.281 & 0.253 & 0.174 \\ 0.212 & 0.139 & 0.368 & 0.240 & 0.042 \\ 0.198 & 0.050 & 0.324 & 0.308 & 0.119 \end{pmatrix} =$$
$$(0.173 \quad 0.087 \quad 0.306 \quad 0.283 \quad 0.152).$$

In terms of fuzzy membership degree, comprehensive evaluation is general, and the comprehensive evaluation scores of hospital management are:

$$0.1779046\times 50 + 0.06538295\times 60 + 0.3063106\times 70 + 0.2908957\times 80 + 0.1595061\times 90 = 71.89.$$

On the basis of the above results, a summary table of first-level indices and comprehensive evaluation scores can be obtained (Table 3).

TABLE 3 First-level indices and comprehensive evaluation

First-level index Ui	Satisfaction membership	Evaluation scores
Administrative management U1	Quite low	69.876
Human resource management U2	General	76.152
Medical management U3	General	71.327
Financial management U4	Quite low	67.612
Logistical support U5	General	71.007
Comprehensive evaluation	General	71.89

5 Conclusions

This study establishes a hospital management evaluation system containing 5 first-level indices (i.e., administrative management, human resource management, medical management, financial management, and logistical support) and 23 second-level indices. It also determines the weights and evaluation methods of indices in this evaluation system.

The hospital management evaluation system in this study can scientifically and rationally solve the problem of irrational evaluation by considering the main quantitative factors in previous evaluation methods.

Empirical analysis of hospital management shows that the comprehensive evaluation score of current hospital management is 71.89 while that of hospital management level needs further improvement. Particularly, administrative and financial managements have low evaluation scores and need to be improved urgently. Meanwhile, human resource management, logistical support, and medical management have general evaluation scores and also need further improvement.

Acknowledgement

Ziqiang Luo and Siyuan Tang are co-corresponding authors. The authors acknowledge A Project Supported by Postdoctoral science foundation funded project of central south university provided by Mei Sun.

References

[1] Zhang Y, Xu L, Pan X 2007 Analysis of Hospital Management and Social Awareness *Chinese Hospital Management* **27**(11) 4-5 *(in Chinese)*
[2] Tan T, Huo M, Feng G, Ou B, Liao T 2007 Investigation of the Present Situation of Medical Staff Resources in Public Medical Institutions of Xindu District in Chengdu *Journal of Military Surgeon in Southwest China* **9**(6) 110-3 *(in Chinese)*
[3] Wang Y 2003 Management Concepts That Hospitals Shall Strengthen during Transition *Jiangsu Health Care Management* **14**(6) 21-2 *(in Chinese)*
[4] Brannigan E T, Murray E, Holmes A 2009 Where does infection control fit into a hospital management structure? *Journal of Hospital Infection* **73**(4) 392-6
[5] Diggs L A, Yusuf J E, De Leo G 2014 An update on out-of-hospital airway management practices in the United States *Resuscitation* **85**(7) 885-92
[6] Bakker A R, Leguit F A 1999 Evolution of an integrated HIS in the Netherlands *Int J Med Inf* **54**(3) 209-24
[7] Zhou D, Yu D 2007 Research on Modernization Index System of Capital PE *Journal of Beijing Sport University* **30**(5) 581-5 *(in Chinese)*
[8] Liu H C 2013 A theoretical framework for holistic hospital management in the Japanese healthcare context *Health Policy* **113**(1-2) 160-9
[9] Guo L, Gao J, Yang J, Kang J 2009 Criticality evaluation of petrochemical equipment based on fuzzy comprehensive evaluation and a BP neural network. *Journal of Loss Prevention in the Process Industries* **22**(4) 469-476
[10] Ahmed A H 2014 Prevention and Management of Hospital-Acquired Anemia *Hospital Medicine Clinics* **3**(1) e71-e84
[11] Zou K, Gong Z 2007 Analysis on Customer Satisfaction Index on Community's Information Service *Library Tribune* **27**(5) 38-42
[12] Lu Z, Shen Y 2011 The Study on Venture Capital Project Appraisal using AHP-Fuzzy Comprehensive Evaluation Methods *International Journal of Advancements in Computing Technology* **3(**8) 50-6

Authors

Mei Sun, born in August 1982, Taiyuan City, Shanxi Province, P.R. China.

Current position, grades: lecturer of Nursing school, Central South University, China.
University studies: Doctor degree at Central South University in China.
Scientific interests: nursing quality control, chronic disease management.
Publications: more than 12 papers.
Experience: teaching experience of 2 years, 4 scientific research projects.

Hui Feng, born in June 1973, Changsha City, Hunan Province, P.R. China.

Current position, grades: Associate Professor of Nursing school, Central South University, PR China.
University studies: Bachelor, Master and Doctor degree in Medicine at Central South University of Hunan in China.
Scientific interests: nursing education and chronic disease management.
Publications: more than 20 papers.
Experience: teaching experience of 21 years, 7 scientific research projects.

Siyuan Tang, born in April 1966, Hengyang City, Hunan Province, P.R. China.

Current position, grades: Professor at Nursing school, Central South University, China.
University studies: Bachelor, Master and Doctor degree in Medicine at Central South University of Hunan in China.
Scientific interests: nursing education, chronic disease management.
Publications: more than 120 papers.
Experience: teaching experience of 26 years, 12 scientific research projects.

Ziqiang Luo, born in April 1962, Changsha City, Hunan Province, P.R. China.

Current position, grades: Professor at the basic medical school, Central South University, China.
University studies: Bachelor, Master and Doctor degree in Medicine at Central South University of Hunan in China.
Scientific interests: respiratory physiology.
Publications: more than 102 papers.
Experience: teaching experience of 30 years, 18 scientific research projects.

Chengdu River status and cause analysis

Jian Zhang, Zhongwei Shen, Hao Shen*

School of Architecture, Southwest Jiaotong University, Chengdu, Sichuan, China

Received 1 September 2014, www.cmnt.lv

Abstract

Focusing on the basic model of the riverside landscape of Chengdu's five major rivers, this paper conducts contrastive analysis on the model characteristics of urban riverside landscape, location of the city where it locates and the urban construction status and explains in detail about the relation between the formation reasons of urban riverside landscape and the level of urbanization development. It discusses the effect of urban development and residents demand on urban riverside landscape design so as to perfect the design thought about it.

Keywords: cause analysis, landscape model, correlation analysis, land property, river status

1 Introduction

Water is a resource closely connected to human life and production. Human being is born to be water-loving and "to live by the riverside" and "to build cities in places surrounded by water" was the basic principle of city construction in the ancient time. In nowadays society, urban river is playing a decisive role in urban establishment and development [1]. Owing to the high correlation of urban river with urban development and the lives of residents, human needs and human participation become particularly important in the design and formation of urban riverside landscape.

Focusing on eight most influential rivers which also spread most widely within the urban area of Chengdu, this paper conducts targeted study on the model difference of the riverside landscape models of these rivers and the forming reason of such difference in varied locations in the urban and suburban area. Paying special attention to the development status of urban riverside landscape under the integrative action of various factors such as level variance of urbanization, intra-city location difference and difference of residential demand for rivers, this paper tries to find out the general rule of urban riverside landscape development during urbanization and in the context that the degree of human participation in landscape development increases, and then discusses the development tendency of urban river landscape and the theory of man-water harmony river landscape construction of modern city.

With rainless spring and winter but rainy summer and autumn, Chengdu enjoys ample rainfalls and has a mean annual precipitation of 1,124.6 millimeters. The amount of Chengdu's annual precipitation varies little and the ratio between the maximum and minimum annual precipitation is about only 2:1. Densely covered by waterways, the city has the Minjiang River system in the southwest and Tuojiang River system in the northeast. There are more than 40 rivers in the city, covering a water area of over 700 square kilometers. The average annual water resources quantity is 30.479 billion cubic meters, of which 3.158 billion cubic meters is underground water and 18.417 billion cubic meters is transit water, can basically satisfy the needs of life, production and construction of Chengdu people and the requirements of river landscape. The Minjiang River, Tuojiang River and other ten main streams and dozens of tributaries have constituted a cross-linked river system of Chengdu of a density as high as 1.22km/sq.km; in addition to the world famous Dujiangyan Irrigation Project, various reservoirs, pools, weirs and canals are spreading all over the city.

2 Research method and definition and division of research scope

Research method: The author mainly uses field investigation, literature consultation and comparative research method to conduct a four-season one-year survey on the eight major rivers in the urban area of Chengdu. The visual language of the river landscape is finally made out of pictures and freehand sketching produced based on field shooting and on-spot section and plane drawing of the landscape pattern.

Definition and division of research scope: In terms of scope, the research involves all the river basins all over Chengdu. As for the definition of urban area, the eastern, western and northern parts are all limited by cycle-city roads. Since the city center is now moving toward south, the boundary of the southern part is where Fuhe River and Jiang'anhe River meets in Huayang Town. Then the research scope is further divided into four survey regions according to the city pattern: the intra-First-Ring-Road-region, First Ring Road-Second Ring Road region, Second Ring Road-Third Ring Road region and extra-Third-Ring-Road region.

* ***Corresponding author's*** e-mail:737006715@qq.com

3 Abstraction and classification of river landscape model

Abstraction of river landscape model: To be specific, the eight major rivers are the Fuhe River, Qingshuihe River, Jiang'anhe River, Shahe River, Dongfengqu River and its tributaries, Xiaojiahe River, Xiaojiahe River and Modihe River. The Fuhe River, Qingshuihe River, Jiang'anhe River, Shahe River, Xijiaohe River, Xiaojiahe River and Modihe River are natural while Dongfengqu River is artificial (illustrated with a drainage map of Chengdu urban area). First of all, according to the spatial difference of the universal space of the riparian green land, the author divides the riverside landscapes of these eight rivers and abstracts the model chart of the landscape types. The model chart includes the planar graph of each landscape type and the sectional drawing of the riverbank [2-13]. 100 models are abstracted in all, including 36 models of the Fuhe River, 24 of the Qingshuihe River, 10 of the Jiang'anhe River, 20 of the Shahe River, 14 of the Dongfengqu River and its tributaries, 4 of the Xijiaohe River, 5 of the Xiaojiahe River and 6 of the Modihe River. The Fuhe River comes from the Shidiyan Sluice of Tuanjie Town, Pixian County. At the upstream, there is the Botiaohe River- the main canal of Dujiang Dam, and the Xuyanhe River-a tributary of Zoumahe River. The water passes through the sluice and is divided into two: the left Pihe River heading the Tuojiang River and the right Fuhe Tributary (the current Fuhe River) of which the discharge capacity is 68m3 / s. The Fuhe River begins to flow southeastward at the Shidiyan Sluice and is divided into the Dongfengqu River by the left in Anjing township of Pixian County then the water flows southward into the Jinniu District of Chengdu. The monthly average discharge of the Fuhe River by the Gaoqiao reach has maintained stable in the past ten years. In Dongzikou, the water is divided to the Shahe River (which is further divided into the right Zhuantou Bam and the left Xiwa Bam) by the left and Fuhe River by the right. The monthly average discharge of the Fuhe River is 17.0 m3/ s, which reduces a bit after being split in Dongzikou. In the southeast direction, the river flows across the northern and eastern part of city center and goes through Mengzhuiwan in the east. By the right side at the Nanhe River Estuary in Hejiangting, the river includes the Nanhe River, which comes from the Qingshuihe River, the downstream of the Zoumahe River; then it passes the Jiuyanqiao Bridge and the Wangjianglou Hydrologic Station, covering an intra-regional drainage area of 505 km2. The Fuhe River has a monthly average discharge of 45.9 m3 /s by the Wangjianglou reach. Before the diversion, the average annual discharge of the Dongfengqu River is 59.1 m3 / s, which reduces to 33.7 m3 / s with 5.5 m of water level amplitude after the diversion, as shown in Equation (1):

$$Q' = \beta Q \,,\ \beta = \sqrt{L/h_a}\sqrt[4]{(2h_a - L)/h_a} \,. \tag{1}$$

Classification of river landscape model. Rivers within Chengdu urban area are classified by region, i.e., rivers of the intra-First-Ring-Road-region, rivers of the First Ring Road-Second Ring Road region, rivers of the Second Ring Road-Third Ring Road region and rivers of the extra-Third-Ring-Road region. River landscape models are compared within the research scope. Rivers with similar landscape models are included into one category, and there are ten categories of river landscape models in Chengdu.

3.1 CATEGORY 1

The typical model of category 1 is as shown in Figure 1. Region: beyond the West Third Ring Road and in the direction of the North Third Ring Road.

Major rivers: Fuhe River, Qingshuihe River & Modihe River.

Status of riparian land: Most lands are villages, towns and farmlands; low level of urbanization.

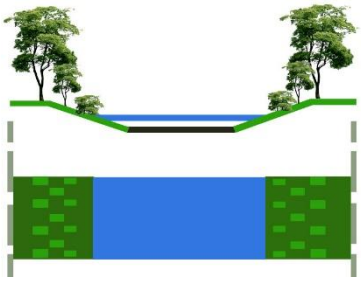


FIGURE 1 The typical model chart of category 1

Status of river: Natural rivers and watercourses; riverbanks are in near-natural status; hardly any manual damming.

Status of landscape: The river landscape is essentially natural. Habitat of the patch core is barely affected by human disturbance, and the landscape patch has high integrity, high species diversity low human participation.

Forming reason for landscape: the status of landscape within this region shows the basic human needs for rivers. Irrigation is the only requirement to be met. Hardly any manual treatment has been exerted upon the rivers.

3.2 CATEGORY 2

The typical model of category of category 2 is as shown in Figure 2. Region: beyond the East Third Ring Road and in the direction of the North Third Ring Road.

Major rivers: Dongfengqu River and its tributaries

Status of riparian land: This region was the industrial and logistics distributing area of Chengdu in the past. Today it has been transformed into a land for residential construction and commercial finance. It also involves some culture and entertainment (in construction).

Higher urbanization level than the western and northern regions beyond the Third Ring Road.

Status of river: Artificial river courses.

Status of landscape: Simple artificial landscape acces-sories and high landscape integrity. Although human

participation has been taken into consideration in the landscape design of some reaches, the level is low.

The 117km Dongfengqu River (with 71km from the Wangjianglou reach to Jiangkou reach) is about 14km long within the city and 28km long out of the city, reaching an average gradient of 1/1000. It is about 100m wide from the Shuangliu Dejiang Temple to the Er'jiangqiao Bridge and about 150m wide from the Erjiangqiao Bridge to the Huanglongxi. The widest part of the Jiandaoyan Bam reach is 265m. The 27km long Zoumahe River, 38km Long Qingshuihe River, 2km long Huanhuaxi River and 6km long Nanhe River, all together, reach a total length of 73km and an intra-regional drainage area of 305 km2. After the integrative waterway regulation of the Dunanhe River, the riverbank has got artificially-built vertical stone embankments and three rubber water dams. The Shahe River is an artificial watercourse built in the industrial zone of the eastern suburbs. After the diversion at Dongzikou, it flows along Chengdu's eastern suburbs toward the southwest and falls into the Fuhe River at around Sanwayao. The reach is about 22km. With a monthly average discharge of 16. 25 m3 / s, the Shahe River is the main source of Chengdu industry and domestic water. It spans 55m at its widest part and only 18.58m at the narrowest part. With an average depth of about 4m, the river can be 2m, 5m, 6m or 7m deep. The average annual discharge is 0.49 billion cubic meters. The total hydrologic discharge is as shown in Equation (2):

$$Q = 1.366KS\left[\frac{L+S}{\lg\frac{2b}{t_0}} + \frac{L}{\lg\frac{0.66L}{t_0} - 0.22arsh\frac{0.44L}{b}}\right]. \quad (2)$$

Forming reason of landscape.

Human needs for rivers: irrigation and human activities.

With the improvement of urbanization, simple artificial landscape begin to appear by the riverbank; human activity is taken into consideration (artificial trails)

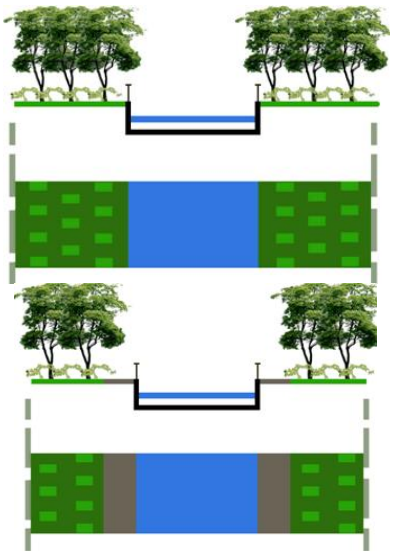


FIGURE 2. The typical model chart of category 2

Due to the factors of regional land property (industrial &logistics distributing area), human requirement for river landscape hasn't been fully considered in the initial landscape setting. Although there is human participation, the level is low.

3.3 CATEGORY 3

The typical model of category 3 is as shown in Figure 3. Region: beyond the West Third Ring Road and in the direction of the South Third Ring Road.

Major rivers: Jiang'anhe Rive.r

Status of riparian land: Most of the city-rounding water runs through farmlands and enters into the city in the direction of southwest before flowing through the residential district. The level of urbanization turns from very low to high gradualll.

Status of riparian land: Natural watercourse with both near-natural and artificially treated reaches.

Status of landscape: transformation from static natural landscape to artificial landscape model.

Forming reason of landscape:Human needs for rivers: irrigation and human activities. With the improvement of urbanization, the riverside landscape transforms gradually from static natural model to artificial model. The degree of human participation increases significantly within the landscape patch when the river flows through the residential district.

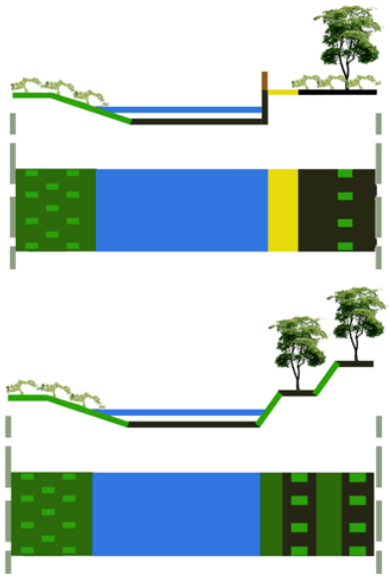


FIGURE 3 The typical model chart of category 3

3.4 CATEGORY 4

The typical model of a category 4 is as shown in Figure 4. Region: beyond the Third Ring Road toward the south of the city.

Major rivers: Fuhe River & Xiaojiahe River.

Status of riparian land: Locating in the High-tech Zone and New Tianfu Zone of Chengdu, it will become the core area of the future urban construction of Chengdu

with integrated function of residence, finance and government administration. Very high level of urbanization.

Status of river: natural rivers and complex artificial landscape after manual treatment.

Status of landscape: Large green area Particularly effective laying in landscape design. Complex landscape structure. Abundant plant species.

Since the landscapes of this region are all artificially developed and the degree of human participation is extremely high, landscape habitat suffers intense interfere. Thus, although the landscape patch covers a large and complete area, the biological diversity is low.

Forming reason of landscape:Human needs for rivers: urban style & human participation;The landscape of this region, playing an important role in city image display, is developed based on strongly goal-directed artificial modification of rivers.

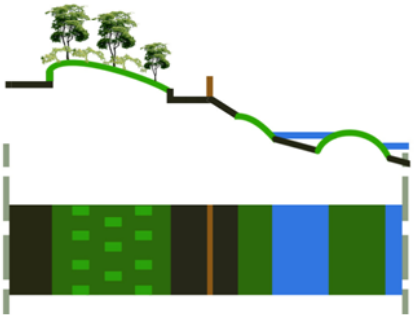


FIGURE 4 The typical model chart of category 4

3.5 CATEGORY 5

The typical model of category 5 is as shown in Figure 5. Region: between the North Third Ring Road and the North Second Ring Road.

Major rivers: Fuhe Rive.r

Status of riparian land: Residential communities increase from the Third Ring Road to the Second Ring Road. Wholesale and distributing zone of commodity and building materials. Very high level of urbanization.

Status of river: natural rivers & artificial river landscape.

Status of landscape: Artificial landscape with low degree of human participation. Human participation hasn't been taken into consideration in the landscape design of some reaches.

Forming reason of landscape: Human needs for rivers: irrigation and human activities. Owing to the goal-oriented property of land, human participation hasn't been taken into consideration in the landscape setting of some reaches. Landscape mod1els are mixed and disorderly with low use ratio. There is no effective maintenance for the rivers and river landscape. After flowing into the residential zone, the landscape models tend to be the same and the degree of human participation in river landscape increases.

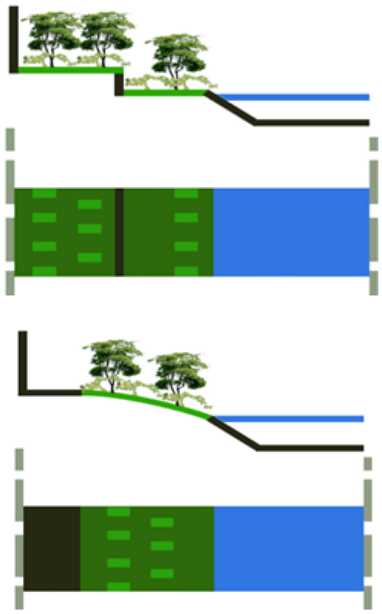


FIGURE 5 The typical model chart of category 5

3.6 CATEGORY 6

The typical model of category 6 is shown in Figure 6. Region: between the North Third Ring Road and the North Second Ring Road & between the East Second Ring Road and the East Third Ring Road.

Major rivers: Shahe River.

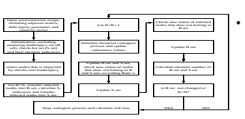
Status of riparian land: Shahe River Source Region Ecological Preservation Area, Wetland parks (Tazishan Park, etc.), Culture and entertainment industry (Eastern Music Park), Residential zone, Very high level of urbanization.

Status of river: natural rivers & artificial river landscape.

Status of landscape: The landscape patch is large, and the green land area is larger than all the other river landscape patches of Chengdu. The degree of human participation is relatively low; it also has rich tree species, the maximum vegetation quantity and high biological diversity.

Forming reason of landscape:Original ecological (river ecological preservation area) and artificial landscapes (residential zone and city park) are cross-linked. In a natural state under artificial protection and treatment, the landscape of this region is artificial transformation based on the ecological preservation of river landscape.

The model of the landscape environment of this region is as shown in Equation (3):



(3)

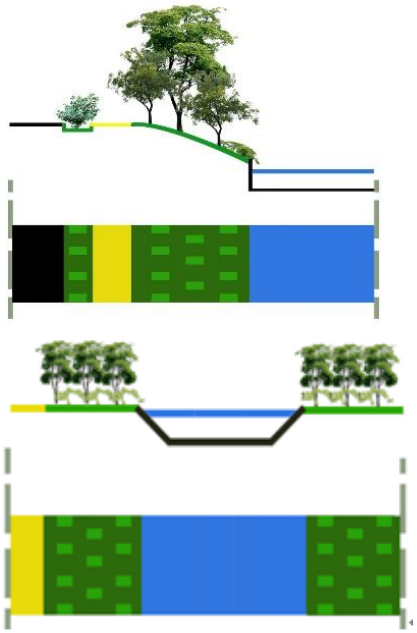


FIGURE 6 The typical model chart of category 6

3.7 CATEGORY 7

The typical model of category 7 is as shown in Figure 7. Region: between the South Third Ring Road and the South Second Ring Road.

Major rivers: Fuhe River.

Status of riparian land: Central urban area (mainly residential communities). Wetland parks. New urban zone (mainly residential communities). Very high level of urbanization.

Status of river: natural rivers & artificial river landscape.

Status of landscape: as described in the region within the First Ring Road: Shahe River Wetland Park and New South Zone.

Forming reason of landscape: Conditions of central urban area (within the First Ring Road). Transition from protected natural state to the strongly artificially treated state beyond the Third Ring Road.

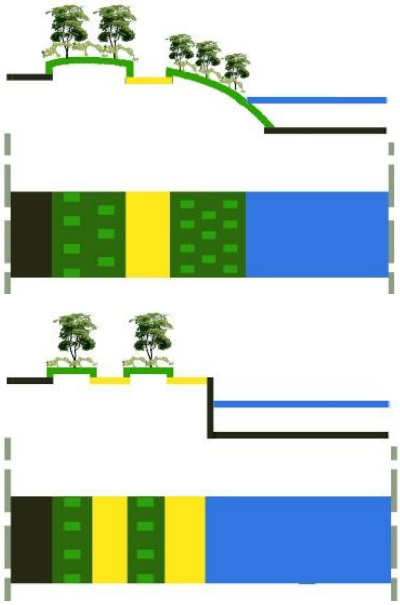


FIGURE 7 The typical model chart of category 7

3.8 CATEGORY 8

The typical model of category 8 as shown in Figure 8. Region: between the Second Ring Road to the First Ring Road.

Major rivers: Fuhe River, Nanhe River (the alternative name of the Qingshuihe River within the Second Ring Road) & Shahe River.

Status of riparian land: Central urban area (mainly residential communities). Very high level of urbanization.

Status of river: natural rivers & artificial river landscape.

Status of landscape: narrow green areas and small patches.

Forming reason of landscape: The landscape setting is mainly human-activity and human-demand oriented. Residential land demand leads. On the premise of maximum residential lands and satisfaction of human needs for riparian green space landscape, the landscape model with small green area in landscape patch and high degree of human participation is formed.

The relation between landscape green & degree of human participation and hydrologic data is as shown in Equation (4):

$$\begin{cases} \frac{\partial}{\partial x}(K_{xx}\frac{\partial h}{\partial x})+\frac{\partial}{\partial y}(K_{yy}\frac{\partial h}{\partial y})+\frac{\partial}{\partial z}(K_{zz}\frac{\partial h}{\partial z})-W+Q_i\delta(x-x_i,y-y_i,z-z_i)=S_i\frac{\partial h}{\partial t}t \\ H(x,y,z,t_0)=H_a(x,y,z)\big|_{t=0} \\ H(x,y,z,t)=H_b(x,y,z,t)\big|_{t\neq 0} \\ K_{xx}\frac{\partial h}{\partial x}\cos(n,x)+K_{yy}\frac{\partial h}{\partial y}\cos(n,y)+K_{zz}\frac{\partial h}{\partial z}\cos(n,z)\big|\Gamma_1=0, t\geq t_0, x,y,z\in\Gamma_1 \\ K_{xx}\frac{\partial h}{\partial x}\cos(n,x)+K_{yy}\frac{\partial h}{\partial y}\cos(n,y)+K_{zz}\frac{\partial h}{\partial z}\cos(n,z)\big|\Gamma_{2,3}=q_1(x,y,z,t), t\geq t_0, x,y,z\in\Gamma_{2,3} \end{cases} . \quad (4)$$

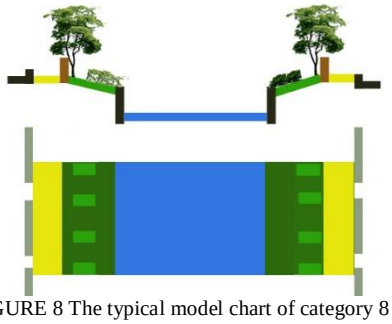


FIGURE 8 The typical model chart of category 8

3.9 CATEGORY 9

The typical model of category 9 is as shown in Figure 9. Region: within the First Ring Road.

Major rivers: Fuhe River & Nanhe River (the alternative name of the Qingshuihe River within the Second Ring Road).

Status of riparian land: Central urban area (mainly residential communities and commercial districts); Very high level of urbanization.

Status of river: natural rivers & artificial river landscape.

Status of landscape: Landscape node accounts the most in landscape setting. It has complex node patches and simpler design of corridor landscapes that connect the nodes.

Forming reason of landscape: the landscape setting is mainly directed at city image display (landscape avenues) and human needs. City style display, landscape avenues, broad greenbelts, exquisite landscape design, abundant vegetation, complex structure and high degree of human participation.

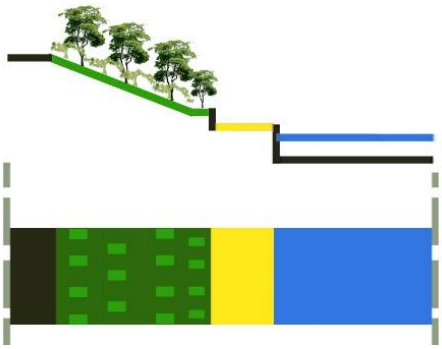


FIGURE 9 The typical model chart of category 9

3.10 CATEGORY 10

The typical model of category 10 is as shown in Figure 10. Region: within the First Ring Road.

Major rivers: Xijiaohe River and its tributaries.

Status of riparian land: Central urban area (mainly residential communities); Very high level of urbanization.

Status of river: natural rivers & artificial river landscape.

Status of landscape: Since both sides of the river landscape are very close from the residential communities, the river landscape is quite narrow. For the landscape setting, there are only very narrow green lands and one or two rows of border tree.

Forming reason of landscape: The primary goal of the landscape setting is to meet the human needs in residence and transportation, thus the width of the riparian landscape is narrow and the landscape is monotonous. The model of this category is as shown in Equation (5):

$$R = \left\{ r_0^2 + 30 \times K \times S^2 (1 + 0.00015 r_0^2) \right\}^{0.5} \quad (5)$$
$$r_0 = \sqrt{A/\pi}$$

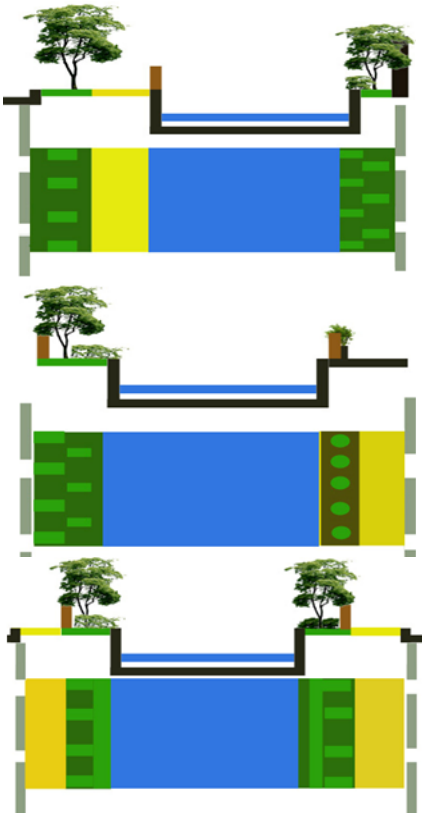


FIGURE 10 The typical model chart of category 10

4 Summary

The landscape model of urban rivers develops gradually with the urbanization process. As city develops, river landscape of areas with low level of urbanization goes through artificial transformation and gradually turns from the original state to a highly artificialized state.

With the improvement of theories of urban planning and urban landscape design, urban river landscape design changes from the old urban center model (Landscape nodes are artificially developed within the First Ring Road, and the links between nodes are weak) to the new city urban center model (large complete landscape).

Based on the correlation analysis between the status of landscape model and the property of the surrounding land, the future river landscape model of a certain region can be predicted.

River landscape models adapting to the property of various urban lands and regions with different levels of economic development can be found out through the study.

5 Conclusion

Based on the correlation analysis between the status of landscape model and the property of the surrounding land, this paper finds out various typical river landscape models under the influence of various land properties and different levels of economic development and explores the formation reasons of these models, so that referable models and practical design thoughts can be provided for the prospective river landscape design directed at different reaches of the city.

Acknowledgement

This paper is subsidized by the NSFC (Natural Science Foundation of China, No. 51208428).

Reference

[1] Chen X 2006 Theoretical research on construction of urban river by promoting people and water harmoniously *China Institute of Water Resources and Hydropower* Peking *(in Chinese)*
[2] Chen X 2010 Spatial analysis of urban river landscape by ecological corridor principle *Chinese Garden* 57-82 *(in Chinese)*
[3] Wu J 2007 Landscape Ecology *Higher Education Press* Peking
[4] Guo W 2008 Planning Course of American Landscape *Landscape Garden* **8**(2) 79-83 *(in Chinese)*
[5] Fu F, Dong L 2010 Analysis of research status on urban river landscape planning and design *Research on Urban Development* (12) 8-11 *(in Chinese)*
[6] Forman R T T, Godron M 1986 Landscape Ecology *John Wiley and Sons* New York
[7] Newbold J D, Erman D C, Roby K B 1980 Effects of logging on macro invertebrates in streams with and without buffer strips *Canadian Journal of Fisheries and Aquatic Science* (37) 1076-85
[8] Large A R G, Petts G E 1996 Rehabilitation of river margins *River Restoration* **71** 106-123
[9] Rohling J Corridors of Green 1998 *Wildlife in North Carolina* **5** 22-7
[10] Deng Yi 2007 *Planning method on urban ecological park design*, China Architecture and Building ,Press:Peking *(in Chinese)*
[11] Toshio 2000 *Application on ecological engineering in civil engineering* Shinyama-sha Cytec Co Ltd:Tokyo
[12] Corporation Legal 2001 Advanced building technology by revetment in river *Shanhai Church*: Tokyo
[13] Li Y, Yang H 2006 Research method on River ecosystem restoration *People Pearl* (2) 16-9 *(in Chinese)*
[14] Fu F 2011 Spatial analysis of Urban river landscape space by ecological corridor principle *Chinese Garden (in Chinese)*

Authors

Jian Zhang, born in October 1987, Chengdu, Sichuan Province, P.R. China.

Current position, grades: Doctor at the School of Architecture, Southwest Jiaotong University, China.
University studies: master's degree in architecture at Southwest Jiaotong University in China.
Scientific interests: human settlement and river landscape.
Publications: 3 papers.
Experience: 4 scientific research projects.

Zhongwei Shen, born in October 1965, Chengdu, Sichuan Province, P.R. China.

Current position, grades: professor at the School of Architecture, Southwest Jiaotong University, China.
University studies: master's degree in Logistical Engineering University of PLA in China.
Scientific interests: human settlement and traffic building.
Publications: more than 20 papers.
Experience: teaching experience of more than 20 years, 6 scientific research projects.

Hao Shen, born in June 1988, Chengdu, Sichuan Province, P.R. China.

Current position, grades: master at the School of Architecture, Southwest Jiaotong University, China.
University studies: bachelor's degree in architecture from Southwest Jiaotong University in China.
Scientific interests: human settlement and river landscape.
Publications: 1 paper.
Experience: 1 scientific research project.

Coal mining subsidence data extraction and verification in a high groundwater area based on Landsat-8 imagery and subsidence prediction

Wu Xiao[1,3*], Guanghua Yang[1], Yaoqi Yang[2]

[1]*Institute of Land Reclamation & Ecological Restoration, China University of Mining and Technology (Beijing), D11 Xueyuan Road, Beijing, China*

[2]*School of Economics, Peking University, 5 Xiheyuan Road, Beijing, China*

[3]*State Key Laboratory of Coal Resources and Safe Mining, China University of Mining & Technology (Beijing), D11 Xueyuan Road, Beijing, China*

Received 16 August 2014, www.cmnt.lv

Abstract

Coal is the main energy resource in China, with its extraction and utilization playing an important role in national economic development. However, coal mining may be causative with respect to critical land subsidence and damage to land. The eastern plain coal mining region of China represents an example of overlapping crop cultivation and coal extraction, and is considered a coal mining area characterized by a high water table. Accordingly, declining ground elevation and seasonal water logging of land due to mining subsidence have become major concerns within the region. Based upon the existing procedure for land reclamation planning, both land damage boundaries and land damage magnitude were determined for the region via subsidence prediction and vertical displacement, respectively. In the current study, a coal mine in Shandong province was employed as a case study area with the following work phases implemented: 1) subsidence prediction was implemented, with land damage magnitude demarcated via the proposed standard; 2) mining induced water area and wetland was extracted via use of remote sensing; 3) comparative analyses of the aforementioned methodologies were undertaken, with a revised methodology proposed for effective provision of improved land damage demarcation. Results indicate that land subsidence at Dongtan coal mines during May 2013 was 1616.70 hm^2, of which categorically mild, moderate, and severe lands were 22.54 hm^2, 257.67 hm^2, and 436.49 hm^2, respectively.

Keywords: Landsat-8, high groundwater, mining subsidence, land damage information, subsidence prediction, damage assessment

1 Introduction

Coal mining induced mining subsidence is a pressing current issue in China as 92% of coal resources are currently extracted from subsurface mining operations [1]. Land characterized by mining induced subsidence in China reached one million hectares by late 2013 and is expected to increase at a rate of 70,000 hectares per year [2]. Regions in eastern China, including Henan province, Shandong province, Jiangsu province and Anhui province, are typified by high water tables and extensive coal mining operations, thus resulting in widespread surface water-logging. The aforementioned provinces produced a total of 230 million tons during the first half of 2014, representing 15.35% of national extraction. It is estimated that half of the land which has undergone subsidence may be entirely inundated due to high groundwater tables in parts of eastern China [3]; subsequent decreases in grain production may result from farmland having undergone subsidence.

Land damage assessment and classification is crucial for effective land reclamation and compensation, with compensatory standards based primarily on vertical displacement or water-logging extent [4,5]. The spatial extent of water-logging may vary due to topographic undulations, however, the use of vertical displacement as a single indicator for demarcation of land damage magnitude is not an appropriately accurate reflection of actual conditions.

With the recent development of remote sensing (RS) techniques, much attention has been given to subsidence boundary extraction based upon multi-source remote sensing data [6]. High resolution imagery is now customarily employed, with RS and geographic information system (GIS) technology also utilized within land subsidence monitoring [7-9]. The data used in these studies are typically high resolution imagery (e.g. Quickbird, D-InSAR or SAR) [10] and characterized by complex workflows. The Landsat-8 satellite was launched in February 2013, thus providing a new data source. Coal mining subsidence data extraction and verification in high groundwater areas based on Landsat-8 imagery and subsidence prediction may thus represent an effective and accurate reference for land compensation and reclamation efforts. Accordingly, the current study objectified the development of methodologies for coal mining subsidence data extraction based on Landsat-8 imagery and land subsidence prediction.

2 Study Area

The current study investigated the DT (Dongtan) coal mine which was developed and is managed by the Yanz-

[*]***Corresponding author's*** e-mail: xiaowuwx@126.com

hou Coal Group. The DT coal mine is located in the southwest of the Shandong province, China (Longitude; 116°50′49″ - 116°56′56″ Latitude; 35°24′11″ - 35°31′25″ (Figure 1). The natural elevation is 42-54 m above mean sea level, in concurrence with a relatively flat pre-mining geomorphology, with a higher elevation to the northwest. The region is characterized by a continental marine transitional climate, with abundant sunshine and four distinct seasons. Annual precipitation occurs within the range 259 - 1263 mm (Annual Mean 715 mm), with the majority occurring during August and September. The area encompasses five districts, spanning three counties within Jining City.

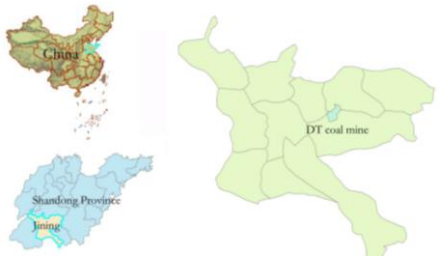


FIGURE 1 Study area: DT coal mine, Shandong province, China, with county boundaries

DT coal mine has an overall spatial extent of 59.96 km^2, and is thus one of the largest underground coal mines in eastern China, with a designed annual level of production equating to 7.5 million tons (Figure 2). By late 2010, the volume of proven coal reserves was 123.85 million tons. Due to the mines location within a highly productive agricultural district within the eastern Chinese plains, the area is characterized as a primary region for both coal and grain production. Mining activities have an inherent potential for damage to adjacent agricultural areas; accordingly, a local conflict of interest between people and land-use may arise, with land protection considered a priority. Thus, in order to effectively implement reclamation and/or compensatory measures, accurate subsidence monitoring and land damage demarcation is imperative.

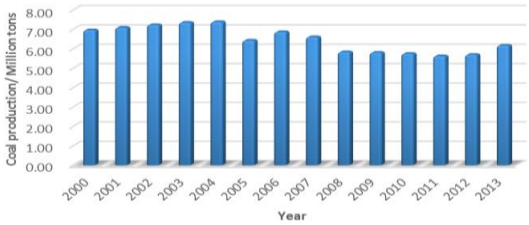


FIGURE 2 Coal production at DT coal mine from 2000 to 2013

3 Land damage magnitude

3.1 LAND DAMAGE MAGNITUDE DEMARCATION

Both surface and subsurface coal mining are characterized by potentially adverse impacts with respect to adjacent land. Subsurface mining may cause significant land subsidence, in addition to coal waste dumping at the surface. Accordingly, numerous environmental challenges arise, including soil erosion, dust and noise production, water pollution and impacts on local biodiversity. Subsided land arising from adjacent subsurface mining operations may undergo tensile and compressive forces, with subsequent horizontal and vertical displacement.

Within eastern China, due to its high mean groundwater table and relatively flat topography, the potentially adverse influence on adjacent agriculture caused by subsidence ponding is considered the most pressing issue. Thus, vertical displacement represents a key indicator for the reflection of local land damage. According to the Land Damage Assessment Rule as set out by Jining city administrative authority, subsided land may be divided into three distinct categories (Table 1), thus enabling classification of the magnitude of land damage.

TABLE 1 Land damage assessment standards in eastern China

Damage degree	Vertical displacement	Description
1 (Mild)	≤2m	Soil erosion, soil and water loss due to the changing inclination of the land
2 (Moderate)	2m~4m	Seasonal land water logging
3 (Severe)	≥4m	Water logging

3.2 DT COAL MINE LAND DAMAGE MAGNITUDE

The probability integration method of mining subsidence is the most frequently used subsidence prediction method in China; according to the probability integration method, the entire mining area is sub-divided into infinite mining units, with adverse influences caused by mine exploittation equal to the sum total of all mining units. The subsidence basins produced by mining units are represented by a normal (Gaussian) distribution and are consistent with the probability density distribution based upon stochastic medium theory. Accordingly, the subsidence profile equation represented by mine exploitation may be expressed as the integral formula of the probability density function. The subsidence basin may be expressed as follows:

$$w_e(x) = \frac{1}{r} e^{-\pi \frac{x^2}{r^2}}, \tag{1}$$

where r is the major influencing radius (primarily related to mining depth and major influencing angle). As shown in Equation (1), the functional form of the subsidence basin is analogous to the normal probability density function [11]. Based on the probability integration method, mining subsidence contours were obtained with respect to coal mining initiated from May 2013, with land damage magnitude demarcation concurrently undertaken (Figure 3).

FIGURE 3 Demarcation of land damage based on predicted vertical displacement

4 Land classification using Landsat-8 data

Landsat data have enabled continuous global monitoring of anthropogenic and natural land cover disturbance since 1972. Recent deteriorations in performance and service intermittence with respect to the Landsat-7 and Landsat-5 sensors have raised concerns surrounding the condition of global observation programs. Nonetheless, Landsat imagery remains a valuable data source for landscape change detection. Landsat-8 data are processed and presented as 185 × 180 km Level-1 terrain-corrected (L1T) products, with a 950 MB compressed GeoTiff file size (over twice that of previous Landsat L1T products). All OLI and TIRS spectral bands are stored as geo-located 16-bit digital numbers within the same L1T file [12]. All 100 m TIRS bands are re-sampled via cubic convolution to 30 m resolution and co-registered with 30 m OLI spectral bands. An associated metadata file stores spectral band gain and offset numbers that may be employed for linear conversion of digital numbers to 'at-sensor' radiance and conversion of OLI digital numbers to 'at-sensor' reflectance (i.e. unitless). Accordingly, it is not necessary to perform non-linear transformation from radiance to reflectance [13].

4.1 MATERIALS

In the current study, contemporary Landsat-8 satellite imagery with OLI Land Imager data was selected for May 21, 2013 (Path/Row No122/35, 122/36). Employed imagery included nine bands, with an associated spatial resolution of 30 m and a 15 m panchromatic band; Landsat-8 imagery increased two-band data, thus resulting in increasingly subtle division, in comparison with ETM+ imagery.

Data were subjected to both radiometric and geometric correction (UTM-WGS84 projection), with pre-treatment of raw data comprising radiometric calibration and atmospheric correction. Original images underwent integration of multi-bands and the panchromatic band, thus improving the resolution of multi-band imagery. The Gram-Schmidt pan sharpening method provided by ENVI 5.1 was employed expressed as follows:

$$HMS_i = LMS_i + w_t(HRP - LRP), w_t = \rho(LMS_i, LRP)\frac{\sigma(LMS_i)}{\sigma(LRP)}, \quad (2)$$

where *LMS* represents multispectral imagery after re-sampling, *HMS* represents multispectral imagery after sharpening, *LRP* represents low-resolution panchromatic imagery, *HRP* represents high resolution panchromatic imagery, w represents spatial detail injection coefficient, ρ represents the correlation coefficient and σ represents standard deviation.

The DT coal mine boundary vector layer was used to obtain the subset image originally acquired from Landsat-8 via overlaying the vector layer with RS imagery.

4.2 CLASSIFICATION

In order to accurately classify land use, a false color composite with Band 5, 4, 3 was selected Figure 4:

FIGURE 4 DT coal mine land use false colour composite with Band 5, 4, 3

With Bayes maximum likelihood classification function utilized for image classification. Based on fieldwork and visual interpretation of the subset image, a supervised classification was performed using a classification workflow tool on the ENVI 5.1 platform. Five land-use classes were identified including farmland, construction land, land under water, wetlands and bare land. Classification accuracy was verified via a stratified random sampling method, with 60 samples randomly distributed into five land-use classifications. Reference data were collated via fieldwork and land-use mapping. Overall classification accuracy was 80.15%. Land cover classification using Landsat-8 imagery is presented in Figure 5.

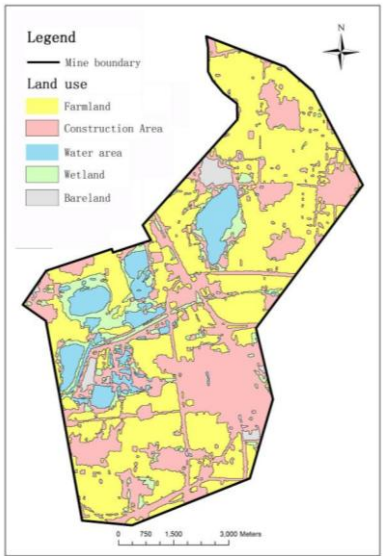


FIGURE 5 Land use classification of DT coal mine based on Landsat-8 imagery

5 Results and Conclusions

5.1 RESULTS

In accordance with the Chinese Law of Land Management, coal mining operations must undertake appropriate requisition and compensation for damaged agricultural land, based upon crop yield reduction due to mining-induced subsidence. According to the Land Reclamation Regulation (LRR), as issued and implemented by State Council, coal mines are required to provide an appropriate land reclamation plan, including identifycation of disturbed adjacent land and reclamation arrangements [14]. Lands requiring reclamation are divided into one of three functional categories (mild, moderate or severe); moderately and severely damaged lands correspond with seasonal and permanent area under water, reflected by wetlands and land under water in Landsat-8 imagery. Due to undulating topography, predicted maps developed as part of the current study cannot reflect true conditions (Figure 6).

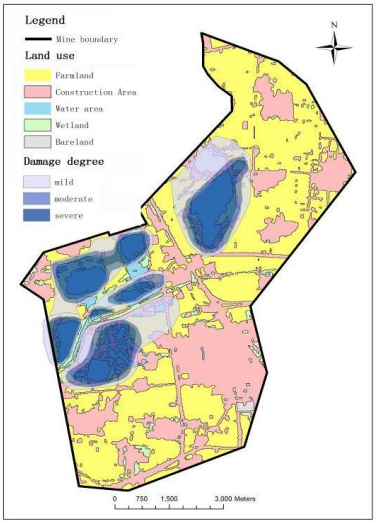


FIGURE 6 Overlapped Landsat-8 imagery and predicted vertical displacement

Results from the current study predict a total subsided area (up to May 2013) of 1616.7 hm^2 within the study area, with categorically mild, moderate and severe land damage predicted to be 622.99 hm^2, 251.44 hm^2 and 742.27 hm^2, respectively. Based upon Landsat-8 imagery extraction, the true conditions equate to 922.54 hm^2 (mild), 257.67 hm^2 (moderate) and 436.49 hm^2 (severe) (Table 2).

TABLE 2 Subsided area comparison between study predictions and Landsat-8 extraction

No	Degree	Predicted/hm^2	Extracted by Landsat-8/hm^2
1	Mild area	622.99	922.54
2	Wet land	251.44	257.67
3	Water area	742.27	436.49
Total		1616.70	1616.60

As shown (Table 2), there were significant differences noted between study predictions and Landsat-8 extractions; accordingly, a comprehensive methodology and associated workflow has been proposed in order to provide increasingly accurate land damage imagery:

1) Implementation of mining subsidence predictions based upon mine layout. Subsidence boundary obtained via 10mm vertical displacement.
2) Land use classification based upon contemporary Landsat-8 imagery, particularly with respect to wetlands and areas under water
3) Overlapping of maps obtained from Steps 1 and 2; determination of damaged land boundary from Step 1, with regions characterized by moderate and severe land damage represented as wetlands or areas under water from Landsat-8 imagery.

Via application of the proposed methodology outlined above, DT coal mine land damage demarcation was undertaken, as shown in Figure 7.

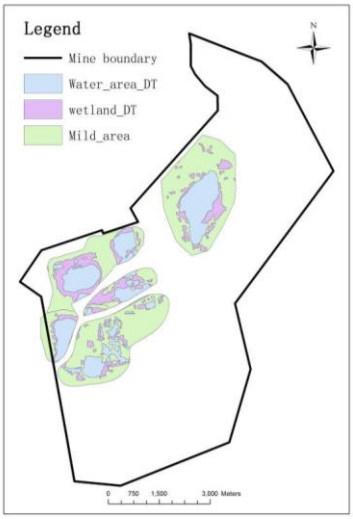


FIGURE 7 Land management map based on vertical displacement predications and Landsat-8 extraction

5.2 CONCLUSIONS

The current study presents a novel methodology for accurate determination and assessment of zones of damaged land for subsequent crop compensation, land requisition and land reclamation in the eastern plain coal mining area of China. The primary study findings and conclusions were as follows:

1) Subsidence prediction was implemented, and land damage magnitude was demarcated via the current proposed standard.
2) Mining induced areas under water and wetlands were extracted by using remote sensing technology.
3) A comparison between these two methodologies revealed that the comprehensive method could provide a more realistic reflection of conditions for subsided land demarcation. Taking DT coal mine as a case study area, the predicted subsided land total was 1616.70 hm^2, comprising 922.54 hm^2 mild, 257.67 hm^2 moderate and 436.49 hm^2 severe.

Acknowledgement

The research was supported by the National Science Foundation of China (Grant No. 41401609), and the State Key Laboratory for Coal Resources and Safe Mining, China University of Mining & Technology (App. No. SKLCRSM12KFB07). The authors would like to extend their thanks for valuable assistance provided by DT coal mine.

References

[1] Hu Z Q, Xiao W 2013 Optimization of concurrent mining and reclamation plans for single coal seam: a case study in northern Anhui, China *Environmental Earth Sciences* **68**(5) 1247-54 *(in Chinese)*
[2] Yun W J, Fan J M 2012 Research on land reclamation in resource-exhausted cities *China Development* **12**(5) 19-23 *(in Chinese)*
[3] Xiao W, Hu Z Q, Chugh Y P, Zhao Y L 2014 Dynamic Subsidence Simulation and Topsoil Removal Strategy in High-Groundwater Table and Underground Coal Mining Area- A Case Study in Shandong Province *International Journal of Mining, Reclamation and Environment* **28**(4) 250-63 *(in Chinese)*
[4] Hu Z Q 1996 Damage characteristics of the cultivated land caused by mining subsidence in the east plain *Coal Mine Environ Protection* **3** 6-10 *(in Chinese)*
[5] Hu Z Q, Xu X L, Zhao Y L 2012 Dynamic monitoring of the land subsidence in mining area from multisource remote-sensing data – a case study at Yanzhou, China *International Journal of Remote Sensing* **33**(17), 5528-45 *(in Chinese)*
[6] Xu X L, Zhao Y L, Hu Z Q, Yu Y, Shao F 2014 Boundary demarcation of the damaged cultivated land caused by coal mining subsidence *Bulletin of Engineering Geology and the Environment* **73** 621-33 *(in Chinese)*
[7] Peng S P, Wang L, Meng Z P, Duan Y N, Bian J L, Wang Y P 2002 Monitoring the seeper subside in coal district by the remote sensing-examples from Huainan coal district *Journal of China Coal Society* **27**(4) 374-378 *(in Chinese)*
[8] Du P J, Guo D Z 2003 The Extraction of Mining Subsiding Land from RS Image Supported by GIS *Journal of Image and Graphics* **8**(2) 231-5
[9] Thompson J A, Lamb D W, Frazier P S, Ellem B 2011 Monitoring the effects of longwall mine-induced subsidence on vineyards *Environmental Earth Sciences* **62**(5) 973-84
[10] Carnec C, Delacourt C 2000 Three years of mining subsidence monitored by SAR interferometry, near Gardanne, France *Journal of Applied Geophysics* **43**(1) 43-54
[11] He G Q 1991 Mining subsidence *China University of Mining and Technology Press* Jiangsu *(in Chinese)*
[12] Roy D P, Ju J, Kline K, Scaramuzza P L, Kovalskyy V, Hansen M C, Loveland T R 2010 Web-enabled Landsat Data (WELD): Landsat ETM+ Composited Mosaics of the Conterminous United States *Remote Sensing of Environment* **114** 35-49
[13] Roy D P, Wulder M A, Loveland T R 2014 Landsat-8: Science and product vision for terrestrial global change research *Remote Sensing of Environment* **145** 154-72
[14] Ming L, Hu Z Q, Li J 2011 Legal construction of land reclamation in China *China Land* **7** 44-6 *(in Chinese)*

Authors

Wu Xiao, born in September 1983, Lianyuan County, Hunan Province, P.R. China.

Current position, grades: instructor at the School of Geoscience and Survey Engineering, China University of Mining and Technology (Beijing), China.
University studies: MSc and DSc in Geodesy and Surveying Engineering at China University of Mining and Technology (Beijing) in China.
Scientific interests: GIS, RS, land reclamation, and land planning.
Publications: more than 20 papers.
Experience: teaching experience of 2 years, 3 scientific research projects.

Guanghua Yang, born in February 1982, Guangan County, Sichuan Province, P.R. China.

Current position, grades: PhD student in Land Resource Management at the School of Geoscience and Survey Engineering, China University of Mining and Technology (Beijing), China.
University studies: MSc in Geographic Information System at Chinese Academy of Sciences in China.
Scientific interests: land reclamation, ecology protection.
Publications: more than 10 papers.

Yaoqi Yang, born in December 1988, LinYi City, Shandong Province, P.R. China.

Current position, grades: Postdoctoral work at the School of Economics, Peking University.
University studies: MSc in MBA from Missouri State University in US. DSc in Land Resource Management at China University of Mining and Technology (Beijing) in China.
Scientific interests: land policy, land economics, land reclamation and ecological reconstruction.
Publications: more than 23 papers.
Experience: town government work experience of 1 years, 2 scientific research projects.

Computational model of implicit interaction for entertainment

Wei Wang*, Xiaodan Huang

School of Information & Electrical Engineering, Hebei University of Engineering, Hebei Handan 056038, China

Received 21 May 2014, www.cmnt.lv

Abstract

Implicit interaction between human and computer is worthy of being researched, especially in the entertainment application. The reason is that in order to be more natural, computers need to interact and collaborate with persons actively. For this purpose, a computational model of implicit interaction is proposed and applied to a computer for entertainment. Firstly, emotional Hidden Markov Model (eHMM) as a part of the computational model of implicit interaction is researched. Then, three parts of ACT-R cognitive architecture are integrated into it to apply for entertainment. Finally, some experiments are carried out with styles of game process recording. Results indicate that the proposed model is helpful to make computers more active and adaptive to persons by adjusting entertainment process, which illustrates a good prospect of application.

Keywords: implicit interaction, ACT-R, affective computing, entertainment

1 Introduction

In order to make the participants more focus on interactive content without interaction devices, human-computer interaction (HCI) need to be expanded from the traditional interactive style, explicit HCI (EHCI), to the ubiquitous one, implicit HCI (IHCI). This new interactive style can reduce the user's cognitive burden. So research on its theories and technologies becomes more and more important.

As a frontier of HCI area, IHCI has draw lots of attention from many research organizations [1]. Nicole Kaiyan in Swinburne University of Technology in Australia proposed the concept of IHCI in 1996, but did not research deeply [2]. Gradually, from 2005, universities and institutes in many countries, such as USA, Germany, China, Australia, and so on, studied on IHCI deeper and deeper. Albrecht Schmidt in the University of Karlsruhe in Germany worked on IHCI theories earlier. He regarded perception and interpretation as the key point of IHCI, considered that context information is extremely important for the interactive process, and modelling the interaction based on extensible markup language (XML) [3]. With computer vision technologies, Andrew Wilson and Nuria Oliver in Microsoft Research of the USA developed four systems to realize IHCI process [4]. In 2007, adaptive vision system was developed by Tao Linmi in Tsinghua University in China, which can detect and understand users' behavior through implicit interactive style [5]. At the same time, Tian Feng in the Institute of Software, Chinese Academy of Sciences, researched on the implicit interaction features from the aspect of post-WIMP [6]. The development of IHCI is sorted as shown in Figure 1.

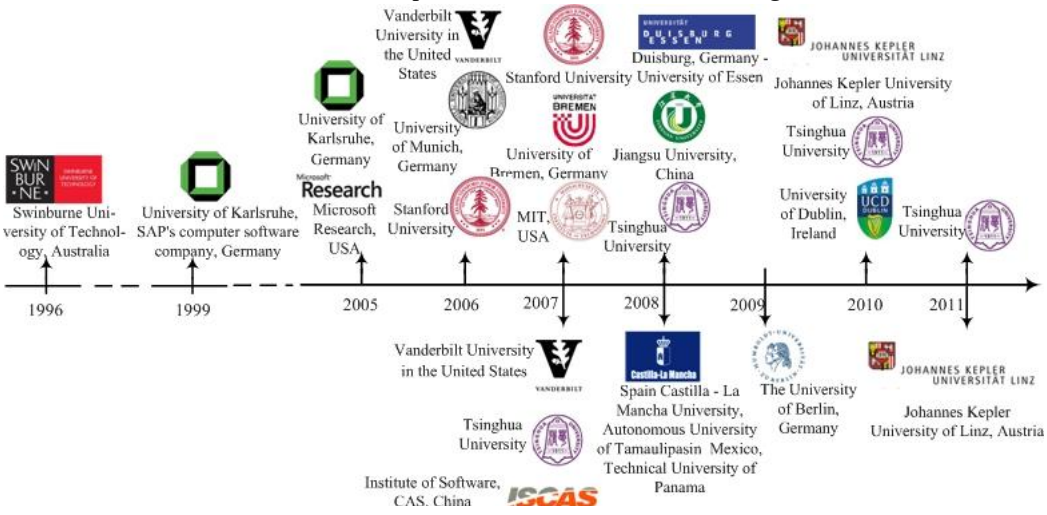


FIGURE 1 Developing process of IHCI

*Corresponding author's e-mail: wangwei@hebeu.edu.com

Entertainments also need the IHCI. With users' context information, such as behavior, emotional states, physiological status, location, and et al, the amusement quality can be improved by context perception and interpretation.

The rest of the paper is organized as follows. In Section 2, a brief literature review is provided on affection modeling and cognitive-affective interactions in entertainment. Section 3 presents a computational model of implicit interaction for entertainment based on adaptive control of thought-rational (ACT-R) cognitive architecture model. Section 4 discusses realizations of a card and a chess playing processes with the model above, analyzes the real experiment results and compare with the processes without the model. Section 5 provides the conclusions and the future work.

2 Related Works

2.1 AFFECTION MODELING

Currently, there are lots of affective models proposed by international and national research institutes. For instance, the OCC affective model is the first one for computing and used more widely [26]. By analyzing various affections connecting with events in the physical world and interactions with another subject, the relations based on rules can be got. Moreover, the theory of OCC affective model also suggests that the reason why affection generates includes event result, agent's action, and the feel to object.

Kismet affective model is used in a robot named Kismet, which is designed by C Breazeal in MIT [27]. It combines environment, inner stimuli with action and includes four parts. They are stimuli, evaluation, arousing and expression. Based on this model, the robot acts differently by considering outside stimuli and inner demand.

Based on the emotional psychology, Euclidean space affection model is proposed by regarding basic emotion as base vector [7]. In the affection space, the author discusses transition between one emotion and the other. But this model is a discrete one. Teng models the affective changing process by using Markov chain and hidden Markov model in probability space [8]. Affective transition can be described well whatever stimuli happen or not.

Moreover, Salt & Pepper model proposed by Botelho [9], affection model of a humanoid robot WE-4R researched in Waseda University [10], and some ones based on random event [11] and self-organization theory [12] are also discussed by researchers with different views.

2.2 COGNITIVE-AFFECTIVE INTERACTIONS IN ENTERTAINMENT

With the development of cognitive and affective computing, related theories and technologies apply in the human computer interactions. There are two ways bringing cognition and affection into entertainment application. One is analyzing the human being's affection [13-15], the other is integrating cognitive-affective model into the application [16-18]. In search of suitable methods for measuring the affective state of video-game players, Jonathan Sykes investigates the hypothesis that the player's state of arousal will correspond with the pressure used to depress buttons on a gamepad. A video game was created that would detect the force of each button press during play. It was found that as the difficulty level of the game increased, players would hit the gamepad buttons significantly harder [13]. Abdullah Al Mahmud propose to incorporate psycho-physiological measurements as a part of the gaming experience, and to integrate a desktop game within its real surrounding (i.e., the entire room) in order to promote more physical activity [14].

Moreover, in AI (Artificial Intelligence) game, authentic behaviours of NPC (Non-Player Character) are great challenges to NPC intelligence. Emotions help to enhance the quality and intelligence of behaviours, contribute to increase entertainment value of game. Zhou and Yu describe several common emotional behaviours of NPC, construct a simple emotion-behaviour model for emotion transition, and simulate in their developing project [15]. The quality and intelligence of NPC's behaviours are improved by the emotion to increase the game entertaining. And Munoz K is focused on a qualitative and quantitative approach to recognizing the learner's achievement emotions. Learners' emotions are inferred from two sources: from observable behaviours and from answers to questions in a game dialogue. The analysis and design involved in the creation of this affective student model are the central focus here. PlayPhysics, an emotional games learning environment, is being implemented for teaching Physics at undergraduate level. When our results are finalized our affective student model will be incorporated into PlayPhysics. To ensure accuracy of the recognition method, a preliminary prototyping study has been conducted. The results from this prototyping phase are presented and discussed [16].

Yannakakis provides taxonomy of PCG algorithms and introduces a framework for PCG driven by computational models of user experience. Personalization of user experience via affective and cognitive modelling, coupled with real-time adjustment of the content according to user needs and preferences are important steps toward effective and meaningful PCG. Games, Web 2.0, interface, and software design are among the most popular applications of automated content generation.

Wang proposed a method to investigate the effects of Cognitive Affective Interaction (CAI) strategy on novices' creative performance in game design. The CAI strategies, including the visualization and the discrepancy strategies, were administrated and served as the experimental treatment [18].

3 Computational Model Generation

3.1 OVERVIEW

IHCI required equipment, hardware and software, to provide active services. Moreover, the users are no longer focused on the task process, and only receive services without notice. So building a computable model of human cognitive behaviours is necessary in order to organize knowledge and produce intelligent behaviours in ICHI. ACT-R (Adaptive Control of Thought-Rational) cognitive architecture, proposed by the American psychologist Anderson in 1976, is widely used to simulate different aspects of the human being's cognition behaviours, such as perception and attention, learning and memory, problem solving and decision making, language processing, intelligent agents, intelligent tutoring systems, and human-computer interaction. Because its characteristics fit needs of the implicit interaction precisely, we can build a computable model of human cognitive behaviours, which is based on ACT-R cognitive architecture for different applications, to realize implicit human-computer interaction and improve interactive quality effectively.

This paper focuses on IHCI in entertainments. By designing ACT-R production rules and bringing in emotional factors, computable model of IHCI for entertainment is established as shown in Figure 2.

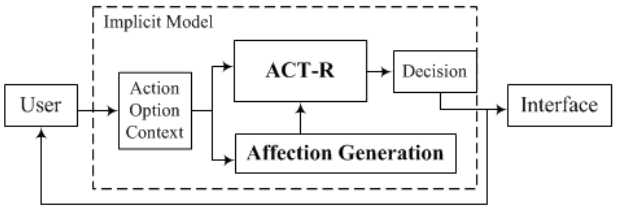


FIGURE 2 Computational model of implicit interaction for entertainment

The devices acquire inputs by users in entertainment, such as behaviours, selections, and situations. Under the influence of emotions, judge user's implicit interaction information after pattern matching and reasoning with ACT-R. And then adjust the interaction process adaptively to improve interaction quality. It is obviously seen that the realization of the model contains two main parts: 1) the generation of machine emotion; 2) the design and implementation of implicit interaction computable model based on ACT-R cognitive architecture in entertainment.

3.2 AGENT AFFECTION

To simulate the emotion generating and changing of human being, an affective model is necessary. As stated above, given the computable feature, we select one named eHMM, for emotional stimuli transferring process proposed in [15] as an emotional engine.

Based on a probability space, Teng regarded an emotional stimuli transferring process as a random one which could be described using hidden Markov model [15]. In other words, a quintuple form determines the model, where N is the number of emotional dimension; M is the number of stimuli type; is an initial emotional state vector; is a state transferring probability matrix, which is calculated by:

$$\hat{A}=\{\hat{a}_{ij}\}_{N\times M}=\begin{bmatrix} \frac{\hat{\theta}\hat{\pi}_1^*-(N-1)}{\hat{\theta}\hat{\pi}_1^*} & \frac{1}{\hat{\theta}\hat{\pi}_1^*} & \cdots & \frac{1}{\hat{\theta}\hat{\pi}_1^*} \\ \frac{1}{\hat{\theta}\hat{\pi}_2^*} & \frac{\hat{\theta}\hat{\pi}_1^*-(N-1)}{\hat{\theta}\hat{\pi}_1^*} & \cdots & \frac{1}{\hat{\theta}\hat{\pi}_2^*} \\ \vdots & \vdots & \cdots & \\ \frac{1}{\hat{\theta}\hat{\pi}_N^*} & \frac{1}{\hat{\theta}\hat{\pi}_N^*} & \cdots & \frac{\hat{\theta}\hat{\pi}_N^*-(N-1)}{\hat{\theta}\hat{\pi}_N^*} \end{bmatrix}, \quad (1)$$

where, $\hat{\theta}$ is a parameter that will be discussed later. $\hat{\pi}^*=[\hat{\pi}_1^*,\hat{\pi}_2^*,\cdots,\hat{\pi}_N^*$ is a limiting probability. $\hat{B}_{N\times M}$ is an observation matrix, supposing that $M=N$ in this paper, it could be deduced that:

$$\hat{B}=\begin{bmatrix}\hat{B}_1\\ \hat{B}_2\\ \vdots\\ \hat{B}_N\end{bmatrix}=\begin{bmatrix} a & b & \cdots & b\\ b & a & \cdots & b\\ \vdots & \vdots & \cdots & \vdots\\ b & b & \cdots & a\end{bmatrix}, \quad (2)$$

Where $\begin{cases} a=\dfrac{r}{N-1+r}\\ b=\dfrac{1}{N-1+r}\end{cases}$, $r>1$, r is a parameter that will also be discussed later.

The user's current behaviour reflects the current level during entertainment. If the current user's behaviour indicates that the entertainment level is higher, the computer treats the entertainment process with caution, and is of positive machine emotion. Otherwise, it indicates that user's entertainment level is not high. The computer generates negative emotions. Based on the eHMM, for emotional stimuli transferring process, emotional value can be got by positive and negative stimulus, which shows the user's current entertainment level.

3.3 MODEL CONSTRUCTION

For the implicit interaction in entertainment applications, three parts of ACT-R cognitive architecture, basic modules, buffering, and pattern matching, are designed as shown in Figure 3. The first part, basic module, has two types of sub-modules. They are a motion perceiving sub-module, which is responsible for interacting with the outside world, and a memory sub-module. The input information of this module is user's mouse or keyboard actions and selecting tendency; and the output is visual and auditory information, including reminders and evaluation for user's entertainment process and feedbacks after entertaining. The second part, buffer, is an interface for generation rules interacting with other basic module. And the contents in buffer identify the current status of ACT-R.

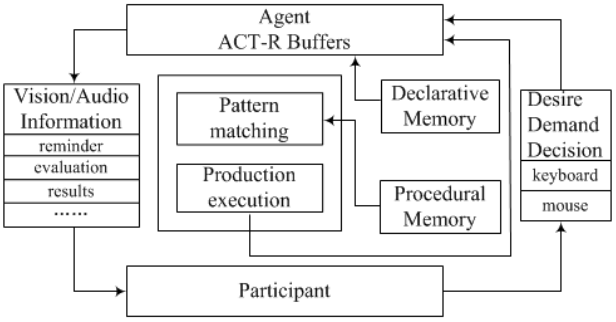


FIGURE 3 Three parts of ACT-R cognitive architecture

The third part, pattern matching, is implemented based on fuzzy inference. Suppose language variables are: the total number of errors X_1, the maximum number of consecutive errors X_2, the average time for thinking X_3, the user's historical entertainment level Y. And the corresponding universes are

$$U_{X1} = U_{X2} = \{x_i | 0 \le x_i \le Sup, x_i \in \aleph\},$$
$$U_{X3} = \{x_3 | 0 < x_3 \le Sec, x_3 \in \Re\},$$
$$U_Y = \{y | 0 < y \le Gra, y \in \Re\},$$

where Sup, Sec and Gra are the maximum number of errors, the maximum time for thinking and the highest user's entertainment level. Their values are determined according to different applications. Linguistic variables are $T(X_1) = T(X_2) = T(X_3) = few + middle + many$ and $T(Y) = low + middle + high$. So the fuzzy sets of the total number of errors X_1 and the maximum number of consecutive errors X_2 are determined as below:

$$F_{xi}^{few} = [few] = \sum \frac{\mu_{Xi-few}(x_j^i)}{x_j^i} = \sum_{j=0}^{Sup} \frac{\left[1 - \frac{1}{Sup} j\right]}{x_j^i}, \quad (3)$$

$$F_{xi}^{mid} = [middle] = \sum \frac{\mu_{Xi-mid}(x_j^i)}{x_j^i} = \sum_{j=0}^{Sup/2} \frac{\left[\frac{j}{Sup/2}\right]}{x_j^i} + \sum_{j=Sup/2}^{Sup} \frac{\left[2 - \frac{j}{Sup/2}\right]}{x_j^i}, \quad (4)$$

$$F_{xi}^{many} = [many] = \sum \frac{\mu_{Xi-many}(x_j^i)}{x_j^i} = \sum_{j=0}^{Sup} \frac{\left[\frac{1}{Sup} j\right]}{x_j^i}. \quad (5)$$

The fuzzy set of the average thinking time X_3 is,

$$F_{X3}^{few} = [few] = \int \frac{\mu_{X3-few}(x)}{x} = \int_{0<x\le Sec/4} \frac{1}{x} + \int_{Sec/4<x\le Sec} \frac{\left[1 + \left[\frac{(x - Sec/4)}{10}\right]^2\right]^{-1}}{x}, \quad (6)$$

$$F_{X3}^{mid} = [middle] = \int \frac{\mu_{X3-mid}(x)}{x} = \int_{0<x\le Sec} \frac{\left[1 + \left[\frac{(x - Sec/2)}{10}\right]^2\right]^{-1}}{x}, \quad (7)$$

$$F_{X3}^{many} = [many] = \int \frac{\mu_{X3-many}(x)}{x} = \int_{0<x\le 3\times Sec/4} \frac{\left[1 + \left[\frac{(x - 3\times Sec/4)}{10}\right]^2\right]^{-1}}{x} + \int_{3\times Sec/4<x\le Sec} \frac{1}{x}. \quad (8)$$

The fuzzy set of the user's historical entertainment level Y is,

$$F_Y^{low} = [low] = \int \frac{\mu_{Y-low}(y)}{y} = \int_{0<y\le Gra/4} \frac{1}{y} + \int_{Gra/4<y\le Gra} \frac{\left[1 + \left[\frac{(y - Gra/4)}{10}\right]^2\right]^{-1}}{y}, \quad (9)$$

$$F_Y^{mid} = [middle] = \int \frac{\mu_{Y-mid}(y)}{y} = \int_{0<y\le Gra} \frac{\left[1 + \left[\frac{(y - Gra/2)}{10}\right]^2\right]^{-1}}{y}, \quad (10)$$

$$F_Y^{high} = [high] = \int \frac{\mu_{Y-high}(y)}{y} = \int_{0<y\le 3\times Gra/4} \frac{\left[1 + \left[\frac{(y - 3\times Gra/4)}{10}\right]^2\right]^{-1}}{y} + \int_{3\times Gra/4<y\le Gra} \frac{1}{y}, \quad (11)$$

The production rules for pattern matching are shown in Table 1.

The inferring part of this ACT-R model infers user's historical entertainment level, which demonstrates the user's whole performance in entertainments, with the total number of errors, the maximum number of consecutive errors and the average thinking time. Moreover, consi-

dering the user's current entertainment level, which is demonstrated by computer's emotion, the computer's entertainment level is determined. The level is able to change dynamically with the user's whole entertainment level to enhance the playability entertainment process and improve the quality of human-computer interaction.

TABLE 1 Conditions - Response production rules for implicit interaction in entertainment

If	Then	If	Then
x_1-few, x_2-few, x_3-few	$y-\text{high}^4$	x_1-many, x_2-few, x_3-few	$y-\text{middle}$
x_1-few, x_2-few, x_3-middle	$y-\text{high}^2$	x_1-many, x_2-few, x_3-middle	$y-\text{middle}$
x_1-few, x_2-few, x_3-many	$y-\text{high}^2$	x_1-many, x_2-few, x_3-many	$y-\text{low}$
x_1-middle, x_2-few, x_3-few	$y-\text{high}^2$	x_1-many, x_2-middle, x_3-few	$y-\text{low}$
x_1-middle, x_2-few, x_3-middle	$y-\text{high}$	x_1-many, x_2-middle, x_3-middle	$y-\text{low}$
x_1-middle, x_2-few, x_3-many	$y-\text{high}$	x_1-many, x_2-middle, x_3-many	$y-\text{low}^2$
x_1-middle, x_2-middle, x_3-few	$y-\text{high}$	x_1-many, x_2-many, x_3-few	$y-\text{low}^2$
x_1-middle, x_2-middle, x_3-middle	$y-\text{middle}$	x_1-many, x_2-many, x_3-middle	$y-\text{low}^2$
x_1-middle, x_2-middle, x_3-many	$y-\text{middle}$	x_1-many, x_2-many, x_3-many	$y-\text{low}^4$

4 Experimental and analysis

In this paper, we use the computational model of implicit interaction in horn chess playing entertainment. Firstly, introduce the method to use it, and secondly, analyze results after several rounds. It is called Horn Chess because the board looks like a horn shown in Figure 4.

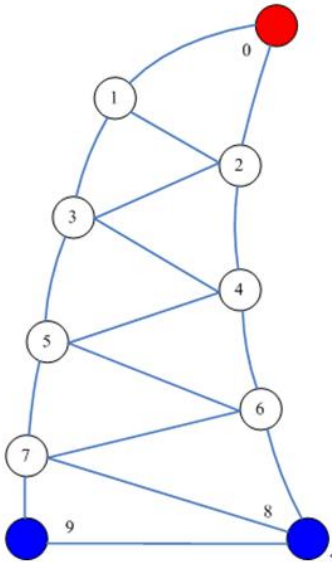


FIGURE 4 Board of horn chess. Positions (0-9) are marked to state conveniently.

On the checkerboard, there are three pieces. One player takes the red, and the opponent takes the blue. The initial pieces position is shown in Table 1. Two players move in turn. And they can only move forwards or backwards one step along lines, rather than no crossing over a piece or moving to the position where there is a piece. If the player taking the blue forces player taking the red to a dead end (position 0 in the board), the former wins. And if the player taking the red could run away to position 8 or 9 of the board, he wins. In order to keep balance, when a new game starts, the player, who takes the red piece, moves first.

4.1 DESIGN OF PLAYING PROCESS WITH IHCI

In this paper, we discussed implicit interaction in Horn Chess gaming based on the affective model for entertainment above. First of all, we suppose the computer has emotion. And in the game process, when the player has an excellent move, or the computer predicts his human opponent's move correctly, the computer's emotion is affected. As we know, it usually influences the computer's next moving strategy to produce excellent or bad move. The game-tree searching algorithm adopts a depth-first mini-max method which is embedded  pruning technique in. Because this paper focuses on the implicit interaction problem of gaming process, the details on game playing algorithm is not discussed more.

In order to manifest the Horn Chess playing process obviously, we named the two chess players, a person and a computer, Alice and Bob respectively. So affecting by the person's entertainment level, the whole process repeats two basic steps. One is Bob's move; the other is Alice's move. Two steps are designed as shown briefly in Figure 5.

The left part of the figure shows that human and the computer move in turn. The process is described as follows:

Step 1: Bob calculates optimal move according to the maximum depth *Maxdeep* based on a game-tree searching algorithm. While searching to the leaf of the game-tree, the best potential situation of Alice's next move is evaluated and recorded, which is signed as $Mark_{Est}$. Then, move to its optimum position. The calculating situation method is $\max\{r\times100+b_1\times10+b_2, r\times100+b_2\times10+b_1\}$, where r, b_1, b_2 are the numbers signed in Figure 4. Their values are integer in domain [0, 9] .

Step 2: Alice thinks and moves according to gaming situation.

Step 3: Bob calculates actual situation $Mark_{Act}$ after Alice moves again.

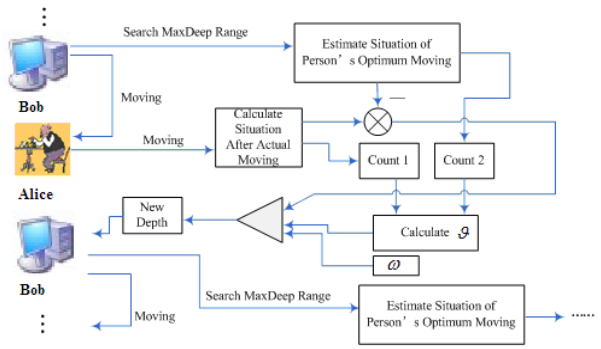


FIGURE 5 Basic steps of Horn Chess gaming affected by the person's entertainment level

Step 4: In the whole process, we need a computational model of implicit interaction for entertainment, especially for Horn Chess Game. To build this model, we can reference the mechanism of the model for Iowa Gambling Task context [19]. Stimuli intensity of emotion is calculated with:

$$I(t) = INT\left(\frac{abs(Mark_{Est} - Mark_{Act})}{\Delta Mark_{Max}} \times I_{Max}\right), \quad (12)$$

where, $\Delta Mark_{Max}$ *s* a maximum difference value of evaluating and actual situations, which could be calculated by $\delta = abs(Mark_{Est} - Mark_{Act})$. While Alice holds the red piece and Bob holds the blue ones, the maximum value $\Delta Mark_{Max}=100$ may occurs. I_{Max} is a legal maximum value of stimuli intensity. In this paper, $I_{Max}=55$. While game is continuing, given that whether $Mark_{Est}$ and $Mark_{Act}$ is equal. If yes, it is explained that Alice's game level is higher. So Bob should play carefully. And His positive emotion is stimulated. Otherwise, for Alice is bad at gaming, Bob is of proud emotion. Negative emotion $\bar{p}_{,Em}$ could be calculated with $I(t)$ based on the statements in the section 3.2 [8]. Equation (12) only shows influence on Bob's emotion caused by Alice's single move, but according to long-memory effect, historical game level scored with *Y* should be considered too.

Considering one situation that Alice plays badly both in a single step and in history, Bob becomes proud to moves unwarily. So there will be much error steps. It can be realized by diminishing the maximum searching depth *Depth* of the game-tree. Update *Depth* as:

$$Depth(t+1) = Depth(t) + \omega \cdot [y(t+1) - y(t)] \cdot \bar{p}_{Em} . \quad (13)$$

Step 5: Run the game-tree searching algorithm again with updated maximum depth $Depth(t+1)$. Repeat step 1.

4.2 EXPERIMENT AND ANALYSIS

Alice and Bob play the Horn Chess game with the rules one by one based on the computational model of implicit interaction above. In order to avoid the searching depth increasing continually, we set the maximum searching depth is 15. As stated above, while stimuli occur and Alice's historical entertainment level is calculated, it could be clearly seen that Bob's emotions is stimulated. And then it influences on his decision-making, such as adjustment of maximum searching depth, width or something else. This paper focuses on the adjustment of maximum searching depth. It changes along with Alice's game level dynamically. Explicit and implicit interactions coexist too.

Alice and Bob carry on a gaming process four times which are Bob moves first and Alice moves first separately, shown in Table 2, 3. To show the experiments result obviously, six aspects, such as the move, the best potential situation of Alice's next move estimated by a Bob $Mark_{Est}$, the actual situation $Mark_{Act}$ after Alice moves, Bob's negative emotion $\bar{p}_{,Em}$, Alice's historical playing level *y* and the self-adjusting maximum searching depth $Depth(t+1)$ are recorded.

In the gaming process recorded in Table 2(b), B represents Bob's moving step, and A represents Alice's moving step. Bob moves first. Taking the first row for example, from row view, B: 0→1 A :(8, 9) → (7, 9) indicates that Bob holds the red piece, and moves from position 0 to position 1 according to Figure 4. Simultaneously, Alice holds blue pieces. One of his pieces moves from position 8 to position 7. And the other keeps still. After Bob's moving, the best potential situation of Alice's next move is estimated, $Mark_{Est}=169$, which is corresponding to Alice's best move. Then, Alice moves, and Bob calculates the current actual situation $Mark_{Act}=179$. While the negative emotion $\bar{p}_{,Em}=0.80624$ and Alice's historical playing level is set to 0.5 initially, we can obtain the maximum searching depth $Depth(t+1)=8$ with the model above.

In addition, the smaller situation is, the more advantaged to Alice's move is according to the min-max game-tree searching theory, when Bob moves first from column view. Alice usually does not move best. So, generally speaking, the actual situation $Mark_{Act}$ is larger than $Mark_{Est}$. That is $Mark_{Est} \leq Mark_{Act}$. When Alice's move is not the best one estimated by Bob, his proud emotion is stimulated. The negative emotion $\bar{p}_{,Em}$ is larger than its initial value (0.5). Moreover, the worse move is, the larger $\bar{p}_{,Em}$ is. When Alice's gaming level is lower, and Bob is proud, Bob reduces the maximum searching depth automatically. Otherwise, the searching depth will be increasing to reflect that emotion and historical game level influences on entertainment process. Similarly, another experiment process in which Bob moves first is recorded in Table 2(a). The analysis to this continued table is the same as statement above.

TABLE 2 Gaming process when Bob moves first

(a) The first time

Record	Moving	Estimating situation	Actual situation	Emotion	Historical entertainment level	Max searching deep
1	B:0→1 A:(8,9)→(6,9)	169	169	0.45653	0.50	8
2	B:1→0 A:(6,9)→(6,7)	49	67	0.92846	0.46	7
3	B:0→1 A:(6,7)→(5,6)	147	156	0.75957	0.41	6
4	B:1→2 A:(5,6)→(4,6)	236	246	0.80624	0.41	6
5	B:2→0 A:(4,6)→(3,6)	26	36	0.80624	0.36	5
6	B:0→2 A:(3,6)→(3,4)	216	234	0.92846	0.32	4
7	B:2→1 A:(3,4)→(2,3)	123	123	0.45653	0.46	6
8	B:1→0 A:(2,3)→(1,2)	123	123	0.45653	0.53	7
9	B: LOSS	-	-	-	-	-

(b) The second time

Record	Moving	Estimating situation	Actual situation	Emotion	Historical entertainment level	Max searching deep
1	B:0→1 A:(8,9)→(7,9)	169	179	0.80624	0.50	8
2	B:1→0 A:(7,9)→(5,9)	59	59	0.45653	0.87	11
3	B:0→1 A:(5,9)→(5,7)	139	157	0.92846	0.74	9
4	B:1→3 A:(5,7)→(6,7)	347	367	0.95854	0.67	8
5	B:3→5 A:(6,7)→(7,8)	547	578	0.99247	0.61	7
6	B:5→6 A:(7,8)→(8,9)	658	689	0.99247	0.55	6
7	B:6→7 A:LOSS	-	-	-	-	-

TABLE 3 Gaming process when Alice moves first

(a) The first time

Record	Moving	Estimating situation	Actual situation	Emotion	Historical entertainment level	Max searching deep
1	A:0→1 B:(8,9)→(6,9)	-	-	-	-	-
2	A:1→2 B:(6,9)→(5,9)	369	369	0.54347	0.50	8
3	A:2→3 B:(5,9)→(5,7)	459	359	0.54347	0.42	7
4	A:3→4 B:(5,7)→(5,6)	457	457	0.45653	0.59	9
5	A:4→2 B:(5,6)→(4,6)	356	256	0.54347	0.51	8
6	A:2→0 B:(4,6)→(2,6)	346	346	0.25728	0.66	9
7	A:0→1 B:(2,6)→(3,6)	126	126	0.45653	0.74	10
8	A:1→2 B:(3,6)→(3,4)	236	236	0.45653	0.82	11
9	A:2→0 B:(3,4)→(2,4)	134	134	0.54347	0.82	11
10	A:0→1 B:(2,4)→(2,3)	124	124	0.45653	0.90	12
11	A: 1→0(LOSS) B: (2,3)→(1,2)	23	23	0.45653	0.97	13

(b) The second time

Record	Moving	Estimating situation	Actual situation	Emotion	Historical entertainment level	Max searching deep
1	A:0→2 B:(8,9)→(6,9)	-	-	-	-	-
2	A:2→4 B:(6,9)→(7,9)	469	469	0.45653	0.50	8
3	A:4→5 B:(7,9)→(7,8)	679	579	0.54347	0.43	7
4	A:5→6 B:(7,8)→(7,9)	678	678	0.45653	0.51	8
5	A:6→8 B:LOSS	879	879	0.45653	0.65	10

When Alice moves firstly, recorded in Table 3, the bigger situation is, the more advantaged to Alice's move is according to the min-max game-tree searching theory, when he holds the red piece from column view. Moreover, Alice usually does not move best. So, actual situation $Mark_{Act}$ is smaller than $Mark_{Est}$. The rest analysis is the same as Table 2.

Moreover, record the detail information of four experiment processes above to contrasting and analyzing. They are total number of errors x_1, the maximum number of consecutive errors x_2, and the average thinking time x_3. With their maximum values shown in Tables 2 and 3, the percentages (TE, MCE, ATT, MaxSD, and MinSD) can be calculated and demonstrated in Figure 6. TE, MCE, ATT, MaxSD, and MinSD means the percentage of total number of errors, the percentage of maximum number of consecutive errors, the percentage of average thinking time, the percentage of maximum searching depth and the percentage of minimum searching depth respectively.

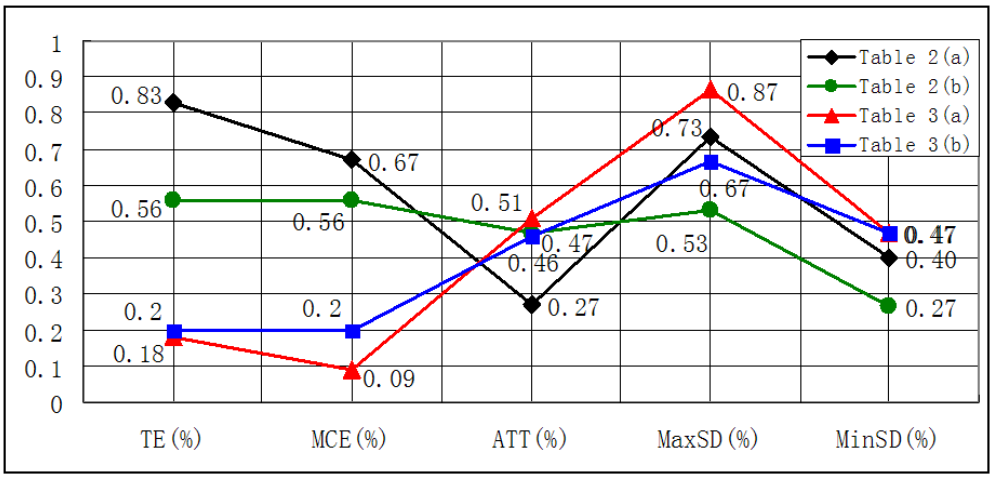


FIGURE 6 Statics of the four gaming processes

In this figure, five percentages of every experiment are put together. No matter Alice or Bob moves firstly, generally speaking, MaxSD and MinSD are bigger when TE, MCE and ATT are smaller, which means Alice is good at gaming and has a high level in entertainment, so Bob must take the game process seriously by thinking deeply. But for the line of Table 2(a), because ATT influences more than TE and MCE, MaxSD and MinSD are getting bigger.

5 Conclusions

In this paper, a computational model of implicit interaction based on emotional Hidden Markov Mode (eHMM) and ACT-R cognitive architecture is devised, which will be used for human computer interaction or corporation, especially for entertainment. We considered the influence of agent's affection and cognition. So eHMM and ACT-R are merging with each other.

The focus of present research is how to construct a hierarchical structure to link agent's affection and cognition. From our point of view, agent's affective state is influenced by current behaviors coming from the inner world and environment. It makes agent react rapidly. Moreover, because of memory and inference, cognition reflects long-term influences. So we integrated them together. No matter in simulation experiments or in practice, there are good results. The proposed model is effective.

Acknowledgements

This work was supported by the Foundation for Young Scholars of Hebei Educational Committee (QN20131152, Q2012070).

References

[1] Schmidt A, Spiessl W, Kern D 2010 *IEEE Pervasive Computing* **9**(1) 85-88
[2] Kaiyan N 1996 Exploratory study of implicit theories in human computer interaction *Proceedings of the 6th Australian Conference on Computer-Human Interaction* 338-9
[3] Schmidt A 2000 Implicit human computer interaction through context *Personal Technologies* **4** (2-3) 191-9
[4] Wilson, Oliver N 2005 Multimodal sensing for explicit and implicit interaction *Proceedings of the 11th International Conference on Human-Computer Interaction* 1-10
[5] Wang G J 2010 Distributed vision system for implicit human computer interaction *Journal of Image and Graphics* **15**(8) 1133-8
[6] Tian F 2007 Research on the implicit interaction characteristic of Post-WIMP user interface *Journal of Frontiers of Computer Science and Technology* **1**(2) 160-9
[7] Wang H 2004 Emotion model based on theory of artificial psychology and numerical simulation *Computer Applications* **22** (S1) 368-70
[8] Teng S D 2006 Research on artificial psychology model applied in personal robot *PhD dissertation Beijing China University of Science & Technology Beijing*
[9] Botelho L M, Coelho H 2001 Machinery for artificial emotions *Cybernetics and Systems* **32**(5) 465-506
[10] Miwa H, Itoh K, Ito K, Takanobu H, Takanishi A 2003 Introduction of the need model for humanoid robots to generate active behavior *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* 1400-06
[11] Wang F, Wang Z L 2005 Affection mathematics model based on the processing of stochastic events *Control & Automation* **21**(3) 101-2
[12] Cheng N, Liu J W 2005 Application of basic emotions theory in construction of artificial psychology model *Computer Engineering* **31**(22) 175-7
[13] Sykes J, Brown S 2003 Affective gaming: measuring emotion through the gamepad *Proceeding of CHI '03 Extended Abstracts on Human Factors in Computing Systems* 732-3
[14] Al Mahmud A, Mubin O, Octavia J R, Shahid S, Yeo L, Markopoulos P, Martens J B, Aliakseyeu D 2007 Affective tabletop game: a new gaming experience for children *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems* 44-51

[15] Zhou C N, Yu X L 2006 Affective computation based NPC behaviours modelling, *Proceedings of 2006 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* 343-6
[16] Muñoz K, Nogues J, Mc Kevitt P, Lunney T, Neri L 2010 Towards an emotional learning model for intelligent gaming *Proceedings of 2010 IEEE Frontiers in Education Conference* T3G-1-2
[17] Yannakakis G N, Togelius J 2011 *IEEE Transactions on Affective Computing* **2**(3) 147-61
[18] Wang L C, Chen M P 2012 The effects of cognitive-affective interaction strategy on novices' creative performance in game design project *Proceedings of 2012 IEEE 12th International Conference on Advanced Learning Technologies* 549-53
[19] Wang W, Wang Z L 2011 Research on the computational model of emotional decision-making *International Journal of Kensei Information* **2**(3) 167-172

Authors

Wei Wang, born in November 1983, Handan, Heibei Province, P.R. China.

Current position, grades: lecturer at the School of Information & Electrical Engineering, Hebei University of Engineering, China.
University studies: MSc in Control Theory and Control Engineering at Jiangnan University in China. DSc at University of Science and Technology Beijing in China.
Scientific interests: human-robot cooperation, implicit interaction.
Publications: more than 30 papers.
Experience: teaching experience of 3 years, 3 scientific research projects.

Xiaodan Huang, November 1983, Handan, Heibei Province, P.R. China.

Current position, grades: lecturer of School of Information & Electrical Engineering, Hebei University of Engineering, China.
University studies: MSc in Electronic Information at University of Science and Technology Beijing in China.
Scientific interests: implicit interaction, robot.
Publications: more than 10 papers.
Experience: teaching experience of 3 years, 2 scientific research projects.

Construction and effect evaluation of the college English multimedia teaching system based on the Blended-learning model

Jingtao Liu*

Foreign Language Department, Heze University, Heze 274000, Shandong, China

Received 13 September 2014, www.cmnt.lv

Abstract

As network technology develops, an increasing number of new technologies and concepts have entered the traditional college teaching system. To improve the education quality and level of domestic college English, as well as to meet the demand for the reform of domestic college English education, the Blended-learning model is introduced to construct a new multimedia teaching system based on the traditional multimedia teaching model of college English. Moreover, a questionnaire analysis is conducted among all staff members that participate in the new multimedia teaching system. Questionnaire analysis indicated that the construction of the Blended-learning model-based traditional multimedia teaching system of college English proposed in this study is practically operable. The effect of this model is significantly better than that of the traditional English teaching model.

Keywords: Blended-learning model, college English, multimedia teaching, system construction

1 Introduction

The first specification on subjects included in the college entrance examination in China is *Provisions on New Students Recruitment of Higher Institutions in the Summer of 1950*, which can be dated back to 1950 [1]. This specification was a programmatic document for English examination instructions in college entrance examinations. Since the reformation and opening up, the national economy has rapidly developed, the public has directed increasing attention toward English [2], initially from senior English at to junior English, primary English, kindergarten English, and even bilingual teaching (refers to English solely) [3], which undoubtedly highlights the importance of English. Since the recovery of the college entrance examination system in 1977 [4], considerable progress has been made in English in terms of continuous learning and reform. Gradually, as a group of freshmen enters university and a wave of graduates enters society, many employers [5], enterprises and scientific research institutes have gradually found that the English level of college students is not direct proportional to English learning time. Many college students learn English just to learning [6]. The starting point seriously deviates from the concept and essence of English teaching. Some students even abandon learning, which has gradually aroused discussion in domestic society and in the education department regarding the reform of English education [7]. Higher institutions should continue to be relevant as well as improve the teaching environment and learning effect through information technology and network technology while relying on education informatization [8]. With this background, in combination with the Blended-learning model that has recently emerged, the author has researched and analyzed the Blended-learning model to construct the existing college English multimedia teaching system innovatively through actual investigation. Meanwhile, all staff members who participated in the construction of this system are involved in the investigation. The actual construction effect is evaluated and analyzed using a questionnaire to provide a preliminary exploration of the construction of the college English multimedia teaching system based on the Blended-learning model.

2 Analysis on Blended-learning model

The blended-learning model was introduced by foreign enterprises, which trained their clerks involved in income and profits. In other words, this model was used to reduce input and make the enterprises gain more profits through training [9]. Domestic educators also call it mixed learning. The key learning objective of Blended-learning is optimization, i.e., the use of appropriate learning skills and methods at appropriate times and with appropriate targets to adapt to and match appropriate learning target habits to convert them into appropriate skills [10].

2.1 BASIC CONCEPT

In domestic education, Professor He Kekang of Beijing Normal University introduced the concept of "Blended-learning." Professor He believes students can only become the subject of learning when new information technology (network or digital or information learning) and the advan-

* ***Corresponding author's*** e-mail: liujingtao22@yeah.net

tages of the traditional learning method are combined [11]. Moreover, the enthusiasm of students can be fully stimulated and their enthusiasm, initiative, and creativity can be included in the process of learning. Teachers can also optimize their roles in guiding, monitoring, and enlightening students when they take the lead. As a result, the process of teaching and learning becomes more saturated and enriched, and the learning effect is optimized. This consensus has been uniformly recognized in the field of domestic education.

2.2 ELEMENTS

In short, the blending-learning model includes face-to-face learning and online learning. However, the two are not single or interdependent. Before the college English multimedia teaching system based on blended-learning was constructed, the components of Blended-learning model must first be introduced and clarified. By doing so, college English multimedia teaching system of Blended-learning model combined with practice can be constructed.

Blended-learning model is divided into two parts: off-line learning and online learning (Table 1).

TABLE 1 `Blended-learning model module classification and comparison of advantages and disadvantages

Module name	**Offline learning (traditional class)**	**Offline learning (E-learning)**
Main characteristics	Teacher oriented, dominated by knowledge impartation	Teaching by internet or in digital form
Advantages	➢ Beneficial to motivating teachers to take the lead ➢ Beneficial to cultivating student cooperation ➢ Beneficial to cultivating student competition awareness ➢ Beneficial to cultivating student's sense of collectiveness ➢ Beneficial for communication between teachers and students and optimizing affectivity in learning	• Wide coverage, flexible transmission means • Learning is not limited by time and space • Rich learning resources • Self-planning, individual learning
Disadvantages	➢ Excessive emphasis on uniformity, rigid ➢ Easily ignores students' independence ➢ Easily ignores students' uniqueness ➢ Excessive reliance on teachers' capability ➢ Limited learning scale	• Teachers and students have less time fact-to-face • Aloof teachers and students • Learners learn alone • Students require strong self-learning and discipline

2.3 LEVEL AND STAGE DIVISION

Blended-learning model can be divided into different learning levels and learning stages, as shown in Table 2.

TABLE 2 learning level division of Blended-learning model

Complexity	**Specific contents and forms**	**Specific roles**
Simple ↓ Complex	Rapid approach to releasing course information and implementing course management	Learning content support
	Method for visiting additional online materials	
	Provides communication between teachers, students, and other relevant personnel	Learning activity and evaluation support (basic knowledge learning)
	Provide computer assisted learning support	
	Have more test exercises and implement summative beginning	
	Provide practical opportunity and extra support for in-school students	Learning activity and evaluation support (senior knowledge learning)
	Student coordination platform based on project learning	
	Transmit complete online courses with completely comprehensive activity	Typical online learning

Although Professor He Kekang has introduced the concept of Blended-learning for some time, it has been applied and popularized in China. His research mainly concentrated on principle, design, and strategy. However, through the domestic researches and according to close combination of learning mode, the application of the blending learning model can be roughly divided into four stages [5] according to the level division above, as shown in Table 3.

TABLE 3 Learning stage division of Blended-learning model

Stage	**1st stage**	**2nd stage**	**3rd stage**	**4th stage**
Representation form	Combination level	Integration level	Coordination level	Expansion level
Description	• Different learning models are simply combined • No correlation between learning modes	• Different learning modes are combined into a correlative structure • Learning modes supplement each other	• The agglomeration method between learning modes is higher • Learning objective oriented	• Learning is combined with practice through various means
Application	• Relatively simple form • Adapts to perfect online learning schools	• Learning contents and methods must be planned in detail • A simple means for activity evaluation must be established	• A more detailed plan must be established • Centered on the learning objective	• Combines learning with practice centered on learning personnel • The system is optimized and various learning methods are combined to promote learning objectives and accurately master the effect of learning

3 Research on college English multimedia teaching system structure based on Blended-learning model

After analyzing the Blended-learning model, we aim to determine how to recognize the blended-learning model and introduce it to the college English multimedia teaching system. By doing so, we can construct a good college English multimedia teaching system.

3.1 RESEARCH TARGET AND INVESTIGATION ANALYSIS

A university in China was selected for investigation. Tracking analysis was conducted from September 2013 for a term. English teachers and students, assisted by relevant education institutions of schools, dominated the investtigation targets.

3.1.1 English teacher investigation analysis

The key factors that influence the construction of the college English multimedia teaching system under blended-learning are found from the preliminary investigation result. By analyzing the actual investigation result, different influential factors are subject to unified optimal management and are organized in a coordinative manner. The factors that disturb the construction of the multimedia teaching system effect are eliminated. Moreover, the proportion of factors beneficial for realizing teaching objectives is increased to realize dynamic adjustment gradually (Figure 1-4).

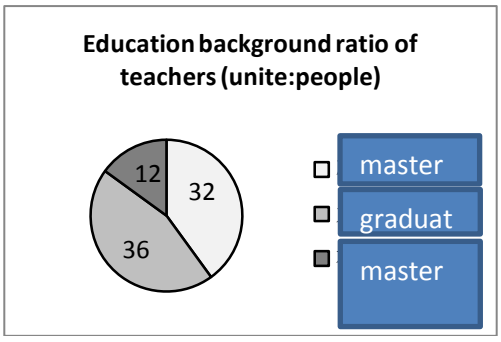


FIGURE 1 Education background ratios of teachers

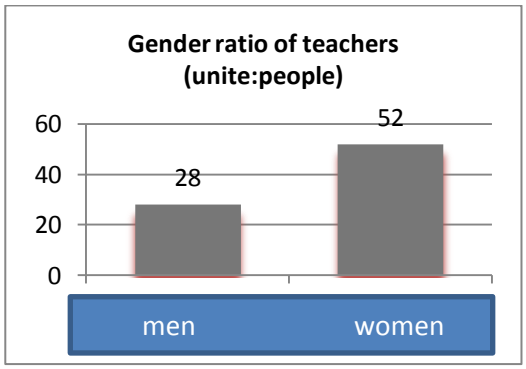


FIGURE 2 Gender ratio of teachers

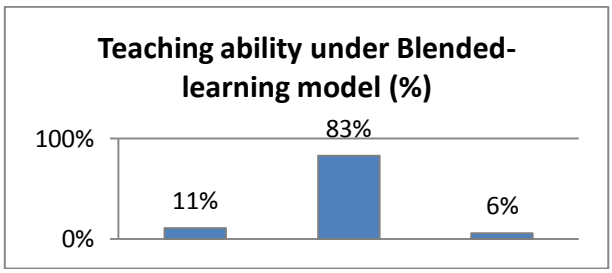


FIGURE 3 Satisfaction of teaching capacity under Blended-learning model

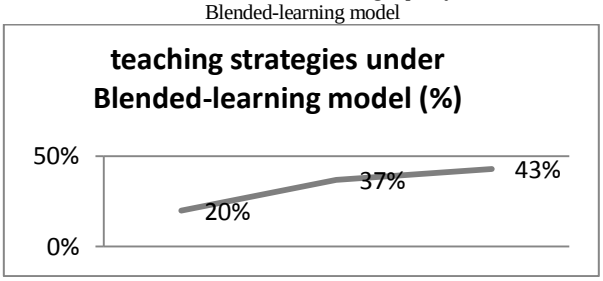


FIGURE 4 Use of teaching strategies under Blended-learning model

The basic information of the teachers selected as research targets can be roughly determined. Many factors from this information influence the success and effect difference of college English multimedia teaching system construction under Blended-learning model.

3.1.2 Investigation analysis on students and relevant teaching reform departments

The students selected were all freshmen that entered the university in September. A total of 60 students from two English major classes were selected, and 60 students from one science and engineering class. For the students that participated in the research, statistical analysis on the factors that might influence the teaching effect under the blended-learning model was conducted through a questionnaire. An English admittance scores examined students' capability to accept fresh concepts, self-discipline capability, self-learning capability, network knowledge acquaintance and multimedia acquaintance. An overall statistical analysis was also conducted. The 120 students were divided into several small classes according to their specific statistical result. The difference among students was minimized for the benefit of uniform planning, teaching, and assessment in the follow-up multimedia teaching system. In addition, relevant English teaching departments that participated in the research were also investigated. The university support for English teaching reform, attitude toward the blended-learning model, as well as their opinions and suggestions on the construction of college English multimedia teaching system based on blended-learning model were mainly investigated.

After the preliminary investigation and analyzing the English teaching reform department, teachers and students, the college English multimedia teaching system,

the Blended-learning model was constructed according to the actual analysis factors in the following.

3.2 CONSTRUCTION OF COLLEGE ENGLISH MULTIMEDIA TEACHING SYSTEM BASED ON BLENDED-LEARNING MODEL

Before an English teaching system is understood, system construction and its meaning should be understood. In our traditional cultural-loaded words, "the system" generally refers to its popular meaning or integrity composed according to a certain order and internal connection within a certain scope, a new system composed of different systems. Such system is an organic integrity with certain functions that is generally established directionally, comprehensively, deeply and even dimensionally for construction.

For the English multimedia teaching system, an English subject must be constructed within in the university. According to the blended-learning model, multimedia teaching resources (including teaching facilities, materials, faculties, education departments, technical forces, etc.) form a highly coordinative, blending, objective, and optimal organic integrity with students as the center. The teacher leads and is assisted by the competent education department in terms of comprehension, multiple angles, and depth.

3.2.1 Construction model and frame

We divided the stages of the model. The specific model, characteristic, and effect of each stage are shown. Stages are countered in the process of constructing the college English multimedia teaching system. We cannot construct this system completely, substantially, and optimally simultaneously. However, our basic model is consistent with model architecture. The specific model architecture constructed is as shown below (Figure 5):

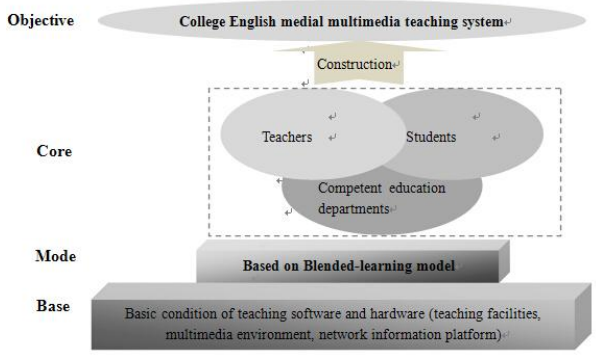


FIGURE 5 System construction model

In this model diagram, the base, model, core, and objective are not constant. Rather, they are in a dynamic adjustment state. The restriction factors should be adjusted according to actual experience and problems. The dynamic change cycle of each department might be different. Such cycles include blending, mixing, and method of students and teachers, and competent education departments, which depends on many factors. However, such blending and coordination will have an important influence on the objective. Therefore, continuous adjustment is required and the blending degree should reach the fourth stage of the blending-learning model or better.

3.2.2 Construction of model core of potential category analysis

The constructing multimedia teaching system model mainly analyzes core modules (i.e. blending of students, teachers and competent education departments). The potential category analysis principle in the statistics is introduced. This principle is established based on probability multivariate analysis so that comprehensive structural equations and linear equations can be used. The correlation among variables with the minimum key latent class is explained, and the statistical thought of local independence is constructed. Through latent class analysis (LCA), the potential factors that influence the construction of multimedia teaching system and mark the theoretical relation between explicit variables and potential variables through the statistical model are used in a questionnaire. Then, the core modules are modeled through analyzing, evaluating, comparing and explaining the latent class model.

The early investigation conditions are selected. These conditions can belong to a certain level of r existing categories in a potential variable x, and each level can be mutually blended. According to the measured external factors, the condition of local independence can be met among them after considering the potential property of the observation data. Such condition can be expressed using a probability function.

If we take students (a), teachers (b), and the school (c) as three external variables, then the core module of latent category model should contain a non-conditional probability that belongs to the latent factor, which we express with π_t^X and constitutes the essential latent factor model together with the three condition probabilities constituted by the latent class structure that influences the core module.

$$\pi_{ijk}^{abc}=\sum_{t=1}^{T}\pi_t^X\pi_{it}^{\bar{a}X}\pi_{jt}^{\bar{b}X}\pi_{kt}^{\bar{c}X}\ , \tag{1}$$

where, π_{ijk}^{abc} represents the joint probability of a latent factor model; π_t^X represents the probability of specific latent class when the observed data belongs to a latent factor x, i.e., $P(X=t),t=1,2,...,T$; $\pi_{it}^{\bar{a}X}$ represents the

probability condition of latent influential factor of model $_t$, i.e., $P(a=i \mid X=t), t=1,2,...,I$, and so on.

According to the probability parameterization of latent class mode, a significant latent variable X has at least two latent classes, which can be expressed as:

$$\pi_{ijk}^{\sum_{M=1}^{N} KX} = \pi_t^X \pi_t^{K1} \pi_t^{K2} \pi_t^{K3} \pi_t^{K4} \pi_t^{KN} X \cdot \quad (2)$$

The second factor is the relative size of each latent class. The larger the scale is, the more important the latent variable is, i.e., the higher the proportion. In LCA, the sum of various latent class probabilities is 1.00:

$$\sum_{t=1} \pi_t^X = 1.00 . \quad (3)$$

In the LCA model, the investigation targets are taken at random, and the probability in the explicit variable is called the condition probability. For the relation and coordination degree among students, teachers, and schools, under different environments of latent variable X, the condition probabilities are respectively marked as $\pi_{it}^{\bar{a}X}$, $\pi_{jt}^{\bar{b}X}$ and $\pi_{kt}^{\bar{c}X}$, according to the model, and the following relation is met:

$$\sum_{t=1}^{T} \pi_{it}^{\bar{a}X} = \sum_{t=1}^{T} \pi_{jt}^{\bar{b}X} = \sum_{t=1}^{T} \pi_{kt}^{\bar{c}X} = 1.00 , \quad (4)$$

where the latent factor within the explicit class and that the sum of condition probabilities of all explicit variables is 1.00. The probability sum of the latent model is also 1.00.

Moreover, in addition to estimating the factor probabilities that influence the construction of the multimedia teaching system, LCA can also be described briefly using a traditional linear model. The LCA estimation model is as follows:

$$\ln(f_{ijkt}^{abcX}) = \eta + \eta_t^X + \eta_t^a + \eta_t^b + \eta_t^c + \eta_{it}^{aX} + \eta_{jt}^{bX} + \eta_{kt}^{cX} . \quad (5)$$

The LCA model contains latent variable factors, and local independence only appears after considering the latent factors for the external variables. Thus, when evaluating the LCA of Equation (5), various items under the second-order term of the latent factors and various external variables should be introduced in pairs to the model for factor estimation.

3.2.3 Model construction optimization

Optimization can be conducted from the following three aspects:

First, English teaching reform departments should provide much support. In addition to software and hardware facilities, corresponding regulations and systems should be established by drawing on the wisdom of the masses. Listening to the opinions of the investigated teachers with successful experience is very important. Moreover, multimedia network environment should be used collect students' personal evaluations and feelings, such that the college English multimedia teaching system based on the blending-learning mode becomes beneficial and optimizes factors that restrict the multimedia teaching system. Then, this system will gradually perfect the English multimedia teaching system of the whole university.

Second, English teachers should be trained to participate in the construction of a college English multimedia teaching system based on blended-learning irregularly. The training form should follow no set form and should be diversified and in depth. The training should mainly be conducted in terms of the following aspects:

1) Cultivating the identity of English teachers in terms of the new teaching system;
2) Participating in the discussion on successful cases of Blending-learning mode for absorptive learning;
3) Unifying the teaching concept "students oriented and teachers guiding";
4) Clarifying and unifying the role of teachers in new multimedia teaching classes;
5) Continuously improving and innovating the English teaching method under the blended-learning model;
6) Improving English teaching ability and strategy under the blended-learning mode;
7) Using scientific research in college English multimedia teaching system based on Blended-learning model;
8) Encouraging and supporting English teachers under the system for multimedia or network technology training, and the exploring and innovating on various learning methods or approaches based on Blended-learning model;

Finally, based on the principle of student oriented, we cannot ignore the role of the students in the whole construction of college English multimedia teaching system based on Blended-learning model because they are the key to inspecting whether the system is perfect. Thus, students must continuously give feedback regarding various difficulties in the new system, discover problems in a timely manner, and communicate with the teachers to gradually perfect and optimize the whole new multimedia teaching system. The process involves joint cooperation, coordination, learning, growth, and the perfection of three parties and thus cannot be accomplished in one action.

3.3 EVALUATION AND ANALYSIS

After a tracing investigation for one year based on investtigating and analyzing the factors influencing construction of college English multimedia teaching system under the blended-learning model. The following findings were obtained through actual research:

First, the construction of a college English multimedia teaching system based on blended-learning improves the enthusiasm of students and incentivizes English learning. The system also continuously cultivates their enthusiasm,

self-discipline, and innovation in the whole process, allowing them to learn many skills and knowledge beyond their language ability.

Second, for teachers, constructing such a small range of college English multimedia teaching system based on the blended-learning model also stimulates the atmosphere for mutual learning among English teachers, which greatly improves the whole English teaching and English teaching of the whole university. Moreover, under such a benign learning and training atmosphere, other majors of the university will also be enlightened and guided, which will popularize the blended-learning concept greatly.

Finally, for the university itself, constructing this college English multimedia teaching system based on the blended-learning model allows English teachers to feel the support and encouragement of the university, which gains the trust of the teachers. With this trust, they continuously innovate and blend new teaching strategies and methods, allowing students to trust the university and their teachers. Thus, under the multimedia teaching system based on the blended-learning model, university leaders and relevant education departments praise and recognize students and teachers are obtained, which provide a new concept for English teaching model reform and innovation.

4 Conclusion

The blended-learning model has a significant influence in many aspects of the domestic education field and is significant to the development of education reform. The principle of Blended-learning model was analyzed, and this model was introduced for the construction of a college English multimedia teaching system. Through actual investigation questionnaires, the factors that influence the construction of college English multimedia teaching system under Blended-learning model were determined. Statistics were used to dynamically adjust the restriction factors in time. The results show that with the support of relevant education departments, on one hand, the teachers are trained for restriction factors. On the other hand, the abilities of students are divided differentially. Thus, professional and non-professional classifications are adopted, limitation is set to the multimedia teaching scale and different students are taught according to their own conditions. The college English multimedia teaching system constructed through this dynamic optimization process is energetic and brings relatively optimal benefits to the university, teachers, and students.

References

[1] Hamilton H J, Geng L, Findlater L, Randall D J 2006 Efficient spatio-temporal data mining with GenSpace graphs *Journal of Applied Logic* **4**(2) 192-214

[2] Li X 2003 A New Clustering Segmentation Algorithm of 3D Medical Data Field Based on Data Mining *JDCTA* **4**(4) 174-81

[3] Borg S 2007 Research engagement in English language teaching. *Teaching and Teacher Education* **23**(5) 731-47

[4] Wai Leong A M, Xi Li J 2012 A study on English teaching improvement based on stakeholders' needs and wants: The case of the Faculty of International Tourism of the Macau University of Science and Technology (MUST) *Journal of Hospitality, Leisure, Sport and Tourism Education* **11**(1) 67-78

[5] Bao X 2004 Research on Quick Sort Methods Based on ID3 algorithm *Modern Electronic Technique* **27**(4) 84-5

[6] Han J, Nishio S, Kawano H, Wang W 1998 Generalization-based data mining in object-oriented databases using an object cube model. *Data & Knowledge Engineering* **25**(1-2) 55-97

[7] Jian Z, Jin X 2004 Research on Data Preprocess in Data Mining and Its Application *Application Research of Computers* **21**(6) 117-8

[8] Wang Y, Zheng L 2012 Endocrine Hormones Association Rules Mining Based on Improved Apriori Algorithm *JCIT* **7**(7) 72-82

[9] Vanci-Osam U, Aksit T 2000 Do intentions and perceptions always meet? A case study regarding the use of a teacher appraisal scheme in an English language teaching environment *Teaching and Teacher Education* **16**(2**)** 255-67

[10] Hayes D 2009 Non-native English-speaking teachers, context and English language teaching *System* **37**(2) 1-11

[11] Fareh S 2010 Challenges of teaching English in the Arab world: Why can't EFL programs deliver as expected? *Procedia – Social and Behavioral Sciences* **2**(2) 3600-4

Author

Jingtao Liu, August 1981, Linyi City, Shandong Province, P.R. China.

Current position, grades: Lecture of Foreign Language Department, Heze University, Heze, China.
University studies: Master Degree in Arts in Education at Capital Normal University, China.
Scientific interests: English language and literature.
Publications: more than 10 papers.
Experience**: teaching experiences of 5 years, 2 scientific research projects.**

Construction quality risk management of projects on the basis of rough set and neural network

Jianbing Liu*, Fang Guo

School of Economics and Management, Jiangxi University of Science and Technology, Ganzhou, Jiangxi, China, 341000

Received 19 August 2014, www.cmnt.lv

Abstract

Construction quality associated with the life of construction enterprises. Risk assessment of construction quality referred to a comprehensive evaluation of the degree of risk confronting construction units during the construction process. Construction units provide a decision-making basis for employers and supervisors, who considered the evaluations of construction quality and accidents in their decisions. The risk factors of construction project quality were classified into personal risk and material risk and machinery and equipment risk and method risk and environment risk. On the basis of these factors, we constructed an index system of project construction quality risk. The risk evaluation model of project construction quality was constructed on the basis of rough sets and neural networks. Finally, a case study residential building projects in the Ganzhou Development Zone and research tools of Rosetta based on rough sets and MATLAB7.0 based on neural networks were used to test model accuracy and reason ability. Empirical results showed that the model has great practical significance.

Keywords: construction quality risk assessment, projects, rough set, neural networks

1 Introduction

The rapid development of the national economy has caused project constructions to exhibit a growing trend. However, such projects are confronted with numerous sun foreseen interference factors during the construction process. These factors give rise to the requirement for project participants to discuss and consider various risks. Uncontrolled risks may pose serious problems to the whole project. Thus, improving the risk management level of project construction quality is a task that project managers must focus on.

The construction quality risk assessment of a project is the most effective management means to improve construction quality and to avoid construction quality accidents. Such assessment is the core link of project quality risk management.

Construction quality risk assessments have yet to establish a unified and comprehensive index system. The lack of effective data sources causes a low construction quality risk management level of a project, such that the project manager is believed to possess a generally weak risk consciousness. These issues must be mitigated as continuous economic globalization has made international competition increasingly fierce.

Given this background, this study employs modern project management theory and incorporates advanced foreign project quality risk management experiences. Considering the inherent nature of construction quality risk management of a project, we propose project quality risk management methods that our suitable for the Chinese construction industry.

The risk factors that influence construction project quality are identified as personal risk, material risk, machinery and equipment risk, method risk, and environment risk. The index system of project construction quality risk was constructed, and the risk evaluation model of project construction quality was established on the basis of rough sets and neural networks.

2 Status of foreign and domestic research

2.1 STATUS OF FOREIGN RESEARCH

Risk management is a rapidly developing integrated discipline. The emergence of risk management can be traced back to ancient human productive labor. As societies constantly progress, enterprises encounter all types of risk during the development process. To help enterprises survive and grow, risk management, as an important part of enterprise management, was developed to reduce the risk of loss. Risk management has been employed to deal with risks, to facilitate stabilization after the occurrence of a financial risk, and to achieve the effective use of enterprise resources.

Early foreign research on risk management has been limited to enterprise risk. However, with the increase in construction engineering projects, the frequency of engineering project risk occurrence and the rapid development of engineering project management technology, risk management has been included in the scope of project management. Thus, project risk management was established.

* ***Corresponding author's*** e-mail:liujb71@sina.cn

Many uncertain factors in project construction may cause projects to fail and economic interests to be lost. Based on the experiences of and lessons from researchers abroad, the quality of project risk management has gradually been enhanced. With the continuous development of engineering project construction, project quality risk management has improved. Since the 1960s, foreign project quality risk management has become increasingly professional, systematic, and scientific, employing professional research personnel and specialized research institutions to address future adverse events.

Project quality risk assessment research abroad has various characteristics. First, most project quality risk research scholars focused on the technical level, including project decision analysis, coping strategies, project risks, and international economic research. Second, with the rapid development and wide application of computer technology, calculation methods have become more accurate and foolproof. The combination of qualitative and quantitative risk analysis methods was the primary approach employed in project quality risk management research [1-5].

2.2 STATUS OF DOMESTIC RESEARCH

Although domestic risk management started late, risk manage men has been widely used in project construction, finance, real estate development, and other fields. Many studies have been conducted on project quality risk, and research results have been beneficial. It has analyzed the causes of engineering quality accidents in detail and used the theory of project management system to discuss how to improve project quality further [6]. It combined fuzzy mathematics and analytic hierarchy process, constructed a quality risk evaluation index system for housing construction project construction, and proposed the housing construction project construction quality risk assessment model [7]. Additionally, it proposed a project risk evaluation model based on cost, time, and quality. They combined the analytic hierarchy process and fuzzy comprehensive evaluation method in a detailed study on risk factors, including cost, time, and quality of influence [8]. It believed that project quality risk exists throughout the whole construction period of a project, such that measures should be implemented at all stages of project quality risk assessment [9]. Steps must be taken to reduce the quality of engineering project risks. Further more, it used analyses of natural risk, material risk, and construction risk to establish a risk evaluation index system of quality and then evaluated the quality of a project risk with the use of the fuzzy evaluation method [10]. It established a project quality risk evaluation index system based on the cause of project quality risk factors, evaluated the quality of the project risk using the fuzzy comprehensive evaluation method, and established a quality risk response according to evaluation results [11]. Finally, it established a project quality risk management model based on Bayesian network and conducted risk identification, risk evaluation, risk diagnosis, and risk control and achieved results using the project quality risk model [12].

3 Identification of construction project quality risk factors

Project construction can be characterized as long-term, outdoors, and large-scale, such that the process is easily influenced by natural and social conditions. Combining the construction technology of production and that of liquid variability and other characteristics reveals that numerous factors affect project quality. These factors may result in construction quality accidents. Applying the principles of system engineering and the Delphi method to a large number of construction quality accident investtigations and data analyses, we identify the main factors influencing engineering project construction quality as man, materials, method, mechanical, and environment. These factors are called 4M1E.

Under normal circumstances, most project construction quality accidents are caused by factors related to 4M1E, such that identifying and analyzing these factors can ensure that the project runs smoothly with high construction quality.

3.1 THE HUMAN FACTORS

Man refers to the project managers, leaders, and operators involved in the project. Man directly and indirectly affects construction quality because of personal, ideological, cultural, business, and quality attributes, among others.

3.2 THE MATERIAL FACTOR

Material refers to raw materials, auxiliary materials, semi-finished products, components, and fittings, among others. The quality of the material determines the level of project quality. If the material quality does not meet the requirements, meeting the project quality standards is impossible.

3.3 MECHANICAL FACTORS

Mechanical, machinery, and equipment generally refer to construction equipment, construction machinery, and construction tools and instruments. With the continuous expansion of project scale, machinery and equipment have become indispensable parts of a facility.

3.4 THE FACTORS OF METHOD

Methods refer to the employed technology programs, processes, construction designs, and construction technology measures during the construction phase. To some extent, choosing the construction program reasonably and correctly directly affects project quality control.

3.5 ENVIRONMENTAL FACTORS

Environmental factors associated with the construction quality of a project include three major factors: the scene of the natural environment, construction quality management, and engineering labor work environment factors. The environmental factors change constantly during construction phase, such that these factors directly affect the project.

4 Quality risk assessment of construction project on the basis of rough sets and neural networks

4.1 CONSTRUCTION OF THE INDEX SYSTEM

Based on the above analysis of the risk factors affecting the quality of the construction phase and combined with the principles and basis of indicators, the quality of risk evaluation index system of the construction phase of a project was built, as shown in Table 1.

TABLE 1 Project quality risk assessment index system of construction stage

Evaluation index	First Indicators	Number	Second indicators	Degrade				
				I	II	III	IV	V
U	U_1 Human factor	1	U_{11} Degree of compliance manager qualifications					
		2	U_{12} Degree of operator qualification standards					
	U_2 Material factors	3	U_{21} Standards of the material quality					
		4	U_{22} Soundness of the material properties					
	U_3 Mechanical factors	5	U_{31} Degree of compliance with the quality of machinery and equipment					
		6	U_{32} Rationality of machinery and equipment selection					
		7	U_{33} Standardized machine operators					
	U_4 Method factors	8	U_{41} The rationality of the construction technology program					
		9	U_{42} Construction technology and construction methods of the advanced and rationality					
		10	U_{43} Rationality of construction detecting methods and construction technical measures					
	U_5 Environmental factors	11	U_{51} Natural environment of construction site					
		12	U_{52} Quality assurance system of construction unit					
		13	U_{53} Quality management system of construction unit					
		14	U_{54} Technical and economic conditions					
		15	U_{55} Construction work environment					

Description: I very good, II good, III fair, IV poor, V very poor.

4.2 PROJECT CONSTRUCTION QUALITY RISK ASSESSMENT ON THE BASIS OF ROUGH SET AND NEURAL NETWOKKS

This study proposes a neural network risk assessment method using a combination of weak coupling methods, namely, rough sets and neural networks. With rough theory as a neural network front-end system, the sample data input of the neural network is first simplified. The simplified sample data are then imported to the neural network, and the mature classification model is produced through training. Finally, the quality of the risk level of the project is evaluated on the basis of the output of the classification model network, which can provide an objective basis for the decision-making of project managers.

Given the basic idea of rough set and neural network risk assessment methods and the actual characteristics of the construction quality risk assessment, a quality risk assessment model of construction projects can be established on the basis of neural networks and rough sets. The basic idea is as follows:

1) Collecting data: by collecting detailed information on the construction site of a project, possible accidents caused by the project construction quality risk factors can be identified;
2) Data processing: discrete data processing is performed;
3) Identified the smallest risk factor set of construction quality of projects: by the means of the theory of rough set attribute reduction of the construction quality of construction project risk evaluation index system, the smallest project quality indicators of the risk factor set can be identified;
4) Data training and establishing network model: training of the neural network data can aid in establishing an appropriate network model;
5) Testing the network model: the network model is tested by using specific project data, which are compared with the actual risk value;
6) Calculating the risk assessment result: the quality risk evaluation index data are used as input to the project construction to be evaluated. The value risk assessment objectives are then derived.

5 Empirical analysis

This study takes the characteristics of project construction as a foundation to establish the quality of the risk assessment model of construction on the basis of rough sets and neural networks. Data collected in the field will be applied and tested.

5.1 DATA COLLECTION

Data were collected at the Ganzhou Development Zone, a high-rise residential building project with two underground floors, 11 floors above ground, and underground parking lot. The project has a total gross floor area of 3,802m^2, a building height of 34.2m, take-situ concrete construction, columns, beams, plates, and concrete strength grade of C30.The project was constructed during summer. The construction site has various facilities, convenient transportation, and good work environment. In accordance with the construction process established in this study and the construction quality of the project risk evaluation index system, we selected the high-rise residential building project layer cast concrete columns for this study.

5.2 CALCULATING THE INDEX WEIGHT AND FORMATION THE SAMPLES SET OF THE RISK FACTORS

Based on the collected data, the number of samples in this article does not meet the required number of samples for the model. Thus, 14 projects of similar nature were added to the sample data for analysis. However, these projects are not listed here.

These projects were in the construction phase. In-site concrete columns and quality risk assessment indicators for each construction stage sample are presented in the Table 2. The symbols "" represent the level of the actual situation, which can be "very good, good, fair, poor, or very poor."

TABLE 2 Actual situation for each sample indicator

Sample	U_1		U_2		U_3			U_4			U_5				
	U_{11}	U_{12}	U_{21}	U_{22}	U_{31}	U_{32}	U_{33}	U_{41}	U_{42}	U_{43}	U_{51}	U_{52}	U_{53}	U_{54}	U_{55}
1	I	II	II	II	II	III	II	II	II	III	III	II	II	I	II
2	II	II	II	II	III	II	III	II	III	III	II	II	II	II	III
3	II	II	II	II	III	III	III	II	III	II	II	II	II	III	IV
4	II	III	II	II	III	II	II	II	II	II	III	II	II	III	III
5	I	II	III	III	III	III	III	II	II	II	IV	II	II	III	IV
6	II	IV	II	II	III	III	II	I	II	II	II	III	III	III	II
7	III	III	III	III	I	II	II	I	I	II	II	II	II	III	II
8	II	I	II	II	III	III	III	II	II	III	II	II	II	II	I
9	II	IV	III	III	II	III	III	II	II	II	I	II	II	II	I
10	III	III	III	III	II	III	III	II	II	III	IV	III	III	III	IV
11	II	III	II	II	II	I	I	II	III	III	Ⅱ	III	III	II	III
12	I	II	II	II	III	III	III	III	II	II	III	I	I	II	III
13	II	III	II	II	II	III	III	II	II	III	II	II	II	IV	II
14	II	I	III	III	II	II	II	I	II	III	II	II	II	II	III
15	II	II	II	II	III	III	II	I	II	II	II	III	III	II	II

On the basis of the original data, we used the AHP-Fuzzy evaluation method was used to calculate the right weight of each site and the quality of the risk of each sample. The risk sample sets of neural networks were then reduced.

First, we determine the weight of paper using AHP. Eight staff members with extensive experience are employed to compare the risk factors in the construction site. These staff members include project managers, construction workers, technicians, security staff, chief super vision engineers, quality engineers, and so on. The judgment matrix is then established. The U_1–U_5 layer-right weight set is: U= (0.271,0.311,0.236,0.104,0.079). Other indexes are calculated in the same manner as the criterion level U_1–U_5. The final weight is determined as follows: U_1 = (0.500, 0.500), U_2 = (0.500, 0.500), U_3 = (0.333, 0.333, 0.333), U_4 = (0.489, 0.291, 0.220) and U_5 = (0.333, 0.111, 0.111, 0.333, 0.111).

Second, each sample of construction quality risk is calculated using the fuzzy comprehensive evaluation method. Similarly, the quality of the risk can be drawn from all the construction site samples, as shown in Table 3.

TABLE 3 Construction quality risk assessment of each site samples

Sample	Risk	Sample	Risk	Sample	Risk
1	0.163	6	0.369	11	0.317
2	0.236	7	0.335	12	0.212
3	0.289	8	0.218	13	0.352
4	0.258	9	0.343	14	0.326
5	0.392	10	0.421	15	0.265

The risk of sample sets of the neural networks was determined, from which the former 10 sample sets were selected for learning neural network training, and the other five sample sets were selected as the test samples for the neural network.

5.3 CONSTRUCTION QUALITY RISK ASSESSMENT PROCESS OF PROJECTS BASED ON ROUGH SET AND NEURAL NETWORK

Construction project quality risk assessment processes were established on the basis of rough set sand neural networks. The index system was simplified with the ge-

netic algorithms in Rosetta software, and the discrete set of post-processing was incorporated into the calculation in the software. The set of simplified data is reduced by the software, as shown in Figure 1.

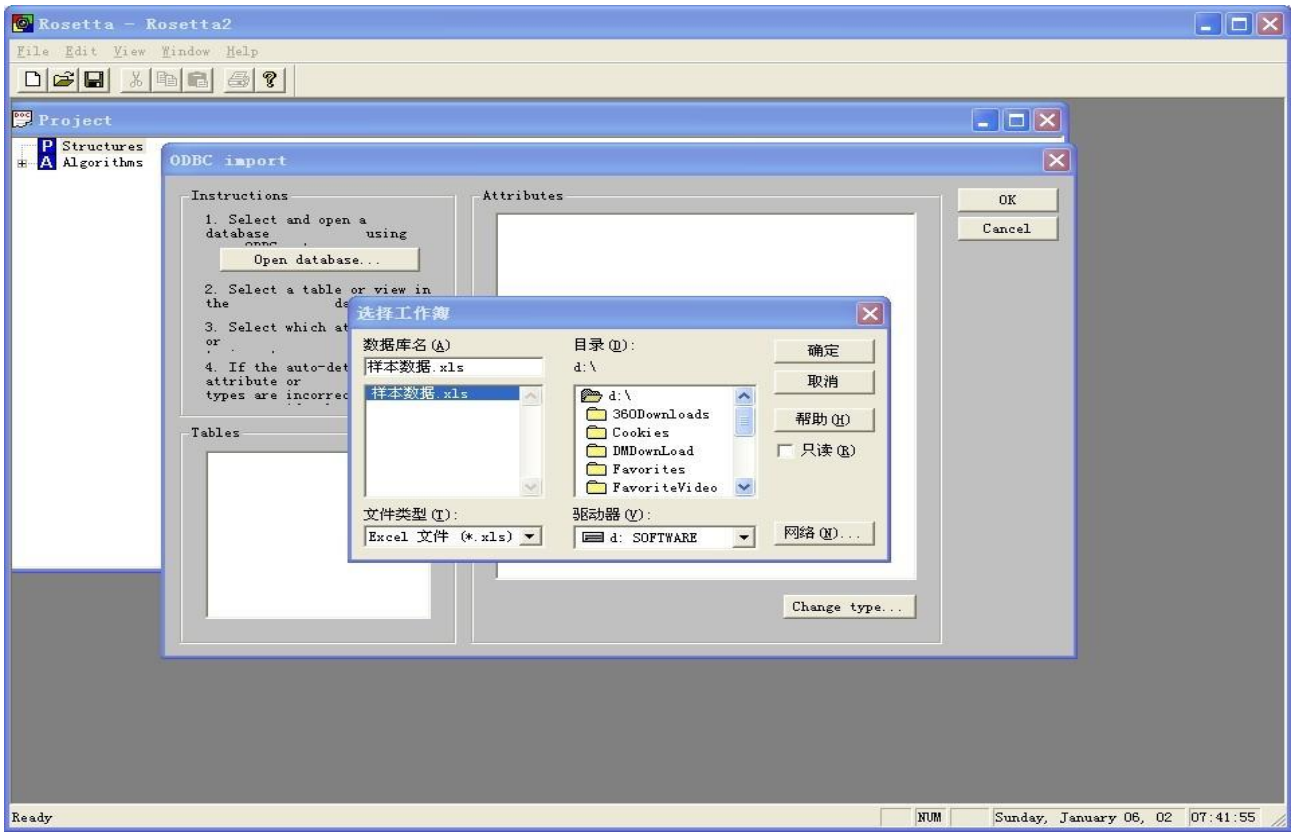


FIGURE1 Rosetta import data manipulation

In this study, considering the support and reduction lengths of the two factors, we selected an optimal set of conditions as follows: U = (U_{11}, U_{12}, U_{21}, U_{22}, U_{31}, U_{33}, U_{41}, U_{54}).

MATLAB7.0 was then used to define, train, and test a back propagation BP neural network. The specific steps are as follows:

Creation of the input and output data. Ten former sets of sample data were selected as the training sample set. The input matrix has a total of 10 lines, representing 10 samples, and a total of eight rows, with eight data points for each row in each sample.The eight data points represent the value attribute of U_{11}, U_{12}, U_{21}, U_{22}, U_{31}, U_{33}, U_{41}, U_{54}. The output matrix t has a total of 10 rows, which represent the value of quality risk attribute in the construction of a project.

Creation of a BP neural network. The specific statement is as follows: net=newff([0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1;],[17,1],{'tansig', 'tansig'},'traingdx'). The range description of each value input unit from the first parameter input is [0,1]. The second parameter [17,1] describes node values of 17 and 1 in the hidden layer and output layer, respectively. The third parameter {'tansig', 'tansig'} illustrates that the hidden layer and output layer transfer functions are all tansig functions. The fourth parameter 'traingdx' illustrates the training function selected for the function of neural networks.

Setting of training parameters. The specific statement is as follows:

net. Train Param. show=50
net.trainParam.lr = 0.09
net. trainParam. epochs = 10000
net. trainParam. goal = 0.001

Start training network. Specific statement is follows as:

[net, tr]=train(net, p', t')

Inspection of data. The five remaining sets of data are used for testing *.pt* is the testing matrix that has a total of five linesand eight rows. The specific statement is as follows:

a=sim (net, pt')

Each of the above statements are used as input for MATLAB7.0 software to obtain the final training results as shown in Figure 2.

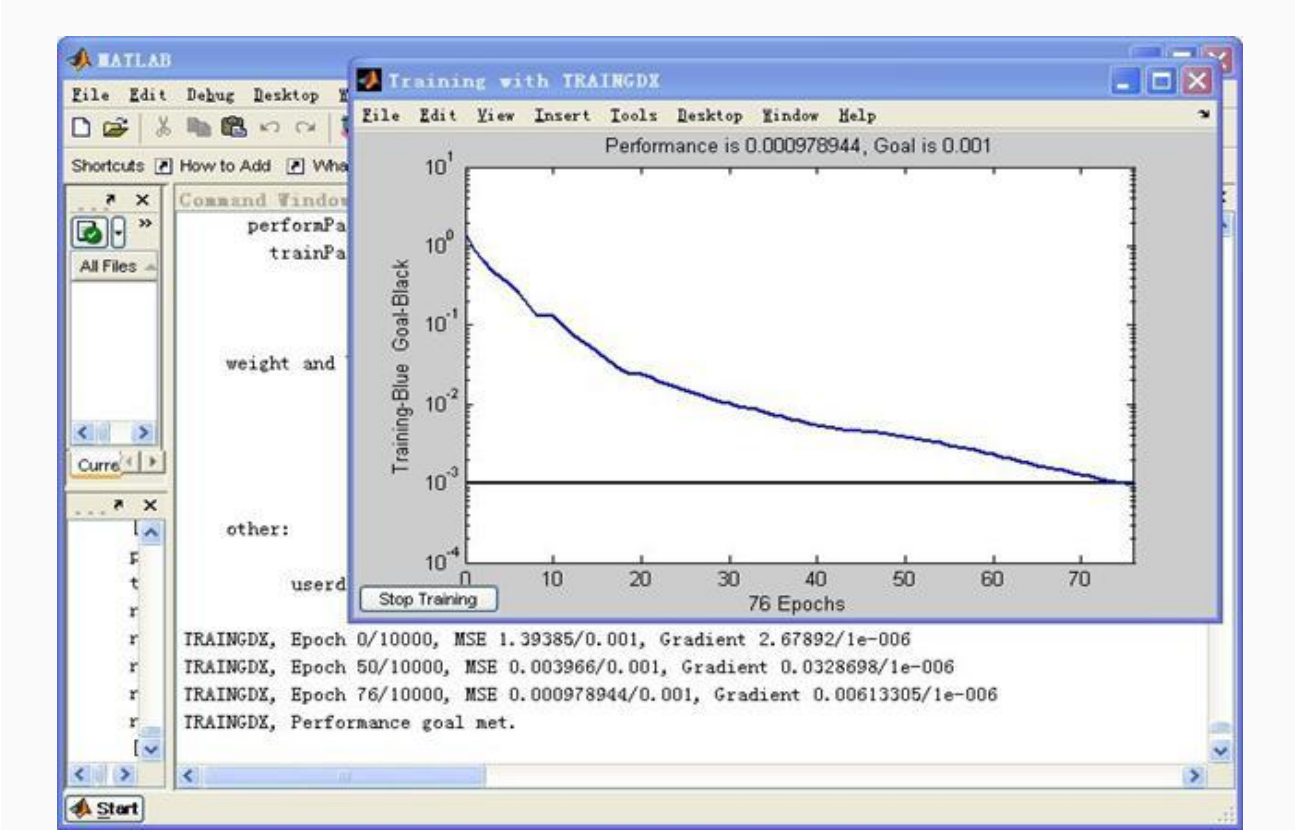


FIGURE 2 Training error curve

As shown in Figure 2, after 76 iterations through the training, the accuracy of the model is less than 0.001. The remaining five groups of samples were used to test the accuracy of the neural network model. The result of the command a = sim (net, pt ') is as follows:

a=sim(net, pt')
a =0.3204 0.2221 0.3433 0.4069 0.2747

The results of the output show that the actual output risk values of sample sites 11 to 15 (i.e., the risk value) were 0.3204, 0.2221, 0.3433, 0.4069, and 0.2747, respectively. The quality risks of construction from the model of site evaluation samples 11 to 15 were smaller, smaller, smaller and smaller, general, and smaller. The results of the model are then compared with the actual value of the output. The difference between the absolute value of the error rate of the output value and the true value is divided by the actual value, as shown in Table 4.

TABEL 4 Output samples

Sample	**11**	**12**	**13**	**14**	**15**
True Value	0.317	0.212	0.352	0.326	0.265
Output Value	0.3204	0.2221	0.3433	0.4069	0.2747
Error rate	1.07%	4.76%	2.47%	24.82%	3.66%

Table 4 shows that the actual output value and the true value of the model are broadly consistent, indicating that the quality risk assessment of the accuracy rate after training and learning of the sample set in project construction is became higher, which is in line with the actual situation of the project. Most error rates were acceptable. The error rate of sample site13is relatively high. However, the average error rate of the five groups of samples was 7.356%, indicating that a higher evaluation model training accuracy results in a more ideal the evaluation.

6 Conclusion

6.1 ESTABLISHING CONSTRUCTION QUALITY RISK ASSESSMENT MAODEL OF PROJECTS ON THE BASIS OF ROUGH SET AND NEURAL NETWORKS

Neural network software programs (e. g., Rosetta and MATLAB7.0) were used as basic research tools in this study. The quality risk assessment model was tested through empirical analysis on the basis of rough sets and neural networks. The results of the empirical analysis show that higher accuracy of model training results in a more ideal evaluation.

6.2 THE INNOVATIVE APPLICATION OF ROUGH SET TO NEURAL NETWORK IN CONSTRUCTION QUALITY RISK ASSESSTMENT

Rough set theory knowledge was taken as a neural network front-end system, and the input data were preprocessed and compared with those derived from the traditional BP neural network, which effectively simplified the complex structure of the neural network, shortened the neural network training time, and improved the accuracy of generalization and training of the neural network.

Rough set and neural network risk assessment model can accurately simulate the decision-making process of construction quality experts on risk assessment of the project, thus artificially reducing uncertainties in the risk assessment process and effectively combining knowledge acquisition, functional expert systems, and fuzzy reasoning.

References

[1] Simister S J 1994 Usage and Benefits of Project Risk Analysis and Management *International Journal of Project Management* **12**(1) 5-8

[2] Arditi D, Gunaydin H M 1997 Total quality management in the Construction process *International Journal of Project Management* **15**(2) 235-43

[3] James P J 1999 Equipment management risk rating system based on engineering endpoints *Biomed Instrument Technology* **33**(2) 115-20

[4] Ward S, Chapman C 2003 Transforming project risk management into project uncertainty management *International Journal of Project Management* **21**(2) 97-105

[5] Soutter M, Alexandreseu M, Sehenk C, Drobot R 2009 Adapting a geographical information system-based water resource management to the needs of the Romanian water authorities *Science and Pollution* **16**(l) 33-41

[6] Chen R, Liu C 2006 The Project Quality Management *Science and Technology Information* **4**(34) 52

[7] Qiao J 2006 Fuzzy comprehensive evaluation of quality risk in hours building engineering construction *Construction quality* **24**(8)14-7

[8] Chen J, Zhang B, Xia J 2007 Based on Cost – time – Quality of project risk management process *Soft Science* **21**(3) 50-4

[9] Huang G 2008 Preliminary project quality risk management based on construction technology using fuzzy evaluation method *construction technology* S2 408-9

[10] Ye J 2008 Risk Management Quality housing construction project *China Construction* **11** 69-70 *(in Chinese)*

[11] He X 2010 Construction project quality risk management of based on fuzzy comprehensive evaluation *Economist* **2** 281-2

[12] Hu S, Mo J, Zhao Y 2013 Based on Bayesian network project quality risk management *Journal of Lanzhou Jiaotong University* **32**(1) 44-8 *(in Chinese)*

Author

Jianbing Liu, born in June 1971,Yongxin Jiangxi, P.R. China.

Current position, grades: School of Economics and Management, Jiangxi University of Science and Technology.
University studies: Master's degree at Jiangxi University of Science and Technology.
Scientific interests: project management, assess valuation.
Publications: more than 30 papers.
Experience: 20 years of experience in project management and assess valuation, more than 50 projects and assess evaluation of more than 20 projects, accomplished 5 provincial scientific research projects.

Fang Guo, born in September 1990, Yiyang Hunan, P.R. China.

Current position, grades: Graduate student of School of Economics and Management, Jiangxi University of Science and Technology.
University studies: BSc in of Civil Engineering at Jiangxi University of Science and Technology.
Scientific interests: project management and risk assessment.

Dynamic modelling of container transport modes between inland terminals and seaports

Lingyun Zhou[1*], Youheng Huang[2]

[1]*School of Traffic Engineering, Huaiyin Institute of Technology, Huaian City, Jiangsu Province, China, 223003*

[2]*School of Transportation and Logistics, Central South University of Forestry and Technology, Changsha City, Hunan Province, China, 410004*

Received 9 August 2014, www.cmnt.lv

Abstract

Container transport has played an important role in international trade. This paper mainly focuses on the container transport modes selection between inland terminals and seaports through a comprehensive comparison of multiple decision factors. According to the multi-object fuzzy mathematics decision-making theory and the container transport processes, the main factors influencing the container transportation mode decision are analyzed. Moreover, a fuzzy decision model of container transport modes between inland terminals and seaports is built by introducing varied weight factors. Thereafter, the proposed model for the final selection of container transport modes between Changsha city and Shanghai harbor is demonstrated through an illustrative example. The results of this example indicate that the model can reflect dynamically importance degrees of related decision factors and different demands of decision makers, and this approach provides a more accurate, effective, and systematic decision support tool for the optimized intermodal mode selection between inland terminals and seaports.

Keywords: container transport, transport mode, route optimization, fuzzy decision making

1 Introduction

The vast majority of liner cargo is containerized – that is, it is carried in sealed metal containers from point of origin to destination. Containers are moved with common handling equipment enabling high-speed intermodal transfers in economically large units [1]. The container, therefore, serves as the load unit rather than the cargo contained therein, making it the foremost expression of intermodal transportation. As an advanced mode of transport [2], container transport has been an important element of not only maritime activity, but also of world trade and of global industrial structure. Since the 1970s, with growing container transport as the main driver, most seaports have evolved into port communities [3].

At present, seaport business is increasingly focused on inland terminals through which the hinterland is served. Seaports are no longer purely considered intermodal transfer centers, but are now becoming comprehensive flow-through areas within logistics chains, which are functionally linked to distribution developments in the hinterland. Seaports and inland terminals belong to a tiered intermodal transport system serving the whole supply chains [4]. Intermodal transport mode involves the use of at least two different modes in a trip from an origin to a destination through an intermodal transport chain, which is facilitated by the use of containers allowing the transport by different modes of transport namely ship, truck and rail [5]. Each transport mode has different economic and technical structures, and provides different quality of transport services. The decision on mode choice is complex. As for how to get the optimum scheme, a suitable method is needed to make a comprehensive, reasonable and comparative study. The wide application of the fuzzy mathematics theory has offered a scientific way to solve the transportation route decision issue.

2 Influence factor analysis of container transport mode selection

2.1 TRANSPORTATION COST

Transport cost is important for transport mode selection. Transport cost includes rates, loading and unloading charges, and special services available (e.g., stopping in transit) from carriers [6]. Transport cost varies from mode to mode owing to the different cost structures of the modes, whereas there are cost variations among carriers within a transport mode because of their dissimilarities in cost structure [7]. While choosing the transport route, different schemes will produce different transport costs, and this factor usually becomes a primary one.

2.2 TRANSPORTATION TIME

In general, transport time includes the time required for pickup, handling, and delivery [8]. Containers transport time between inland terminals and seaports should be linked up with the lading time of arrived liner ships, and more fees such as storing fees, keeping fees should be paid if containers are delivered more ahead of time.

[*] ***Corresponding author's*** e-mail: zlycsu@163.com

2.3 TRANSPORTATION SECURITY

In the container transport industry, the importance of adopting technology for enhancing transport security has been well acknowledged [9]. To guarantee the security of the containers, delivered goods and the bills enclosed during the process of transporting is the basic requirement for transportation, and also it is a primary condition that must be considered while drafting all the transportation schemes that are prepared to be chosen.

2.4 TRANSPORTATION RELIABILIGY

Transport reliability refers to the carrier's capability and accessibility for providing the service over the route in question, and it determines whether a particular carrier can physically perform the transport service desired. While choosing the transportation route schemes, it is necessary to make an investigation and analysis about the possibility degree of gaining transportation capacity and the punctual conditions of transportation.

3 Decision making issue description of container transport modes

Consignors want to choose the optimal container transport mode between inland terminals and seaports, and the consignors have specified the nature of goods and transportation requirements [10]. After a detailed market has been conducted to identify the potential transport modes, m schemes available are educed through the initial screening. As for the m schemes available such as $b_1, b_2, \cdots, b_m$ between inland terminals and seaports. There are n evaluation factors such as $a_1, a_2, \cdots, a_n$ (transport time, transport cost, transport security, transport reliability, etc.), respectively corresponding to the varied weight of decision factors $w_1, w_2, \cdots, w_n$. Every scheme $b_j (j = 1, 2, \cdots m)$ in reserve has one index eigenvector $f_{ij} (i = 1, 2, \cdots, n; j = 1, 2, \cdots, m)$ corresponding to n evaluation factors $a_i (i = 1, 2, \cdots n)$ and the membership degree of its eigenvalue is r_{ij} to individual appraisal factor "excellent". To utilize a method that weighted relative warp interval is the minimum to choose the optimum scheme. Thereby, it is necessary to calculate the ideal scheme. The standard value vector of the ideal scheme is:

$$f^0 = (r_{11} \vee r_{12} \vee \cdots \vee_{1m}, r_{21} \vee r_{22} \vee \cdots \vee r_{2m}, \cdots, r_{n1} \vee r_{n2} \vee \cdots \vee r_{nm}) = (f_1^0 \wedge f_2^0 \wedge \cdots \wedge f_n^0) \quad . \tag{1}$$

The eigenvector of each available scheme that is the most close to the ideal scheme is illustrated as: $R_j = (r_{1j}, r_{2j}, \cdots, r_{nj})^T$, the weight of decision factors:

$$W = ((1+\varepsilon_1)w_1, (1+\varepsilon_2)w_2, \cdots, (1+\varepsilon_n)w_n)^T .$$

Among them: $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ and $w_1, w_2, \cdots, w_n$ are respectively the varied weight factor and constant weight items of each factor index. Hamming closing degree with weight is used to describe the quality degree of the scheme in reserve [11], namely:

$$N(f^0, R_j) = 1 - \sum_{i=1}^{n} W_i (f_i^0 - r_{ij}) . \tag{2}$$

If $T_j = \max[N(f^0, R_j)]$, $1 \le j \le m$, scheme b_j is an optimum scheme.

4 Building the fuzzy decision model of container transport modes

4.1 CALCULATING MEMBERSHIP DEGREE OF QUANTITATIVE FACTOR INDEX

When the index value f_{ij} of factor index i corresponding to scheme j is a quantitative index, the membership degree of the factor index is calculated based on the comprehensive decision-making method as:

$$r_{ij} = \begin{cases} 0.1 + \dfrac{f_{i\max} - f_{ij}}{d} & \text{if } f_i \text{ is a minus index} \\ 0.1 + \dfrac{f_{ij} - f_{i\min}}{d} & \text{if } f_i \text{ is a plus index} \end{cases} . \tag{3}$$

In the Equation (3), d is a grading value $d = \dfrac{f_{i\max} - f_{i\min}}{1 - 0.1}$, r_{ij} is the membership degree of factor item i to scheme j in reserve. In $n \times m$ dimensional space, n evaluation values of m schemes constitute a fuzzy evaluation matrix R.

4.2 CONFIRMING MEMBERSHIP DEGREE OF QUALITATIVE FACTOR INDEX

As for the confirmation of the membership degree of qualitative factor index, a comprehensive decision method is adapted to make an evaluation, when every factor index value f_{ij} is a qualitative index, the evaluation of the fuzzy matrix R can be made by experts. The specific way is that the factor index is to be divided into seven grades (worst, worse, bad, general, good, better, best), and the evaluation value can be get according to the standard shown in Table 1.

TABLE 1 The table of evaluation standard

Remark	Worst	Worse	Bad	General	Good	Better	Best
Evaluation value	0.05	0.20	0.35	0.50	0.65	0.80	0.95

4.3 CONFIRMING THE WEIGHT OF EVERY FACTOR INDEX

It is necessary to confirm the importance degree of every factor to the container transport schemes according to the true environment [12]. And the importance degree, namely, is the corresponding factor weight value of n evaluation factors: $w_1, w_2, \cdots, w_n$. In order to dynamically reflect the different importance degrees of each factor

index to each transportation route scheme, the varied weight factors of every factor index $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ should be given a corresponding value [11]. The weight set with varied weight of each factor, namely $W = ((1+\varepsilon_1)w_1, (1+\varepsilon_2)w_2, \cdots, (1+\varepsilon_n)w_n)^T$, can be gained, which shows the dynamic features of the evaluation factors.

4.4 EDUCING THE OPTIMUM SCHEME

Make a calculation according to Equation (2), and get the result $T_j = \max[N(f^0, R_j)]$. Therefore, scheme b_j is an optimum scheme.

5 Case study

5.1 TRANSPORT MODE INVESTIGATION AND SCHEME ANALYSIS

The land-based transportation section of the containers, sent from the Changsha city and exported via Shanghai Harbor, is chosen as a case of route scheme. The container transport routes between Changsha and Shanghai Harbor are illustrated in Figure 1.

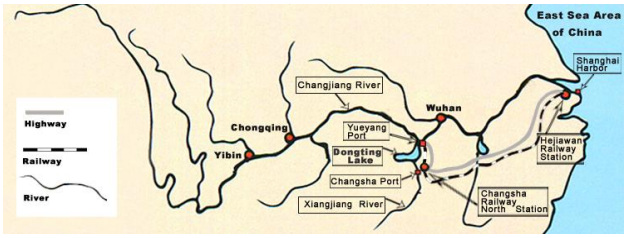


FIGURE 1 Container transport routes between Changsha and Shanghai Harbor in China

Scheme b_1: the railway + highway transport scheme, namely after the exported containers being sent at Changsha North Station, then they are transported to Hejiawan Station of Shanghai by railway, then transshipped on container trucks. Again they are transported to Shanghai Harbor by the short-distance road transport [13].

Scheme b_2: the whole highway transport scheme, namely after exported containers are loaded on container trucks in Changsha and transported to Shanghai Harbor by highway.

Scheme b_3: the whole inland water transport scheme, namely after exported containers are loaded into the ship in Changsha Port and transported through the Xiangjiang River, by way of Tongting Lake and Yueyang, to Changjiang River, then transported to Shanghai Harbor through the Changjiang River.

After Scheme b_1, Scheme b_2 and Scheme b_3 are designed based on the multi-modal freight transport network theory [14], the three schemes in reserve are chosen to be compared with each other. The index values of schemes in reserve have been educed in Table 2.

TABLE 2 The index values of schemes in reserve

	Scheme b_1	**Scheme** b_2	**Scheme** b_3
Transportation costs	￥3100 RMB /TEU	￥8300 RMB/TEU	￥2500 RMB /TEU
Transportation time	4 days	3 days	7 days
Transportation security	Better	Between "better" and "good"	Good
Transportation reliability	Worst	Best	Better

5.2 CALCULATING THE MEMBERSHIP DEGREE OF EVERY FACTOR INDEX

According to Table 2, the two indexes of "transportation costs" and "transportation time" are quantitative and minus indexes, so their membership degrees are calculated according to the Equation (3). The two indexes of "transportation security" and "transportation reliability" are qualitative indexes, so their membership degrees are calculated with corresponding assessed values of fuzzy comments, then the fuzzy matrix *R* is got:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \end{bmatrix} = \begin{bmatrix} 0.91 & 0.10 & 1 \\ 0.78 & 1 & 0.10 \\ 0.80 & 0.7 & 0.65 \\ 0.20 & 0.95 & 0.80 \end{bmatrix}.$$

According to the Equation (1), the index standard value vector of an ideal scheme is f^0 = (1, 1, 0.80, 0.95).

5.3 CONFIRMING THE WEIGHT OF EVERY FACTOR INDEX

With Delphi method, through the consultation investigation of two experts engaged in the Third Party Logistics, four experts engaged in international freight agency business and four experts engaged in transportation for imports and exports enterprises, and they are invited to evaluate every index weight. The result that the constant weight values w_1, w_2, w_3, w_4 of the 4 factor indexes of transportation costs, transportation time, transportation security and transportation reliability are 0.41, 0.22, 0.11, 0.26 respectively after the data of assigning the weight are collected and processed. After introducing varied weight factor, the weight value is:

$$W = ((1+\varepsilon_1)w_1, (1+\varepsilon_2)w_2, (1+\varepsilon_3)w_3, (1+\varepsilon_4)w_4)^T = (0.41(1+\varepsilon_1), 0.22(1+\varepsilon_2), 0.11(1+\varepsilon_3), 0.26(1+\varepsilon_4)).$$

5.4 CALCULATING THE HAMMING CLOSING DEGREE TO GET AN OPTIMUM SCHEME

Calculate Hamming closing degree of every scheme with Equation (2), the result is got as following:

$$b_1: N(f^0, R_1) = 1 - \sum_{i=1}^{n} W_i (f_i^0 - r_{i1}) = 0.720 - 0.037\varepsilon_1 - 0.048\varepsilon_2 - 0.195\varepsilon_4 ,$$

$$b_2: N(f^0, R_2) = 1 - \sum_{i=1}^{n} W_i (f_i^0 - r_{i2}) = 0.620 - 0.369\varepsilon_2 - 0.011\varepsilon_3 ,$$

$$b_3: N(f^0, R_3) = 1 - \sum_{i=1}^{n} W_i (f_i^0 - r_{i3}) = 0.746 - 0.198\varepsilon_2 - 0.017\varepsilon_3 - 0.039\varepsilon_4 .$$

When the weight of every factor index is constant weight and varied weigh factors are not considered, namely if $\varepsilon_i = 0 (i = 1, 2, 3, 4)$, then:

$$T_j = \max[N(f^0, R_j)] = \max[0.720, 0.620, 0.746] = 0.746 .$$

Thereby, scheme b_3 is the optimum scheme. If varied factors are evaluated, there are differences to some extent in the gained result of the scheme compositor.

6 Conclusions

The major contribution of this paper lies in the development of a comprehensive methodology for the planning of transport modes between inland terminals and seaports. According to the multi-object fuzzy mathematics decision-making theory and the centralized and evacuated transportation features of the containers, a fuzzy decision model of container transport modes between inland terminals and seaports is constructed, which is under the restriction of multiple qualitative and quantitative decision factors and based on varied weight factors. The introduction of varied weight, which can reflect dynamically importance degrees of related decision factors, adapts the decision model to the dynamic changes of the decision factors. This study raises several important issues that warrant further research. Evaluation and refinement of the model using additional field studies may prove beneficial, and its applicability may be expanded to other similar decision problems. Further, the intelligent software based on the methodology may also be developed.

Acknowledgments

The research work was supported by Science Research Project of Ministry of Housing and Urban-Rural Development of the People's Republic of China under Grant (2012-R2-5), and Qing Lan Project of JiangSu Province.

References

[26] Hayuth Y 1981 Containerization and the load center concept *Economic Geography* **57**(2) 8-16
[27] Cudahy B J Box Boats 2006 *How Container Ships Changed the World Fordham University Press New York*
[28] Notteboom T E 2004 Container shipping and ports: an overview *Review of Network Economics* **3**(2) 86–106
[29] Bergqvist R, Falkemark G, Woxenius J 2010 Establishing intermodal terminals *World Review of Intermodal Transportation Research* **3** (3) 285-302
[30] Rodrigue JP, Comtois C, Slack B 2009 *The Geography of Transport Systems Routledge: New York*
[31] Anderson J, Van Wincoop E 2004 Trade costs *Journal of Economic Literature* **42**(3) 691–751
[32] Gorman M F, Conway D G 2005 Logistics costs based estimation of freight transportation demand for capacity planning *Journal of the Transportation Research Forum* **44**(1) 141–55
[33] Crainic T G 2000 Service network design in freight transport *European Journal of Operational Research* **122**(2) 272-99
[34] Lun YHV, Wong WYC, Lai KH, Cheng TCE 2008 Institutional Perspective on the Adoption of Technology for Security Enhancement of Container Transport *Transport Reviews* **28**(1) 21–33
[35] Caschili S, Medda F R 2012 A review of the maritime container shipping industry as a complex adaptive system *Interdisciplinary Description of Complex Systems* **10**(1) 1–15
[36] Zhang J F, Deng BR 1988 *Application Fuzzy Mathematics Geological Press: Beijing* (*in Chinese*)
[37] Pinder D, Slack B 2004 Shipping and ports in the twenty-first century: globalization, technical change and the environment *Routledge London*
[38] Jiang B, Li J, Mao X Y 2012 Container ports multimodal transport in China from the view of low carbon *The Asian Journal of Shipping and Logistics* **28**(3) 321–44 (*in Chinese*)
[39] Wieberneit N 2008 Service network design for freight transportation: a review *OR Spectrum* 30(1) 77–112

Authors

Lingyun Zhou, born in March, 1980, Huaian City, Jiangsu Province, P. R. China.

Current position, grades: the associate professor of School of Traffic Engineering, Huaiyin Institute of Technology, China.
University studies: PhD in management in Beijing Jiaotong University in 2011.
Scientific interest: transportation engineering, logistics management.
Publications: 12 papers.
Experience: teaching experience of 9 years, 5 scientific research projects.

Youheng Huang, born in October, 1964, Changsha City, Hunan Province, P. R. China.

Current position, grades: the professor of School of Transportation and Logistics, Central South University of Forestry and Technology, China.
University studies: PhD in engineering in Beijing Jiaotong University in 2009.
Scientific interest: container transport, data mining.
Publications: 30 papers.
Experience: teaching experience of 27 years, 10 scientific research projects.

Location of competitive basketball athletes based on RSR comprehensive evaluation method

Xiaohua Ma*

Institute of Physical Education, Taishan University, Taian City, Shandong Province, China, 271000

Received 6 September 2014, www.cmnt.lv

Abstract

Rank sum ratio (RSR) positions each research index to superior or inferior index, conducts rank allocation of these indices in all evaluation objects according to rank principles, sorts them, and carries out grading according to the RSR value. In this way, the superior and inferior indices can be clearly determined. For complex open sports, good sport decisions can make each member give play to their strong points to achieve good results. However, most existing research methods belong to non-quantitative research methods. Research on the location of competitive basketball athletes is rare. This study adopts RSR comprehensive analysis method to assess examination indices for 15 members in a basketball team of a college to select the members that are most suitable for five locations in basketball competitions and corresponding benches and candidates. A general method is provided to decide the location of commercial basketball athletes.

Keywords: sport decision-making RSR comprehensive evaluation method, commercial basketball athlete

1 Introduction

Rank sum ratio (RSR) was proposed by Tian Feng (a famous scholar in China and former professor in Chinese Academy of Preventive Medicine) in 1998 [1]. RSR is a statistical analysis method that integrates the advantages of classical parametric statistics and modern nonparametric statistics [2]. This method was mainly applied in medical field in the early stage. RSR comprehensive analysis method is applicable to statistics and analysis of form and measurement data [3]. The main thought is as follows: in a form including n samples and m characteristics, dimensionless statistical magnitude RSR is gained through rank conversion calculation, and evaluation objects with RSR values are sorted; grading (many comparable groups) is then conducted according to the comparable groups, or a credibility interval of RSR square root arc sine transformation values (few comparable groups) is obtained [4].

Sport decision making means the process that athletes feel information, understand information, and take actions under sport situations [5]. The major differences of sport decision making from general decision-making process are the heavy mental stress in limited time, inconsiderable available decision-making information, and uncertain results [6]. Sport decision making mainly includes cognitive decision making dominated by logical thinking and intuitive decision making dominated by intuitive thinking [7]. For complex open sport events, sport decision making has vital influences on athlete performance and final results [8]. Thus, the study of sport decision making is an important issue in sport field. Basketball sport is a confrontational sport mainly based on shooting, lay-up, and slam dunk [9]. Both teams own five players. The team able to shoot the basketball into the basketball hoop can gain the score. Thus, in basketball competitions, selection and decision of strategies and tactics are vital. In recent years, research on competitive basketball tactic system mainly includes the following two trends:

1) the differences in the technologies are compared to discuss the differences in offensive and defensive techniques of the teams;

2) the scores gained and lost are predicted to conclude the factors that influence the team gaining and losing the score to analyze the offensive and defensive strategies of the teams [10]. The former stresses "to know others," whereas the latter stresses "to know you." As a saying goes, if you know others and know yourself, you will not be imperiled in every battle [11].

Based on the perspective of "knowing yourself," the best location for basketball athletes can be gained through analyzing each index of team members according to RSR comprehensive analysis method [12]. This study introduces classical RSR comprehensive analysis method in competitive basketball decision making, conducts statistical analysis according to the given sample data of the team members to be tested, obtains corresponding conclusions, and provides methods to make decisions about the location in competitive basketball sport [13]. This study has important practical significance and certain innovative significance.

2 RSR comprehensive analysis methods

2.1 BASIC THOUGHT OF RSR

Samples with a large RSR value are generally considered as good. Therefore, the superior and inferior indices during

* ***Corresponding author's*** e-mail: mxh8682067@163.com

ranking must be confirmed. The examination level of students is considered for example. High scores mean good performance. Score is thus the superior index. A long duration to answer questions leads to poor performance. The duration to answer questions is thus the inferior index.

Apart from simply differentiating superior and inferior indices, slightly superior indices and slightly inferior indices should be set in some complex problems. Ranking should generally be considered according to the problem itself.

2.2 GENERAL CALCULATION STEPS OF RSR

The basic steps of RSR are as follows: each research index is first positioned to superior or inferior index through survey, statistics, and professional knowledge; ranking allocation of these indices in all evaluation objects is then conducted according to ranking principle and sorted; they are finally filed according to RSR value. Figure 1 shows the specific flow chart for calculations.

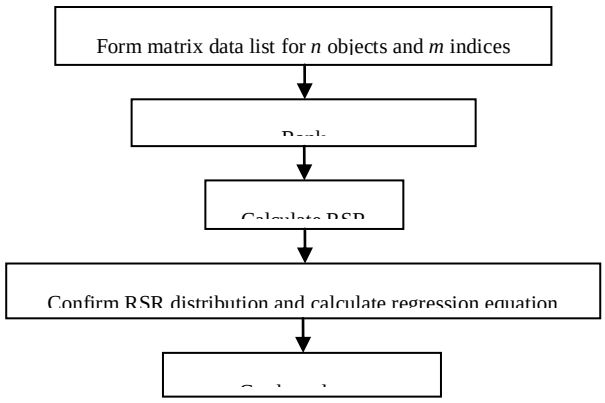


FIGURE 1 Flow chart of RSR calculation

The specific steps are as follows.

2.2.1 Formulation of data table

Formulation of data table is the calculation foundation of the whole RSR method. Follow-up calculation and grading can be effectively conducted only through selecting proper indices or characteristics and measuring them with data. n characteristics of m examples will generally be chosen. They are then ranked to $m \times n$ data table.

2.2.2 Ranking

Each index must be confirmed to superior or inferior index according to experience. The ranking principle is that superior indices are sorted in a descending order, whereas inferior indices are sorted in an ascending order. If the order is the same, the mean value is obtained. Equation (1) is the ranking calculation formula of superior indices. Equation (2) is the ranking calculation formula of slightly superior indices. Equation (3) is the ranking calculation formula of inferior indices. Equation (4) is the ranking calculation formula of slightly inferior indices. This study selects superior and inferior indices to rank for calculation convenience.

$$\text{Ranking of superiorindexes} = \frac{(\text{superiorranking} + \text{n}/2 + 0.5)}{2}, \tag{1}$$

$$\text{Ranking } \textit{ofslightly} \sup \textit{eriorindexes} = \frac{\{(\sup \textit{eriorranking} + \text{n}/2 + 0.5)/2 + \text{n}/2 + 0.5\}}{2}, \tag{2}$$

$$\textit{Rankingof} \inf \textit{eriorindexes} = \frac{(\inf \textit{eriorranking} + \text{n}/2 + 0.5)}{2}, \tag{3}$$

$$\textit{Rankingofslighly} \inf \textit{eriorindexes} = \frac{\{(\inf \textit{eriorranking} + \text{n}/2 + 0.5)/2 + \text{n}/2 + 0.5\}}{2}. \tag{4}$$

2.2.3 Calculation of RSR

RSR is a statistical magnitude with rich connotation and shows the comprehensive level of multiple indices with different measuring units. Calculation of RSR is often conducted according to row or line. The three main calculation methods are shown below.

In Formulas 5 and 6, m is the number of indices, and n is the number of samples. Equation (7) is the RSR calculation formula with weight, where w is the weight. Thus, this study adopts Equation (5) as the calculation foundation.

$$RSR = \sum_{1}^{m} \frac{R}{m \times n}. \tag{5}$$

$$RSR = \sum_{1}^{n} \frac{R}{m \times n}. \tag{6}$$

$$RSR = \sum \frac{RW}{n}. \tag{7}$$

2.2.4 Confirmation of RSR distribution and calculation of regression equation

One row is a group. Ranking R of each group of RSR is confirmed. The probability unit is $profit = \frac{R}{n} \times 100\%$. The regression equation of RSR is calculated. Equation (8) is the regression equation. Equations (9)-(11) are the linear fittings conducted according to least square method. Equation (9) is the augmented matrix of profit, and Equation (10) is the fitting equation of least square method.

$$\hat{RSR} = w_0 + w_1 \times profit. \tag{8}$$

$$A^T = \begin{pmatrix} 1 & 1 & \ldots 1 \\ profit_1 & profit_2 & \ldots profit_n \end{pmatrix}. \quad (9)$$

$$b^T = \left(R\hat{S}R_1 \; R\hat{S}R_2 \quad \ldots R\hat{S}R_n \right). \quad (10)$$

$$A^T A \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = A^T b. \quad (11)$$

2.2.5 Grading and sorting

Each sample is evaluated and sorted according to optimal principle and calculated in accordance with 95% CI. The computational formulas are Equations (12) and (13).

$$y = \sin^{-1} \sqrt{RSR}. \quad (12)$$

$$Filingvalue = y \pm \sqrt{\frac{820.7}{m \times n}}. \quad (13)$$

2.2 STEPS FOR APPLICATION OF RSR IN COMPETITIVE BASKETBALL SPORT

The detailed steps for application of RSR in competitive basketball sport are as follows:

1) a data structure statistic table similar to Table 1 is established;

2) the indices are positioned to superior or inferior indices according to experience and basketball knowledge;

3) they are ranked according to line, and numerical values of the samples of each index are gained;

4) RSR value is calculated, i.e., $RSR = \sum \frac{R}{m \times n}$;

5) RSR distribution is confirmed, and site selection, analysis, and decision making are conducted according to optimal principle.

TABLE 1 General RSR statistic table

	L1	L2	…	RSR
Player 1				
Player 2				
…				
Player 15				

3 Five locations in competitive basketball sport and analysis of relevant examination indices

Basketball sport location was only divided according to the locations of rear guard, forward, and center during offensive and defensive in the past. In fact, global basketball development trend gradually divides the five locations according to the distinction of "technical functions" of basketball athletes on the court. Each location is introduced one by one as follows. Superior and inferior indices are also selected.

3.1 CENTER

The center should own a tall figure, explosive power, balance, and the ability to resist collision. The center plays important work in scoring in the forbidden zone, passing, grasping the rebound, and defense.

First, the center should own strong ability to seize the rebound.

Second, the forbidden zone is a place of strategic importance. One team should not let the opponents to enter the forbidden zone easily. Thus, the ability of offensive resistance and block shot is essential.

Finally, during offensive, the center often has the opportunity to catch the ball in the forbidden zone near free-throw line. The center should therefore have favorable ball-leading ability and deliver the ball at the proper corner.

To sum up, eight important indices should be considered when selecting the center, such as inside shot, close shot, low-post shot, free-throw, block shot, backboard, height, and strength.

3.2 POINT GUARD

The point guard primarily protects ball handling over the half court and drives the pace of the entire team to attack and defend. He should be capable of rapid traverse, ball handing, assisting offense, having a high range of three point line and commanding the pace for offense and defense.

Therefore, the eight most important considerations for selecting the point guard are: three-point shot, ball-controlling ability, passing ability, ball handling, shooting, stealing, physical agility, and speed.

3.3 SHOTTING GUARD

The shooting guard is the second ball controller and the one who initiates offense. The shooting guard is responsible mainly for scoring. He should have good three-point shot rate and stability.

Moreover, the shooting guard must be fast enough to seek the gap for a three-point shot. He sometimes has to seek gaps or opportunities for conducting singles because his shooting distance is usually quite remote. Thus, his shooting rate will not be high.

Therefore, when selecting the shooting guard, the factors mainly considered are: three-point shot, speed, perimeter shot, lay-up, close shot, passing, and physical agility.

3.4 SMALL FORWARD

First, the small forward should have a good bodily form and speed. Second, he should be capable of controlling and passing the ball well. Third, the small forward should handle the ball to enter, or enter empty-handed and catch the ball to score a lay-up. Fourth, he should be capable of long shooting. Fifth, if necessary, he must know how to conduct singles in the forbidden zone and have certain destruction ability.

In fact, the fundamental requirement for the small forward is scoring through long shot. His shot rate is not necessarily high.

In selecting the small forward, the following should be considered: perimeter shot, lay-up, slam dunk, free throw, block shot, backboard, height, and strength.

3.5 POWER FORWARD

The power forward should be tall and capable of grabbing the rebound. Meanwhile, he should be flexible, well-coordinated, and assist offense, defense, and covering. He also should be capable of perimeter shots and low-post singles.

Generally speaking, the indices to be considered in selecting the power forward are inside shot, close shot, perimeter shot, defensive, block shot, backboard, height, and strength.

4 Data statistics and analysis

In summary, the present paper selects 35 members of a college basketball team as the samples and the center of a competitive basketball game as basis for data, with each index scored by the coach (0-100 scores). The scores represent the mean performance of each member in basketball games. Table 2 is the table of center selection decision-making.

It can be seen from Table 2 that, in terms of the inside shot, No.35 performs most outstandingly; members No.17, No.22, No.13, No.25, and No. 32 perform relatively well. In the aspect of close shot, No.35 still performs best; members No. 10, No.17, No.9, and No.22 perform relatively well. In the perimeter shot, No.35 member still performs best, followed by members No. 15, No.13 and, No. 25. No.35 performs best in free throws, followed by members No.13, No.16, No.22 and No.33. For block shots, No.33 performs most outstandingly this time, and No.13, No.24 and No.35 perform favorably. In terms of grabbing the rebound, No.35 member performs best again; No.13, No.17, and No.22 are relatively excellent. No.17 is the tallest player, followed by No.13, No.25, No.32 and No.35 members, whereas the strongest member is No.35 followed by members No.33, No.22, No.24, and No.25 members.

TABLE 2 Table of center selection decision-making

Member No.	Inside shot	R1	Close shot	R2	Perimeter shot	R3	Free throw	R4	Block shot	R5	Backboard	R6	Height	R7	Strength	R8	RSR
1	59	19.0	52	15.0	68	12.5	65	13.5	48	25.5	44	21.5	66	15.5	62	16.5	0.50
2	54	21.0	45	20.0	70	11.0	73	7.5	66	13.0	44	21.5	49	24.5	55	18.5	0.49
3	32	28.0	34	25.0	38	30.5	50	26.0	44	28.5	25	28.0	35	31.0	15	34.0	1.14
4	61	15.5	54	12.5	71	10.0	76	6.0	56	22.0	59	8.0	58	19.0	62	16.5	0.39
5	11	34.0	13	35.0	9	35.0	16	35.0	16	35.0	0	35.0	30	33.0	13	35.0	0.99
6	28	30.0	32	26.0	48	26.0	46	27.0	43	30.0	29	26.0	44	28.0	45	24.5	0.78
7	52	23.0	35	24.0	54	20.5	36	29.5	58	19.5	33	25.0	46	26.0	53	20.0	0.67
8	23	33.0	22	28.0	38	30.5	36	29.5	20	34.0	19	30.0	40	29.0	16	33.0	0.88
9	67	12.0	73	4.0	62	17.0	66	12.0	61	16.0	45	19.5	67	13.5	65	13.5	0.38
10	65	13.0	79	2.0	75	5.5	64	16.0	71	8.5	56	12.5	71	6.5	64	15.0	0.28
11	32	28.0	26	27.0	52	23.0	37	28.0	45	27.0	18	31.0	13	35.0	30	31.0	0.82
12	48	24	40	22.0	39	29.0	52	24.5	48	25.5	49	16.5	45	27.0	41	27.0	0.70
13	82	5.0	50	16.0	78	3.0	77	4.0	77	2.5	72	4.0	79	2.0	76	11.5	0.17
14	79	8.5	43	21.0	66	14.5	64	16.0	68	10.5	45	19.5	68	11.5	66	11.5	0.40
15	42	26.0	21	29.5	49	25.0	30	32.0	44	28.5	9	33.0	32	32.0	43	26.0	0.83
16	53	22.0	65	6.0	75	5.5	78	2.0	64	14.0	49	16.5	68	11.5	52	21.0	0.35
17	86	2.0	72	5.0	75	5.5	72	9.0	59	18.0	70	5.0	83	1.0	71	7.5	0.19
18	59	19.0	39	23.0	45	28.0	53	23.0	49	24.0	41	23.0	53	23.0	45	24.5	0.67
19	32	28.0	16	31.0	32	32.0	33	31.0	25	33.0	27	27.0	36	30.0	32	30.0	0.86
20	26	31.0	46	19.0	56	19.0	73	7.5	71	8.5	37	24.0	67	13.5	65	13.5	0.49
21	72	11.0	63	8.0	52	23.0	65	13.5	63	15.0	57	11.0	63	17.0	55	18.5	0.42
22	83	3.0	77	3.0	72	8.5	77	4.0	73	5.5	79	2.5	66	15.5	86	2.0	0.16
23	25	32.0	12	32.0	22	33.0	20	34.0	38	31.0	21	29.0	49	24.5	36	28.0	0.87
24	76	10.0	63	8.0	63	16.0	70	10.0	74	4.0	79	2.5	69	10.0	79	4.0	0.23
25	82	5.0	62	10.0	80	2.0	62	19.0	73	5.5	62	7.0	72	5.0	78	5.0	0.21
26	61	15.5	54	12.5	72	8.5	64	16.0	72	7.0	68	6.0	54	21.0	46	23.0	0.39
27	64	14.0	48	18.0	60	18.0	62	19.0	53	23.0	58	9.5	71	6.5	51	22.0	0.46
28	79	19.0	53	14.0	68	12.5	62	19.0	60	17.0	52	14.5	54	21.0	71	7.5	0.44
29	80	7.0	21	29.5	46	27.0	61	21.5	58	19.5	56	12.5	70	8.5	79	9.0	0.48
30	60	17.0	55	11.0	66	14.5	61	21.5	67	12.0	48	18.0	70	8.5	67	10.0	0.40
31	5	35.0	10	33.0	20	34.0	23	33.0	37	32.0	2	34.0	17	34.0	17	32.0	0.95
32	82	5.0	63	8.0	54	20.5	67	11.0	68	10.5	58	9.5	75	3.0	77	6.0	0.26
33	79	8.5	49	17.0	75	5.5	77	4.0	78	1.0	52	14.5	54	21.0	80	3.0	0.27
34	45	25.0	5	34.0	52	23.0	52	24.5	57	21.0	11	32.0	62	18.0	35	29.0	0.74
35	87	1.0	80	1.0	93	1.0	85	1.0	77	2.5	82	1.0	73	4.0	96	1.0	0.04

Consequently, the primary conclusion is that member No.35 gains good scores in each index and may be the best candidate for the center. No.17 and No. 24 members are also excellent and can be considered as the center bench.

To make decisions more accurately, the 35 members are graded using the data in Table 2.

Table 3 shows the RSR value distribution of Table 2.

TABLE 3 Table of RSR value distribution

RSR	$\frac{\bar{R}}{n}\times 100\%$	profit
0.04	2.9	3.1043
0.16	5.7	3.4195
0.17	8.6	3.6342
0.19	11.4	3.7945
0.21	14.3	3.9331
0.23	17.1	4.0498
0.26	20.0	4.1584
0.27	22.9	4.2579
0.28	25.7	4.3474
0.35	28.6	4.4349
0.38	31.4	4.5155
0.39	37.1	4.6708
0.40	40.0	4.7467
0.42	42.9	4.8211
0.44	45.7	4.8920
0.46	48.6	4.9649
0.48	51.4	5.0351
0.49	54.3	5.1080
0.49	57.1	5.1789
0.50	60.0	5.2533
0.67	62.9	5.3292
0.70	68.6	5.4845
0.74	71.4	5.5651
0.78	74.3	5.6526
0.82	77.1	5.7421
0.83	80.0	5.8416
0.83	82.9	5.9502
0.86	85.7	6.0669
0.87	88.6	6.2055
0.88	91.4	6.3658
0.95	94.3	6.5805
0.99	97.1	6.8957

For the data in Table 4 is obtained using least squares fitting, and the linear regression equation of RSR is: $\hat{RSR} = 0.256\,profit - 0.787$.

TABLE 4 The four categories for 35 members

	Most suitable (RSR<0.109)	Bench (0.109<RSR<0.493)	Candidate 0.493<RSR<0.877	Inappropriate RSR>0.877
Member No.	35	2, 4, 9, 10, 13, 14, 16, 17, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 32, 33	1, 6, 7, 11, 12, 15, 18, 19, 23, 34	3, 5, 8, 31

Among 35 members, No. 35 member the most suitable center. The members listed in the second line of Table 4 are suitable for being the center bench. When members are lacking, members can be selected from the third line. Members in the last line are not suitable for being the center.

5 Conclusions

Traditionally, RSR is mainly used in the medical field. However, as a visual quantitative classification statistics method, RSR can be introduced in decision-making in sports. The present paper introduces classical RSR comprehensive statistics method in basketball decision-making. 35 members in a basketball team were chosen as the research samples. They were scored according to different indices and the RSR value of their scores were calculated and used as bases for selecting the optimal center player and benches.

The feasibility of using RSR comprehensive analysis in decision-making in basketball is proven through data. The ability of each member in different aspects is visualized through data, which is beneficial to targeted training and cultivation. In conclusion, this paper determines the athletes' location according to RSR comprehensive analysis, which makes the members harness their strong points in practical competitions and gain the best state and the optimal result. Thus, RSR has certain practical significance.

References

[1] Yan C, Zheng H 2008 Effects of sport decision-making on the speed and accuracy of basketball athletes *Journal of Beijing Sport University* **31**(11) 1569-82 *(in Chinese)*

[2] He W 2012 Study of defense ability of the strong team in 2012 European Football under the research view of RSR *Journal of Sports Adult Education* **28**(6) 90-2

[3] Wang M, Li J, Yan C 2007 Optokinetics study of sport decision-making accuracy and speed difference of basketball athletes *Journal of Beijing Sport University* **30**(6) 774-6 *(in Chinese)*

[4] Ma Z, Shan D 2009 Research and development of decision-making testing system A1.0 of basketball athletes *Journal of Tonghua Teachers College* **30**(10) 64-7 *(in Chinese)*

[5] Fu Q 2004 Overview of researches of sport decision-making. *Journal of Beijing Sport University* **27**(6) 863-5 *(in Chinese)*

[6] Parashar A K, Parashar R 2012 Innovations and Curriculum Development for Engineering Education and Research in India *Procedia-Social and Behavioral Sciences* **56**(1) 685-90

[7] Lin P, Wang W, Zhong Q, Chen Q 2009 Realization of RSR calculation and evaluation with excel *Journal of Mathematical Medicine* **22**(2) 183-6

[8] Tian F 1993 Grading problem in RSR *Chinese Journal of Health Statistics* **10**(2) 26-8 *(in Chinese)*

[9] Tian M, Liu S 1998 A new type of fuzzy neural network and its application *Fuzzy system and mathematics* **12**(4) 56-61

[10] Cantwell J D 2004 The Physician Who Invented Basketball *The American Journal of Cardiology* **93**(4) 1075-7

[11] Fromm E G, Quinn R G 1989 An experiment to enhance the educational experience of engineering students *Engineering Education* **79**(3) 424-9

[12] [Parhami B 1986 Computer science and engineering education in a developing country: the case of Iran *Education and Computing* **2**(4) 231-242

[13] [Yeomans S R, Atrens A 2001 A methodology for discipline-specific curriculum development *International Journal of Engineering Education* **17**(6) 518-24

Authors

Xuahua Ma, born on April, 1976, Taian County, Shandong Province, P.R. China.

Current position, grades: lecture at the Institute of Physical Education, Taishan University, China.
University studies: BSc in Sports Management at Qufu Normal University of Shandong in China. MSc at Hunan Normal University in China.
Scientific interests: sports psychology and sports law.
Publications: more than 8 papers.
Experience: teaching experience of 12 years, 1 scientific research projects.

Management information system for college track and field games on the basis of infrared radio-frequency technique

Wenxin Xu[1*], Jiwei Yao[2]

[1]*Institute of Physical Education and Sport Science, Fujian Normal University, Fuzhou 350007, Fujian, China*

[2]*Physical Education Institute, Hunan University of Science and Technology, Xiangtan 411201, Hunan, China*

Received 9 September 2014, www.cmnt.lv

Abstract

College track and field sports involve numerous participants and events. Moreover, recording and announcement jobs are complex. The requirements for accuracy and precision are high. The traditional manual management mode cannot adapt to development needs. Based on the advantages of Radio Frequency Identification (RFID) such as no-barrier read, remote penetration, speed scanning, large memory space, anti-pollution capacity, durability, diversified shapes, reusability, and good security, an RFID-based infrared radiofrequency technique is designed to achieve automatic identity verification for athletes and to construct a complete score management information system. An automatic check mode is used to identify athletes by combining an ID card, a card reader, and a computer. This mode can rapidly identify the personal information of athletes and prevent cheating. This system has important significance for improving management level and reducing manpower.

Keywords: RFID, track and field games, management information system, card reader

1 Introduction

Radio Frequency Identification (RFID) uses inductive space coupling (inductive or electromagnetic coupling) to automatically identify objects [1]. RFID is a non-contact automatic identification technology. This technology can identify specific objectives through a radio signal, and read and write relevant data without mechanical contact or optical contact between the identification system and the specific objective [2]. Common techniques include low frequency (125 k–134.2 k), high frequency (125 k–134.2 k), super high frequency, and microwave. In this study, infrared light is selected as the radio frequency signal source [3]. Common barcode technology has unparalleled obvious advantages in the following aspects: amount of information, read-write capacity, confidentiality, intelligence, photopathy, environmental adaptation, speed identification, running speed, reading distance, service life, and multi-tag identification [4]. RFID technology is used to achieve the fully-automatic management of college track and field sports; it is used to construct an automated score management information system, which can simplify management process and improve management efficiency [5].

Track and field is an important means to improving health [6]. Many methods and means in track and field sports are adopted by other sports and serve as the main means to improving athletic performance and promote metabolism, improve the functions of visceral organs, and comprehensively develop physical quality [7]. Track and field is an effective way to inspecting the effects of the proposed system. In various track and field sports, grassroots track and field sport has the strongest fundamentality and the widest generalization performance [8]. Given that national fitness consciousness is continuously increasing, physical education for college students enters a period of rapid expansion [9]. College track and field sport as an important link in college physical education is characterized by numerous participants and events, complex recording and announcement work, and high accuracy and precision [10]. College track and field sports are traditionally managed manually or artificially. Such management mode involves heavy workloads and a large amount of manpower. Moreover, errors may easily occur [11]. Thus, it cannot meet the requirements of college track and field sports.

The advantages of RFID (such as no-barrier reading, remote penetration, speed scanning, large memory space, anti-pollution capacity, durability, diversified shapes, reusability, and good security) have received much attention from many fields [12]. This study analyzes college track and field sports to confirm the functional objectives of this system as well as to design functional software, select hardware, and complete the design of score management system for college track and field sports.

2 RFID

2.1 RFID SYSTEM COMPOSITION

Given that RFID systems have different applications, their compositions also differ. Typical FRID is mainly

[*]***Corresponding author's*** e-mail: wenxin20142014@163.com

composed of an e-tag, a reader, RFID middleware and application system software. Table 1 shows the functions of each part and their composition.

TABLE 1 RFID system composition

Composition	Function	Composition
E-tag	Identify objects/articles; own data storage mechanism; capable of receiving electromagnetic field modulating signal of the reader and return data carrier of corresponding signals	Active tag and passive tag; low-frequency/high-frequency/superhigh-frequency e-tags; active/semi-active/passive tags
Reader	The device writes or reads data for RFID tags through radio-frequency coupling	Carries out the read-write function through integrated or split-type antenna
RFID middleware	Hardware management, data collection, data processing and data transmission; the "nerve center" of RFID systems	The software that connects RFID hardware, leads and controls data collection, filtering, and application; between RFID devices (reader and labeling machine) and backend application system
Application system software	Control coordination work of the tag and reader through middleware; processes all data collected using the RFID system; computes, stores, and transmits data	Application software developed according to the specific needs of different industries; can effectively control the reader when reading and writing e-tag information; carries out concentrated statistics and processing of target information collected

2.2 REALIZATION PROCESS OF RFID SYSTEM FUNCTIONS

When the tag enters the radiation range of reader aerial, it receives the radio-frequency signal sent by the reader. The passive tag sends the data stored in the tag chip through the energy gained from induced current; the active tag actively sends the data stored in the tag chip. The reader is generally equipped with middleware with certain functions. The middleware can read data, decode, directly perform simple data processing, and send the data to the application system. The application system judges the legality of the tag according to a logical operation and conducts corresponding processing and control in allusion to different settings.

3 Design of score management system for track and field athletics

3.1 ANALYSIS OF CUSTOMER BASE DEMAND OF SCORE MANAGEMENT SYSTEM FOR TRACK AND FIELD ATHLETICS

The main customer base of the score management system for track and field athletics includes athletes, in-charge persons of each department, referees, and management committees. System demand analysis for each type of users is shown in Table 2.

TABLE 2 Demand analysis for each type of customers of track and field athletics

Customer composition	Customer demand
Athletes	Registers, competes, and inquires about the scores
In-charge persons of each department	Athlete registration management; submit registration information to the management committee
Referees	Types in, modifies, and inquires about the cores of the items
Management committee	Works before the application: prepare sport events, sets competition schedules, sets competition events, divides the work between referees (1) Before the competition: athlete information entry, grouping, competition schedule preparation, preparing and printing the competition sequence (2) In the competition: athlete registration, score input, score modification (3) After the competition: competition result announcement, score printing, prize presentation

3.2 OVERALL DESIGN OF SCORE MANAGEMENT SYSTEM FOR TRACK AND FIELD ATHLETICS

3.2.1 Overall system objective

The overall system objective is to realize computer-automated management of college track and field sports, establish a perfect track and field athletics management system, change semi-labor management to computer automated management, and realize the following functions: athlete information tracking, automatic pre-competition reminding, historical score comparison, automatic score release, etc.

The development of this system can free athletes, in-charge persons of each department, referees, and management committees from heavy repeated data entry and statistics, realize information network sharing, and lay the foundation for information automation.

3.2.2 Functional design

According to customer demand analysis and the overall system objective mentioned above, the score management system mainly achieves the following functions: user management, athlete information management, score management, and database management, as shown in Table 3.

TABLE 3 Design of main functions of management system

Function	Object	Inclusion
User management	System administrator	Maintains uniform management of the management system customer base; maintains software security and stability
Athlete information management	In-charge persons of departments, management committees	Adds, modifies, and deletes personal information of athletes
Score management	Referees, management committees	Inputs, modifies, deletes, and publishes, scores
Database management	System administrator	Inputs current useful data; calls out of the database; maintains data security

In traditional college track and field sports, due to laggard means, athlete registration, i.e., personal information checking before the competition and score input after the competition, takes much time.

1) Athlete registration before the competition. This activity gives referees a very heavy workload and long waits also affect the state of athletes. Moreover, traditional registration modes cannot effectively prevent cheating, which affects competition fairness. In such case, RFID-based automatic registration can greatly shorten registration time as well as effectively prevent cheating and ensure competition fairness.

2) Post-competition Score management. In the traditional score input mode, referees must first confirm the list of athletes according to the scores and then input the data, a process that consumes much time and energy.

If athletes can be automatically identified, pre-competition registration and post-competition score input will be very simple. The following processing can be carried out:

1) Before the sports meeting starts, according to athlete information submitted by each department, the management committee uniformly manages athlete information. The information includes the name, ID number, gender, department, grade, photo, etc., of the athletes.

2) Before every competition, the administrative staff calls out the information of the athletes participating in this competition and inputs their IC cards based on RFID technology. Before registration, IC cards are distributed to athletes participating in registration.

3) During registration, the athletes wear the IC card at the registration site and the card reader is used for identification. Through scanning the IC card, the computer displays the registration information of the athlete including the photo. The referee can directly contrast the photo for registration and judge for consistency and cheating behavior.

4) When the competition is over, the card reader reads the IC card of the athlete. Then, the referee can complete score input.

After the competition is over, ID cards are collected uniformly. They can be used in the next competition again according to the above process. The specific application process is shown in Figure 1.

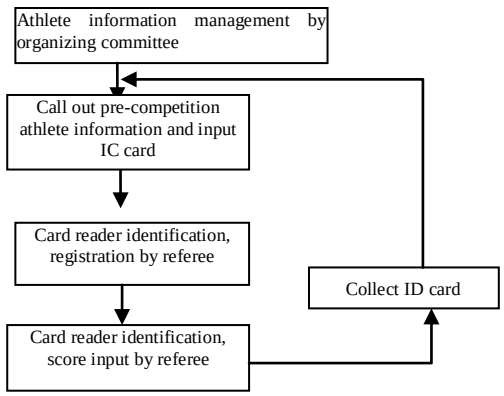


FIGURE 1 RFID-based ID card automatic athlete identification and application process

3.2.3 Network design

RFID uses a radio-frequency signal, space coupling, and the transmission characteristics of radio frequency signal to automatically identify static or mobile articles and data exchange. This technology uses a radio frequency to conduct non-contact both-way communication for identification and data exchange. Thus, the network design of track and field athletics site is also an important link of this management system.

Usually, optional network types include Wireless Local Area Network (WLAN) and wireless LAN.

LAN connects computers distributed in different places in the specified range. Using network software, computer resource sharing or communicative network system can be realized. Wired LAN usually uses copper wire or a cable as the main transmission medium of computer networking.

WLAN is the extension of fixed LAN that uses radio frequency technology to connect to a network without being limited by a cable and replaces traditional LAN made using twisted-pair copper wire. For users, it is completely transparent. Users can clearly establish the network meting their demands according to the simple architecture of WLAN.

TABLE 4 Design of main functions of management system

Type	Advantages	Disadvantages
Wired LAN	Work stably; fast communication speed	Fixed place, complex wiring. Complex work
WLAN	Move flexibly; easy t extend; convenient to use; easy to plan and adjust	Unstable signal, poor communication quality

Through comprehensive consideration, wired LAN and WLAN are combined to construct the overall network of the score management system.

Wired connection: 1) LED large screen and control computer; and 2) control computer and score input computer.

Wireless connection:

1) control computer and registration computer;

2) RFID card reader and registration computer; and

3) RFID card reader and score input computer.

The wired network connection depends on planning for wiring to completely establish a wired network. Wireless network establishment is based on the school wireless network. The router is set up based on the original wireless network to establish LAN. Through LAN settings, the computer on the platform is set to the mainframe to realize information sharing between the mainframe and extension. Thus, the mainframe can obtain competition information, which is convenient for comprehensive organization and uniformly printing information for sports meetings.

3.2.4 Analysis of system design rationality

After the system objective setting, functional design, confirmation of the ID card identification process, and

network design are completed, the Delphi method is used for opinion collection and modification to ensure effective and rational system design. The Delphi method uses predicted questions and background materials to create objective and scientific questionnaires, which are sent to the experts through door-to-door delivery or mail. The experience and knowledge of experts are used to make predictions. Through synthesis, concluding, and much feedback, opinions can basically remain consistent to improve the accuracy of predictions.

This study chooses 20 people including physical exercise experts, large-scale competition organizers, management experts, and software scholars to fill in the questionnaires, as shown in Table 5.

TABLE 5 Interview expert composition

Type	Title	Number	Proportion
Physical exercise experts	Associate professor	7	35%
Large-scale competition organizers		6	30%
Management experts	Professor	4	20%
Software scholars	Associate professor	3	15%

Expert consultation is divided into three rounds. Questionnaire collection and statistics are shown in Table 6.

TABLE 6 Results of expert consultation

Round	Number of questionnaires not collected	Number of questionnaires connected	Number of questionnaires issued
The first round	2	18	20
The second round	1	19	20
The third round	2	18	20

The questionnaire evaluation experts of 20 experts are shown in Table 7.

TABLE 7 Results of expert evaluation

Type	Effective	Basically effective	Invalid
Number	4	16	0

Table 6 shows that, to experts, the functional design of this management system is basically effective and rational. They also find that it can be used for score information management in college track and field sports.

4 Realization of information management system software for track and field athletics

4.1 DEVELOPMENT TOOL AND TECHNOLOGY

ASP.NET operates on Windows platforms under the .NET framework. It is a part of the Microsoft .NET strategy and mainly promotes the technology of Microsoft. ASP.NET is completely based on modules and components. ASP.NET has good expandability and customizability. It is established on the basis of CLR (Common Language Runtime), and mainly used to develop Web applications with strong functions in the server.

As a part of NET Framework, ASP.NET can use all services provided by this framework, including networking, data access, security, and other services. Thus, ASP.NET can create richer Web application programs and constructs building blocks needed by the application program conveniently and swiftly the logic of the application program can be focused on. Visual Basic 6.0 (vb6.0 download) is an event-driven programming language developed by Microsoft. This software includes assistance development environment. VB has the largest number of users. VB6.0 is derived from BASIC programming language. Visual Basic 6.0 has a graphical user interface (GUI) and a rapid application development (RAD) system. Thus, it can easily use DAO, RDO, and ADO to connect the database or easily create ActiveX control.

This study comprehensively applies VB language and ASP.NET technology for he management information system software for track and field athletics.

4.2 OVERALL BLOCK DIAGRAM OF SYSTEM SOFTWARE

System software design is the core content of the whole information management system for sports meetings. The system must complete backstage management interface and system functions must be completed. Thus, software design is an important constituent part of a system. VB language was selected to program the system software. Three major contents are included: backstage management, identity identification, and data information management. The overall block diagram of system software is shown in Figure 2.

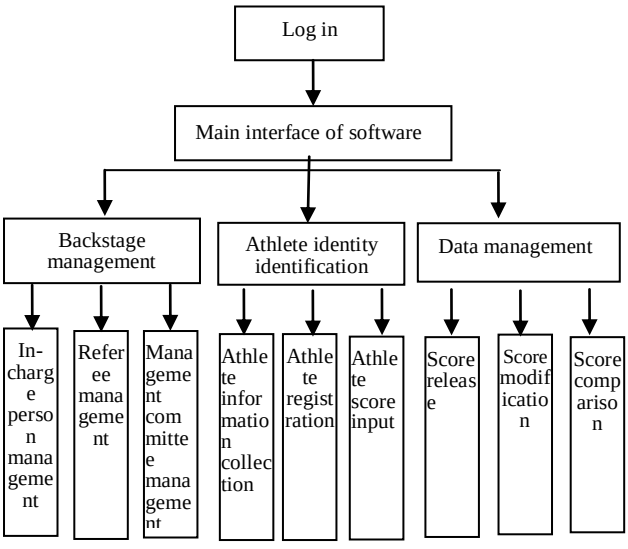


FIGURE 2 Overall block diagram of system software

4.3 REALIZATION OF USER LOG-IN FUNCTION

Use log-in process is shown in Figure 3. Specific codes written by VC are as follows:

```
Set adoRS = New Recordset
adoRS.Open "SELECT * FROM user where user name="& Textl.Text &"", db, adOpenStatic, adLockOptimistic
If adoRS.EOF Then
    Label4.Caption =""
Else
   If Textl.Text""Then
        Label4. Caption = adoRS.Fields("privilege")
        If Trim(Text2.Text) = adoRS.Fields("password")Then
            Unload Me
            db.Close
            MDIForml.Show
        Else
            MsgBox " wrong password"
            Text2.Text = ""
        End If
    End If
End If
```

If user information has been added in the corresponding user database, users can log in through the above process. After new user information is added, the following codes can be used to timely update backstage database, such that new users can promptly log in. The code is as follows:

```
Adodcl.ConnectionString -
"Provider=MicrosoftJet.OLEDB.4.0;Persist Security Info=False;Data Source=" & App.Path & "\name.mdb;Jet OLEDB :Database Password=abc;"
Adodcl .RecordSource = ”SELECT * FROM user”
Adodcl .Refresh
MSHFlexGndl.Refresh
```

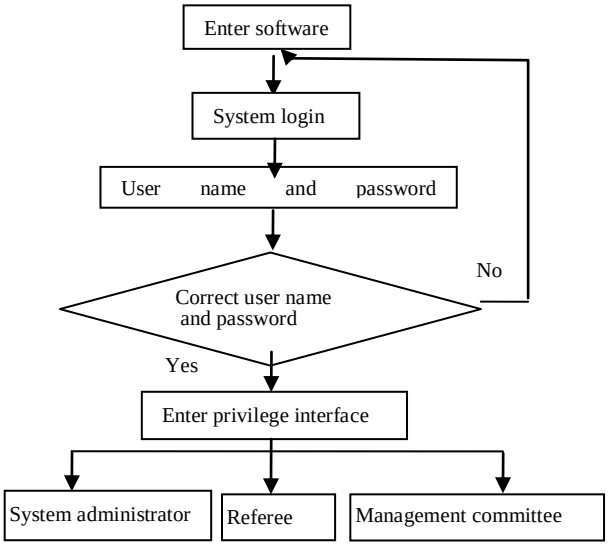


FIGURE 3 Flow diagram of user login function

4.4 REALIZATION OF ATHLETE IDENTITY IDENTIFICATION FUNCTION

Athlete identity identification is the core of the whole management system and was realized through the mutual recognition of IC card and card reader, wireless network connection to the computer, and data transmission. The reader serves as the slave computer, whereas the computer serves as the principal computer. The two conduct serial port communication through a RS-485 serial port.

Before the serial port is used for communication, an initialized setting of the parameters of the serial port is needed: the communication port number must be selected and the size of the buffer, the number of characters of the transmission buffer, the parameters of Baud rate, odd-even check, data bit, and stop bit must be set. The specific process is shown in Figure 4.

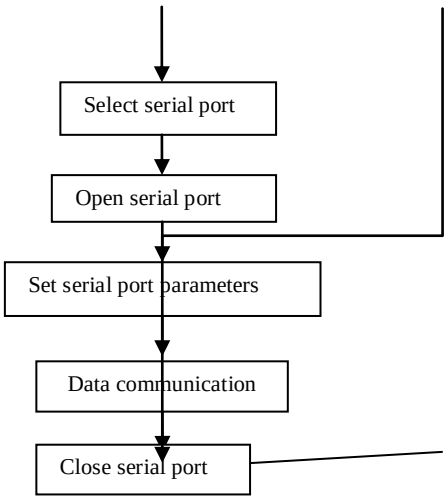


FIGURE 4 Flow diagram of serial port communication

The realization of specific codes is as follows:

```
MSComml.CommPort = 1                                    //set COMl
MSComm 1 .Settings = "9600, N, 8, r"
MSComm l.InputLen = 0// tell the control to read the whole buffer
MSComml .PortOpen = True //open serial port
MSComml.RThreshold = 16 // set to receive a byte to generate OnComm event
//Rem judge whether communication port is open (used in the button of “open serial port”)
If MSComml .PortOpen = False Then
MSComml .PortOpen = True //open communication port
If Err Then
                                                        // error handling
MsgBox "invalid serial port communication”
End If
```

4.5 REALIZATION OF SYSTEM PRINTING FUNCTION

The main functions of print module include printing athlete files for each event, competition results of each event, department scores and total scores, as well as competition arrangements. To realize this function, the score management computer is connected to the printer. The code is as follows:

```
Printer.Print
```

The system interface also has other functions, such as introducing competition rules of the sports meeting, which are not difficult to realize. We can click the toolbar button and apply VB code to connect the interface and WORD document and then click the event code of the button:

```
Dim wordObj
    Set wordObj = CreateObject("Word.Application") //create a word object
    With wordObj.Documents.Open("C:\Documents and Settings\Administratoi\ desktop\sports meeting new \new\ sports meeting rules.doc")
End With
```

5 Hardware design for score management system of track and field athletics

The hardware architecture of score management system for college track and field sports is shown in Figure 5.

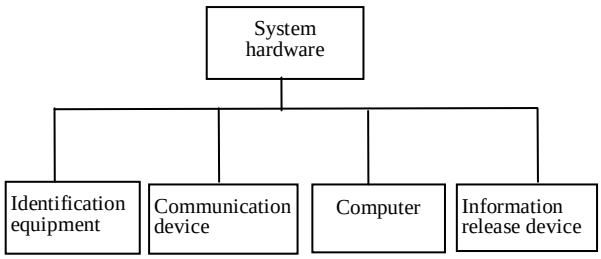


FIGURE 4 System hardware architecture

The card reader is a main part of the identification equipment and connects ID cards to the computer. EM-100T radio-frequency card reader is selected. EM-100T is a non-contact radio frequency IC card reader with a working frequency of 125kHz. The embedded micro-controller is designed with a radio-frequency receiving line and receives a 64-bit compatible ID card uein4100 combined with a decoding algorithm. The microcontroller has high receiving sensitivity, a small working current, and a single DC power supply.

Technical parameters of EM-100T are as follows:

Working frequency: 125 kHz

Working voltage: 9 VDC

Working current: 200 mADC

This system uses a close-loop wrist-strap ID card, which has high resilience and is easy to wear without affecting athlete performance. The watch-shaped card includes a non-contact chip that can send infrared signal to identify the athlete.

The hardware list for constructing the whole system is as follows: 4 computers including 1 host computer, 1 computer for releasing information, and 2 computers for registration; a large LED screen located on the platform of the sporting meeting; 8 card readers including 2 fixed-type card readers and 6 handheld card readers; 80 watch-type ID cards; and extra 20 IC cards for standby applications. The hardware list is shown in Table 7.

TABLE 7 Results of expert evaluation

Hardware equipment	Quantity
Computer	4
LED	1
Card reader	8
ID card	100
Router	4
RS serial port	2

6 Conclusions

This study uses college track and field sports as an example. Then, it starts from practical application and combines RFID technology, database technology, and network technology to establish a RFID-based sport information management system. This study focuses on the features of college track and field sports, such as numerous competition events, many athletes, tight time and heavy workload, and proposes an automatic checking mode for athlete identification combined with ID cards, an card reader, and a computer. Given that RFID technology is unique, it can rapidly identify personal information of athletes, prevent cheating, and improve sport management level. The completion of this system serves a promotional function for the development of college sports. Moreover, using RFID technology in college sports also significantly promotes sports as a whole.

References

[1] Feng Li, Rusong Wang, Juergen Paulussen, et al. 2005 Comprehensive concept planning of urban greening based on ecological principles: a case study in Beijing China *Landscape and Urban Planning* **72**(4) 325–36

[2] Caspersen O H, Konijnendijk C C, Olafsson A S 2006 Green space planning and land use: An assessment of urban regional and green structure planning in Greater Copenhagen *Danish Journal of Geography* **106**(2) 7-20

[3] Bentsen P, Lindholst C, Konijnendijk C C 2010 Reviewing eight years of Urban Forestry & Urban Greening: Taking stock, looking ahead. *Urban Forestry & Urban Greening* **9**(4) 273-80

[4] Deng Y, Shen X, Liu P 2006 A research of the management pattern of traditional settlements landscape under GIS support *Economic Geography* **9**(4) 693-7

[5] Wu Y 2006 Research status of RFID technology and development prospect *Microcomputer Information* **11**(2) 234-5

[6] Yang X 2001 Development of management system for college

track and field sports *Microcomputer Application* **17**(10) 21-3
[7] Yuan X, Yang B, Li B 2003 Development and application of information management system for track and field sports *Microelectronics and Computer* **22**(11) 67-9
[8] Preecha Noiumkar, Thawatchai Chomsiri 2012 Web's Dynamic Session IDs: Design and Analysis *Journal of Convergence Information Technology* **7**(2) 83-91
[9] Feng XU, Guijie QI, Yanan S 2012 The risk analysis of software projects based on Bayesian Network *Journal of Convergence Information Technology* **7**(5) 158-66
[10] Zhang ZF, Ren J, Hao Y 2002 Based GIS component Nanjing environmental pollution emergency monitoring GIS *Environmental Monitoring Management and Technology* **14**(4) 18-20
[11] Deng M, Zhao B, Xu Z, et al. 2011 GIS space targets distance expression and analysis *Computer Engineering and Applications* **47**(1) 35-9
[12] Li W, Zhang X, Jiang B, et al. 2013 A Web-based model for integrative and intelligent geographic information services technology *International Journal of Advancements in Computing Technology* **5**(1) 70-8

Author

Wenxin Xu, born on March, 1975, Fuzhou County, Fujian Province, PR China.

Current position, grades: associate professor of School of P.E. and Sport Science, Fujian Normal University, PR China.
University studies: Doctor Degree of Science in Education from Fujian Normal University in PR China.
Scientific interest: the science of physical culture and sports.
Publications: 15 papers.
Experience: teaching experience of 16 years, 6 scientific research projects.

Jiwei Yao, born in Septembri, 1976, Shaoyang City, Hunan Province, PR China.

Current position, grades: associate professor of institute of physical education, Hunan University of Science and Technology, PR China.
University studies: Doctor Degree of Science in Education from Fujian Normal University in PR China.
Scientific interest: the science of physical culture and sports.
Publications: 20 papers.
Experience: teaching experience of 16 years, 6 scientific research projects.

Models for measuring and predicting value creation during merger and acquisitions: a study of bank industry

Guo-yi Chen*

Chongqing three gorges university, Wanzhou City, Chongqing, China, 404120

Received 6 September 2014, www.cmnt.lv

Abstract

This paper employs event study methodology with a 36-day event window to assess the value effects of the US bank mergers occurring between 1994 and 2003. A 38-transaction sample is chosen from the top fifty US bank mergers (according to the assets of targets) occurred during the period from 1994 to 2003. Through analysis, result indicates that the average cumulative abnormal return of the bidders in the chosen sample is negative (-0.99%), while the targets and combined firms are both positive (15.07% and 2.57% respectively). Significance testing also verified that the negative bidder return is confirmed to be insignificant, whilst the positive return of the target and combined firm are both significant. Combined together, It indicates that the 3,517 US bank mergers occurred between 1994 and 2003 create insignificantly negative value for bidders, whilst benefit the target and the integrated banks with significant positive gains.

Keywords: mergers, acquisitions, event study, value effects, measurement model

1 Introduction

The US bank sector has experienced a couple of merger waves. The first one occurred in the 1960s and the second started from the late 1980s and has not finished yet [1]. During the second wave, till 1998, the number of US bank mergers accounted for 15 percent of all public mergers compared to 8 percent before [2]. Such outstanding proportion encourages researchers to search for the reasons of the popularity of M&A in the US bank sector, particularly in terms of the value effects, i.e. whether M&A creates value for the participants and its magnitude.

This study chooses stock market reaction approach as the method to assess the value effects of US bank mergers. A 38-transaction sample is chosen from the top fifty mega-bank mergers occurring from 1994 to 2003, which has not been analysed comprehensively by previous researchers. A 36-day event window is adopted for the event study considering the trade-off between the market efficiency and avoiding the impacts of irrelative noisy events. The stock market data are collected through software Datastream Advance 3.5. This paper observes negative average cumulative abnormal return for bidders and positive ones for both targets and their combinations. Further, t-test, sign test and Wilcoxon signed rank test are used to test the statistical significance of the above results.

2 Establishing value creation measurement model

The primary methodology of this research is event study. Event study is an empirical study of prices of an asset just before and after some event, like an announcement, of merger or dividend. It contains three main steps, namely identifying the event, calculating the average cumulative abnormal return and testing its statistical significance [3]. This paper follows Becher's (2000) 36-day event window (30 trading days prior to and 5 trading days after the announcement date) to test the value effects of bank mergers [4]. 5-day after the announcement date takes account of the under reaction of the stock market and, on the other hand, beneficial for avoiding the unwanted noisy events. 30-day prior to the event aims to involve the advance market reaction caused by the semi-strong form efficiency of the realistic stock market.

2.1 EVENT IDENTIFICATION

The sample used in this study initially contains fifty US bank mergers occurred between 1994 and 2003. They are the top fifty US bank mergers according to the ranking of the targets' assets. Through searching for relative data, including announcement date, share price, market value and S&P 500 Composite Index, there are eventually 38 transactions whose required data are completely available. The rest 12 cases are abandoned due to the lack of data of either the merging or merged firm. In this case, in Datastream Advance 3.5, which is the main source of stock market data for this study, the record under the current name actually reflects the historical data of the merging bank before and after the combination. The record of the acquired bank has thus been covered and cannot be found any more. This situation is severe especially for those banks that experienced several mergers and changed their name in the preceding way. The typical examples include Firstar Corporation and JPMorgan Chase. The main reason of choosing a sample comprised of large bank mergers lies in that it represents the most influential bank mergers in this ten-year period. The distribution of the sampled 38

* ***Corresponding autor's*** e-mail: cgy_22@126.com

bank mergers across years is illustrated by the following Table.

TABLE 1 Number of sampled transactions of each year

Year	Number
1994	2
1995	3
1996	7
1997	5
1998	8
1999	2
2000	1
2001	7
2002	2
2003	1

2.2 CUMULATIVE ABNORMAL RETURNS

After data collection, the next step is to calculate abnormal returns. As explained by Frame and Lastrapes, abnormal return is calculated by subtracting estimated normal return from actual return during the event period [5]. Estimated normal return may be attained through three most widely used methods, namely single-index model, Capital Asset Pricing Model (CAPM), and Market Model [6]. This study follows Becher (2000) to employ single-index model in estimating normal return. The standard and poor (S&P) 500 Composite Index is adopted in this study. S&P 500 Composite Index is a market value-weighted index containing 500 widely held common US stocks that measures the general performance of the stock market. Through observing its daily figures over the event period, the estimated normal return can be obtained by calculating the difference of the S&P 500 Composite Index between a certain day (n) and its previous day (n-1), and then dividing the difference by the index of the previous day (n-1),

$$NR_n = \frac{I_n - I_{n-1}}{I_{n-1}}, \tag{1}$$

where NR_n is the normal return of day n, I_n is the S&P 500 Composite Index of day n and I_{n-1} is that of day (n-1). The similar Equation is used to compute the change of share price (the actual return).

$$\Delta P_n = \frac{P_n - P_{n-1}}{P_{n-1}}, \tag{2}$$

where ΔP_n refers to the change in share price of day n from that of its previous day (n-1); P_n is the share price in day n and P_n-1 is that in day (n-1). Based on these calculations, the abnormal return (AR_n) can be attained through subtracting the normal return from the change of share price.

$$AR_n = \Delta P_n - NR_n. \tag{3}$$

The daily data of both S&P 500 Composite Index and stock price are collected through the software Datastream Advance 3.5. The cumulative abnormal return (CAR) can be computed using either arithmetic or geometric process. The arithmetic process computes the sum of the daily abnormal returns.

$$CAR = \sum_{n=-30}^{5} AR_n . \tag{4}$$

Compared to the simply summing, geometric process multiplies the abnormal return of a certain day (n) with the geometric CAR of its previous day (n-1), and then subtract 1 (Becher 2000).

$$CAR_n = (1 + AR_n) \times (1 + CAR_{n-1}) - 1. \tag{5}$$

Therefore, the geometric CAR over the selected event window, i.e. the geometric CAR of day (+5), is actually the product of every daily abnormal return minus 1.

$$CAR = \prod_{n=-30}^{5} (1 + AR_n) - 1. \tag{6}$$

In order to calculate the cumulative abnormal return of the combined bidder and target (CCAR), this study follows the method of Houston and Rygaert (1994), which is also employed by Becher (2000). The Equation is as below:

$$CCAR = \frac{MV_b \times CAR_b + MV_t \times CAR_t}{MV_b + MV_t}, \tag{7}$$

where MV_b and MV_t are respectively the market value of the bidder and target 30 days prior to the announcement date, which are also collected from Datastream Advance 3.5. The CCAR is the market value-weighted CAR of the combined bidder and target. It takes account of not only their separate CAR over the event window, but also their joint contributions towards the integrated CAR from the stock market. Being weighted by their market values at the beginning of the 36-day event window, the combined cumulative abnormal return reflects the cumulative abnormal increase in terms of the combined market value. This is an effective representative of the market reaction to the bank merger.

From the mathematic perspective, this method exactly calculates the abnormal increase of the combined market value over the event period. It can be proved through the following calculations. Firstly, the change of market value (ΔMV) equals the market value of the last day minus that of the first day of the event window.

$$\Delta MV = MV_1 - MV_0, \tag{8}$$

where MV_1 represents the market value at the end of the event window (in this case the day +5), and MV_0 represents that at the beginning (day –30). Subtracting the normal increase of the market value (NMV, the increase derived from the normal return) from it, ΔMV can be transferred to abnormal increase of market value (AMV), that is:

$$AMV = MV_1 - MV_0 - NMV \,. \tag{9}$$

Equation (9) indicates that the abnormal increase of market value is only composed of the cumulative abnormal return of all issued common shares.

$$AMV = n \times NAR \,, \tag{10}$$

where *n* refers to the number of common shares.

As the *CAR* here is a proportional increase of share price, to attain the net abnormal return per share (*NAR*), we need to multiply *CAR* with the share price of the beginning of the event window (P_0):

$$CAR = \frac{NAR}{P_0} \,,$$

$$NAR = P_0 \times CAR \,. \tag{11}$$

Combining Equation (10) and Equation (11), the next Equation is available:

$$AMV = n \times P_0 \times CAR \,. \tag{12}$$

Then it assumes the number of shares retains unchanging over the event window (through analyzing the sampled banks, none of them issued new shares during the event period), i.e. n1 = n0, then Equation (12) can be further transferred to Equation (13).

$$AMV = n_0 \times P_0 \times CAR = MV_0 \times CAR \,. \tag{13}$$

The proportional abnormal increase of the market value, i.e. the cumulative abnormal increase of market value (*CAMV*), is dividing abnormal increase of market value by the beginning market value.

$$CAMV = \frac{AMV}{MV_0} \tag{14}$$

When the bidder and target are treated as a whole, its beginning market value is the sum of their respective beginning market values (MV_b+MV_t), and the total abnormal increase of market value equals the sum of their individual abnormal increase of market value. Thereby, the cumulative abnormal increase of the combination's market value is:

$$CAMV = \frac{MV_b \times CAR_b + MV_t \times CAR_t}{MV_b + MV_t} \,. \tag{15}$$

Equation (15) is identical with Equation (8) which is used to determine the cumulative combined abnormal return. Therefore, the combined cumulative abnormal return determined by weighting their CARs is equivalent to the combined abnormal increase of market value, which reflects the integrated market reaction to the bank merger. Alternatively, this relationship can also be proved by simply combining Equation (13) and Equation (14):

$$CAMV = \frac{AMV}{MV_0} = \frac{MV_0 \times CAR}{MV_0} = CAR \,. \tag{16}$$

Most of these sampled transactions occurred at different time. Even if being announced at the same day, such as NationsBank' merger with BankAmerica and BankOne acquiring First Chicago NBD, the transactions did not dependent on each other. From this perspective, all the transactions in the sample can be regarded as independent and therefore the average combined cumulative abnormal returns of the sample ($\overline{CCAR}$) can be expressed by their arithmetic mean.

$$\overline{CCAR} = \frac{1}{N} \sum_{n=1}^{N} CCAR_n \,. \tag{17}$$

where *N* stands for the sample size.

2.3 SIGNIFICANCE TESTING

The final stage is to test the statistical significance of the sample mean ($\overline{CCAR}$). There are 3,517 bank mergers occurring in America from 1994 to 2003 [7]. What is chosen is thus a small sample relative to the population. After calculating the mean of the accumulative abnormal returns of all the 38 transactions, it is necessary to test the statistical significance of the result. This study applies t-test, sign test and Wilcoxon signed rank test in this stage. The reasons of choosing these three methods and their respective procedures are demonstrated as follow.

The t-test is the standard and most sensitive test for interval data [8]. It deals with the issues associated with the inference based on small samples. However, the t-test is subject to the assumption of a normal distribution. According to the central limited theorem that the mean of a sample is normally distributed if the size of the sample is over 30 [9], the 38-transaction sample in this study is large enough to consist with the required assumption and hence the result of the t-test should be valid.

The standard Equation of t-statistic is:

$$t = \frac{\bar{x} - u}{s} \sqrt{n} \,, \tag{18}$$

where $\bar{x}$ is the sample mean and u is the population mean; *n* is the sample size and s is the standard deviation [10]. Another parameter is p-value, which indicates how accurately the regression coefficient has been estimated. The p-value can be obtained through checking the Confidence Limits Table according to the t-statistic.

These two parameters above will assist the hypothesis testing. The null hypothesis will be set as that the population mean equals zero:

$H_{0:}$ u = 0.

Using the t-test and the corresponding p-value, we can decide whether to reject or accept the null hypothesis as well as how accurate the decision is. If the t-statistic is located within the confidence interval (under the condition of 37, (*N*-1), degrees of freedom and 5 percent significance level, i.e. p≥5%, we will accept the null hypothesis [11]. Otherwise, if p<5% the null hypothesis is unacceptable, i.e. u≠0. The p-value will provide the exact degree of significance of this decision. To calculate the p-value, according to Watsham and Parramore (1997), this paper firstly identify that the t-statistic can be interpolated in the interval between β_1 and β_2 [12].

Through checking the Confidence Limits Table, this interval corresponds to a interval of significance level between a_1 and a_2. Then the calculation of p-value can be based on the following Equation:

$$p = \alpha_1 - (t - \beta_1) \times (\alpha_1 - \alpha_2). \tag{19}$$

In this study, both the t-test and p-value will be calculated with the assistance of the statistical software Stata-8.

In order to ensure the accuracy of the statistical significance testing, this study also employs two nonparametric statistics, Wilcoxon signed rank test and sign test. Although the 38-transaction sample can be treated to consist with the normal distribution assumption, the actual distribution is still unknown. Under this circumstance, it is necessary to apply nonparametric statistics which are most appropriate when the sample is small. In contrast with t-test, sign test is not subject to any assumption on distribution and hence treated as the most insensitive test and also the most convincing and easiest test to be applied. It only uses the signs to test the null hypothesis that the number of + signs (the probability of observing a + sign is denoted as p, + sign represents the positive abnormal return) is equal to the number of – signs (the probability of observing a – sign is denoted as q, - sign represents the negative abnormal return), i.e. H0: p=q=0.5. The relationship between p and q consists with binomial distribution because the possibility of observing a + sign plus that of observing a – sign is equal to 1, i.e. p+q=1. Therefore we can employ the Equation of binomial distribution to calculate the p-value of each probability of observing + signs (x, x=n, n-1,…,0) in this sample:

$$p\{x\} = \frac{x!}{x!(n-x)} p^x q^{n-x}. \tag{20}$$

We choose a certain p_1 which makes $p\{x \geq p_1\}$, i.e. $p\{n\}+p\{n\text{-}1\}+\cdots+p\{p_1\}$, the closest probability to 5 percent level(degree of significance). If the p-value of real number of + signs in the sample ($p\{x \geq p\}$) falls in the critical region, i.e. $p\{x \geq p\} < p\{x \geq p_1\}$ (or <5%), we reject the null hypothesis. Otherwise, if $p\{x \geq p\} \geq p\{x \geq p_1\}$ (or ≥5%) the null hypothesis will be accepted.

Wilcoxon test takes account of not only the signs but also the magnitude of each sign through calculating the rank sums. Thereby, it is regarded as a more sensitive test than sign test. Moreover, it is more sensitive than t-test for small samples with unknown distributions [13] as it does not require estimating the parameters of the values (such as the standard deviation of the population) and the distribution of the sample.

In Wilcoxon test, the elements are ranked according to their absolute values (ascending sort). Then the ranks of the positive elements are summed (the sum is denoted by V). The null hypothesis (H0: u=0) here indicates that there is no difference between V and the rank sum of the negative elements (W), i.e. $V = W = \frac{1}{2}\sum_{n=1}^{N} n$ (N is the sample size). The mean (u_v) and standard deviation (σ_v) of V are calculated by the following Equations:

$$u_v = \frac{1}{2}\sum_{n=1}^{N} n = \frac{N(N+1)}{4}, \tag{21}$$

$$\sigma_v = \sqrt{\frac{N(N+1)(2N+1)}{24}}. \tag{22}$$

The z-value is computed by the Equation below:

$$z = \frac{V - u_v}{\sigma_v}. \tag{23}$$

Compared the z-value with the critical values under the condition of 5 percent significance level and 37, (38-1), degrees of freedom, if the z-value falls in the critical region the null hypothesis will be rejected, otherwise we accept it. The p-value is available from the table of Rank-Sum Critical Values, and its precise value can be calculated by the same method as that in t-test (Equation (19)). In this study, Stata-8 will calculate the corresponding p-value of each z-value. We will compare it with 5 percent significance degree: if p≥5%, we will accept the null hypothesis; Otherwise, if p<5%, the null hypothesis will be rejected.

According to the hypothesis testing assisted by t-test, sign test and Wilcoxon signed rank test, the mean of the cumulative abnormal returns will be justified to be acceptable or not as well as at what significance level. These three methods are all used for testing small samples. However, the nonparametric statistics sign test and Wilcoxon signed rank test relax the normality assumption of the t-test. Moreover, the Wilcoxon signed rank test take the magnitude of the signs into account and thereby can further confirm the result of sign test.

3 Empirical analysis of US bank M&A

3.1 SAMPLE MEAN

Through the operations based on the Equations (1-7) and assisted by Microsoft Excel, the cumulative abnormal returns and combined cumulative abnormal returns are computed from the collected data about the sampled 38 transactions. The results of the cumulative abnormal returns of the bidders (CAR(b)) and targets (CAR(t)) and their CCARs are summarized and illustrated by the following Figure 1.

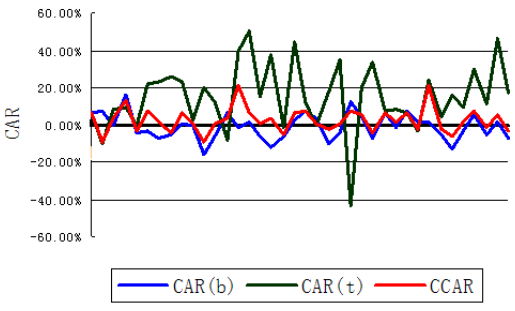


FIGURE 1 CAR Distributions

Figure 1 shows that most of the sampled mergers provide positive cumulative abnormal returns to the targets. There are two outstanding exceptions that characterised by negative target returns but positive bidder returns. One is No. 13---NationsBank (CAR: 6.34%) acquires Boatmen's Bancshares (CAR: -8.44%); the other is No. 24---Wells Fargo & Company (CAR: 12.04%) acquires First Security Corporation (CAR:-43.62%). However, their combined cumulative returns maintain positive.

The average cumulative abnormal returns are listed in the following Table 2.

TABLE 2 Means of CAR and CCAR

	CAR(b)	CAR(t)	CCAR
Mean	-0.99%	15.07%	2.57%

Therefore, within the chosen 38-transaction sample, the observed value effect of the US bank mergers on the bidders is negative (-0.99%) but positive on targets (15.07%) and combined banks (2.57%). Using Stata-8, the statistical significance of the above results can be tested by t-test, sign test and Wilcoxon signed rank test respectively.

3.2 RESULTS OF SIGNIFICANCE TESTING

The process of Stata-8 based significance testing is recorded and analyzed in this part. The following analysis interprets the result of the significance testing.

3.2.1 t-test

As explained in Methodology, the null hypothesis of t-test is that the population mean of the cumulative abnormal returns is zero ($H_{0:\,u}$=0). The result of t-test is demonstrated by the Table 3 below.

TABLE 3 t-test

	t	p	Accept H_0?
CARb	-0.863	0.394	Yes
CARt	5.181	0.000	No
CCAR	2.334	0.025	No

On the bidder side, its t-vale (-0.863) corresponds to a p-value of 39.4%, which is much bigger than the selected degree of significance, 5%. Thereby, this t-vale is located in the acceptance region that means the null hypothesis cannot be rejected. Under this circumstance, the negative sample mean is insignificant for the population. Based on the result of t-test, it is concluded that the value effect of the US bank mergers on bidders is negative, but insignificantly.

In contrast to the situation of the bidder, both the t-statistics of target (5.181) and combined bank (2.334) fall in the critical region, as their p-values, 0.0% and 2.5% respectively, are smaller than 5%. The H0 for the target and combined firm are therefore rejected. As the conclusion, the value effects of the US bank mergers on target and combined firm are significantly positive.

3.2.2 Sign test

There are 17 bidders with observed positive cumulative abnormal returns whilst 21 targets with negative ones. The null hypothesis here is p=q=0.5, i.e. the probabilities of observing a positive and a negative cumulative abnormal return are equivalent. According to the results of sign test shown in Table 4, $p\{x\geq 21\}$ is 62.7% which is much larger than 5%, which means that the null hypothesis should not be rejected. Consists with that of t-test, the corresponding conclusion is that the value effect on bidder is insignificant negative.

TABLE 4 sign-test

	P	Accept Ho?
CARb	0.627	Yes
CARt	0.000	No
CCAR	0.034	No

As for targets, the number of observed + signs is 31 while that of - signs is 7. $p\{x\geq 31\}$=0.0% which is smaller than 5% and H0 therefore can be reject even at 1% level. Similarly, there are 26 pluses and 12 minuses observed within combined banks. In this case, $p\{x\geq 26\}$=3.4% that is also smaller than 5%. The H0 is therefore can be rejected at 5% significance level. Based on the couple of rejections of H0, the sign test obtains the same results as t-test: the positive value effects on targets and combine firms are both significant.

3.2.3 Wilcoxon signed rank test

The result of Wilcoxon test is demonstrated by the Table 5 below.

TABLE 5 Wilcoxon signed rank testt

	z	p	Accept H_0?
CARb	-0.812	0.417	Yes
CARt	4.343	0.000	No
CCAR	2.110	0.035	No

According to the result attained by Stata-8, the z-values of the CAR's rank sums of the bidder, target and combined firm are –0.812, 4.343 and 2.110 respectively. Their corresponding p-values are 41.7%, 0.0% and 3.5%. Therefore, the z-value of the bidder is located in acceptance region while that of the other two fall in critical region. It indicates that the null hypothesis, H0: $V = W = \frac{1}{2}\sum_{n=1}^{N} n$, should be accepted for the bidder while rejected for both the target and the combined firms. As the conclusion, the same as that of t-test and sign test, the negative CAR of bidders is insignificant while the positive CAR of targets and combined firms are both significant.

4 Research finding

The significance testing through t-test, sign test and Wilcoxon signed rank test thus draw the same conclusion, that is, over a 36-day event window the US bank mergers occurring between 1994 and 2003 create insignificantly negative gains for bidders, and significant positive value for target banks and their combinations.

This conclusion firstly supports the previous studies that find positive gains for targets. Such studies include

those conducted by Becher (2000), Houston and Ryngaert (1994), Zhang (1995) and other six studies quoted by Becher (2000) [14], which calculate 20.48% of target bank gains on average. However, although a negative bidder gain is observed (-0.99%) in the 38-transaction sample, it is insignificantly different from zero. This is different from most of the previous studies. Two-third of studies conducted before 1994, found significant negative gains of acquiring banks. Through analysing the bank mergers in the 1990s, Becher (2000) calculate significantly positive bidder returns over a 36-day event window that is similar as what has been used by this study, while significant negative bidder returns are found over an 11-day event window. As for the combined firms, this study documents a significant positive return which supports the conclusion of Becher (2000). Nevertheless, more previous studies suggest the insignificant effect of bank mergers on combined firms due to the offset of negative bidder returns against the positive targets gains. Such studies include those conducted by Hannan and Wolken (1989) and Houston and Ryngaert (1994) [15].

According to Becher's (2000) criteria of clarifying merger motivations, the results of the event study consist with the characteristics of mixed hypotheses of hubris and synergy--- positive target returns, negative bidder returns and positive combined returns. However, the negative bidder return is insignificant, therefore, its attribute is much closer to synergy hypothesis, which is characterised by non-negative bidder returns compared to the mixed hypothesis which implies negative bidder returns. In this case, the US bank mergers occurring between 1994 and 2003 are incline to be motivated by synergy, rather than hubris.

5 Conclusions

As a popular business strategy, M&A has significant merits, including enabling companies to enter new markets more efficiently and smoothly compared to establishing new subsidiaries, as well as requiring only managerial skills in stead of entrepreneurial ones. However, in practice, merger activities are motivated by a variety of factors based on practical situations.

This Paper applies event study methodology with a 36-day event window (30 trading days before and 5 trading days after the announcement date). A 38-transaction sample is chosen from the top fifty US bank mergers (according to the assets of targets) occurred during the period from 1994 to 2003. Through the calculations based on the collected data, this study finds that the average cumulative abnormal return of the bidders in the chosen sample is negative (-0.99%), while that of the targets and combined firms are both positive (15.07% and 2.57% respectively). With the assistance of significance testing, including t-test, sign test and Wilcoxon signed rank test, the negative bidder return is confirmed to be insignificant, whilst the positive return of the target and combined firm are both significant. It indicates that the 3,517 US bank mergers occurred between 1994 and 2003 create insignificantly negative value for bidders, whilst benefit the target and the integrated banks with significant positive gains. Combining with the criteria of the couple of hypothesizes concerning bank merger motives, the conclusion is incline to support the synergy hypothesis, which claims that the US bank mergers occurring from 1994 to 2003 are more likely driven by synergy motive.

References

[1] Chronopoulos DK, Girardone C, Nankervis JC 2013 How do stock markets in the US and Europe price efficiency gains from bank M&As *Journal of Financial Services Research* **43**(3) 243-63

[2] Hagendorff J, Collins M, Keasey K 2008 Investor protection and the value effects of the bank merger announcements in Europe and the US *Journal of Banking and Finance* **32**(7) 1333-48

[3] Gerald P, Dwyer J 2001 The use of event study in finance and economics http://www.dwyerecon.com/pdf/lecteven.pdf-2001/ 10 Sep 2014

[4] Becher DA 2000 The value effects of bank mergers *Journal of Corporate Finance* **20**(6) 189-214

[5] Frame WS, Lastrapes WD 1998 Abnormal returns in the acquisition market: the case of Bank Holding Companies 1990–1993 *Journal of Financial Services Research* **14**(2) 145-63

[6] Dasgupta S, Laplante B, Mamingi N 1997 Capital Market Responses To Environmental Performance In Developing Countries http://www.worldbank.org/nipr/work_paper/market/index.htm-1997/ 10 Aug 2014

[7] Pilloff SJ, Santomero AM 1997 The value effectts of bank mergers and acquisitions *The Wharton Financial Institutions Centre* ***97(7)*** 101-7

[8] IFA 2005 Statistical tests http://www.fon.hum.uva.nl/Service/Statistics.html - 2005/ 26 Aug 2014

[9] Wheelock DC, Wilson PW 2004 Consolidation of US banking: Which banks engage in mergers *Review of financial economics* **13** 7-39

[10] Hoel PG, Jessen RJ 1982 Basic statistics for business and economics (3rd edition) *Canada: John Wiley & Sons, Inc*

[11] Gujarati DN 1999 Essentials of econometrics (2nd edition) *Columbus: McGraw-Hill Companies, Inc*

[12] Watsham TJ, Parramore K 1997 Quantitative methods in finance *London: International Thomson Business Press*

[13] Neter J, Kutner M, Nachtsheim C, Wasserman W 1996 Applied linear statistical models (4th edition) *McGraw-Hill, Irwin*

[14] Beccalli E, Frantz P 2009 M&A operations and performance in banking *Journal of Financial Services Research* **36**(2) 203–26

[15] Berger AN, Bonaccorsi di Patti E 2006 Capital structure and firm performance: a new approach to testing agency theory and an application to the banking industry *Journal of Banking and Finance* **30**(4) 1065–102

Author

Guo-yi Chen, born on Jauary 20, 1982, Wanzhou district, Chongqing Province, P.R. China.

Current position, grades: the lecturer of business administration, Chongqing three gorges University, Wanzhou district, Chongqing city, China.
Scientific interest: business management and regional economy.
Publications: 8 papers.
Experience: teaching experience of 5 years, 6 scientific research projects.

Structural equation model of college foreign language writing and classroom teaching quality from perspective of teacher evaluation

Fumeng Gao*

School of Foreign Languages, Fuyang Teachers' College, Fuyang City, Anhui Province, China, 236037

Received 8 September2014, www.cmnt.lv

Abstract

Existing teaching quality evaluation systems have insufficient comprehensive indices reflecting teaching quality: index repetition, complication, quantification difficulty, and difficulty in judging the relationship. To make classroom teaching quality evaluation consistent with teaching effect, element weight is confirmed based on a structural equation model. The present paper takes a college foreign language writing class as the object of study and sets up a structural equation model for classroom teaching quality evaluation. Effect coefficients of each variable are calculated, and the weight of each element of classroom teaching quality evaluation is then obtained. Such method avoids artificially scoring elements during the confirmation of element weight, reduces evaluation subjectivity, and makes the evaluation results more accurate, rational, and credible.

Keywords: structural equation model, teacher evaluation, classroom teaching, teaching quality

1 Introduction

The structural equation model originated from the path analysis concept introduced by Swell Wright in the 1920s. Structural equation model is also called simultaneous equation model and causal model etc. A great breakthrough in the development process of the structural equation model is the development of the concept of latent variable. It is the result of the joint development of latent variable and economics, etc. [4]. The structural equation model in this paper uses the latent variable. Traditional multivariate analysis method can only test the relationship between single independent variable and independent variable at the same time. Factor analysis can reflect the relationship among variables, but cannot further analyze causal relationships among variables. Although path analysis can analyze causal relationships among variables, complying with the following fundamental assumptions is difficult: zero fundamental assumption, unrelated residual error, and single-way causal relationship. A composite model which can synthesize these models is urgently needed. Structural equation model integrates path analysis, confirmatory factor analysis, and general statistical test method, and can analyze mutual causal relationship among variables. It contains advantages of factor analysis and path analysis. Meanwhile, it makes up for disadvantages of factor analysis, considers the error factor, and is not limited to assumed conditions of path analysis.

At present, the enrollment scale of Chinese higher education expands continuously. Problems caused by sharp expansion of enrollment scale include scarcity of teaching resources and reduction of teaching quality. College classroom teaching quality thus becomes a social problem. No stable student pool under the condition where talent cultivation quality receives increasing attention without efficient classroom teaching quality. Thus, colleges will lose core competitiveness. Teaching quality is a multi-layer concept and a vital problem in the development of colleges. Since teaching quality involves the input, process, and results of teaching activities, as well as diversified public, teaching quality is a multi-dimensional evaluation index. Teaching quality refers to students' growth in knowledge, ability, and values, and is the result of comprehensive function of the whole teaching system. Scientific and effective classroom teaching quality evaluation is significant in achieving total teaching quality management and quality monitoring, and plays an irreplaceable role in improving teaching quality. On the other hand, in recent years, English writing ability of college students have received great attention. Unsatisfied employers noted that only 5% of college students have good writing and expression ability, whereas 37% of college students fail to reach the standards and are even poor in terms of writing ability. Thus, many colleges pay more and more attention to foreign language study and with a focus on writing training and cultivation. Many schools now strive to explore other methods and improve the evaluation of foreign language writing teaching results. During course teaching quality evaluation, students are the evaluation subjects, as they know whether the teaching method is effective. However, college course teaching is academically dependent, and final evaluation of teaching quality by students cannot really reflect actual situations.

* ***Corresponding author's*** e-mail: gaofumeng@yeah.net

On the contrary, if students grasp the evaluation of course quality based on interaction between students and teachers this will have direct negative impacts on constructing harmonious student-teacher relationships. Thus, other subjects must be chosen. Teachers participate in evaluation index design, which reflects colleges respect for teachers' autonomous right. In this way, teachers can be aware of teaching evaluation significance and evaluation scheme bases, mentally accept evaluation indices, and provide professional evaluation from an objective perspective. Therefore, teacher evaluation is significant in classroom quality evaluation.

The features of structural equation model are as follows: many research variables cannot be directly be observed and measured like teaching quality. The concept of latent variable is applied when we need to study such variables, and corresponding dominant variables for a research are selected. For example, when we study classroom teaching quality, teaching quality is, to a great extent, reflected in final scores, which can ce measured. Thus, structural equation model with latent variables is applied to classroom teaching quality. Structural equation model with latent variables have two basic modes: structure pattern and measurement pattern. Computational formula system approach is used to verify college foreign language writing teaching. Latent variable factors and their mutual causal relations can be explored through measuring and forming the system. The present paper sets up the structural equation model for college classroom teaching quality evaluation, expands research perspectives and research methods for college teaching quality on the basis of teacher evaluation, and establishes proper teaching quality evaluation indices to estimate teaching quality of foreign language writing class in a college.

2 Theory summary

2.1 STRUCTURAL PATTERN

Structural pattern states causal relationship between latent exogenous variables and latent endogenous variables. The structural equation is expressed as:

$$\eta=B\eta+\Gamma\xi+\zeta, \quad (1)$$

where ξ is the latent exogenous variable (latent independent variable) matrix, η is the latent endogenous variable (latent dependent variable) matrix. Γ is the structure coefficient matrix that represents the effects of ξ on η in the structural model. B is structure coefficient matrix that represents the mutual effects of component factors of η in the structural model, and ζ is the residual error matrix of the structural equation.

2.2 MEASUREMENT PATTERN

Measurement pattern explains the relationship between latent variables η and ξ, and measured variables y and x.

The measurement equation in terms of X is $X=\Lambda X\xi+\delta$, where X is the measured variable matrix of ξ, ΛX is the measurement coefficient matrix that represents the relationship between η and measured variable X and δ is residual error matrix of the measurement equation.

The measurement equation in terms of Y is $Y=\Lambda Y\eta+\varepsilon$, where Y is the measured variable matrix of η; ΛY is measurement coefficient matrix that represents the relationship between η and measured variable Y, and ε is the residual error matrix of the measurement equation.

3 Construction of structural equation model of college foreign language writing classroom teaching quality from perspective of teacher evaluation

3.1 THEORETICAL MODEL FOR TEACHING QUALITY

The structural equation model does not offer the general pattern to set up the theoretical model. Theoretical model construction is based on the research each relevant field and the study of index system establishment. The present paper starts from the perspective of teacher evaluation. Given that teachers tend to be more processional and objective when evaluating course quality, 4 latent variables and 16 observed variables are set (Table 1) The structural model is composed of 3 exogenous latent variables (teaching attitude, teaching content and teaching method) and 1 endogenous latent variable (teaching effect).

Teaching attitude. Teaching attitude involves strictly following teaching discipline, putting sufficient time and energy to prepare lessons and improving teaching methods after class, actively shortening distance with students, communicating with students freely and creating harmonious student-teacher relationship. Teaching attitude reflects the teachers' degree of valuing classroom teaching. Teachers' teaching attitude can directly impact teaching quality as the first key index. Teaching attitude is reflected through 4 indices, i.e. I11, U12, U13 and U14 (Table 1).

Teaching content. Teaching content is the core of classroom teaching, the basic link of knowledge hierarchy construction by students, and the core of classroom teaching quality evaluation. Only when teaching content is correctly planned can the accuracy of classroom teaching quality evaluation be guaranteed. Teaching content refers to the knowledge imparted to students by teachers in the knowledge hierarchy. For most colleges, course teaching content has no absolute uniform requirement. Each college can order teaching materials as needed, and teachers can expand the scope of knowledge on the basis of the teaching material. Evaluation should include indices U21, U22, U23, and U24 (Table 1).

Teaching method. Teaching method refers to the method and process used by teachers to impart knowledge to students. Teaching methods mainly include tools and methods, which directly affect the improvement of teaching quality. Teaching method mainly consists of indices U31, U32, U33, and U34 (Table 1).

TABLE 1 Latent variables and observed variables in the structure

Type of latent variable	Latent variable	Type of observed variable	Observed variable
	Teaching attitude		U11: carefully and responsibly prepare lessons; teaching documents are complete U12: serve as a model for students; impart knowledge and educate students U13: carefully teach; strong enthusiasm for teaching U14: pursue studies rigorously; give strict requirements
Exogenous latent variable (ζ)	Teaching content	Exogenous observed variable	U21: familiar with teaching materials and teaching plans; proficient in teaching U22: concepts are correct; highlight key points and difficult points; demonstration is precise and expression is correct U23: link theory and practice; reflect cutting-edge knowledge of the subject U24: teaching progress is moderate; class period is allocated rationally
	Teaching method		U31: vivid and concise language; logical; accurate examples U32: flexible and diversified teaching methods; pay attention to ability and quality cultivation U33: properly apply various modern teaching supplementary means U34: rational blackboard-writing layout; clear and tidy writing
Endogenous latent variable (η)	Teaching effect	Endogenous observed variable	U41: teaching is attractive; lively classroom atmosphere U42: point out reference materials, arrange after-class review; good classroom discipline U43: students have strong learning interest U44: students' mastery of basic knowledge, theory and skills

Teaching effect. Teaching effect directly reflects classroom teaching quality and classroom teaching quality evaluation indices. It means the approval of course necessity. Teaching effect is mainly reflected by indices U41, U42, U43, and U44 (Table 1).

3.2 DETAILED MODEL OF FOREIGN LANGUAGE WRITING CLASS

3.2.1 Identification of structural equation

For foreign language writing class in particular, exogenous latent variables in Table 1 also contain "teacher quality." Teacher quality is mainly reflected in two aspects: "teachers' pronunciation is standard" and "teachers have strong listening, speaking, and comprehension skills." These are two corresponding exogenous observed variables. Aside from this, endogenous s latent variable (i.e. teaching effect) can be divided into "students' ability to listen to and comprehend a foreign language" and "students' ability to speak and express a foreign language." Thus, the parameters (t) that need to be estimated include 22 path coefficients $\lambda(x)$ and $\lambda(y)$ from latent variables to dominant variables. A total of 18 residual errors δ and ε of dominant variables, 8 path coefficients $\lambda(\zeta)$ and $\lambda(\eta)$ among latent variables, 14 exogenous dominant variables (p), and 4 endogenous dominant variables (q) exist. The freedom degree of the path diagram of the model can be expressed as:

$$df = [(p+q)(p+q+1)/2] - t = 149 > 0.$$

Thus, the model can be identified.

3.2.2 Fitting of structural equation model

The present paper utilizes AMOS software, selects CMIN/DF, RMSEA, GFI, AGFI and NFI) as evaluation indices, and calculates the fitting degree [8]. Table 2 shows the calculation results.

TABLE 2 List of fitting coefficients

Index	CMIN/DF	RMSEA	GFI	AGFI	NFI
Data	2.702	0.077	0.836	0.891	0.854

Bain et al., considering CMIN/DF < 3, show that the overall effect of the model is good. Lyle et al. note that if RMSEA < 0.08 and the value of GFI, AGFI and NFI is between 0.8 and 0.9, then the mode exhibits a good fit. According to Table 2, the explanatory ability of structural equation model for foreign language writing teaching quality evaluation is strong.

3.2.3 Parameter estimation and test

This paper adopts maximum likelihood method and applies AMOS software to analyze factor loading and path coefficient of each dominant variable, then normalization processing is carried out. Finally, normalized factor loading and normalized path coefficient are obtained.

In accordance with the principle of maximum likelihood method, when factor loading exceeds 0.4, the mode has a good explanatory ability. The normalized factor loading of the above dominant variables exceeds 0.4, which indicates that each factor has strong explanatory ability for measurement model. For the first exogenous latent variable teaching attitude $\xi 1$, the loading of the observed variable X_1 "U11" is the largest. This shows the first exogenous latent variable has the greatest contribution, followed by X_2, X_3 and X_4. Similarly, dominant variable X_6 "U22", X_{10} "U32", X_{14} "teachers have strong ability in listening, speaking and comprehension", Y_1 "U41", and Y_3 "U43" have the greatest contribution to the exogenous latent variables "teaching content", "teaching method", "teaching quality", "students' ability to listen to and comprehend a foreign language", and "students' ability to speak and express a foreign language", respectively.

TABLE 3 Normalized factor loading

Dominant variable	Normalized factor loading	Dominant variable	Normalized factor loading
$X_1(\xi_1)$	0.6023	$X_{12}(\xi_3)$	0.4936
$X_2(\xi_1)$	0.4877	$X_{13}(\xi_4)$	0.5824
$X_3(\xi_1)$	0.4721	$X_{14}(\xi_4)$	0.6305
$X_4(\xi_1)$	0.5862	$Y_1(\eta_1)$	0.7082
$X_5(\xi_2)$	0.6417	$Y_2(\eta_1)$	0.5399
$X_6(\xi_2)$	0.7135	$Y_3(\eta_1)$	0.4768
$X_7(\xi_2)$	0.6628	$Y_4(\eta_1)$	0.4693
$X_8(\xi_2)$	0.4176	$Y_1(\eta_2)$	0.6270
$X_9(\xi_3)$	0.5034	$Y_2(\eta_2)$	0.5345
$X_{10}(\xi_3)$	0.6209	$Y_3(\eta_2)$	0.6547
$X_{11}(\xi_3)$	0.5470	$Y_4(\eta_2)$	0.5384

TABLE 4 Normalized path coefficients

Latent variable path	Normalized path coefficient	Significance at 0.0001 level	Verification result of Hypothesis
$\xi_1 \cdots \rightarrow \eta_1$	0.6288	Significant	Support H_1
$\xi_2 \cdots \rightarrow \eta_1$	0.8406	Significant	Support H_2
$\xi_3 \cdots \rightarrow \eta_1$	0.8073	Significant	Support H_3
$\xi_4 \cdots \rightarrow \eta_1$	0.7592	Significant	Support H_4
$\xi_1 \cdots \rightarrow \eta_2$	0.6317	Significant	Support H_5
$\xi_2 \cdots \rightarrow \eta_2$	0.8129	Significant	Support H_6
$\xi_3 \cdots \rightarrow \eta_2$	0.7965	Significant	Support H_7
$\xi_4 \cdots \rightarrow \eta_2$	0.8248	Significant	Support H_8

According to path coefficients of latent variables in the table, overall effects of exogenous latent variables ξ_1, ξ_2, ξ_3 and ξ_4 on endogenous latent variable η_1 are 0.688, 0.8406, 0.8073 and 0.7592, respectively, and their effects on endogenous latent variable η_2 are 0.6317, 0.8129, 0.7965 and 0.8248, respectively. This implies that if teaching attitude, teaching content, teaching method and teacher quality improve one standard unit, students' ability to write and comprehend a foreign language will increase by 0.688, 0.8406, 0.8073, and 0.7592 standard units, respectively, and students' ability to speak sand express a foreign language will increase by 0.6317, 0.8129, 0.7965, and 0.8248 standard units, respectively.

Table 3 and Table 4 show that teaching attitude, teaching content, teaching method, and teacher quality have significant effects on two indices of foreign language writing teaching quality: "students' ability to write and comprehend a foreign language" and "students' ability to speak and express a foreign language". Thus, they can estimate teaching quality. Observed variables X_1 "U11", X_6 "U22", X_{10} "U32", X_{14} "teachers have strong ability in listening, speaking and comprehension", Y_1 "U41", and Y_3 "U43" have the greatest contributions to each latent variable, with each contribution signify-cantly different from the others.

Therefore, the structural equation model forms a reliable quantification measurement scale with certain reliability and validity, and provides a weight coefficient and good theoretical support for displaying foreign language writing teaching quality.

4 Conclusions

Based on the perspective of teacher evaluation, the present paper investigates foreign language writing teaching quality of a college and presents a structural equation model for classroom teaching quality through analyzing and testing sample data. The application of structural equation model strengthens the objectivity of the weight of each element of teaching quality evaluation, and overcomes the subjectivity of scoring by experts using existing point-factor method. Its application on classroom teaching enhances the comparability of teaching quality and keeps evaluation results consistent. The model addresses the difficulty in directly measuring teaching attitude, teaching content, teaching method, and teacher quality, which have direct positive effects on teaching, and promotes studies on the cause and effect influencing foreign language writing teaching. Nowadays, colleges should improve teaching content and approach to enhance education and encourage the continuous improvement of colleges.

References

[1] Aspy DN, Roebuck FN, B Black. 1972 The relationship of teacher-offered conditions of respect to behaviors described by Flanders' interaction analysis *The Journal of Negro Education* **41**(4) 370-8
[2] Zhang L, Wang Y, Pan Y 2011 Case Study on Classroom Teaching Features of IT Expert Teachers On Basis of FIAS *E-education Research* **19**(7) 83-8
[3] Chen T, Zhou J, Pu J, et al. 2007 Application of FIAS in Classroom Teaching *Vocational Education Research* **14**(4) 111-2
[4] Ulker Vanci-Osam, Tijen Aksit 2000 Do intentions and perceptions always meet? A case study regarding the use of a teacher appraisal scheme in an English language teaching environment. *Teaching and Teacher Education* **16**(2) 255-67
[5] Zimmerman BJ. 1990 self-regulated learning and academic achievement: an overview *Educational Psychologist* **25**(1) 3-17
[6] Hayes D 2009 Non-native English-speaking teachers, context and English language teaching *System* **37**(2) 1-11
[7] Bao X 2004 Research on Quick Sort Methods Based on ID3 algorithm *Modern Electronic Technique* **27**(4) 84-5
[8] Li X2003 A New Clustering Segmentation Algorithm of 3D Medical Data Field Based on Data Mining. *Journal of Digital Content Technology and its Applications* **4**(4) 174-81
[9] Ratchagit K, Phat VU N, Niamsup P 2011 The Novel Sufficient Condition for Stability of Discrete-Time Control System of Neural Networks. *International Journal of Applied Mathematics and Statistics* **21**(J11) 25-32
[10] Karimov A, Moharrami S. 2010 Automatic Classification with Neural Networks Using New Decision Rule. *International Journal of Applied Mathematics and Statistics* **19**(D10) 90-6
[11] Hu M 2013 Existence and stability of anti-periodic solutions for an impulsive neural networks on time scales *International Journal of Applied Mathematics and Statistics* **47**(17) 61-9

Author

Fumeng Gao, born in December 1979, Huoqiu County, Anhui Province, P.R. China.

Current position, grades: the lecturer of School of Foreign languages, Fuyang Teachers' College, China.
University studies: BSc in English Education from Fuyang Teachers' College, MSc from Shanghai University in China.
Scientific interest: busniess english and english teaching methodology.
Publications: 6 papers.
Experience: teaching experience of 13 years, 3 scientific research projects.

The empirical analysis on mode of developing the rural areas with the aid of the urban areas

Chen Yan[1], Lan Nan[2,3*], Liu Yunlang[4], Huang Rong[5]

[1]*School of Distance Learning and Continuing Education, China University of Geosciences, Wuhan, Hubei, China*

[2]*Key Laboratory of Legal Evaluation Project, Ministry of Land and Resources, Wuhan, Hubei, China*

[3]*School of Public Administration, China University of Geosciences, Wuhan, Hubei, China*

[4]*School of Environment, China University of Geosciences, Wuhan, Hubei, China*

[5]*School of Economics, Huazhong University of Science and Technology, Wuhan, Hubei, China*

Received 3 September 2014, www.cmnt.lv

Abstract

Developing the rural areas with the aid of the urban areas, which was summarized through the long practice in China, is a new mode of promoting the coordinated development of urban-rural areas. The evaluation of the impact of this mode on promoting the coordinated development of urban-rural areas has great significance for reference for the urbanization development in other countries around the world. This paper takes Chenggong New District in Yunnan, China as an example, calculates the coordination degree of its urban and rural development in 1999, 2004 and 2009 with the use of urban and rural development coordination degree model, then makes clustering analysis of the urban and rural development coordination degree in various districts in Chenggong New District in 2009 with the use of Ward's method. Results indicate that a) the urban and rural development coordination degree of Chenggong New District is 0.359 in 1999, 0.545 in 2004, and 0.504 in 2009. The increase in the coordination degree year by year shows the mode of developing the rural areas with the aid of the urban areas has obvious effects on the coordinated development of urban-rural areas; b)the mode of developing the rural areas with the aid of the urban areas has obvious coordination impact on the aspects of infrastructure construction and basic livelihood guarantee, but has not enough coordination impact on the aspects of environmental governance, medical treatment and education; c) when the mode of developing the rural areas with the aid of the urban areas is applied, the coordinated development of industries and especially the agricultural industrialization development should be paid attention to.

Keywords: the mode of developing the rural areas with the aid of the urban areas, the coordinated development of urban-rural areas, Chenggong New District, China

1 Introduction

With the impetus of economic globalization and the rapid economic development, China has summarized the urban-rural integration mode of developing the rural areas with the aid of the urban areas under the guidance of scientific development concept [1]. The mode of developing the rural areas with the aid of the urban areas, which means the integration of urban and rural areas and aims at developing the rural economy with the help of the urban economic development and promoting the urban development by the potential development of rural areas at the same time, is a virtuous cycle development mode that urban and rural areas, as important resources and markets to each other, serve each other and commonly share modern civilization. China has made developing the rural areas with the aid of the urban areas as a national strategy for the integrated development of urban and rural areas since 2001, but there are now still huge differences in inf*rast*ructures and incomes between rural and urban areas [2]. Therefore, the mode of developing the rural areas with the aid of the urban areas requires scientific understanding and evaluation.

The researches on urban-rural integration have a long history, the much more famous of which are the Urban-rural Fusion Theory from Marx and Engels [3], Urban and Rural Development Theory from Theodore W. Schultz [4] and Central Place Theory proposed by W. Christaller and A. Lsch [5]. The 1957 Nobel Laureate Karl Gunnar Myrdal proposed Geographical Dual Economy Theory, making up for the flaw of the Growth Pole Theory [6]. Research on the coordination of urban and rural development evaluation focused on two aspects [7,8]. The first is the studies on the construction of the index evaluation of the urban and rural coordinated development. (Yang Zhenning (2008), Li Qin, Zhang Yuanhong, Zhang Jun et al (2009)) [9,10], the second is the empirical coordination evaluation research (Deng Ling, Wang Binbin (2008), Xue Hongxia, Liu Juxian, Luo Weiling (2010), Xiao Shien, Li Xianshi (2009), Zhang Deliang, Jiang Yue (2008)) [11-14]. It can be seen that scholars have conducted quite a few researches on the urban-rural integration, which provide related

[*] ***Corresponding author's*** e-mail: lannan7534@163.com

theoretic basis for this paper. In China, there are mature index system and corresponding model for the evaluation of the coordinated development of urban and rural areas, but quantitative evaluation of the mode of developing the rural areas with the aid of the urban areas is rarely involved in. Based on this situation, this paper takes Chenggong New District in Yunnan, China as an example, calculates the coordination degree of its urban and rural development in 1999, 2004 and 2009 with the use of urban and rural development coordination degree model, then makes clustering analysis with the use of Ward's method. The aim is to show the effectiveness and limited factors of promoting the urban-rural integration by the mode of developing the rural areas with the aid of the urban areas, and then come up with some policy suggestions when the mode is undertaken.

TABLE 1 The process of the urban-rural integration in China

Period	Stage	Characteristic
1949-1957	Infancy stage of urbanization	Urban-rural morphological characteristics: Unequal agricultural trade, the restrictive flow of urban-rural labor force, unequal rights between urban residents and farmers, enforceability, promoting urban areas with rural areas
1958-1965	Ups and downs stage of urbanization	
1966-1978	Stagnation stage of urbanization	
1979-1984	Recovery of urbanization	Weakened enforceability, the relatively flow of production factors, capital and interest groups oriented marketization
1985-1992	Stable development of urbanization	
1993-2000	Rapid development of urbanization	
2001 to ---	Scientific Development	Developing the rural areas with the aid of the urban areas under the guidance of the scientific development concept, the strategy understanding of urban and rural areas has been changed from "divide and conquer", "vigorously promote urbanization" to " developing the rural areas with the aid of the urban areas", the urban-rural policy has extended from the economic aspect to political, economic, social, cultural and other aspects

Source:
Sheng Guangyao, Research on urbanization patterns and its changes, China Social Sciences Press, 2008 (6), Beijing, 37-39
Zhao Qunyi, Exploration of the strategic transformation of the relationship between urban and rural areas and the urban-rural integration planning in the new era, Urban Planning Journal, 2009 (6), 47-49

2 The study area

Chenggong New District locates at the southeast the main city of Kunming in Yunnan province. Formerly it was named Chenggong Town, which included Quan County, Longcheng Town, Luoyang Town, Dounan Town, Dayu Village, Majinpu Village, Wujiaying Village and Qidian Village and was famous for its location near Tian Lake. On May 30, 2003, Yunnan government made the strategy of constructing "modern new Kunming". The construction of "Modern New Kunming" firstly started in Chenggong New District. In 2007-2008, Chenggong New District was under administrative redivision, and it has been divided into seven districts (see Figure 1). From the perspective of its development process, Chenggong New District has experienced the development from township to the new district. Before 2003, it was a town; after 2003, it was planned to be a new district, forming the mode of developing the rural areas with the aid of the urban areas. In addition, Chenggong New District borders Tian Lake which coupled with the increasingly serious pollution problems, and its main characteristic industry is the flower industry. Combining these three aspects, it is obvious that Chenggong New District is a typical area to evaluate the effectiveness and the limiting factors for the mode of developing the rural areas with the aid of the urban areas.

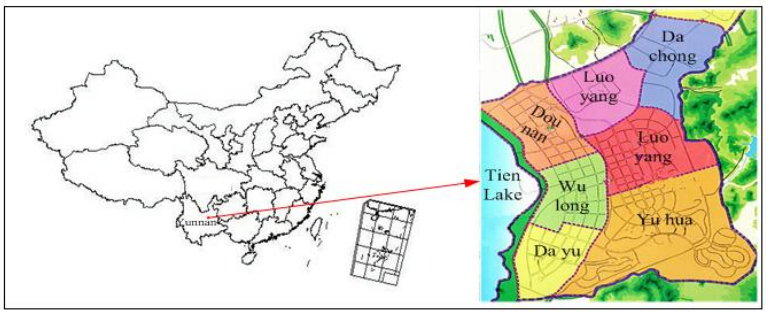


FIGURE 1 Chenggong New District location and zoning

3 Method and data

3.1 THE EVALUATION INDEX

Combining the indexes designed by previous scholars [15] and following the index selecting principles of comprehensiveness, representativeness, operability and comparability, this paper divides the coordinate development index system of urban-rural areas into four layers: A is the target layer, mainly reflecting the comprehensive coordinated development of urban and rural areas; B is the systematic layer, mainly reflecting the coordinated development of the three subsystems of economy, society, environment and space, which refines the target layer; C is the control layer, which further refines the system level indicators; D is the operating index layer, which can be used to measure the specific data to reflect the indicating level of the control layer. Those four index layers have formed a complete hierarchical structure of urban and rural coordinated development evaluation system (as shown in Table 2).

TABLE 2 The evaluation index system of the urban and rural coordinated development

Layer A	Layer B	Layer C	Layer D	Indicator variables
Integrated and Coordinated development of urban and rural areas	Urban and rural economy	Living standards	Per capita GDP (yuan)	X1
		Economic structure	The proportion of agricultural output accounted for the three industries output (%)	X2
		Revenue	Revenue (million)	X3
		Balanced economic benefits	Engel coefficient	X4
	Urban and rural society	Health	Number of beds (sheets) per thousand people	X5
			Number of doctors per thousand (people)	X6
			Financial health care expenditures (yuan)	X7
		Social Security	Minimum living security coverage (%)	X8
		Culture and Education	Educational services expense ratio (%)	X9
		Changes in population	Natural growth rate (‰)	X10
	Urban and rural environment and space	Ecological Construction and Environmental Governance	Urban forest coverage rate (%)	X11
			Sewage treatment rate (%)	X12
		Infrastructure	Car increasing ratio (%)	X13
			Villages traffic rate (%)	X14
			Fixed telephone penetration rate (%)	X15

3.2 MODEL

3.2.1 The evaluation model of the coordination degree of the urban and rural development

Existing studies have shown that the coordination degree models all make the coordinated evaluation on the basis of the comprehensive evaluation of the urban and rural areas. So the evaluation model of the integrated development of urban and rural areas should be determined firstly. To simplify the analysis of the problem and prevent the contained information overlap or cross, this paper uses the principal component analysis to reduce the dimension of the variables and classifies the highly linear original indicators related as a class, turning the original 15 indicators into a few mutual independent main ingredients, which reflect most of the information of the original variables. Principal component analysis takes advantage of the idea of using the least squares estimation to estimate the score of each principal component, which is to express the principal component as a linear combination of indicator variables, namely:

$$F_{ij} = b_{i1} \cdot X_1 + b_{i2} \cdot_{X2} + \cdots b_{in} \cdot X_n \quad (i = 1,2,3\cdots), \tag{1}$$

where F_{ij} is the principal component of the i-th, j-th region, b_{i1} is the sub-factor of the i-th principal component, X_n (n=1,2,...,15) is the nth indicator variables. And then summarize the weight αij of each principal component (the proportion of variance contribution rate of the principal component counting the total variance contribution rate) and come up with the comprehensive score of each urban and rural area in Chenggong New District. Normalize the score F_j and then we can get the score F of each urban and rural integrated development degree, which shows that the higher the score is, the higher the development level will be. The model of urban and rural comprehensive development degree is as follows:

$$F_j = \sum_{i=1}^{m} a_{ij} F_{ij}, \tag{2}$$

$$F = \frac{F_j - F_{j\,min}}{F_{j\,max} - F_{j\,min}} (i = 1,2,3\cdots). \tag{3}$$

After determining the integrated development model of urban and rural areas, we can come up with the coordination degree model of the urban and rural development. Through summarizing current research methods on coordination degree, we can classify them into two categories:

the distance coordination degree and the change coordination degree. Distance coordination degree aims at evaluating the degree primarily through measuring the static distance between each system and the concentration or dispersion degree of coordination between the systems; while change coordination degree aims at evaluating the degree primarily through measuring the consistency of relative changing degree between the systems [16]. According to the actual urban-rural integration development of Chenggong New District, the distance degree coordination is more suitable for its characteristics, so the paper selected this method to determine the urban-rural coordination development degree of Chenggong New District.

At present, the model of distance degree coordination contains two main types. One is to build the coordination degree model based on the dispersion degree between urban and rural development, namely:

$$B = 1 - \frac{S}{\overline{F}}, \tag{4}$$

where B ($0 \le B \le 1$) is the coordination degree of the urban and rural development; $\overline{F}$ is the average of the urban and rural integrated development degree; S is the standard deviation of development degree. The larger the value of B is, the better the degree the coordinated development between the various urban and rural areas will be, and vice versa.

The other is to build the model based on the overall and composite efficiency between urban and rural areas, the theoretical basis of which is the balance theory and utility theory. The balance theory means to maintain a balanced relationship between urban and rural areas, that is to increase A's efficiency without decreasing B's. The utility theory refers to the simultaneous development of urban and rural areas, making the maximum overall benefit [11]. The model is as follows:

$$B = \left\{ \frac{F_1 \cdot F_2}{\left(\frac{F_1 + F_2}{2} \right)^2} \right\}^k, \tag{5}$$

where B is the coordination degree, $0 \le B \le 1$; F_1 and F_2 represent the integrated development degree of the urban and rural areas; K is the adjustment coefficient, $K \ge 2$. Theoretically speaking, the smaller the deviation between F_1 and F_2 becomes, the higher the coordination degree between urban and rural areas will be, and vice versa.

Model (4) is applicable to evaluating multiple (two or more) urban and rural development coordination degrees. Based on this model, we can evaluate the coordination degree between the comprehensive development of each urban area and that of the rural area. Besides, we can also further determine the coordination degree of various subsystems including economic development level, the social status, the environment and space in urban and rural areas. In the end, we can find out the factors limiting the coordinated development of urban and rural areas and provide an important reference for the strategic decisions for the urban and rural coordinated development. Model (5) is applicable to the coordination evaluation between two subsystems, namely, the urban subsystem and the rural subsystem, which uses a whole to represent the development of different rural areas. This may lead to the question whether the development index of the rural area selected is typical and comprehensive. Therefore, the paper selects model (4) as the coordination degree model.

In summary, Equations (1), (2), (3), (4) constitute the evaluation model of the coordination degree of the urban and rural development.

According to previous studies on the coordination degree [12,13], and combining practice, the criteria of the coordination degree of the urban and rural development in this paper is shown in Table 3.

TABLE 3 The criteria of the coordination degree of the urban and rural development

Coordination degree	Very coordinated	Coordinated	Primary coordinated	Less coordinated	Not coordinated
B' value	B≥0.8	0.6≤B≤0.8	0.4≤B≤0.6	0.2≤B≤0.4	B≤0.2

3.2.2 Ward's method

This method was proposed by Ward. Divide n samples into k categories: $G_1, G_2 \ldots G_k$, $X_i^{(t)}$ represents the i-th sample in G_t, n_t represents the sample number of in G_t, $\overline{X}_i^{(t)}$ is the gravity of G_t, and then the sum of squared deviations of the samples in G_t is as follows:

$$S_t = \sum_{i=1}^{n_t} (X_i^{(t)} - \overline{X}_i^{(t)})'(X_i^{(t)} - \overline{X}_i^{(t)}), \tag{6}$$

The sum of squared deviations of the samples in k categories is as follows:

$$S = \sum_{t=1}^{k} S_t = \sum_{t=1}^{k} \sum_{i=1}^{n_t} (X_i^{(t)} - \overline{X}_i^{(t)})'(X_i^{(t)} - \overline{X}_i^{(t)}). \tag{7}$$

3.3 DATA SOURCES

The units of measurement are changeable and the levels of the variable values are very different from each other. In order to eliminate its effects on the analysis of development and coordination degrees of urban and rural areas, the paper uses the statistical analysis software SPSS16.0, and applies dimensionless method to the raw data in 1999, 2004, 2009 from Chenggong New Town Statistical Year-

book and Yunnan Provincial Statistical Yearbook with the use of *Z*-score method, namely

$$Z = \frac{x_i - \bar{x}}{S}. \quad (8)$$

4 Results

4.1 EXTRACTION OF PRINCIPAL COMPONENTS AND WEIGHTS

From the output results of the principal component analysis of the index data of Chenggong New District in 2009 by the software SPSS16.0, it is obvious that there is a high correlation between the 15 indicator variables, indicating that it is very necessary to reduce the dimension of the data and take the principal components out. (as shown in Table 4) in accordance with the principle that the eigenvalue is greater than 1. The paper extracted four principal components respectively represented by F_{1j}, F_{2j}, F_{3j} and F_{4j}. After calculating, the accumulated variance contribution rate of the four principal components reached 91.236%, which can reflect most of the information of the 15 indicators used in this paper. According to the above calculations, we can get the weights of the four principal components as follows: 28.5%, 26.47%, 25.58% and 19.45% respectively.

TABLE 4 Total variance explained

Serial number	The initial eigenvalue			Square load rotation sum		
	Total	Variance %	Cumulative%	Total	Variance %	Cumulative %
1	7.203	48.019	48.019	3.902	26.014	26.014
2	2.962	19.750	67.768	3.623	24.151	50.165
3	2.096	13.972	81.740	3.500	23.335	73.500
4	1.424	9.496	91.236	2.660	17.736	91.236
5	.931	6.207	97.443			
6	.308	2.052	99.495			
7	.076	.505	100.000			
8	5.287E-16	3.525E-15	100.000			
9	3.337E-16	2.225E-15	100.000			
10	2.291E-16	1.527E-15	100.000			
11	1.751E-16	1.167E-15	100.000			
12	8.875E-17	5.917E-16	100.000			
13	6.764E-18	4.509E-17	100.000			
14	-2.216E-16	-1.477E-15	100.000			
15	-3.985E-16	-2.657E-15	100.000			

4.2 DETERMINING PRINCIPAL COMPONENT FACTORS

From Table 5 it can be seen that the load value of the first principal component is relatively large in sewage treatment (X12), Engel's coefficient (X4), the number of one thousand people owning beds (X5) and the number of one thousand people having doctors (X6), as a result F1 mainly reflects the environmental and healthy level; The second principal component is relatively large in the forest coverage in urban and rural areas (X11), villages traffic rate (X14), car increasing ratio (X13) and per capita GDP (X1), so F2 mainly reflects the infrastructure and environmental indicators; The load value of the third principal component is larger in the natural growth rate (X10), fiscal health care expenditure (X7), fixed telephone penetration rate (X15), and minimum living security coverage (X8), and they are all positive correlations, so F3 mainly reflects the social development indicators of the urban and rural areas; The load value of the fourth principal component is larger in the educational services expense ratio (X9) and revenue (X3), so F4 mainly reflects the urban and rural economic and educational development.

TABLE 5 Rotated component matrix

	Factor			
	1	2·	3	4
X12	.957	.054	.247	.051
X4	-.860	-.419	.036	.119
X5	.769	.095	-.006	.628
X6	.743	.396	.117	.486
X11	-.230	-.875	.360	.152
X14	.257	.868	.333	.213
X13	-.009	.844	.295	.405
X1	-.565	-.700	-.294	-.019
X2	-.161	-.647	-.491	.530
X10	-.165	.218	.879	.241
X7	.310	.113	.867	.129
X15	.508	.053	.651	.046
X8	.420	.219	.599	.472
X9	.241	.147	.278	.833
X3	-.152	-.076	.619	.711

4.3 EVALUATION OF URBAN AND RURAL INTEGRATED DEVELOPMENT

According to Equations (1), (2), (3) and the data in Table 4 and 5, the integrated development degree can be calculated for each urban and rural area in Chenggong New District in 2009 (as shown in Table 6).

TABLE 6 Urban and rural development degree of Chenggong New District in 2009

	F1	F2	F3	F4	F
Quan District	0.165	0.624	1.000	1.000	1.000
Longcheng Street	1.000	0.791	0.495	0.224	1.000
DounanStreet	0.361	0.714	0.000	0.776	0.547
Wujiaying Street	0.660	0.476	0.905	0.074	0.823
Luolong Street	0.102	0.842	0.618	0.270	0.639
Wulong Street	0.000	1.000	0.612	0.023	0.589
Yuhua Street	0.119	0.955	0.520	0.000	0.570
QidianStreet	0.063	0.000	0.403	0.062	0.000

Based on the above methods, this paper makes the principal component analysis to the indicator data in 2004 and 1999 respectively, and ultimately gets the comprehensive development degree of Chenggong New District in 2004 and 1999 (see Table 7), which is used as the calculation basis of urban and rural coordination degree.

TABLE 7 Urban and rural development degree of Chenggong New District in 1999 and 2004

	The whole county	**Longcheng town**	**Luoyang Town**	**Dounan Town**	**Dayu Village**	**Majinpu Village**	**Wujiaying Village**	**Qidian Village**
F1999	1.000	0.878	0.550	0.608	0.228	0.503	0.339	0.000
F2004	1.000	0.892	0.838	0.832	0.581	0.587	0.777	0.000

4.4 ANALYSIS OF URBAN AND RURAL INTEGRATED DEVELOPMENT

According to Equation (5), a comprehensive calculation of the coordinated development of urban and rural areas in 1999, 2004 and 2009 is conducted, and the results are shown in Table 8.

TABLE 8 The coordination degree of Chenggong New District in 1999,2004,2009

Year	**1999**	**2004**	**2009**
Coordination degree	0.359	0.545	0.504
Coordination phase	Less coordinated	Primarily coordinated	Primarily coordinated

Meanwhile, the author calculated the coordination degrees of the four principal components of Chenggong New District based on the scores of each principal component of urban and rural areas in 2009. And then we got the coordination degree of the principal components in sequence (see Table 9). That the development of urban areas and the development of rural areas are uncoordinated in the first and the fourth principal components is the major factor for limiting the coordinated development of urban and rural areas in Chenggong New District.

TABLE 9 Coordination degree of four principal components

Principal component category	F1	F2	F3	F4
Coordination degree	-0.13	0.52	0.46	-0.24
Coordination phase	Uncoordination phase	Primary coordination phase	Primary coordination phase	Uncoordination phase

To analyze the integrated development of each urban and rural area in 2009 more directly, the paper chooses four principal components as the indicator variables of the cluster analysis and classifies various urban and rural levels. The cluster analysis tree is as follows.

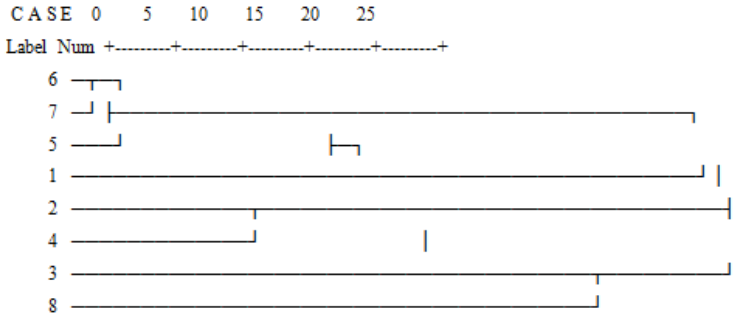


FIGURE 2 Cluster analysis tree

Note: 1:Quan District; 2:Longcheng Street; 3:Dounan Street; 4:Wujiaying Street; 5:Luolong Street; 6:Wulong Street; 7:Yuhua Sreet; 8 Qidian Street

5 Discussion and results

The mode of developing the rural areas with the aid of the urban areas is valid. The coordination degrees of urban and rural development of chenggong new district in 1999, 2004 and 2009 are respectively 0.359, 0.545, and 0.504. So it shows that the urban and rural areas developed rapidly and coordinately from 1999 to 2004, and the urban areas drove the development of rural areas in this short time. The result matches the year 2003 when the Provincial Party Committee and Yunnan government determined the overall strategy of "Modern New Kunming", with Chenggong as a key development area, and the development strategy of "developing the rural areas with the aid of the urban areas" through re-planning. The implementation time consistency of the mode of developing the rural areas with the aid of the urban areas indicates the development strategy mode is valid.

The mode of developing the rural areas with the aid of the urban areas has obvious coordination impact on the aspects of infrastructure construction and basic livelihood guarantee, but has not enough coordination impact on the aspects of environmental governance, medical treatment and education. According to Table 5, the first principal component reflects the level of environmental management and health care, the second represents the infrastructure, the third represents the basic livelihood guarantee, and the fourth reflects the development index of economy and education of urban and rural areas. According to Table 7, the second and the third principal components are coordinated while the first and fourth principal components are uncoordinated. The mode of developing the rural areas with the aid of the urban areas has obvious coordination impact on the aspects of infrastructure construction and basic livelihood guarantee, but has not enough coordination impact on the aspects of environmental governance, medical treatment and education.

Based on the results from Cluster analysis in Figure 3, it can be seen that Dounan Street and Qidian Street are in the last category. Dounan District locates in the northwest of the New District, which mainly relies on Dounan international flower auction trading center and is a flower industry dominated district. Qidian Street is dominated by agriculture and doesn't combine well with the industrialization and urbanization. Besides, each development index is still relatively backward. As a result, developing the rural areas with the aid of the urban areas means that urban areas should service the rural areas well. In addition, the coordinated development of industries and especially the agricultural industrialization development should be paid attention to. Only in this way can it play a better role.

Acknowledgements

Fund Project: Evaluation of Land and Natural Resources Engineering, Key Laboratory Legal Fund (CUGFP-1406); Central Universities Fundamental Research Funds for the project (CUGW120240)

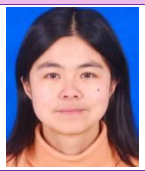
References

[1] Zhao Q 2009 Exploration on the strategic transformation of the relationship between urban and rural areas and urban and rural planning in the new era *Urban Planning Forum* (6) 47-9 (*in Chinese*)
[2] Liu Q 2009 Studies on the pathway of integration of urban and rural areas in the new rural construction process *Hebei Agricultural Sciences* (12) 86-7 (*in Chinese*)
[3] Fei L2010 Integration of urban and rural industries: Contemporary horizons of the urban-rural fusion of Marxist ideology *heory Research* (1) 62-65 (*in Chinese*)
[4] Nerlove M 2001Transforming Economics: Theodore W. Schultz, 1902–1998 in Memoriam *Economic journal* (12) 726-48
[5] Hottes R, Christaller W 1983 Annals of the Association of American *Geographers* (3) 51-4
[6] Yu H 2005 Innovative thinking about the system of overall planning urban and rural Development in Shanghai *Shanghai Academy of Social Sciences Press* 330-1 (*in Chinese*)
[7] Xia A, Xu X, Xue D 2003 Research review of Chinese Urban and Rural Coordinated Development Research *Geography* (5) 56-60 (*in Chinese*)
[8] Song H, Jiang C, Du Y 2004 Research review of overall planning urban and rural development *Research in the world* (10) 34-7
[9] Yang Z 2008 Research on index of urban and rural development - Based on the timing ordinal number *Economic Management* (11) 35-6
[10] Li Q,Zhang Y, Zhang J, etc 2009 Evaluation system of urban and rural development: Research overview and vision *China Rural Survey* (5) 2-10 (*in Chinese*)
[11] Deng L, Wang B 2008 Research on evaluation index System of Urban and Rural Development - Based on application of Wenjiang District in Chengdu *Southwest University of Nationalities (Humanities and Social Sciences Edition* (4) 80-4 (*in Chinese*)
[12] Xue H, Liu J, Luo W 2010 Research on coordinated development degree of urban and rural areas in Guangzhou *China Land Science* **24**(8) 39-45 (*in Chinese*)
[13] Xiao S, Li X 2009 Analysis of the coordinated development of urban and rural economy in Hebei *Technology & Management Research* (5) 113-5 (*in Chinese*)
[14] Zhang D, Jiang Y 2008 Empirical analysis of urban and rural overall development in Yunnan Province *East China Economic Management* (11) 42-6 (*in Chinese*)
[15] Jing PQ 2003 Progress and dynamic of urban-rural integration *Urban Planning* (6) 30-5 (*in Chinese*)
[16] Lai L W-C 2004 Spontaneous Catallaxis in Urban & Rural Development Under Planning by Contract in a Small Open Economy: The Ideas of Hayek and Mises at Work in Town & Country Planning in Hong Kong *The Review of Austrian Economics* **17**(2-3) 155-86

Authors

Yan Chen, born in October 1974, Jingmen County, Hubei Province, P.R. China.

Current position, grades: the assistant professor of School of China University of Geosciences, China.
University studies: MEc from China University of Geosciences in China.
Scientific interest: environmental system, economic policy.
Publications: 25 papers.
Experience: teaching experience of 10 years, 6 scientific research projects.



Nan Lan, born in May 1978, Jingshan County, Hubei Province, P.R. China.

Current position, grades: the associate professor of School of China University of Geosciences, China.
University studies: ML from Wuhan University in China, J.D from China University of Geosciences in China.
Scientific interest: environmental law.
Publications: 60 papers.
Experience: teaching experience of 11 years, 8 scientific research projects.

Yunlang Liu, born in October 1991, Chongqing City, P.R. China.

Current position, grades: the postgraduates of School of China University of Geosciences, China.
University studies: LLB in Law from China University of Geosciences in China.
Scientific interest: environmental planning, environmental assessment.
Publications: 10 papers.
Experience: completed 3 scientific research projects.

Rong Huang, born in February 1986, Xiantao County, Hubei Province, P.R. China.

Current position, grades: Ph.D. student in Huazhong University of Science and Technology, China.
University studies: MEc from Xiamen University in China.
Scientific interest: development economics, environmental economics.
Publications: 3 papers.
Experience: teaching experience of 4 years, 2 scientific research projects.

Two-sided matching considering the preferences of agents and intermediary

Qi Yue[1], Yuhua Li[1,2*]

[1]*School of Information Management, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China*

[2]*School of International Trade and Economics, Zhongshan Jiahede Investment Company, Zhongshan, Guangdong 528400, China*

Received 10 September 2014, www.cmnt.lv

Abstract

This paper proposes a novel method for solving the two-sided matching problem, in which the preferences provided by agents are ordinal numbers, and the preferences provided by intermediary is the expense standard on ordinal numbers. In this paper, the description of the considered two-sided matching problem is given. Then the concepts of satisfaction degrees and expense are introduced. Furthermore, a multi-objective optimization model can be set up consider satisfaction degrees of agents and expense of intermediary. The method of weighted sums based on membership function is used to convert the multi-objective optimization model into a single-objective model. The matching result is obtained by solving the model.

Keywords: two-sided matching, ordinal number, satisfaction degree, expense

1 Introduction

There are plenty of two-sided matching problems in many fields of real life, such as marriage assignment [1], college admission [2], employee selection [3], personnel assignment [4] and trading partner selection [5]. Therefore two-sided matching is a research topic with extensive application backgrounds.

Study on the two-sided matching problem originates from the problem of college admission and marriage assignment [6]. Gale and Shapley initially investigate the concept, existence, optimality and solution algorithm of stable assignment [7]. Then, Roth explicitly proposes the concept of two-sided matching [3]. Following that, various methods, techniques and algorithms have been proposed for solving the two-sided matching problem with different forms of information. For example, Teo et al. derive the optimal cheating strategy in Gale-Shapley stable marriage model, and then apply it to Singapore school admissions problem [8]. Ehlers studies truncation strategies in matching markets using the deferred acceptance algorithm (i.e., Gale-Shapley algorithm), and show that truncation strategies are also applicable to all priority mechanisms and all linear programming mechanisms [9]. Azevedo gives a simple equilibrium model of an imperfectly competitive matching market, in which an infinite number of firms is matched to a continuum of workers [10]. Teo and Sethuraman study the classical stable marriage and stable roommate problems using a polyhedral approach [11]. Fleiner gives a linear characterization of the bipartite stable b-matching polytope [12]. Manlove et al. give a 2-approximation algorithm for the stable marriage problem with incomplete lists and ties of finding a stable matching of maximum or minimum size [13]. Iwama et al. give a $(2-c/\sqrt{N})$ -approximation algorithm to solve the stable marriage problem of finding a stable matching of maximum size when both ties and unacceptable partners are allowed in preference lists, where c is an arbitrary positive constant that satisfies $c \le 1/4\sqrt{6}$, and N is the instance size [14].

Prior studies have made significant contributions to solving the two-sided matching problems. However, on the one hand, most of the existing studies focus on obtaining stability matching(s). However, in some cases, comparing with stability matching, each agent is more concerned about his/her own satisfaction degree over his/her partner. On the other hand, in real world problems, the determination of the stability matching(s) is usually suggested by intermediary [5, 15, 16]. The intermediary could also be the type of pursuing the benefit, and yet the existing studies seldom consider this type of intermediary. Therefore, how to consider the satisfaction degree of agents and the profit of intermediary in the two-sided matching problem is a valuable research topic. These are the motivation of this paper.

The purpose of this paper is to propose a novel method for solving the two-sided matching problem considering the preferences of agents and intermediary. In the method, the concepts of satisfaction degree and expense are introduced. Furthermore, a multi-objective optimization model is built, where the considered sub-objectives are to maximize the satisfaction degree of each agent and expense of intermediary. By solving the multi-objective optimization model, the matching result can be determined.

The rest of this paper is provided as follows: Section 2 describes the two-sided matching problem considering the preferences of agents and intermediary. Section 3 gives the concepts of satisfaction degrees and expense. Section 4 builds a multi-objective optimization model considering satisfaction degrees and expense. Section 5 develops an

* ***Corresponding author's*** e-mail: liyuhuajx@sina.com

algorithm for solving the model. Section 6 summarizes and highlights the main features of the proposed method.

2 The problem

The related concept and notation on two-sided matching can be found in [17, 18]. The two-sided matching problem considering the preferences of agents and intermediary is described below.

For convenience, let $M=\{1,2,\ldots,m\}$, $N=\{1,2,\ldots,n\}$, and suppose $m\le n$. Let $A=\{A_1,A_2,\ldots,A_m\}$ be the set of agents of side A, where A_i denotes the i-th agent of side A, $i\in M$; $B=\{B_1,B_2,\ldots,B_n\}$ be the set of agents of side B, where B_j denotes the j-th agent of side B, $j\in N$. Let $R_i=(r_{i1},r_{i2},\ldots,r_{in})$ be the ordinal number vector provided by agent A_i over the agents of side B, where r_{ij} denotes that agent A_i ranks B_j in the r_{ij}-th position, $r_{ij}\in N$. Let $T_j=(t_{1j},t_{2j},\ldots,t_{mj})^T$ be the ordinal number vector provided by agent B_j over the agents of side A, where t_{ij} denotes that agent B_j ranks A_i in the t_{ij}-th position, $t_{ij}\in M$. The expense standard of agents is usually determined by intermediary. Here, the intermediary refers to a single person, an organization or a decision system that makes a matching between agents of sides A and B.

The problem concerned in this paper is how to obtain the reasonable matching result based on ordinal number vectors $R_i\ (i\in M)$ and $T_j\ (j\in N)$ such that the satisfaction degrees of each agents and the profit of intermediary are as large as possible.

3 The satisfaction degrees and expense

In the two-sided matching problem, without loss of generality, if agent A_i ranks B_j in the first position, then the satisfaction degree of agent A_i over B_j is the highest; if agent A_i ranks B_k in the last position, then the satisfaction degree of agent A_i over B_k is the lowest. Usually, satisfaction degrees of one agent over potential partners are in interval [0, 1]. Hence, we give the following definitions of satisfaction degrees.

Definition 1. Let $r_{ij}\in N$ be the ordinal number provided by agent A_i over B_j. Then the function l_i used to obtain the satisfaction degree of agent A_i over B_j (noted as α_{ij}) is defined as:

$$l_i: N\to[0,1],$$

$$l_i(r_{ij})=\alpha_{ij},$$

where l_i is a monotone decreasing function and satisfies $0\le l_i(n)$, $l_i(1)\le 1$.

Definition 2. Let $t_{ij}\in M$ be the ordinal number provided by agent B_j over A_i. Then the function g_j used to obtain the satisfaction degree of agent B_j over A_i (noted as β_{ij}) is defined as:

$$g_j: M\to[0,1],$$

$$g_j(t_{ij})=\beta_{ij},$$

where g_j is a monotone decreasing function and satisfies $0\le g_j(m)$, $g_j(1)\le 1$.

The expressions of functions l_i and g_j could be different due to the different consideration of agents. The typical expressions of functions l_i and g_j are expressed by:

$$l_i(r_{ij})=\left[\frac{n+1-r_{ij}}{n}\right]^2,\ i\in M \tag{1}$$

and

$$g_j(t_{ij})=\left[\frac{m+1-t_{ij}}{m}\right]^2,\ j\in N. \tag{2}$$

By Equations (1) and (2), the satisfaction degree matrices $\tilde{\mathrm{A}}=[\alpha_{ij}]_{m\times n}$ and $\tilde{\mathrm{B}}=[\beta_{ij}]_{m\times n}$ are constructed, respectively.

Definition 3. Let $r_{ij}\in N$ be the ordinal number provided by agent A_i over B_j. Then the function $y_A(r)$ used to obtain the expense that agent A_i provides to the intermediary if agent A_i is matched with agent B_j (noted as λ_{ij}) is defined as:

$$y_A(r): N\to R,$$

$$y_A(r_{ij})=\lambda_{ij},$$

where $y_A(r)$ is a monotone decreasing function and satisfies $y_A(n)>0$.

Definition 4. Let $t_{ij}\in M$ be the ordinal number provided by agent B_j over A_i. Then the function $y_B(t)$ used to obtain the expense that agent B_j provides to the intermediary if agent B_j is matched with agent A_i (noted as π_{ij}) is defined as:

$$y_B(t): M\to R,$$

$$y_B(t_{ij})=\pi_{ij},$$

where $y_B(t)$ is a monotone decreasing function and satisfies $y_B(m)>0$.

Usually, the expressions of function $y_A(r)$ and $y_B(t)$ are various. In this paper, we consider the following formats:

$$y_A(r)=\begin{cases} p_1, & r=1, \\ p_2, & r=2, \\ \ldots, & \\ p_n, & r=n, \end{cases}, \tag{3}$$

$$y_B(t)=\begin{cases} q_1, & t=1, \\ q_2, & t=2, \\ \ldots, & \\ q_m, & t=m, \end{cases}, \tag{4}$$

where $p_1, p_2, \ldots, p_n$ are the expense values and satisfy $p_1 > p_2 > \ldots > p_n > 0$, and $q_1, q_2, \ldots, q_m$ are also the expense values and satisfy $q_1 > q_2 > \ldots > q_m > 0$. By Equations (3) and (4), expense matrices $\Gamma=[\lambda_{ij}]_{m\times n}$ and $\Pi=[\pi_{ij}]_{m\times n}$ are constructed, respectively.

4 The two-sided matching model

Let x_{ij} be an 0-1 variable, where $x_{ij}=0$ denotes $\mu(A_i)\neq B_j$, $x_{ij}=1$ denotes $\mu(A_i)=B_j$. To maximize the satisfaction degrees of agents and the expense of intermediary, a multi-objective optimization model (5) can be established as follows:

$$\max \quad Z(A_i)=\sum_{j=1}^{n}\alpha_{ij}x_{ij},\ i\in M, \tag{5a}$$

$$\max \quad Z(B_j)=\sum_{i=1}^{m}\beta_{ij}x_{ij},\ j\in N, \tag{5b}$$

$$\max \quad Z(T)=\sum_{i=1}^{m}\sum_{j=1}^{n}(\lambda_{ij}+\pi_{ij})x_{ij}, \tag{5c}$$

$$\text{s.t.} \quad \sum_{j=1}^{n}x_{ij}=1,\ i\in M, \tag{5d}$$

$$\sum_{i=1}^{m}x_{ij}\le 1,\ j\in N, \tag{5e}$$

$$x_{ij}=0 \text{ or } 1,\ i\in M,\ j\in N. \tag{5f}$$

In the model (5), the meaning of Equation (5d) is that agent A_i must match only one agent of side B. The meaning of Equation (5e) is that agent B_j matches at most one agent of side A.

Generally, each agent of one side has equal priority, thus model (5) can be further transformed into the following optimization model (6):

$$\max \quad Z(A)=\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_{ij}x_{ij}, \tag{6a}$$

$$\max \quad Z(B)=\sum_{i=1}^{m}\sum_{j=1}^{n}\beta_{ij}x_{ij}, \tag{6b}$$

$$\max \quad Z(T)=\sum_{i=1}^{m}\sum_{j=1}^{n}(\lambda_{ij}+\pi_{ij})x_{ij}, \tag{6c}$$

$$\text{s.t.} \quad \sum_{j=1}^{n}x_{ij}=1,\ i\in M, \tag{6d}$$

$$\sum_{i=1}^{m}x_{ij}\le 1,\ j\in N, \tag{6e}$$

$$x_{ij}=0 \text{ or } 1,\ i\in M,\ j\in N. \tag{6f}$$

5 The algorithm

In order to solve model (6), the method of weighted sums based on membership function is adopted [19]. In the followings, we give the detailed analysis on the solution of model (6). Firstly, let $Z_{\min}(A)$ and $Z_{\max}(A)$ be the minimum value and the maximum value for objective function $Z(A)$. Let $Z_{\min}(B)$ and $Z_{\max}(B)$ be the minimum value and the maximum value for objective function $Z(B)$. Let $Z_{\min}(T)$ and $Z_{\max}(T)$ be the minimum value and the maximum value for objective function $Z(T)$. Then, three membership functions can be respectively described with the followings:

$$\mu(Z(A))=\frac{Z(A)-Z_{\min}(A)}{Z_{\max}(A)-Z_{\min}(A)}, \tag{7}$$

$$\mu(Z(B))=\frac{Z(B)-Z_{\min}(B)}{Z_{\max}(B)-Z_{\min}(B)}, \tag{8}$$

$$\mu(Z(T))=\frac{Z(T)-Z_{\min}(T)}{Z_{\max}(T)-Z_{\min}(T)}, \tag{9}$$

obviously, $0\le\mu(Z(A))\le 1$, $0\le\mu(Z(B))\le 1$ and $0\le\mu(Z(T))\le 1$.

Let w_A, w_B and w_T be the weight of objectives functions $Z(A)$, $Z(B)$ and $Z(T)$, respectively, such that $w_A+w_B+w_T=1$, $0<w_A, w_B, w_T<1$, then model (6) is transformed into the single-objective optimization model (10):

$$\max \quad Z=w_A\mu(Z(A))+w_B\mu(Z(B))+w_T\mu(Z(T)), \tag{10a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n}x_{ij}=1,\ i\in M, \tag{10b}$$

$$\sum_{i=1}^{m}x_{ij}\le 1,\ j\in N, \tag{10c}$$

$$x_{ij}=0 \text{ or } 1,\ i\in M,\ j\in N, \tag{10d}$$

where weight w_D $(D=A,B,T)$ reflects the importance

degree of objective function $Z(D)$ in practical decision problems. Usually, to guarantee the fairness of the agents of sides A and B, we have $w_A = w_B$.

Model (10) the can be solved using the existing mathematical optimization software. Then the matching result is determined based on the obtained optimal solution.

In sum, we give an algorithm for solving the two-sided matching problem and its steps are presented as follows:

Step 1: Set up the satisfaction degree matrices $\tilde{\mathrm{A}} = [\alpha_{ij}]_{m\times n}$ and $\tilde{\mathrm{B}} = [\beta_{ij}]_{m\times n}$ by Equations (1) and (2).

Step 2: Set up the expense matrices $\Gamma = [\lambda_{ij}]_{m\times n}$ and $\Pi = [\pi_{ij}]_{m\times n}$ by Eqs. (3) and (4).

Step 3: Set up the multiple-objective optimization model (5) based on satisfaction degree matrices $\tilde{\mathrm{A}}$ and $\tilde{\mathrm{B}}$, and expense matrices Γ and Π.

Step 4: Determine $Z_{\min}(D)$ and $Z_{\max}(D)$, $D = A, B, T$.

Step 5: Transform model (6) into model (10) by Eqs. (7)-(9).

Step 6: Determine the matching result by solving model (10).

6 Conclusion

The two-sided matching problem arises from a wide range of real-world situations. Although many researchers have paid attention to the two-sided matching problem, there are few methods considering the satisfaction degrees of agents and the profit of intermediary. This paper has presented a novel method for solving the two-sided matching problem with ordinal numbers.

Comparing with the existing methods, the proposed method has two characteristics as discussed below. Firstly, the agents' satisfaction degrees and intermediary's profit are considered. This is usually absent in the existing methods. Secondly, the proposed method is simple and is a supplement or extension of the existing methods. In terms of future research, the proposed method can be extended to support situations in which the information data are in other formats.

References

[1] Irving R W, Manlove D F, O'Malley G 2009 Stable marriage with ties and bounded length preference lists *Journal of Discrete Algorithms* **7**(2) 213–9
[2] Klaus B, Klijn F 2006 Median stable matching for college admissions. *International Journal of Game Theory* **34**(1) 1–11
[3] Roth A E 1985 Common and conflicting interests in two-sided matching markets *European Economic Review* **27**(1) 75–96
[4] Korkmaz I, Gökçen H, Çetinyokuş T 2008 An analytic hierarchy process and two-sided matching based decision support system for military personnel assignment *Information Sciences* **178**(14) 2915–27
[5] Bloch F, Ryder H 2000 Two-sided search, marriage and matchmakers *International Economic Review* **41**(1) 93–115
[6] Alkan A, Gale D 2003 Stable schedule matching under revealed preference *Journal of Economic Theory* **112**(2) 289–306
[7] Gale D, Shapley L 1962 College admissions and the stability of marriage *American Mathematical Monthly* **69**(1) 9–15
[8] Teo C P, Sethuraman J, Tan W P 2001 Gale-Shapley stable marriage problem: Revisited strategic issues and applications *Management Science* **47**(9) 1252–67
[9] Ehlers L 2008 Truncation strategies in matching markets *Mathematics of Operations Research* **33** (2) 327–35
[10] Azevedo E M 2014 Imperfect competition in two-sided matching markets *Games and Economic Behavior* **83**(1) 207-23
[11] Teo C P, Sethuraman J 1998 The geometry of fractional stable matchings and its applications *Mathematics of Operations Research* **23**(4) 874–91
[12] Fleiner T 2003 On the stable *b*-matching polytope. *Mathematical Social Sciences* **46**(2) 149–58
[13] Manlove D F, Irving R W, Iwama K, Miyazaki S, Moritab Y 2002 Hard variants of stable marriage *Theoretical Computer Science* **276**(1-2) 261–79
[14] Iwama K, Miyazaki S, Yamauchi N 2008 A $(2-c/\sqrt{N})$ – approximation algorithm for the stable marriage problem *Algorithmica* **51**(3) 342–56
[15] Ehlers L 2007 Von Neumann–Morgenstern stable sets in matching problems *Journal of Economic Theory* **134**(1) 537–47
[16] Kang N, Han S Y 2003 Agent-based e-marketplace system for more fair and efficient transaction *Decision Support Systems* **34**(2) 157–65
[17] Echenique F 2008 What matchings can be stable? The testable implications of matching theory *Mathematics of Operations Research* **33**(3) 757–68
[18] Hałaburda H 2010 Unravelling in two-sided matching markets and similarity of preferences *Games and Economic Behavior* **69**(2) 365–93
[19] Chen Y W, Wang C H, Lin S J 2008 A multi-objective geographic information system for route selection of nuclear waste transport *Omega* **36**(3) 363–72

Authors

Qi Yue, born in November 1983, Jiangxi, China.

Current position, grades: lecturer at Jiangxi University of Finance and Economics, China.
University studies: MSc at Guangxi University of Nanning in China. DSc at the Northeast University of Liaoning in China.
Scientific interests: decision analysis and operations research.
Publications: more than 40 papers.
Experience: teaching experience of 3 years, 5 scientific research projects.

Yuhua Li, 1984.6, Shangdong, China.

Current position, grades: lecturer at Jiangxi University of Finance and Economics, China.
University studies: BSc and MSc at the School of Business Administration at Northeastern University of Shenyang in China. DSc at Kyushu University of Fukuoka in Japan.
Scientific interests: multinational banking and foreign direct investment.
Publications: more than 20 papers.
Experience: teaching experience of 1 years, 4 scientific research projects.

VAR dynamic analysis of the impact of population structure on urban residential land price: the case of Zhengzhou

Zhaoxia Si[1,2*], Ningli Chen[2]

[1]*School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou, 221116, China*

[2]*Henan Polytechnic University, Jiaozuo, 454000, China*

Received 8 January 2014, www.cmnt.lv

Abstract

Based on the vector autoregression model, this paper focuses on Zhengzhou and uses the 1994–2013 population structural variables and relevant data on commercial and residential land price to analyze the dynamic relationship between population structure and residential land price through impulse response function and variance decomposition. Results show a co-integration between Zhengzhou residential land price and three variables, namely, urbanization rate, per capita disposable income, and household size. The short-term variation in land price is mainly caused by the residential land price itself, and the three population substructural variables show long-term effects on the land price with a certain time lag. Among these variables, per capita disposable income has the shortest positive effect on residential land price, whereas urbanization rate has the longest and most remarkable positive effect.

Keywords: population, land price, VAR model, Zhengzhou

1 Introduction

All sectors of society have expressed great concern about the increasing land price in large and medium-sized cities of China. In recent years, scholars have attempted to explain the increase from different angles, such as economic growth, social policy, monetary supply, speculative demand and motivation, and housing price. Population is an intrinsic factor that influences housing demand and is less vulnerable to business cycles than other factors, and thus it is more likely to influence housing and land price. However, the influence of population structure on the variation of residential land price has not been studied enough.

Foreign scholars have studied the relationship between population structural variable and housing price and arrived at varied findings. For instance, David [1] examined the Dutch commercial housing price from the aspect of household disposable income and found that disposable income is positively related to and significantly affects housing price. Wolfgang Maennig [2] studied the relationship between the population fluctuation of 98 major German cities and housing price and suggested that population decline in large cities exerts a more significant effect on housing price than population growth. Vansteenkiste and Hiebert [3] studied seven European countries using the vector autoregression (VAR) model to determine the relationship between real per capita income and housing price and found that real per capita income contributes to the rise of real housing price. Ramesh Kumar Jain [4] examined relevant data on major Indian cities and found that the deepening urbanization and urban population accretion promote local economic development and increase regional housing price. Bischoff and Oliver [5] studied the regional difference in the relationship between housing price and income in Germany and found a positive interactive effect between income and housing price. Based on these studies, population development is the most important factor that causes the disparity between housing price and income in all regions.

Most of the literature published by Chinese scholars about the relationship between population and housing price focus on the influence on housing price caused by four factors: population age structure, per capita disposable income, urbanization, and population migration. The consistently drawn conclusion is that per capita disposable income, which signifies a lower correlation index than other influencing factors, is not the primary factor affecting housing price; rather, the ongoing process of urbanization improves urban housing demand and subsequently leads to the fluctuation of housing price. Ha Jiming [6] showed that urbanization and changes in population structure promote economic growth and affect housing price, and that population structure directly affects housing price. Wu Tao [7] demonstrated that gross domestic product and per capita disposable income indirectly affect housing price by influencing CPI. Kuang Weida [8] showed that cities with a higher population growth rate exhibit more volatile housing prices than cities with a relatively small population. Xue Liwei et al. [9] analyzed the data on Beijing, Wuhan, and Xi'an and found a positive correlation between population structural

Corresponding author's e-mail: sizhaoxia@hpu.edu.cn

variables and housing price income. Guo Jian[10] analyzed the influence of the urbanization of the Jiangsu population on commercial housing price and concluded that the extent of population aggregation is the primary factor that affects urban housing price. Conversely, Xun Fang [11] studied Jiangxi and found that level of urbanization could significantly influence housing price. He Fan [12] examined the dependency ratio of population and housing price data of 19 countries and found a negative correlation between dependency ratio of population and housing price fluctuation in most countries; however, the proportionate rise in the aging population of China, rather than a price fall, causes an active demand for commercial housing. Du Benfeng and Zhang Yu [13] used the grey relational degree to study the correlation between the Chinese comprehensive demographic factor and home sales price index and found that population is highly correlated with home sales price index.

The literature review shows that foreign scholars have conducted extensive research about the influence of population structure on housing price fluctuation. Their studies, although recent, have covered a wide range of concerns about theoretical exploration and used the econometrical approach for a quantitative analysis of the correlation between population structure and price fluctuation. By contrast, Chinese scholars have used foreign research methods and the true state of China to focus on typical Chinese regions and found that population structural variables affect the volatility of housing price. Studies involving the correlation between population structure and urban residential land price are rarely conducted. This paper benefits from the method used to study the influence of population structure on the volatility of housing price. It analyzes the dynamic influence of Zhengzhou's population structural variables on residential land price. The findings provide reference for similar research on other regions and contribute to the government's decision making in the development of diverse regulatory policies for local residential land price.

2 Data resource and research method

This paper takes the urban residential land price indexes and population structural variables of Zhengzhou as analysis data. Without affecting the results, price indexes lower data heteroscedasticity while increasing sequence stationarity. The data selected covered the period of 1994–2013 and were taken from the China Urban Land Price Dynamic Monitoring Network, Henan Statistical Network, and Zhengzhou Statistical Network.

This study mainly discusses the extent of influence of Zhengzhou's population structural variables on the fluctuation of residential land price. Analysis software Eviews 6.0 was used for building the VAR model. Co-integration analysis, impulse response function, variance decomposition, and other approaches were used to analyze the correlative mechanism between Zhengzhou population structural variables and land prices.

The AR model is a simultaneous form of the autoregression model. If y_{1t} is related to y_{2t} and the two autoregression models are built separately, the relationship between these two variables will not be captured:

$$y_{1t} = f\left(y_{1,t-1}, y_{1,t-2}, \ldots\right), \tag{1}$$

$$y_{2t} = f\left(y_{2,t-1}, y_{2,t-2}, \ldots\right). \tag{2}$$

Once the simultaneous form is taken, the relationship between the two variables can be determined. Taking the VAR model of one-period lagged variables y_{1t} and y_{2t} as an example, the VAR model can be expressed as:

$$y_{1t} = c_1 + \pi_{11.1} y_{1,t-1} + \pi_{12.1} y_{2,t-1} + u_{1t}, \tag{3}$$

$$y_{2t} = c_2 + \pi_{21.1} y_{1,t-1} + \pi_{22.1} y_{2,t-1} + u_{2t}. \tag{4}$$

3 Empirical analysis

3.1 VARIABLE SCLECTION

Based on demography, population structure generally covers the natural population structure, the socio-economic structural features of the population, and the regional structure of the population. This approach is adopted in this study on population structural variables. By selecting the variables according to the three aspects, this study attempts to discuss Zhengzhou's residential land price from the perspective of population structure.

On the basis of the literature review, this study tentatively selects seven population structural variables: sex ratio (*XB*), labor resource structure (*LP*), dependency ratio (*FY*), household size (*JG*), employment structure (JY), per capita disposable income (*SR*), and urbanization rate (*CZ*).

3.2 DATA STATIONARY TEST

Stationarity was the first variable to be tested. Time series stationarity indicates that statistical laws on time series do not change with time; i.e., the features of the stochastic process during which the variable time series data are generated are time-invariant. Logarithm operation was conducted on these time series variables to reduce heteroscedasticity. The ADF approach was adopted for the unit root test of the same variables. The results indicate that variables do not form a stationary series, that urbanization rate is the only stationary series after the first-order difference, and that the rest of the variables are stationary series after the second-order difference (Table1).

TABLE 1 Unit root test on the time series of land price and population structure variable in Zhengzhou City

Variate	Test Type (c,t,n)	ADF statistical magnitude	5% critical value	Inspection results	Variate	Test Type (c,t,n)	ADF statistical magnitude	5% critical value	Inspection results
ln *CZ*	c, 0, 0	–0.69	–3.25	Non-stationary	D^2 ln *CZ*	c, 0, 1	–3.45	–3.32	stationary
ln *FY*	c, 0, 0	–2.48	–3.25	Non-stationary	D^2 ln *FY*	c, 0, 1	–3.84	–3.52	stationary
ln *JY*	c, 0, 0	–1.94	–3.25	Non-stationary	D^2 ln *JY*	c, 0, 1	–3.58	–3.52	stationary
ln *LP*	c, 0, 0	–2.46	–3.25	Non-stationary	D^2 ln *LP*	c, 0, 1	–3.89	–3.52	stationary
ln *SR*	c, 0, 0	3.33	–3.25	Non-stationary	D^2 ln *SR*	c, 0, 1	–6.61	–3.52	stationary
ln *XB*	c, 0, 0	–3.08	–3.25	Non-stationary	D^2 ln *XB*	c, 0, 1	–7.23	–3.52	stationary
ln *JG*	c, 0, 0	–0.74	–3.25	Non-stationary	D^2 ln *JG*	c, 0, 1	–4.06	–3.40	stationary
ln *HF*	c, 0, 0	–0.31	–3.25	Non-stationary	D^2 ln *HF*	c, 0, 1	–3.88	–3.52	stationary

Notes: (c, t, n) are the constant term, trend term, and lag order, respectively, in the unit root test.

3.3 JOHANSEN CO-INTEGRATION RELATIONSHIP TEST

The Johansen test is used to determine the existence of several non-zero characteristic roots for the vector matrix of co-integration test, which suggests a co-integration among various time series. The test was conducted with Trace Statistic and Max-Eigen Statistic. The data generation test proved that the system shows no linear trend and that the co-integration equation is a random walk with drift. The Trace Statistic and Max-Eigen Statistic approaches used by the Johansen co-integration relationship test indicate that, below the level of 5%, no non-zero characteristic root exists between Zhengzhou residential land price and dependency ratio, labor resource, employment proportion, and sex ratio, thus indicating the absence of co-integration. Moreover, no non-zero characteristic root exists between Zhengzhou residential land price and urbanization rate, per capita disposable income, and household size, thus indicating a co-integration relationship. Table 2 shows the results of the co-integration relationship test.

TABLE 2 Johansen test between land price and population structure variables in Zhengzhou City

	Original hypothesis co-integration equation	Eigenvalue	Trace Test	Trace Test C.V.	Maximum Eigenvalue	Maximum Eigenvalue C.V.
ln *HP* and ln *CZ*	None*	0.98	35.11	15.49	34.50	14.26
	At most 1	0.07	0.62	3.84	0.62	3.84
ln *HP* and ln *SR*	None*	0.87	21.89	15.49	16.54	14.26
	At most 1	0.49	5.34	3.84	5.34	3.84
ln *HP* and ln *SR*	None*	0.83	16.62	15.49	14.21	14.26
	At most 1	0.26	2.40	3.84	2.39	3.84

Notes: "*" is that refused to the original assumption under the 0.05% significant level

3.4 VAR MODEL ESTIMATION

The co-integration relationship test suggests a long-term equilibrium relationship between Zhengzhou residential land price and urbanization rate, per capita disposable income, and household size. The VAR model, which is used to analyze the correlation between variables, is an unstructured equation model based on data statistics and is completely built on endogenous variables.

Lagged orders, which significantly influenced the test results, need to be set to build the VAR model. Low orders disregard some important variables, whereas high orders lower the freedom of the model, lead to a considerable standard deviation of parameter estimation, and reduce accuracy. This study determines the lag period according to the lag length criteria. As shown in Table 3, the smallest lag periods of the five evaluation indexes are represented by "*". According to the minimum AIC and SC criteria, the optimal lag period is set to 2 and a VAR (2) model is determined for building.

TABLE 3 Criteria to determine VAR lag length and appropriate lag order number

Lag	LogL	LR	FPE	AIC	SC	HQ
0	–321.67	NA	6.10e+10	36.18	36.38	36.21
1	–211.70	15.72	15637203	27.52	29.30	27.76
2	–227.42	136.14*	10850357*	27.49*	28.48*	27.62*

Eviews 6.0 was used to build the four-variable VAR model of Zhengzhou urbanization rate, per capita disposable income, household size, and residential land price. The model estimation results are as follows:

$$\begin{bmatrix}\ln CZ\\ \ln DJ\\ \ln JG\\ \ln SR\end{bmatrix}=\begin{bmatrix}-4.52\\ 124.14\\ 6.24\\ 188339.77\end{bmatrix}+\begin{bmatrix}1.06 & -0.06 & 4.12 & -6.11\\ 2.54 & 1.55 & -60.67 & 5.43\\ 0.05 & 0.02 & -0.32 & 1.42\\ 17274.21 & -1278.27 & 84524.99 & -0.54\end{bmatrix}\begin{bmatrix}\ln CZ(-1)\\ \ln DJ(-1)\\ \ln JG(-1)\\ \ln SR(-1)\end{bmatrix}+\begin{bmatrix}0.09 & 0.04 & -4.06 & -3.45\\ -2.96 & -0.40 & 31.04 & 6.59\\ -0.12 & -0.01 & 0.36 & 1.80\\ -15478.00 & 1274.89 & 150190.05 & -0.22\end{bmatrix}\begin{bmatrix}\ln CZ(-2)\\ \ln DJ(-2)\\ \ln JG(-2)\\ \ln SR(-2)\end{bmatrix}.$$

The stationary test for the model, all characteristic roots in the model shown in Figure1 are in the unit cycle; therefore, the VAR model has a stable system. After the test for the residuals of the VAR (2) model, the model residuals were found to follow a normal distribution and

to have no autocorrelation or heteroscedasticity. Estimation results also suggest that the joint parameter survey is significant. The model determination coefficient R is between 0.955 and 0.996; therefore, the statistical property of VAR (2) is favorable.

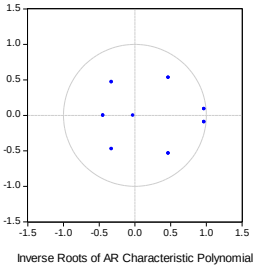


FIGURE 1 Stability of the VAR model
of Zhengzhou residential land price and population structural variables
(AR characteristic polynomial root diagram)

3.5 IMPULSE RESPONSE ANALYSIS AND CONTRIBUTION ANALYSIS

The conduct of impulse response analysis and variance decomposition for the VAR model was necessary to study the change rule of the long-term equilibrium relationship within a certain period after the influence.

3.5.1 Impulse response analysis

Impulse response analysis is usually carried out on the basis of the graph observation of the impulse response function, which describes the response of error change. The random disturbance term is assumed to be affected by a standard deviation, and the variables would be affected both in the current and later years; i.e., the change in each endogenous variable can influence itself and all other endogenous variables. Figure 2 shows the disturbance when one standard deviation is given. The response route, which revealed the influence of variables on land price within 10 predictive periods, was observed. The full line denotes the change orbit of land price after the effects, and the dotted line indicates the resulting value of land price plus or minus the double standard deviation, thus showing the maximum range of possible effects, i.e., the upper and lower limits to be reached.

As shown in Figure 3, when a random disturbance of one standard deviation is observed from the household size against the residential land price, a strong influence appears immediately. The current impulse curve falls gradually after going up by approximately five grades; at the eighth period approaching zero, the influence gradually disappears. This trend shows that the development of land price is stable and that short-term variations of household size can cause fluctuations in residential land price. However, the fluctuation range reduces with time. Land price increases immediately when a positive effect of one standard deviation is exerted on itself, but the increase range shrinks gradually and immediately, and the price begins to fall by the second period until the final recovery. When a positive effect of one standard deviation from the urbanization rate is exerted on the residential land price, the effect is gradually seen. The price begins to gradually decline after reaching the second grade by the second predictive period, falls to first grade by the third predictive period, begins to gradually increase by the fourth predictive period, and remains at the third grade. No significant response is shown in land price when a disturbance of one standard deviation from the per capita disposable income is exerted on the residential land price. Based on the effect of population structural variables, the per capita disposable income has the shortest positive effect on residential land price. Given the high price of residential land, income increases may have no significant influence on the increasing land price. By contrast, the positive effect of urbanization rate on residential land price is the longest and strongest. Therefore, new housing demands caused by urbanization can actively promote the residential land price market of Zhengzhou and significantly improve the land price of this region.

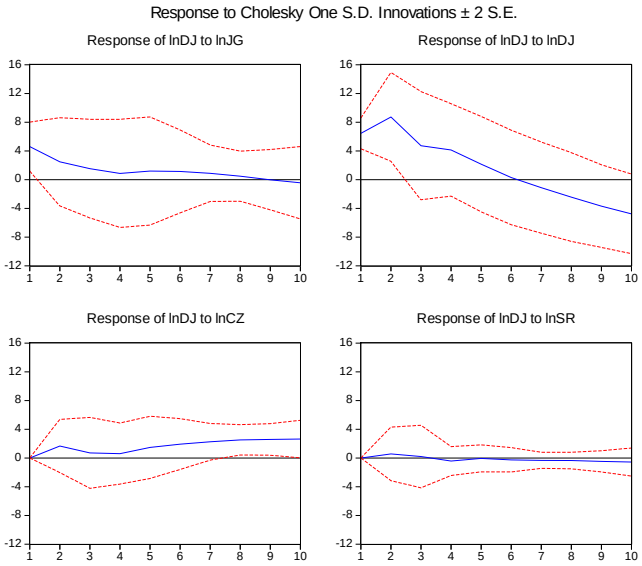


FIGURE 2 Impulse response analysis of Zhengzhou residential land price

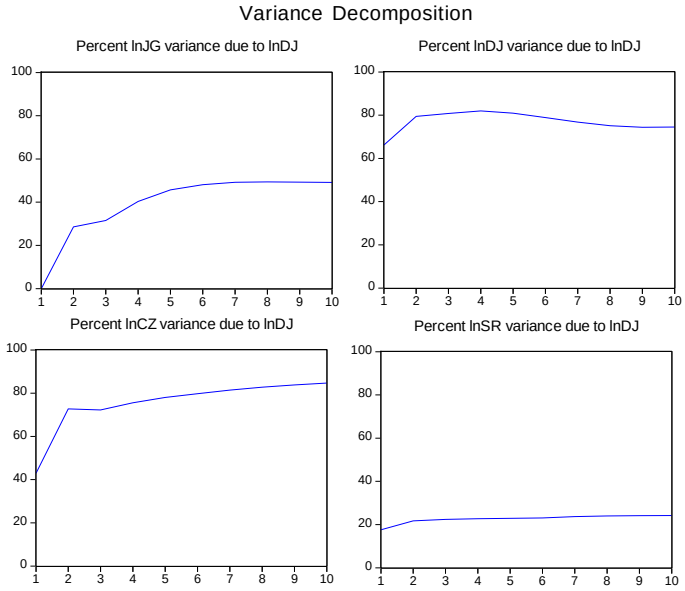


FIGURE 3 Variance decomposition of population structural variables against residential land price in Zhengzhou

3.5.2 Contribution analysis

Variance decomposition means that interactions between variables can be expressed by percentages of predication error variances after a variable in the system suffers one unit of effect. Variance decomposition reflects the relative importance of random information. That is, the variation of each endogenous variable in the system is decomposed into all constituent parts relevant to random disturbance terms to determine the relative importance of innovation to the endogenous variables of the model. Variance decomposition analysis was conducted to study the influence of the Zhengzhou population structural variables on the fluctuation of residential land price.

In Figure 3, the horizontal axis represents the periods after various effects, and the vertical axis represents the contribution of the four variables (including land price itself) to the increase in land price. As shown in Figures 3 and Table 4, land price has an effect on itself. The contribution rate reaches 66.22% by the first predictive period and increases gradually. It reaches approximately 80.91% by the fifth period, declines, and then tends to stabilize until it is maintained at 75%. Household size (lnJG) has no contribution to Zhengzhou land price by the first period. However, the contribution rate increases to 28.58% by the second predictive period, escalates, reaches 48.09% by the sixth predictive period, and then stabilizes. The contribution rate of urbanization rate (InCZ) to Zhengzhou residential land price is only 42.99% by the first predictive period and increases fast thereafter. It reaches 72.75% by the second predictive period and escalates until the 10th predictive period when the contribution rate reaches 84.66%. The contribution rate of per capita disposable income to Zhengzhou residential land price is maintained at a low level all the time at around 23.00%.

TABLE 4 Variance decomposition of population structural variables against residential land price in Zhengzhou

Period	Percent LNDJ variance due to LNDJ (%)	Percent LNSR variance due to LNDJ (%)	Percent LNCZ variance due to LNDJ (%)	Percent LNJG variance due to LNDJ (%)
1	66.22	17.58	42.99	0.00
2	79.47	21.70	72.75	28.58
3	80.81	22.38	72.26	31.52
4	81.97	22.69	75.59	40.34
5	80.91	22.89	78.01	45.66
6	78.92	23.07	79.79	48.09
7	76.83	23.66	81.39	49.19
8	75.15	23.97	82.76	49.38
9	74.42	24.09	83.82	49.26
10	74.53	24.17	84.66	49.15

The contribution analysis shows that different population structural variables have varied effects on the fluctuation of residential land price. Among all the variables, urbanization rate causes the greatest effect, followed by household size. Per capita disposable income causes the smallest effect. In the short term, land price fluctuation is mainly caused by its own influence. Nevertheless, as the predictive period progresses, the effect of population

structural variables on Zhengzhou residential land price becomes stronger and gradually becomes the leading influencing factor. This result suggests that the effect of population structure factors on land price is long standing and stable.

4 Results and discussion

4.1 RESULTS ANALYSIS

This study examines the correlation between the 1994–2013 land prices of Zhengzhou and its population structural variables. Co-integration test, impulse response function, variance decomposition, and other approaches are used to analyze the influence of population structural variables on land price. The findings are as follows:

1) Co-integration test results show that no co-integration relationship exists between Zhengzhou land price and the population structural variables of dependency ratio, labor resource, employment proportion, and sex ratio. However, urbanization rate, per capita disposable income, and household size show significant effects on Zhengzhou land price.
2) The positive effect of Zhengzhou's per capita disposable income on residential land price is the shortest. As income increase may have no significant effect on the increase in residential land price, it is not the primary influential factor on land price. The positive effect of urbanization rate on residential land price is the longest and strongest. Therefore, new housing demands caused by urbanization can actively promote the residential land price market of Zhengzhou and significantly improve the land price of this region.
3) Different population structural variables exert various effects on residential land price fluctuation. Among all the variables, urbanization rate causes the greatest influence, followed by household size. Per capita disposable income causes the smallest effect.
4) In the short term, land price fluctuation is mainly caused by its own influence. Nevertheless, as the predictive period progresses, the effect of population structural variables on Zhengzhou residential land price becomes stronger and gradually becomes the leading influencing factor. This finding suggests that the effect of population structure factors on land price is long-standing and stable.

4.2 DISCUSSION

Through a correlation study between Zhengzhou population structural variables and urban residential land price, this paper examines the impact degree of population structural variables on residential price. Two policy suggestions are suggested:

1) Taking measures to increase resident income should be continued. The empirical tests conducted by this study show that per capita disposable income is not significantly related to residential price. Income growth does not necessarily mean that it can remarkably promote commercial residential land price. The real estate market shows a trend of excessive speculation. Therefore, the basic solution to the housing problem is to increase resident income.
2) Conducting rational control over urbanization should be maintained. Data analysis shows that the increase in and the concentration of urban population are principal factors that increase the price of Zhengzhou commercial residential land. Urbanization, an inevitable process in urban development, greatly influences the development prospect of the real estate market. However, excessively rapid growth can accelerate the increase in urban residential land price, and it will be further promoted under the market force. Therefore, the government should conduct rational control over it. fluctuation in Tianjin caused by a land price fluctuation in Beijing is lower than that in Beijing caused by a land price fluctuation in Tianjin city. This is consistent with the quantitative analysis on the land price fluctuation of the two cities.

Acknowledgments

The authors are grateful for the financial support provided by the National Foundation of China (No. 40971074).

References

[1] Hoffman D, Nadal De Simone F, Walsh M 2005 Kingdom of the Netherlands-Netherlands: Selected Issues *IMF Country Report* 05/225
[2] Maennig W, Dust L 2008 Shrinking and growing metropolitan areas asymmetric real estate price reactions? The case of German single-family houses *Regional Science and Urban Economics* **38**(1) 63-9
[3] Vansteenkiste I, Hiebert P 2011 Do House Price Developments Spillover acrossEuro Area Countries? Evidence from a Global VAR. *Journal of Housing Economics* **20**(4) 299-314
[4] Kumar J R 2011 Houses Turn Gold as Prices Skyrocket *Money Today*
[5] Bischoff O 2010 Explaining Regional Variation in Equilibrium Real Estate Prices and Income *Journal of Housing Economics*, **21**(1), 1-15
[6] Ha J M Population structure and urbanization promote real estate development *Real Estate Information of China* (7) 28-9 *(in Chinese)*
[7] Wu T 2009 The research on the influence factors of urban residential market based on VEC model *Statistics And Decision* (17) 96-8 *(in Chinese)*
[8] Xue L W, Zhao X J, Xu J 2010 Analysis on the Factors Affecting the Housing-Price-Income Ratio *Zhejiang Social Science* (3) 17-20 *(in Chinese)*
[9] Dong F 2011 The Correlation analysis and research of urbanization on urban housing prices based on the panel data model *Jiangxi University of Science and Technology (in Chinese)*
[10] Kuang W D The fluctuate research of expectations, speculation and China urban house price *Economic research* **45**(9) 67-78 *(in Chinese)*
[11] Guo J, Sun W 2010 Quantitative analysis of the impact of urbanization on housing prices *Commercial Times* (5) 122-3 *(in Chinese)*
[12] He F Looking for demographic factors behind rising house prices. *China Foreign Exchange* (17) 63 2011 *(in Chinese)*
[13] Du B F, Zhang Y 2011 Gray Correlation Analysis on Population Factors and HSPI *Population Journal* (6) 11-7 *(in Chinese)*
[14] Lang D G 2013 An Empirical Study between population structure and residential land prices in Wuhan city *Huazhong Agricultural University (in Chinese)*

[15] Zhang X T 2009 Eviews User Guide and Case *Machinery Industry Press* Beijing China *(in Chinese)*
[16] Gao T M 2009 Econometric analysis method and modelling-Eviews application and instance *Qinghua University press* Beijing China *(in Chinese)*
[17] Lei S, Huang Y 2012 VAR Dynamic analysis of the population structure affecting on urban housing prices *Qinghai Social Sciences* (4) 27-31 2012 *(in Chinese)*
[18] He D 2011 An Empirical Study between population structure and residential prices in Wuhan city *Huazhong Agricultural University* China *(in Chinese)*
[19] Hu X L 2010 The research of population structure and housing demand in Xi'an city *Xian university of science and technology* China *(in Chinese)*
[20] Zhou Y, Cheng F 2013 The Impact of population structure on the housing price – Based on empirical research in Beijing *Reformation & Strategy* **29**(11) 78-81 *(in Chinese)*

Authors

Zhaoxia Si, born in March 1980, Gaoping County,Shanxi Province, P.R. China.

Current position, grades: PhD student at the School of Environment and Spatial Informatics, China University of Mining and Technology, China.
Lecturer at Henan Polytechnic University, China.
University studies: BSc at Taiyuan Normal University in China. MSc at Sun Yat-Sen University in China.
Scientific interests: land economics, land monitoring and valuation
Publications: 3 papers.

Ningli Chen, born in April 1988, Zhoukou County, Henan Province, P.R. China

Current position, grades: the Master graduate student of Henan Polytechnic University, China.
University studies: MSc at Xinyang Normal University in China.
Scientific interests: land monitoring, ecological evaluation.
Publications: 2 papers.
Experience: teaching experience of 2 years, participated in 2 scientific research projects.

Research on the application of ARMA model in China's trade economic development

Xue Wu*

Business Administration College, Chongqing University of Science and Technology, Chongqing city, China, 401331

Received 20 October 2014, www.cmnt.lv

Abstract

The slowing down development of China's foreign trade brings significant impact to its regional economy, which makes many middle and small-sized export enterprises facing with business crisis and even bankruptcy. In this paper, a deep analysis on the development status of China's trade economy was conducted with ARMA model. The short-term development trend and causes of China's trade economy were predicted and analyzed. Results demonstrated that China's imports will decrease while export will keep increasing. The balance of trade in China will decrease and the total volume of foreign trade will decline. This indicates the depression of China's trade economy, which is against China's economic development. To improve the export structure and facilitate the rapid development of trade economy, China is suggested to adopt various measures, such as optimizing fiscal policy, promoting technology upgrading, enhancing export management and perfect import-export infrastructures. Meanwhile, China shall expand domestic demand vigorously to offset impact of trade economic slowdown on regional economy.

Keywords: ARMA model, trade economy, application research, decision reference

1 Introduction

Although the world economy recovers after the financial crisis, the international trade market still remains in downturn with grim situation. Trade economy not only is an important goal of opening to the outside world, but also can facilitate economic development and increase production level of China. With a big population but few land resources, China is in severe shortage of many resources, such as energy which is the basis of economic development. Therefore, developing trade economy is an important measure to offset China's resources and economic deficiencies. In 2013, the total volume of imports and exports of China reached 4,160 billion dollars, ranking the first in goods export in the world. It accounted for 45.39% of GDP. Among them, 24.11% was contributed by goods export. This reflects that trade economy has become very important to China's economy. It is of important practical significance to study the trade economic development.

2 Development status of China's trade economy

Although the world economy recovered gradually in 2013, the international market still remains in downturn, thus resulting in the difficult foreign trade development. In 2013, the total volume of imports and exports in China was 4,160 billion dollars, with 7.6% growth. China became the first trade power in the world. As the first country whose total volume of goods trade exceeds 4,000 billion dollars, China is the miracle in the world trade development history. The exports were 2,210 billion dollars, showing 7.9% growth. It accounted for 11.8% of global exports and was 0.7% higher than that in 2012. China's exports ranked first for five successive years in the world. The imports valued 1,950 billion dollars, showing 7.3% growth. It accounted for 10.3% of global imports and was 0.5% higher than that in 2012. China's imports ranked second for five successive years in the world. The trade surplus was 259.75 billion dollars, occupying 2.8% of GDP. Moreover, the general foreign trade of China in 2013 increased by 9.3% to 2,200 billion dollars, 52.8% of the total volume of foreign trade and 0.8% higher than that in 2012. The processing foreign trade was increased by only 1.1% to 1,360 billion dollars, 32.6% of total the total volume of foreign trade and 2.2% lower than that in 2012. This implies that China's guidance on attention shift from processing trade to technology trade achieved positive progress. Among the processing foreign trade, 11.9% was contributed by middle and west China, showing 2.1% growth than that in 2012 and 8.9% growth than that in 2008. Other foreign trade increased by 18.3% to 610 billion dollars, taking up 14.5% of the total volume of foreign trade.

Facing with the tightening demand and fierce market competition, import and export enterprises adopt transformation actively. They increase investment to technology R&D, brand building and quality management. This further increases technology content and value added of export commodities, thus providing strong support to the stable export growth of commodities. Exports of 7 kinds of labor-intensive products, including textile, clothing, bag, footwear, toy, furniture and plastic products, reached 461.8 billion dollars, increasing by 10.3% (2.4% higher than the overall export growth). Exports of mechanical and elec-

* ***Corresponding author's*** e-mail: wuxue400020@163.com

trical products were 1,265.5 billion dollars, increasing by 7.3% and accounting for 57.3% of the total exports. Viewed from the internal export structure of mechanical and electrical products, automatic data processing equipment (ADPE) export reduced slightly, but appliance export increased slowly and some mechanical equipment export increased quickly, becoming a new growth point of mechanical and electrical products export. Specifically, aircraft exports increased by 76.1%, accompanied with 13.5% growth in mechanical handling equipment and parts as well as 12.4% growth in textile machinery and parts. High-tech product exports increased by 9.8% to 660.3 billion dollars, 29.9% of the total exports and 0.5% up compared to that in 2012.

Generally, China's foreign trade with developed countries increases slowly. In 2013, China's foreign trade with America, EU and Japan increased by 2.1% to 1,392.6 billion dollars, which was equal to 33.5% of the total foreign trade volume, 1.7% lower compared to that in 2012. To be more specifically, China's foreign trades with America and EU were 559.1 billion dollars and 521 billion dollars, increasing by 2.1% and 7.5%, respectively. China's foreign trade with Japan was 312.6 billion dollars, decreasing by 5.1%. On the contrary, China's foreign trade with emerging economies generally maintained a quick growth. However, China's foreign trade with some emerging economies declined to a certain extent after August and September. This is caused by the financial market turmoil and economic slowdown of these emerging economies. In 2013, China's foreign trade with ASEAN, South Africa, Brazil and Russia increased by 10.9%, 8.6%5.3% and 1.1% respectively, but China's foreign trade with India decreased by 1.5%.

Middle and west China plays an increasing important role in trade economy. In 2013, middle and west China contributed 562,6 billion dollars of the total foreign trade volume, increasing by 14.3% (6.7% higher than the growth of total foreign trade volume). Its proportion in the total foreign trade volume increased to 13.5% from 12.7% in 2012. The exports of middle and west China increased by 17.1%, 9.2% higher than the total exports. Yunnan, Ningxia, Guizhou, Gansu, Chongqing and Henan achieved more than 20% growth in exports. Export growth of Yunnan and Ningxia even reached as high as 59.3% and 55.5%. The foreign trade of East China was 3,597.7 billion dollars, increasing by 6.6%. Exports of east China achieved 6.9% growth.

3 Application of ARMA model in China's trade economic development

3.1 THEORY OF ARMA MODEL

ARMA model, or called as autoregression moving average model, was a precise time series forecasting method established by BOX and Jenkins. It believes that some time series are a group of random variables depending on time. Although a single variable composing time series is uncertain, the whole series vary according to certain law. As a result, the future development of such time series can be forecasted. Due to the forecast error, it is impossible to forecast continuously based on the predicted value, but only can make short-term forecast.

1) Autoregression model (AR model): if the time series y_t is the function about its former time series and stochastic term:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... + \varphi_p y_{t-p} + \mu_t ,$$

y_t is the *p*-order autoregression sequence. φ_1, φ_2 ,…, φ_p are autoregression coefficients. Let B^k to be the lag operator and then:

$$B^k y_t = y_{t-k} .$$

Then, the AR model can be rewritten as:

$$y_t = \varphi_1 B y_t + \varphi_2 B^2 y_t + ... + \varphi_p B^p y_t ,$$

$$\varphi(B) = \varphi_1 B + \varphi_2 B^2 + ... + \varphi_p B^p .$$

It can be further simplified into:

$$\varphi(B) y_t = \mu_t .$$

2) Moving average model (MA model): if y_t is the linear function of its former series and former error term:

$$y_t = \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - ... - \theta_q \mu_{t-q} ,$$

y_t is called as *q*-order moving average model. θ_1, θ_2 , , θ_q are moving average coefficients. Similarly, let B^k to be the lag operator and then:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q .$$

It can be further simplified into:

3) ARMA model: if y_t is the function of the error of its current and former stochastic terms as well as former series:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... + \varphi_p y_{t-p} + \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - ... - \theta_q \mu_{t-q} .$$

y_t is called as the (*p*, *q*)-order ARMA model.

3.2 APPLICATION OF ARMA MODEL

Firstly, ARMA model has to analyze its lag orders to get *p* and *q* (Table 1). AR (7) MA (2) can be implemented to variables. Secondly, the least square method is implemented by taking variables as dependent variables and AR(1), AR(2), AR(3), AR(4), AR(5), AR(6), AR(7), MA(1) and MA(2) as independent variables. The calculated parameters are listed in Table 2. ARMA has no strict requirements on *T* and tail probability (*P*). Therefore, there are only three main parameters, namely,

correlation coefficient, AIC and SC. Finally, dependent variables are forecasted by the ARMA model. Results are shown in Figures 1-4.

TABLE 1 *p* and *q* for model analysis

Orders	Imports & Exports	Exports	Imports	Balance
p	7	7	7	6
q	2	2	2	2

TABLE 2 Basic parameters for model forecast

Parameters	Balance	Exports	Imports	Imports & Exports
Adjusted R-squared	0.9054	0.9862	0.9890	0.9832
Akaike info criterion	18.3686	20.2353	19.7202	21.6793
Schwarz criterion	18.4600	20.6597	20.1445	22.1037

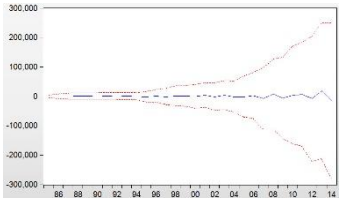


FIGURE 1 Balance of trade forecast in 2014

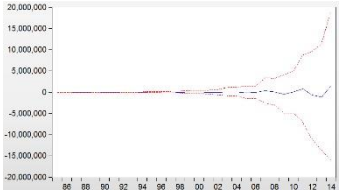


FIGURE 2 Export forecast in 2014

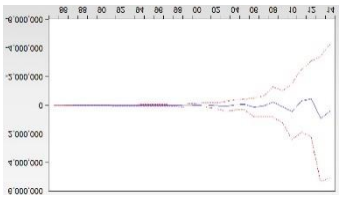


FIGURE 3 Import forecast in 2014

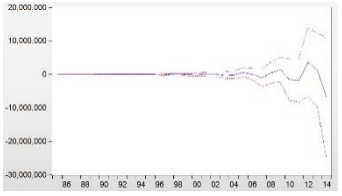


FIGURE 4 Foreign trade forecast in 2014

In Figure 1, the center line represents the forecasted balance of trade, which indicates that the balance of trade will decline in 2014. This agrees with China's trade economic development logic. China has maintained trade surplus for a long time, which brought a lot of trade conflicts with many developed countries. Main causes of balance reduction of trade are analyzed. Firstly, Chinese government is committed to coordination to eliminate trade conflicts and adopts various measures to reduce balance of trade and inhibit the increasing trade deficit. Secondly, China's economy has experienced rapid development and now enters into stable development. This will slow down the trade economic development, thus decreasing balance of trade. Thirdly, the chronic trade deficit intensifies the trade conflicts between China and developed countries. Developed countries will take various measures against China's export development, thus influencing the balance of trade finally.

China's export forecast is shown in Figure 2. The center line reflects that China's export shows a growth trend. There are three causes of China's export growth trend. Firstly, although China's economy enters into stable development, its growth rate is still higher compared to the world economic development. The rapid development of production speed can facilitate export to a certain extent. Secondly, although China pays close attentions on trade balance, it still seems partial to export development, because export can drive region economic development significantly and China still faces with excess capacity in many industries. Thirdly, China is improving the pattern of production continuously. It takes a certain time to transform from labor-intensive product export to technology-intensive product export. Therefore, export will increase continuously.

China's import forecast is shown in Figure 3, which presents a declining trend. There are four main causes of import reduction. Firstly, the world economic slowdown decreases the world's production capacity, thus weakening export capacity of other countries. Secondly, China depends less on import with the enriching production. Thirdly, China, though has entered into the WTO, still has various protection policies to domestic enterprises, which are great barriers for the entry of foreign products. Fourthly, China's policy of promoting domestic demand goes against import. China's subsidy program for rural purchases is only available to domestic products, which is against marketing of foreign products in China.

China's foreign trade forecast is shown in Figure 4, which presents a declining trend. This indicates the gap between China's export growth and import reduction. This is the reason of the reduced balance of trade in China. Such declining trend of China's foreign trade reflects that China's trade economy will face with recession caused by three reasons:

1) the world economic recession;
2) effect of China's economic development strategies;
3) China's economic development trend.

4 Conclusions and suggestions

Based on above analysis, China's import will decline, while export will keep increasing. The balance of trade in China will decrease and the total volume of foreign trade will decline, indicating China's trade economy downturn.

This is against China's economic development. Favorable trade economic development shall include rapid development both import and export, reduced balance of trade, but continuous growth of total foreign trade volume. To achieve this goal, China shall:

1) Facilitate import development through financial policy adjustment. Firstly, import tariff of some commodities shall be adjusted. According to China's current economic development need, provisional tax rates can be adopted to reduce import tariff of energy sources. This can facilitate imports and offset our energy shortage. Additionally, import tariff of rare living goods in China but closely related to people life shall be lowered appropriately. This can improve the living standard in China and increase imports. To improve China's production level and product quality, import tariff of some advanced equipments and parts shall be adjusted appropriately. Key attentions shall be paid to lower import tariff of some products which are needed in primary energy sources and emerging industries but couldn't be produced in China or poor quality of domestic-made ones. China is encouraged to give certain tariff preference to commodities importing from underdeveloped countries. This not only can win their economic and political supports, but also can facilitate import and improve our balance of trade. Tariff reduction to underdeveloped countries shall be accelerated and expanded. Meanwhile, imports from members of free trade zone can be accepted according to practical situations of the free trade zone. Secondly, China can promote imports through financial support. China has set special funds for foreign trade. To promote imports, Chinese government is suggested to increase financial support, provide subsidies on interest payment to encouraged imports, and expand or adjust supports appropriately. Moreover, government shall guide and establish various business platforms with expanded import function, encourage to providing public services (e.g. import communication meeting), and continuously increase support to imports from developing countries.

2) Promote product upgrading and improve export structure. Firstly, China shall propel transformation and upgrading of processing trade. Government shall keep stable policies for processing trade, improve macro environment for technology and industrial transformation, increase confidence of investors, control trade companies with high energy consumption, high pollution but low value added, and guide transformation and derivation of processing toward high-end industries. These can improve China's trade structure comprehensively and help to produce high-quality products. Secondly, China shall establish and perfect domestic trading platform, cultivate enterprises with good reputation and conditions, encourage enterprises to building self-owned brands, and create good domestic marketing channels. Actually, government also can improve export structure and promote product upgrading through tax policies. On one hand, favorable tax treatment shall be offered to high-tech products. On the other hand, additional taxes shall be collected from enterprises with low technology level and value added as well as high pollution risks. Thirdly, China shall attach high attention on service and technology export development. According to experiences of foreign countries, technology and service export can increase value added of exports, protect our resources and prevent drain of our rare resources.

3) Enhance foreign trade management. Firstly, China shall optimize import management, including:

- **a.** eliminating unreasonable restrictions and measures;
- **b.** lowering trade cost, reducing catalogues concerning automatic import licensing commodity management;
- **c.** promoting online application actively;
- **d.** accelerating online verification between electronic data of automatic import licensing and the customs;
- **e.** increasing online verification efficiency, and implementing scientific and effective regulation.

Secondly, China shall impel smooth connection between imports-exports and domestic goods, including:

- i encouraging domestic circulation enterprises to participate in international trade;
- ii supporting qualified enterprise to integrate imports and domestic circulation business;
- iii encouraging domestic commercial enterprises to sell foreign brand articles of consumption as agencies and develop platforms for below-the-line promotional activities to break monopoly and realize full competition;
- iv following standard international practices to perfect relevant laws and regulations, and support development of departure business enjoying tax exemption;
- v increasing import ports to special commodities (e.g. medicine) appropriately and expanding related product export;
- vi accepting imported commodities with inspection and quarantine certificates completely without needing additional test after entered into the Chinese market.

4) Improve infrastructure and improve clearance efficiency. China shall:

- **a.** further improve infrastructures in border ports, check and supervision facilities as well as infrastructures in border economic cooperation zones, establish a modern logistics system integrating cargo transportation, storage and processing, and increase handling capacity of ports;
- **b.** improve supporting facilities in border trade places, fully implement polices about promoting economic and trade development in border

regions, intensify economic and trade contacts with surrounding countries and regions;

c. propel the "Construction Engineering for Improving Work Efficiency of Ports" continuously and accelerate construction of electronic ports;

d. accelerate standardization and electronization of trade documents greatly, facilitate sharing of trade documents and supervision information among different departments, and realize "integrated input but independent declarations" of exports and imports;

e. perfect consulting service platform of technical regulations and conformity assessment of import commodities;

f. make full use of the leading role of local people's government and support to establish local information service platforms.

References

[1] Meng Y 2009 Study on China-ASEAN trade balance trend – an empirical analysis based on ARMA model *Journal of Xi'an University of Finance and Economics* **03** 110-6 *(in Chinese)*

[2] Liu X 2009 The analysis of time series of China's general trade import and export amount based on ARMA model *Statistics and consulting* **04** 20-1 *(in Chinese)*

[3] Shen J 2011 Trade present situation and Prospect of China in Post Financial Crisis Era *Foreign Investment in China* **12** 20 *(in Chinese)*

[4] Zhao J 2011 China's trade surplus prediction and trend analysis based on ARMA model *China Business and Trade* **25** 228-9 *(in Chinese)*

[5] Zhang J, Qi C 2007 A study on the root and the development tendency of Chinese trade surplus *Journal of Finance and Economics* **08** 28-40 *(in Chinese)*

[6] Dai L, Zhang Y 2010 China's tradepolicy options under the financial crisis *Journal of Changsha University of Science and Technology (Social Science)* **02** 32-7 *(in Chinese)*

[7] Yang D 2010 The cause analyses of our country's rapid growth in trade surplus *Future and Development* **10** 39-42 *(in Chinese)*

[8] Wang W 2012 Analysis of measures and trade status in China *Business* **05** 17 *(in Chinese)*

Author

Xue Wu, January 198, Chongqing city, P.R. China.

Current position, grades: lecturer at the Chongqing University of Science and Technology, Chongqing, China.
University studies: PhD student at the National Institute of Development Administration, Thailand, in China.
Scientific interests: the study of international trade and regional economic.
Publications: more than 20 papers.
Experience: teaching experience of 6 years, 5 scientific research projects.

Transmission power control strategy based on partially observable Markov processes for IEEE802.11 WLAN

Yanpeng Feng*, Hongyan Zheng, Bo Wu

Education Technology and Information Center, Liuxian Road, Shenzhen Polytechnic, Shenzhen, China

Received 1 November 2014, www.cmnt.lv

Abstract

With the limitation of independent channels in IEEE802.11 WLAN, co-channel AP interferes with each other in repeat coverage area. As traditional AP launched a fixed power in sending data packages, which ignores the differences and mobility of STA. According to the mobility of STA and the partially observable feature of AP-STA link, this paper analyses the internal relations among the link state, transmission power and channel interference of co-channel AP, and proposes a single-link transmission power control (TPC) algorithm based on the Partially Observable Markov Decision Processes (POMDP). Firstly, the single AP-STA link POMDP-TPC model is constructed, and the neural network learning model is established to describe the observation function of POMDP-TPC. Secondly, the algorithm constructs reachable belief searching trees to obtain the approximate optimization, which implements the dynamic creating and on-line updating of the power consumption policy. Finally, under the experiment environment in OPNET and IEEE802.11b, the results demonstrate the algorithm can reduce the AP power consumption efficiently and improve the network throughput greatly.

Keywords: wireless local area networks, access point, POMDP, transmission power control, anti-interference

1 Introduction

The rapid rise of wireless LAN makes the deployment of wireless access point (AP) become increasingly intensive; however, WLAN has limited non-interfering channels, especially the 2.4GHz band only uses three non-overlapping channels (1, 6 & 11). Same channel used by different APs may cause hidden nodes, making channel interference and transmission conflict become the main issues that limit network throughput [1,2]. Thus, the link transmission power control (TPC) is becoming one of the key technologies to solve this problem [3]. It ensures AP coverage, reduces the channel interference and improves the network throughput.

Accordingly, Javier del Prado Pavon took 802.11b as the object of study and proposed a link-based power control algorithm, which calculated packet loss rate through SNR and BER, increased AP throughput [4] and lowered energy consumption. However, the algorithm involved changes in protocols and physical layer, so that it is difficult to promote in large scale. Daji Qiao took 802.11a/h as the object of study, and proposed an optimal energy transfer policy (miser strategy) [5]. This method can effectively conserve AP power, but it is difficult to handle changes in STA position. Wei Li proposed a PCAP (Power Control for AP Performance enhancement) algorithm [6], which used heuristic method to achieve balance between network throughput and AP proportional fairness. Jing Nie proposed a pccf protocol [7]. RTS transmission power is based on the carrier sensing threshold, reception threshold and the maximum transmission power, and the data transmission power is calculated according to the SINR (signal to interference noise ratio) threshold. The above-mentioned algorithms have effectively promoted the development of TPC algorithm, but still have the following problems:

1) The existing algorithms mainly take AP reference transmission power as the control object, and control AP coverage by adjusting the transmission power of beacon frame to achieve the purpose of reducing interference; however, the differences between STA position and performance are not considered. It is like "talking with many people in a place". If talking with all persons with the same volume, the objects far away or with weak hearing could not hear clearly, while the objects in near place or with good hearing will feel deafening.

2) The existing algorithms lack consideration of uncertainty in wireless environment. Most algorithms assume that the environment is known and unchanging, but the transmission environment has a great deal of uncertainty in practice due to multiple-operator deployment and illegal AP structure.

To solve the above-mentioned problems, this paper introduces the thinking of probability theory, proposes Partly Observable Markov [8] Decision Processing – Transmission Power Control (POMDP-TPC), takes single-link data packet transmission power as the control object and builds POMDP six-tuple model to analyze the correlation of the strength, SNR, BER and link status of received signals, establish the observation function learning model based on neural network and use reachable belief state online search to achieve dynamic planning and real-time decision-making of strategies, and achieve the

* ***Corresponding author's*** e-mail: ypfeng@szpt.edu.cn

purposes of reducing AP energy consumption and improving network throughput.

2 TPC model based on partly observable Markov

2.1 SYSTEM MODEL

WLAN operating frequencies are 2.4GHz and 5GHz. At present, most operators only support 2.4GHz. For large WLAN network of high-density, although the channels are arranged alternately, the same channel still needs to be used by different APs, and an AP of the same channel can't detect the AP of the opposite end; however, STA in overlapping coverage areas will encounter the risk of conflict (i.e. hidden nodes) in the process of data transmission and reception. Therefore, the link transmission power should be controlled to ensure AP coverage, reduce channel interference and improve the network throughput.

In this paper, co-channel AP interference is the object of study, and the following assumptions are made:

1) AP physical layout and channel planning are basically reasonable, and there is no complete coverage or blind zone;

2) STA has been connected to AP, and this algorithm only controls data transmission power.

If AP and STA can exchange data at a high rate after connecting, they have better wireless channel. Under the premise of high-speed transmission, AP should gradually reduce the transmission power of STA to reduce energy consumption and reduce the interference range of other transmission links. Conversely, if the wireless channel becomes worse and results in speed reduction or packet loss, AP should increase its transmission power. Therefore, the power control algorithm should be specific to each AP-STA link, adjust transmission power in real-time according to the link status and achieve the optimal balance between transfer rate and transmission power.

2.2 SYSTEM FRAMEWORK

The wireless transmission environment has randomness and uncertainty. On the one hand, there are differences between different STAs, including the location of AP from STA and STA sensitivity; on the other hand, the same STA has mobility. Therefore, the transmission power control requires link status observation and assessment for each AP-STA link and then selecting the optimal power adjustment strategy. Its essence is the optimal decision in a random environment. POMDP is an ideal mathematical model to solve such problems. The structure of single link POMDP-TPC is shown in Figure 1.

POMDP allows describing $< S, A, T, R, Z, O >$ with a six-tuple. *S*, *A* and *E* represent state set, action set and observation set respectively; $T: S \times A \to \prod(S)$ is state transition function, which represents the probability of executing a under the state s to state s', recorded as $T(s, a, s')$; $R: S \times A \to R$ represents the strategy evaluation function; $O: S \times A \to \prod(Z)$ is an observation function, which represents the probability of getting observation z when executing a to state s', recorded as $O(s', a, z)$. Evaluate the current link status through six-tuple to get the belief state b. Due to the partial observability and randomness of the link, belief state is actually a probability distribution of state set. The belief state search and value iteration method are used to obtain the optimal strategy, i.e. the optimal transmission power of single link.

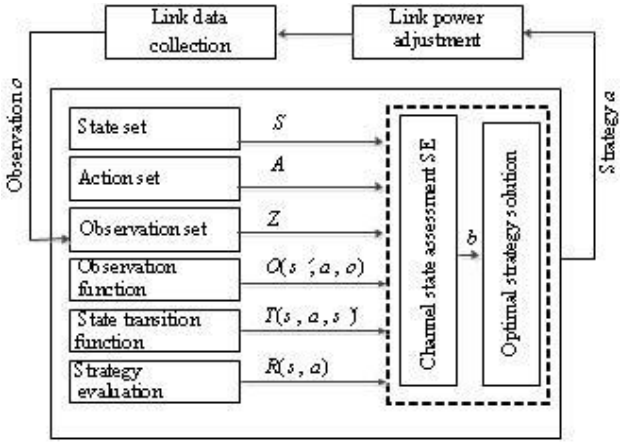


FIGURE 1 POMDP-TPC system framework

POMDP-TPC controller is the core of the system. It takes single AP-STA link as the object. To achieve self-adaption of AP to STA in the data transmission process, a POMDP decision model should be established for each link to obtain the optimal transmission power of the link in real-time and achieve high throughput and low power consumption of the AP.

3 Establishment of POMDP-TPC model

3.1 DEFINITION OF STATE SET, ACTION SET AND OBSERVATION SET

In POMDP-TPC model, the settings of state set, action set and observation set directly affect the effectiveness of control strategies and the complexity of strategy solution. In this paper, the state set, action set and observation set are set as follows through a large number of experiments and correlation analysis:

1) Single AP-STA link status $S_l = S_{ack} \times S_p \times S_{sta}$. S_{ack} indicates whether the message is sent successfully. In IEEE802.11, each packet transmission requires ACK response, and therefore, whether correct ACK packet is received indicates whether the message is sent successfully. S_p indicates the transmission power of the link, $s_p = p_{link} \cdot S_{sta}$ includes STA state, which includes two factors: the distance of the region from AP and whether there is interference in the region. Therefore, the AP transmission power is increased in increments of 0.5 dBm, and its coverage is divided into m regions. Assume that

the maximum transmission power and minimum transmission power of AP are $P_{\max}$ and $P_{\min}$ respectively, then $m = (P_{\max} - P_{\min})/0.5$. In an unknown environment, each region has the possibilities of interference and non-interference, and therefore AP coverage area contains 2m cases, and STA state can be expressed by a two-tuple: $\{(s_{sta\_area}, s_{sta\_inter}) \mid s_{sta\_area} \in (1,m), s_{sta\_inter} \in (0,1)\}$, of which s_{sta_area} indicates the region that STA locates and s_{sta_inter} indicates whether there is interference in this region.

2) The action set A represents a set of power adjustment actions that can be executed by AP. Theoretically, the power adjustment may be any value within the limited range. In fact, AP power adjustment needs to be able to quickly respond to the real time changes in the wireless environment, and should meet the adaptation of opposite end STA to the changes in transmission power. Therefore, the adjustment of transmission power is set to four levels: $\{0.5dBm, 1dBm, 2dBm, 4dBm\}$, so that the power is adjusted to variable step size, which can respond quickly to changes in the environment and can be tuned according to the link status. Depending on the STA, AP transmission power can be increased, reduced or remained unchanged.

3) Observation set Z represents the set of link parameters that can be obtained by AP. STA varies in sensitivity and position and is variable in real time, S_{link} can't be obtained directly, and thus should be inferred by measurable link parameters. The selected parameters should reflect the environment of the reaction medium and the link quality and analyze the association of each parameter and the link state. The calculation is as shown in Equation (1).

$$Correl(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2}}; i, j = 1,2,...,10, \quad (1)$$

where *n* represents the number of samples, $\bar{x}$ and $\bar{y}$ represent the sample mean.

By analyzing the correlation of wireless communication data in January of a district of Shenzhen, Guangdong Province, the strength *q*, signal-to-noise ratio (SNR), bit error rate *b* and transmission rate (DTR) of the signals of the correlation coefficient greater than 0.7 are selected as the members of observation set.

3.2 STATE TRANSITION FUNCTION

For each AP-STA link, the state transition function represents the changes of link status in two adjacent time slices. Status *s* includes s_{ack}, s_p and s_l, which are independent of the conditions, and therefore the state transition function can be written in the form of conditional probability:

$$T(s,a,s') = Pr(s'_{ack}, s'_p, s'_l \mid a, s) = Pr(s'_{ack} \mid a, s) Pr(s'_p \mid a, s) Pr(s'_l \mid a, s), \quad (2)$$

s_l only associates with the transmission medium and STA position, and doesn't change with the AP transmission power, so that the Equation (2) can be rewritten as:

$$T(s,a,s') = \begin{cases} 0 & s'_l \neq s_l \\ 0 & s'_p \neq s_p + \Delta p \\ Pr(s'_{ack} \mid a, s_{ack}, s_l, s_p) & \text{others} \end{cases}, \quad (3)$$

$Pr(s'_{ack} \mid a, s_{ack}, s_l, s_p)$ is obtained from historical data.

3.3 OBSERVATION MODEL

Under different link status, executing different power adjustment strategies will obtain different observations, namely to observe the model in different states, the possibility of obtaining the corresponding observation under different statuses and policies, which is expressed as follows with probability:

$$O(s',a,o) = Pr(o \mid s', a). \quad (4)$$

It is inferred from Bayesian rule that:

$$Pr(o \mid s', a) = \frac{Pr(s' \mid o, a) Pr(o \mid a)}{Pr(s' \mid a)} = \frac{Pr(s' \mid o, a) Pr(o \mid a)}{Pr(s' \mid a, s)} = \frac{Pr(s' \mid o, a) Pr(o \mid a)}{T(s,a,s')}, \quad (5)$$

where the denominator is the state transition function, which is obtained from Equation (3); in the numerator, $Pr(o \mid a)$ can be regarded as a normalization factor; $Pr(s' \mid o, a)$ represents the probability to obtain status s_{ack} by executing policy a under current observation. As s_{ack}, s_p and s_l are independent of the conditions, so that:

$$Pr(s' \mid o, a) = Pr(s'_{ack} \mid o, a) Pr(s'_p \mid o, a) Pr(s'_l \mid o, a). \quad (6)$$

Obtain $Pr(s'_{ack} \mid o, a)$, $Pr(s'_p \mid o, a)$ and $Pr(s'_l \mid o, a)$ as follows:

1) For $Pr(s'_{ack} \mid o, a)$, whether each data packet is transmitted successfully is determined by whether ACK packet is received. Therefore, the probability of successful transmission of a single packet = probability of data transmission failure * probability of ACK packet transmission failure, as shown below:

$$Pr(s'_{ack}=1|o,a)=(1-Pr(e\_data|o,a))\times(1-Pr(e\_ack|o,a)) \quad (7)$$

The study of the [3] shows that the probability of packet transmission failure can be calculated through BER, SNR and transmission rate:

$$Pr(e\_data|o,a)=1-(1-Pr(b|snr,dtr))^{8L}, \quad (8)$$

where L indicates the packet length. The length of ACK packet is fixed at 14bytes, so

$$Pr(e\_ack|o,a)=1-(1-Pr(b|snr,dtr))^{112}. \quad (9)$$

2) For $Pr(s'_p|o,a)$, assume that the AP power transmitting module always can specify the transmission power, that is, $Pr(s'_p|o,a)$ is independent of the observation and strategies. For all observations, $Pr(s'_p|o,a)=1$.

3) For $Pr(s'_l|o,a)$, link status is strongly correlated to signal strength, SNR, bit error rate and transmission rate strongly correlated. However, the parameters have complex nonlinear relationship, and it is difficult to derive through a simple mathematical model. In this paper, we propose a link status inference model based on BP neural networks, and obtain s' of corresponding observation and strategies through network training and inference.

With three layers BPNN structure, the neurons of input layer and output layer are 4 and 2 respectively. Through the correlation analysis in section 3.1, the link status neural network model inputs are determined as follows: signal strength *q*, SNR, DTR and transmission power *p*. The model output is link status s_l. The number of neurons in the hidden layer calculated in accordance with the empirical equation is identified as 10, and the inference model neural network structure:

$$s_l=\sum_{i=1}^{10} w_{i,o}^{ack}\,\text{tansig}(w_{i,1}^{ack}q+w_{i,2}^{ack}snr+w_{i,3}^{ack}lt+w_{i,4}^{ack}dtr), \quad (10)$$

$w_{i,o}^{ack}$ is the corresponding weight of the *i*-th neuron of the hidden layer to the output neuron, $w_{i,j}^{ack}$ is the corresponding weight of the *j*-th neuron of the input layer to the *i*-th neuron of the hidden layer, and tansig(·) nonlinear Sigmoid activation function is used.

3.4 POLICY-RETURN FUNCTION

If the current link is free of interference and the packet transmission is successful in the process of AP packet transmission, the transmission power should be reduced to the power of STA minimum coverage as soon as possible, which can not only ensure successful data transmission, but also reduce energy consumption of AP; conversely, if the current link is free of interference but the packet isn't sent successfully, the transmission power should be increased to be equal to or slightly greater than the minimum coverage power of STA as soon as possible to increase the probability of success and avoid repeated failures and retransmission:

$$R(s,a)=\begin{cases}\dfrac{1}{|s_p+\Delta p-s_{l\_area}*0.5|} & s_{l\_inter}=0, s_{ack}=1\\ \dfrac{1}{|s_p+\Delta p-s_{l\_area}*0.5-\eta|} & s_{l\_inter}=0, s_{ack}=0\\ \dfrac{1}{|\Delta p|} & s_{l\_inter}=1, s_{ack}=1\\ \dfrac{1}{|s_p+\Delta p-s_{l\_area}*0.5-\mu|} & s_{l\_inter}=1, s_{ack}=0\end{cases}, \quad (11)$$

where Δp is the amount of power change, η and μ are anti-interference margin set according to experience. In particular, when the denominator is zero, $R(s,a)=2(1/0.5)=4$.

4 Optimal control strategy solution of single link POMDP-TPC

4.1 ALGORITHM PROCESS

Traditional Markov solution algorithms usually use offline planning. With the increase in the number of iterations, the complexity of algorithm time and space expands rapidly, which is a typical NP-hard problem. To ensure real-time and dynamic adaptability of TPC algorithm, online algorithm [9] is used to divide the entire strategic planning and policy execution into several small plans and executions, start from the current belief state to build reachable belief status search tree, control the spatial scale of reachable belief state and achieve fast solution, as shown in Figure 2.

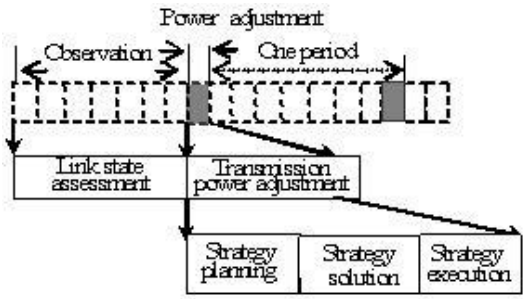


FIGURE 2 Online Solution of AP Link Power Control Algorithm

During system operation, the entire process is run at a fixed period, each period is divided into several small time slices, the former time slices are the observation phase for data collection, filtering and link status assessment, and the power calculation and adjustment are carried out in the last time slice, including strategic planning, strategy solution and strategy execution. After the power is adjusted, repeat observation, assessment of link status and adjustment of

transmission power in the new cycle, and thus ensure that the link transmission power can dynamically adapt to the changes in STA and the environment.

4.2 OPTIMAL STRATEGY SOLUTION

The TPC algorithm of the entire link can be divided into off-line training, online strategy planning, online strategy solution, and strategy execution. The objective of offline training is to obtain the corresponding weight of each neuron of the neural network in gradient descent method. The steps of online strategy planning and online strategy solution are as shown in Table 1.

TABLE 1 TPC online planning and solution algorithm

Algorithm: TPC online planning and solution algorithm
Definition: b_c , current belief state; *T*, belief state search tree;
//Online strategic planning
1): Obtain the current belief state b_t according to observation;
2): Build belief state search tree T with current belief point b_t as the root;
// Online strategy solution 3): Calculate the iterative return value function of belief point in *T* through edge node upward propagation algorithm;
4): Obtain the optimal power control strategy a^* under b_t ;
// Online strategy execution 5): Adjust the packet transmission power according to the power control strategy;
6): Collect the link data, and update belief state b_c according to a^* and *z*;
7): Return to 1) and reconstruct belief state search tree;

Here the belief state b is an assessment by observing the current link status. Due to the partial observability and uncertainty of link, it is the probability distribution on a real state set, determined by the action and observation at the moment of 0-t and described:

$$b_t = Pr(s_t \mid a_t, z_t, a_{t-1}, z_{t-1}, ..., a_0, z_0, s_0). \quad (12)$$

To effectively prevent strategy jitter, it is necessary to consider the long-term effects of the current strategy, that is, to construct the reachable belief state search tree according to the current belief state. The root node of the search tree is the current belief state, and the leaf node is the reachable belief state point after one or several steps of power adjustment. Assume that $b(s')$ is the new belief state point, and $b(s')$ can be obtained from the father belief point, observation function and state transition function:

$$b_t(s') = \frac{\sum_{s' \in S} O(s', a^{t-1}, z^t) T(s, a^{t-1}, s') b^{t-1}(s)}{P(z^t \mid b^{t-1}, a^{t-1})}. \quad (13)$$

Assume that the depth of search tree is *D*, and obtain to reachable belief state search tree through iteration, which indicates all possible intermediate states and end states starting from the current belief state after power adjustment for *D* times.

Convert partly observable Markov strategy solution process into belief state based Markov solution and strategy evaluation function $\rho(b,a) = \sum_{s \in S} b(s) R(s,a)$. Calculate the function *V* with Bellman equation; as shown in Equation (15), γ is the discount factor.

$$V^{t+1}(b) = \max_{a \in A} \left[\sum_{s \in S} b(s) R(s,a) + \gamma \sum_{z \in Z} P(z \mid b,a) V^t(b') \right], \quad (14)$$

$$\pi^{t+1}(b) = \arg\max_{a \in A} \left[\sum_{s \in S} b(s) R(s,a) + \gamma \sum_{z \in Z} P(z \mid b,a) V^t(b') \right]. \quad (15)$$

Effectively estimate the long-term impact of different power control strategies on future link status through constructing belief state search tree and value iteration, select the strategy π with the maximum cumulative return value as the optimal strategy to achieve the balance among AP energy consumption, minimum interference and packet transmission success rate, and avoid jitter of control strategies effectively.

5 Experiments and results analysis

Use OPNET simulation platform to build IEEE802.11b wireless network test environment. The simulation parameters are shown in Table 2. FTP traffic flow of normally distributed packet length is used to simulate the STA business environment of each site.

TABLE 2 Parameters of simulation environment

Parameter	Value
Region size	300m*300m
AP number	Four, evenly distributed arrays
STA number	20, randomly and uniformly distributed
AP channel	1
P_{max}	27dBm
P_{min}	5dBm
P_{beacon}	21dBm

The network is shown in Figure 3. Four APs are recorded as AP1-AP4, and 20 STAs are randomly distributed, of which 9 can detect SSID of two or more APs. All STAs establish a connection with AP1, AP2, AP3 and AP4 respectively, of which AP1 has the heaviest load and AP4 is relatively light.

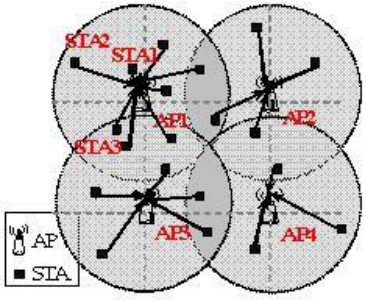


FIGURE 3 Experimental environment network structure

5.1 AP THROUGHPUT AND ENERGY CONSUMPTION

Keep the STA position, and transmission power unchanged and carry out the experiments in three groups: none AP runs this algorithm, AP1 and AP2 run this algorithm, and all APs run this algorithm. Repeat each group of experiments 10 times and evaluate the average. Run each experiment 30 minutes, and compare the average throughput and average energy consumption of AP1, AP2, AP3 and AP4, as shown in Figures 4a and 4b.

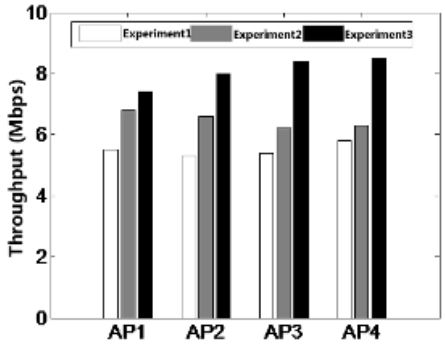


FIGURE 4a Throughput comparison chart

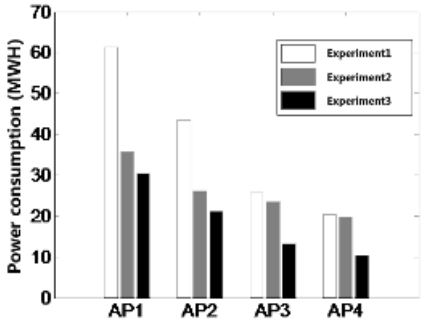


FIGURE 4b Power consumption comparison

Figure 4a shows that the throughputs of four APs have increased when AP1 and AP2 run this algorithm, and AP1 and AP2 are superior to AP3 and AP4, which fully shows that the algorithm can effectively transmit conflict. Although AP3 and AP4 didn't run this algorithm, the throughput was significantly improved because the transmission conflict in repeated coverage region was reduced, and thus the throughput of the entire network was improved effectively; when none AP run the algorithm, all the throughputs were significantly improved, but AP throughputs were the smallest, indicating that although the proposed algorithm can reduce the transmission conflict to some extent, it can't eliminate the conflict. AP1 had the maximum load, so that the possibility of transmission conflict was still greater than other APs, and therefore the throughput was the smallest. Figure 4b well verifies the effectiveness of the algorithm in energy saving. The power consumption of AP is reduced significantly because the algorithm enables AP transmitting data in appropriate power in accordance with the link state.

5.2 DYNAMIC ADAPT ABILITY

Move STA1 position away from 1m from AP1, and move STA2 position close gradually at a distance 80m from AP1. Keep the speed at 5m/s, STA3 position unchanged, and test the transmission power and packet loss rate of these three AP-STA links.

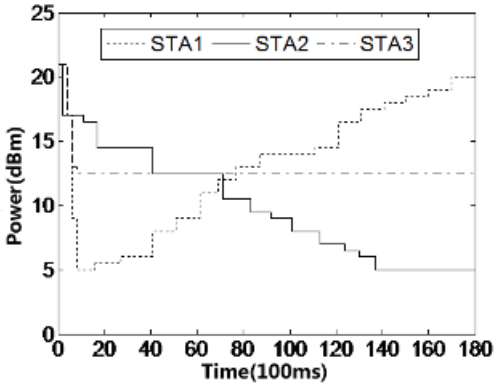


FIGURE 5a Transmission power changes with time

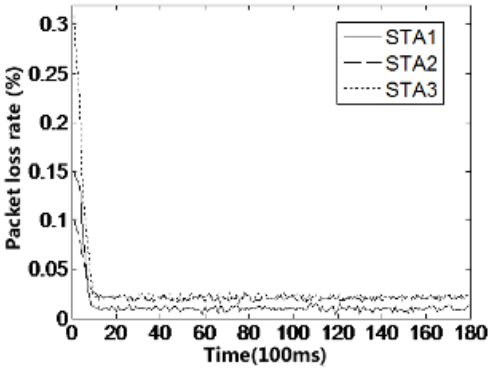


FIGURE 5b Packet loss rate changes with time

Figures 5a and 5b show the relationship of the changes in AP transmission power and packet loss rate. It is known that the adjustment frequency of transmission power is faster in the initial 10 control periods, and the packet loss rate also decreases rapidly. If STA position is unchanged, the transmission power tends to be stabilized and the packet loss rate is reduced to a minimum in the 10^{th} control period; if STA position changes, AP transmission power gradually increases or decreases with the change of STA position, and the packet loss rate can be maintained at a lower level, which fully illustrates that POMDP-TPC has excellent dynamic adaptability.

6 Conclusion

A single link transmission power control method is proposed according to the characteristics that STA position and transmission environment have differences in the control of transmission power. By introducing partly observable Markov theory, establish six-tuple model of single AP-STA link TPC optimal control, construct the

observation function and state transition function of link state, obtain the optimal strategy through belief state search and value iteration, achieve maximized expected value of reward function and achieve the objective of reducing power consumption and co-channel interference. The results of OPNET based simulation experiment show that the algorithm can better optimize AP energy consumption and throughput online, and reduce the interference between co-channel APs effectively. At present, the strategy reward function is set according to expert experience, and machine learning method will be studied to optimize the reward strategy function and correct experience setting.

References

[1] Chen J Y, Olafsson S, Yang Y, Gu X Y 2009 Joint Distributed Transmit Power Control and Dynamic Channel Allocation for Scalable WLANs *Wireless Communications and Networking Conference* London 1-6

[2] Douros G D, Polyzos G C2011 Review of some fundamental approaches for power control in wireless networks *Computer Communications* **34**(13) 1580-92

[3] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81

[4] del Prado Pavon J, Choi S 2003 Link Adaptation Strategy for IEEE 802.11 WLAN via Received Signal Strength Measurement *IEEE International Conference on Communications 2003 ICC '03* 1108-13

[5] Zhang C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34

[6] Li W, Cui Y, Cheng X, Al-Rodhaan M A 2011 Achieving Proportional Fairness via AP Power Control in Multi-Rate WLANs. *IEEE Transactions on Wireless Communications* **10**(11) 3784-92

[7] Nie J, Parr G, McClean S 2011 Power control of control frames in IEEE 802.11 networks *AEU – International Journal of Electronics and Communications* **65**(3) 165-72

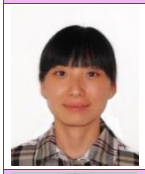
[8] Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103

[9] Ross S, Pineau J, Paquet S, Chaib-draa B 2008 Online Planning Algorithms for POMDPs. *Journal of Artificial Intelligence Research* **32** 663-704

Authors

Yanpeng Feng, born in December 1980, Shenzhen City, Guangdong Province, P.R. China.

Current position, grades: associate research fellow of Education Technology and Information Center, Shenzhen Polytechnic, China.
University studies: BSc in Computer Science and Technology at University of Science and Technology China. MSc at University of Science & Technology China.
Scientific interests: intelligent decision, wireless network and big data.
Publications: more than 20 papers.
Experience: teaching experience of 9 years, completed 5 scientific research projects.



Hongyan Zheng, born in January 1983, Shenzhen City, Guangdong Province, P.R. China.

Current position, grades: Senior Engineer of Education Technology and Information Center, Shenzhen Polytechnic, China.
University studies: BSc in Electrical information engineering at Central South University in China. MSc at Central South University in China.
Scientific interests: wireless network, fuzzy control.
Publications: more than 10 papers.
Experience: teaching experience of 4 years, 2 scientific research projects.

Bo Wu, born in December 1979, Shenzhen City, Guangdong Province, P.R. China.

Current position, grades: Associate Professor of Education Technology and Information Center, Shenzhen Polytechnic, China.
University studies: BSc in Computer Science at Central South University in China. MSc and PhD at Central South University in China.
Scientific interests: decision-marking, machine learning and big data.
Publications: more than 40 papers published in various journals.
Experience: teaching experience of 10 years, completed 7 scientific research projects.

An analysis framework for building commodity futures market simulation model based on heterogeneous traders

Tan Li, Weiwei Wu*

[1]*School of Management, Harbin Institute of Technology, 13 Fa Yuan Street, Harbin, China*

Received 6 June 2014, www.cmnt.lv

Abstract

Futures price volatility is always the hot topic for academic researchers and traders in futures market. For exploring the rules of futures price fluctuation, we try to develop a new analysing framework from the angle of the heterogeneous traders. We describe heterogeneous traders as four respects: trading motive, predict styles, risk-return preference, reaction speed, which can accurately describe the heterogeneous traders in futures market. According to the categories of heterogeneous traders, we construct traders' internal model, which is used in a commodity futures market simulation model to verify the framework, and the results show the framework is useful.

Keywords: bounded rationality, commodity futures markets, simulation model, heterogeneous traders

1 Introduction

The rules of futures prices movement has been the hot topic that academic researchers and traders have been interested in all the time. The deep understanding of this question not merely facilitates the futures price forecasts, but also favourable to the formulation of the futures market supervision rule, guarantee the market runs effectively. About the rules of futures prices volatility academic writing and actual operators have totally different understanding. In futures market, it is numerous traders that do a deal according to individual decision, forms market price. Studying actual operators' behaviour in futures market is a feasible route to understand price dynamic.

We make analysis to trader's behaviours in the market, and form a dynamic framework to analyse futures market price volatility. Then this framework is used for building a commodity futures market simulation model. The simulation results can be adopted to verify the validity of this analysis framework.

2 Review of commodity futures price theory

There are three main kinds of theoretical research of futures price volatility: (i) all these kinds of research beginning with 1930 of Keynes, he regarded the futures price as the function of the storage cost. The central idea of theory of price of storage is to analyse the composition of the futures price. Inherit the Keynes futures price theory, Working and Hicks developed the cost-of-carry theory to explain the relationship between spot price and futures price. (ii) the second kind of theoretical research mainly relate to core of the Capital Asset Pricing Model, risk premium problem, and think hedgers group in commodity futures market must pay risk premium to speculators group to make them enter position on the contrary to total hedgers group. (iii) Because of remarkable development in econometrics in recent years, almost each new technology in econometrics is rapidly applied to research price fluctuates in various financial markets. A lot of econometrics models are developed to analyse the relationship between spot price and futures price. The results through these models generally show the lead-lag relationship and the long-term forecast result is better, but short-term forecast is relatively bad.

The above (i) and (ii) items can partly explain futures price phenomenon, such as backwardation and contango. Many econometrics models which describe the spot price and futures price relationship can be used to predict the futures price, though the short-time forecast result is not satisfied. But the intrinsic character of the futures price dynamic is still not clear. The key problem is hypothesis of rationality. Many researchers generally used the hypothesis that traders in the market are completely rational and homogeneous. Since the presence of Simon's bounded rationality theory, this hypothesis is constantly questioned by many scholars.

Since 1980s', two representative schools' opinion about bounded rationality have been appeared. First is financial behaviour. The representatives are Robert Shiller and Daniel Kahneman. They have inherited Simon's bounded rationality theory partly. Based on the results got from psychological and human subjective experiments, they studied the relationship between traders' bounded rational behaviour and price dynamics. They developed theories including prospect theory, noise trading model, etc. Behaviour finance theory broke the traditional assumption about rational traders which was dominated in academic cycles in the past, and studied trader's decision under the uncertain conditions, and offered a feasible road for exploring the intrinsic characters of price dynamics in the financial market. But behavioural finance cannot build a unified

* ***Corresponding author's*** e-mail: hit_wu@126.com

frame to explain all the price anonyms appeared in the financial market. It is an obstacle to further develops for behavioural finance. Actual traders always adopt fundamental or technical analysis as their forecasting approaches. In the noise trading model, the traders of technical analysis generally are viewed as noise traders or call as speculators. Noise is defined as a process leading to shadow information, i.e. market mechanisms which blur observations of the market [1]. Noise traders, however, do not transcend the categories of fundamentalists and chartists. Recently, some scholars adopted the questionnaire among the traders in foreign exchange market, and find the majority of traders adopt the technical analysis method for short-term forecasting. As the extension of time horizon changes, these traders take fundamental analysis as the main method gradually.

Other scholars take the method of experimental economics, simulate financial market through the computer program, study traders' strategies and the influence to financial market price, and call it agent-based computational finance. Santa Fe of USA is famous in this field, their artificial financial market can really recur the statistics characteristic of the data on real financial markets. In their artificial financial market model, traders are heterogeneous. Agent-based computational finance so is a feasible road to study the financial market price dynamic characteristic. Studies of Behaviour finance and agent-based computational finance have one important common characteristic to think traders of financial market are bounded and heterogeneous. But in agent-based computational finance field, the empirical study on heterogeneous traders is still lack. In this field, traders are classified as fundamental and chartist / technical ones according to perception, not to first-hand data. Several results from investigations about heterogeneous expectations are not analysed systematically. Just like what Goodhart said, Economists cannot just rely on assumptions and hypotheses about how speculators and other market agents may operate in theory, but should examine how they work in practice, by first-hand study of such market [2]. Some researchers used questionnaire and interview to study foreign exchange market traders' expectation. The further work is to use these methods for the research of the commodity futures markets, and use the results of investigations to construct artificial futures markets simulation model. The following part of the paper is arranged like this, section two will analyse the meaning of trader's heterogeneity, and evidence for it. Section four will describe all kinds of traders' heterogeneity in the market carefully, and construct the internal model of the traders used in the artificial futures markets simulation model. Section five analyses the futures market price dynamic based on the hypothesis of heterogeneous traders. In the last section, summarize the futures market price dynamic analysis frame of heterogeneous traders and propose the future research work.

3 Definition of heterogeneity of traders and evidences

3.1 DEFINITION OF HETEROGENEOUS TRADERS

One of the reasons to explore trader's heterogeneity is to explain anomies in financial market. These anomies generally include aggregational Gaussianity, clustered volatility, fat-tailed return distributions, and long-term memory. Before define traders' heterogeneity, we will first discuss traders' rationality in financial market.

The bounded rationality theory first appears in Simon's "Management Behaviour". Simon thought, in broad sense, "rationality" refers to a behavioural manner: (i) Suitable for realizing the goal; (ii) And within the designated constraints. Economists use the word "rationality" to describe the attribution of action plans, not of decision process [3]. Kahneman and Tversky thought that people pay more attention to the amount of changes in their wealth, instead of their wealth final quantity. People incline to take the risk of gambling while facing the losses with suitable terms, and while facing the profit with suitable condition, incline to accept the determinacy and make profits. The happiness of making profits brings less than the agony that equal amount of losses. Yang Xiao-Kai summarizes Simon's bounded rationality theory to 3 main points: non-complete information cost of process information and some non-traditional goal function of decision makers. But Yang thought the 3 points of Simon' bounded rationality mentioned did not touch the nature of bounded rationality conception. Knight pointed out the foundation of bounded rationality lay in fundamental uncertainty, and this cannot be equal to non-complete information.

In this paper, traders' bounded rationality is the description of the decision-making process attribute. Traders' bounded rationality means two respects: (i) What traders faced is a complicated, uncertain market environment, this kind of uncertainty includes endogenous and exogenous uncertainty, endogenous uncertainty is caused by numerous trader's decision, endogenous uncertainty still exists even in such situation that not exogenous uncertainty; Exogenous uncertainty causes the uncertainty of the external factor of the financial market of various influence. So traders do not have complete information, have information asymmetry; (ii) Trader's computing capability and cognitive ability to the environment are limited, this one that includes to understanding that has already acquired knowledge and computing capability is limited.

Financial economics and new classical economics usually view individual traders as homogeneity, there are common risk preferences and beliefs between traders, or further assume traders have common knowledge. Under market equilibrium frame, several researchers study non-homogeneous traders' influence to market equilibrium. These researchers described heterogeneity of traders as heterogeneous beliefs or heterogeneous risk preferences, or both of them. According behavioural decision theory, a robust discovery is that the individual traders are heterogeneous in beliefs and risk-preferences [4-7].

The traders in the real financial market are bounded rational, and have different risk preferences, beliefs and information sets. According to the above-mentioned analysis, In this paper, trader heterogeneity of futures market refers to, the bounded rational traders, because the individual's risk preference, computing capability, information quantity and quality hold, difference level of affections by other traders, form different beliefs on futures price.

3.2 EVIDENCES FOR HETEROGENEITY OF TRADERS

Tests for heterogeneous traders mainly concentrate on the test for heterogeneous expectations in the futures market, and generally have two kinds of methods. One is "indirect test", while being so-called, because the method to estimate, receive through parameter trader have among the distribution or market that heterogeneity it is expected on the basis of constructing all kinds of and expecting the model heterogeneously Only allow reason to expect the proportion that traders account for verifies the existence of anticipated heterogeneity. Baak added bounded rational traders in the linear programming model that the standard reason of a financial market expects. In the model, bounded rational traders think the price obeys the AR(1) process. Finding market trend and traders' anticipated form in the market have close relations [8]. According to the calculation, when bounded rational traders' proportion "n" value exceeds a certain critical value in the market, the price can't converge to the equilibrium state. Frechette has adopted "indirect test" to examine heterogeneous expectations too, use the time series data of the commodity futures to test method validity [9]. One common point of the method is to divide traders into two or three groups, including fundamental and technical analysis traders, or rational expectation traders and bounded rational expectation traders, or is divided into naive traders, quasi-rational expectation traders and rational trader according to forecasting methods [10,11]. In addition, Hommes and Sorger assumed bounded rational traders form expectation according to the observed value of price time series, and this is similar to the traders who adopt the technical analysis to forecast in real market. Results of these research support market have heterogeneous expectations, but this kind of traders' categorized method is according to the general consciousness, not empirical investigation and statistical analysis, does not have convincingness, the accuracy of the result is queried.

Another test method for heterogeneous expectations can call "direct test". Oberlechner adopted questionnaire and interview method, studied the forecasting approaches used by European exchange market practical traders and financial journalists. His results confirm that most traders use both forecasting approaches (fundamental and technical analysis method), and that the shorter the forecasting horizon, the more important chartist/technical analysis is. In the past dozens of years, the importance of technical analysis in the foreign exchange market rises remarkably [12]. In addition, Lui studied the Hong Kong foreign exchange market dealers, and found technical analysis is considered slightly more useful in forecasting trends than fundamental analysis, but significantly more useful in predicting turning points and move average or other trend follow system is the most useful technical analysis method [13]. Cheung investigated traders' views on determination of exchange rate [14]. Cheung found a large part of traders thought that purchasing power parity (PPP) can be used to measure an foreign exchange rate's the fundamental value, however a much smaller percentage of traders would trade in such a way as to move exchange rates closer to PPP levels. It means the fundamental value criterion of exchange rate that traders admit is inconsistent with their standards of estimating adopted at the time of real trade. Cheung mentions, he has experienced the divergence between academic writing and traders' views. The above researchers' questionnaire investigation result s are enough to prove traders are not homogeneous in the foreign exchange market, including forecasting approaches at different forecasting horizon, the admitted factors of fundamental values, traders' reaction speed, traders' motivations, etc.. The difference between traders is so huge, and homogeneity and rationality are unable to describe all traders. Have not found that carry on the studying of this kind of investtigation to commodity futures markets traders, according to the materials collected at present, but believe that will get the similar result.

4 Developing heterogeneous traders' internal model

According to above-mentioned researches, there are sufficient evidences to support traders' heterogeneity, how to accurately describe traders' heterogeneity in constructing the artificial market simulation model is very important. De la Maza and Yuret first researches artificial futures market price dynamic, put forward information set, algorithm set, model set, constraint set to describe heterogeneity, but they fixed algorithm set and model set in the model [15]. This kind of 4 sets of de la Maza and Yuret describes constructs a rough frame for trader's internal model. According to the reasons, describe and affection for traders' heterogeneous, this paper constructs heterogeneous traders' internal model.

4.1 REASONS FOR HETEROGENEITY OF TRADERS

Researchers hold different views on what lead to trader's heterogeneity. Chavas' opinion was there were fundamental analysis traders (rational) and technical analysis traders (bounded rational) and different traders received and different experience caused trader's difference in knowledge [16]. Frechette attributed traders' heterogeneity to different traders' priori beliefs and filtration way in information, regarded individual's expectations difference as the expression of trader's heterogeneity. Oberlechner found trader's different prediction styles were related to individual background (such as age, gender, professional rank, years of work experience), but have no relations with trading locations. Author thought this because of in will obtain more information and training, etc. close to global transaction place of banking centre. We divide reasons into the internal factors and external factors. Internal factor refers to trader's education and training experience, individual risk-preference, predict style, etc. which are determined by individual characteristic. External factors refer to the differences between individual caused by individual trading motivation. According to the differences, we can divide traders group into several sub-groups.

According to the external factors, traders can be divided into hedgers, risk management traders, speculators.

1) Hedgers generally hold stock for production, or themselves are stock manufacturers. The aim of hedgers is just to hedge the position of the stock using futures contracts. In theory, hedgers will not adjustment position frequently, so they might not intraday traders. Hedgers including every industry's manufacturers, such as Jiangxi copper industry group in China.

2) Risk management traders refer to futures broker who engage agency business and clear their position in a day for risk avoidance need, they deal in order to guarantee their own position minimum risk in market. Three domestic futures exchanges (not including Hong Kong and Taiwan areas) implement the member system at present. There are two kinds of members, self-operation member, managers, and no comprehensive type. Therefore, in our country at present there is no market maker whose purpose is to minimize the risk by adjusting intraday position.

3) Speculators refer to who do not need to hold the stock for production, may hold stock for speculation, trading purely with the purpose of making a profit in speculation among futures market. This kind of traders can adjust their position for maximizing profit at any time. All hedge funds which invest in derivative markets belong to this kind of traders. Some domestic funds raised privately belong to this group at present.

According to the internal factors, individual traders can be distinguished as hedgers, risk management traders, speculators. Each trader among his group, because internal factor cause individual difference trade behaviour of the course reflect according to heterogeneous traders' behaviour.

4.2 REPRESENTATION OF HETEROGENEITY OF TRADERS

Sager and Taylor researched on foreign exchange market and got the results that heterogeneity of traders included information asymmetries, different reaction speed to information innovations, diverse trading opportunity sets and risk-return expectations [17]. Klitgaard and Weir used private information and public information to distinguish heterogeneous traders [18]. The private information here is analogous to traders and form all contents included price expectations. De la Maza and Yuret's research adopted information set, algorithm set, model set, constraint set, to describe traders' heterogeneity. We use forecast style, risk-return preference, and reaction speeds, totally three respects to describe traders' heterogeneity.

1) Forecast style. Oberlechner adopted clustering analysis method to obtain four forecast styles: chartist, ascending (technical analysis method is more popular in short-term, as time horizon is lengthened, change direction and fundamental analysis gradually becomes more popular when time horizon become longer than one year), fundamental, ascending (adopt the basic analytical method more than the technical analysis of rising in short-term, as the time horizon is lengthened, change direction and rely mainly on the fact that the foundation is analysed gradually, level become extreme foundation analysis trader longer than one year, in time), constant chartist (no matter time level grow section, pay attention to technical analysis even more, does not change much, this and Cheung, Chinn and Marsh, 2004 research conclusion the same), inverse middle (adopt the foundation to analyse more in a short time, but turn in long-term forecast to and adopt the technical analysis method finally). Through interview traders, Oberlechner found traders view the charts as "visual representations of mass psychology".

2) Risk-return preference. Generally speaking, hedgers should carry on the middle or long-term trading in the futures market, and belong to risk neutral traders, because their trading motive lies in hedging stock value, but not in order to earn the great amount of profit through futures trading. Risk management traders generally do intraday transaction, and the purpose is for voiding business risk. It is generally acknowledged, speculators do short-term transaction, and their purpose to buy or sell futures contracts is just for earning great amount of profit. According to modern finance theory, more return with more risk, congenial profit-making traders, just traders have offered flow ability in order to hedge. But Haigh, Hranaiova and Overdahl studied speculators in futures market. They found the original common consensus that speculators do short-term transaction was not true, such as hedge fund mostly did middle or long-term transaction, but hedgers more frequently changed their positions than speculators did [19].

TABLE 1 Risk preference of three groups

Time Level	Hedgers	Risk Manager	Speculators
Short	RISK AVERSION	RISK AVERSION	RISK AVERSION
Middle or Long	RISK NEUTER	–	RISK APPETITE

Cheung, Chinn and Mars investigated traders in foreign exchange market and found that traders in market thought the predictability of price movement increased with time horizon extended. They also found the next day or intraday trade has little predictability. This result proved these traders thought next day or intraday trade was very risky, but interview results showed these traders felt little risk in the next day or intraday. These results appear to contradict each other, and the reason may be that at least for part of traders, the risk measurement index may not be the variance, but the changes of assets. Because intraday or next day transaction even though makes loss, change percentage is very small for total amount of assets. Although it have more predictability for a long-time horizon, for an exact time level, traders think the predictability is extremely limited too. So we should adopt a new risk measurement method for descripting trader's risk-return preference, which is different from the past to adopt price variance of change estimate the risk. And what traders think is too risky to bring the higher income maybe not higher income. Such as risk management traders do intraday transaction, but they are risk-aversion. Three groups of risk management traders, hed-

gers and speculators are divided according to time horizons (see Table 1).

3) Reaction speed. Sager and Taylor thought some hedge funds had the fastest reaction speed to information innovations. But in Haigh, Hranaiova, and Overdahl's research hedge funds adjusted their positions after hedgers' position changed, that was to say hedgers have a faster reaction speed than hedge funds. It is still unclear what causes this each other contradictory result, but the fact is that traders really have different reaction speeds to new information. So reaction speed can become one aspect of trader's heterogeneity.

The traders are influenced by market sentiment mentioned in some literatures, and this is a kind of behaviour of heterogeneity too. Find through interview trader think "chart masses psychology but view ", can think traders adopt the technical analysis method to include his receiving the mood effect of the market while predicting. Using trading motive, predict styles, risk-return preference, reaction speed, totally four respects to describe heterogeneous traders.

4.3 IMPLICATIONS OF TRADERS' HETEROGENEITY

Trader's heterogeneity influence is divided into two parts, including the influence on producing between traders, i.e. the interaction between traders, and trader's overall influence on producing dynamically at the market price.

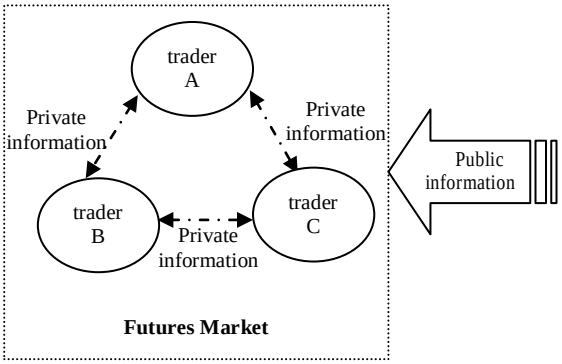


FIGURE 1 Relationship between internal model and outside environment

Because of difference predict style, deal motivation, risk-return preference, reaction speed, heterogeneous traders constantly influence each other. Traders who adopt technical analysis experience the mood in market by chart, influenced indirectly by mood in market. But as for individual, the influence received is the common influence which other heterogeneity traders produce in the market, traders react in the market price according to the decision that technical analysis is made, thus have an influence on other traders. So the heterogeneity traders in the market are the interaction, influence each other. Because of traders' heterogeneity, individual traders' decision exerts an influence on the price commonly at the same moment.

The reasons, representation and influence, totally three respects form commodity futures market trader internal model of artificial model of heterogeneity. The internal model exists in the whole market system, there is exchange of information and energy with the whole market, influence each other (Figure 1).

5 Commodity futures markets simulation

Since middle period of the 1990s last century, a lot of scholars have adopted the heterogeneous actor to suppose the research of carrying on artificial modelling to the financial market trend already. For instance Hommes, Le Baron, Lux, et al. adopt agent-based financial markets model to study a lot of market prices which accord with the result of study of experience count the characteristic. Carry on to commodity futures market price artificial research still relatively getting little dynamically[20], futures market and stock market are in the trade mechanism, the participant in the market, the structure of market, and the decision of basic value has a lot of difference. In the research of de la Maza and Yuret, adopt and fix model collecting and algorithm collection on two scholar's treatment to trader's internal model, observe trader information set restrain situation to collect change from, evolution trends of the market. Discover the traders with fund advantage and relative information advantage are easier to obtain in the market, two scholar in propose, should break through to some extent in constructing the more careful model of participant in the market to the further investigation of field here in their article.

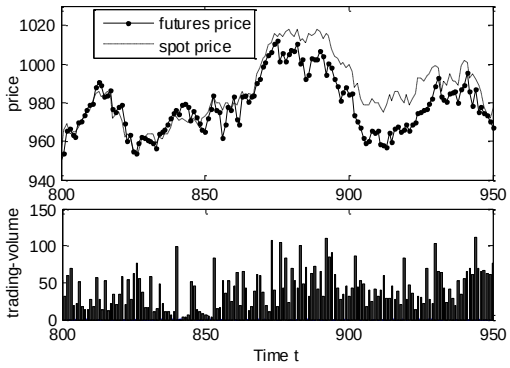


FIGURE 2 Market price and trading-volume in simulation

We apply the above framework to construct a simulation model to certificate the framework is useful. The software of Matlab is used for simulation analysis in the paper and Figure 2 shows the output. We use the actual spot price and simulation, and then we get the futures price series. Finally, we use the futures price series to test whether or not the framework can explain the anomies in futures market.

In the simulation, we choose real wheat spot price and the data can be obtained from wheat database on the website http://www.czce.com.cn.

From the simulation, we got the futures price and trading-volume which is showed in Figure 2. The test results have listed in Table 2. The simulation generates the futures price series have these characters: aggregational Gaussianity, clustered volatility, fat-tailed return distributions, and long-term memory. These results can partly verify the framework which is constructed in this paper (see Table 2).

TABLE 2 Futures prices series tests results

Statistics index	Futures price	Spot price
Sample size	960	960
Max (Min)	0.05(-0.04)	0.07(-0.04)
Standard variance	0.01	0.01
Skewness	0.60	3.76
Kurtosis	12.28	51.16
JB statistic	3477.86 [0.00]	94521.02 [0.00]
Q-statistic	98.62 [0.00]	52.66 [0.00]
Q^*-statistic	185.27 [0.00]	29.31 [0.00]
Hurst	0.69	0.64
Hurst*	0.50	0.52
LM test	6.76 [0.009]	0.12 [0.73]

In Table 2, Hurst index can be calculated by R/S analysis. Hurst* can be calculated by R/S analysis with using rearranged original sequence; Q is Box-Pierce Q-statistic used for autocorrelation test, Q* is Box-Pierce Q-statistic with using absolute return; LM test is a test for heteroskedastic in time series. All the values in square brackets are the probability of relevant test.

6 Conclusion

This paper summarizes the current literature about heterogeneous expectation test in financial markets and about financial market price dynamics simulation based on heterogeneous behaviour hypothesis, and analyses the reasons for heterogeneity, the representation of heterogeneous traders, as well as the influence of heterogeneity on the market, put forward the analysis framework of commodity futures market price dynamics based on heterogeneous traders.

We will further adopt direct examine method, use questionnaire and interview method to investigate traders' heterogeneity in the commodity futures market. The investigation may include all kinds of traders' risk preferences, forecasting technology, trade level, judgment of contract value, reaction speed, learning capability, etc. The results will be used for representation of heterogeneous traders in commodity futures market simulation model. Using artificial method to study micro pricing mechanism in futures market will be an effective route. This is the content that we will study further.

References

[1] Fisher B 1986 Noise *Journal of Finance* **41** (3) 529-44
[2] Goodhart C 1988 The foreign exchange market: a random walk with a dragging anchor *Economica* **55** (11) 437-60
[3] Simon H. A 1960 *The New Science of Management Decision* New York: Harper and Row
[4] Timothy C, Thomas J S, Viktor T 2014 Wealth Dynamics in a Bond Economy with Heterogeneous Beliefs *Economic Journal* **124** (575) 1-30
[5] Cvitanić J, Jouini E, Malamud S, Napp C 2011 Financial markets equilibrium with heterogeneous agents *Review of Finance* **16** (1) 285-321
[6] Chien Y, Lustig H 2010 The market price of aggregate risk and the wealth distribution *Review of Financial Studies* **23**(4) 1596-650
[7] Bhamra H S, Uppal R 2013 Asset prices with heterogeneity in preferences and beliefs *CEPR Discussion Papers*.
[8] Saang J B 1999 Tests for bounded rationality with a linear dynamic model distorted by heterogeneous expectations *Journal of Economic Dynamics and Control* **23** (9-10) 1517-43
[9] Frechette D L, Weaver R D 2001 Heterogeneous expectations of traders in speculative futures markets *The Journal of Futures Markets* **26**(5), 429-446
[10] Frechette D L 1999 The Supply Of Storage Under Heterogeneous Expectations *Journal of Agricultural and Applied Economics* **31**(3) 461-74
[11] Weaver R D, Zhang Y 1999 Volatility of market prices: the role of heterogeneous expectations *The Annual Meetings of the American Agricultural Economics Association* Nashville TN
[12] Oberlechner T 2001 Importance of Technical and Fundamental Analysis in the European Foreign Exchange Market *International Journal of Finance and Economics* **6**(1) 81-93
[13] Lui Y, Mole D 1998 The use of fundamental and technical analyses by foreign exchange dealers: Hong Kong evidence *Journal of International Money and Finance* **17**(3) 535-545
[14] Cheung Y-W, Chinn M D, Marsh I W 2004 How do UK-based foreign exchange dealers think their market operates? *International Journal of Finance & Economics* **4**(4) 289-306
[15] de la Maza M, Yuret D 1994 A futures market simulation with non-rational participants *Artificial life IV* MIT Press 325-30
[16] Chavas J P 1999 On the economic rationality of market participants: the case of price expectations in the U.S. pork market *Journal of Agricultural and Resource Economics* **24**(1) 19-37
[17] Sager M J, Taylor M P 2006 Under the microscope: the structure of the foreign exchange market *International Journal of Finance & Economics* **11**(1) 81-95
[18] Klitgaard T, Weir L 2004 Exchange rate changes and net positions of speculators in the futures market *Economic Policy Review Federal Reserve Bank of New York* May 2014 17-28
[19] Haigh M, Hranaiova J, Overdahl J 2005 Price Dynamics, Price Discovery and Large Futures Trader Interactions in the Energy Complex *Staff Research Report Commodity Futures Trading Commission* Washington DC
[20] Ussher L 2004 An agent based model of a speculative futures market: do margin requirements and transaction taxes affect price volatility? *PhD thesis*

Authors

Tan Li, August 1978, Harbin City, Heilongjiang Province, P.R. China.

Current position, grades: lecturer at the School of Management, Harbin Institute of Technology, China.
University studies: BSc, MSc and DSc in Management at Harbin Institute of Technology in China.
Scientific interests: agent-based computational finance, risk management, behavioral finance
Publications: more than 7 papers
Experience: Teaching experience of 5 years, 7 scientific research projects.

Weiwei Wu, November 1978, Taiyuan City, Shanxi Province, P.R. China

Current position, grades: associate professor at the School of Management, Harbin Institute of Technology (HIT), China.
University studies: BSc, MSc and DSc in Management at Harbin Institute of Technology in China.
Scientific interest: innovation management, agent-based computational finance.
Publications: more than 20 papers.
Experience: Teaching experience of 7 years, 6 scientific research projects.

GEM-based evaluation of competitiveness of enterprise cluster

Jingxian Tang[1*], Yan Yu[2], Haitao Zhou[2]

[1]*Changchun University of Science and Technology Changchun, Jilin Province, China, 130600*

[2]*Beijing Normal University, Beijing City, China, 100875*

Received 28 October 2014, www.cmnt.lv

Abstract

Enterprise cluster development can improve development capacity of enterprises and industrial competitiveness by lowering cost, stimulating innovations, increasing efficiency and intensifying competition. It is one key industrial development trend in the future and an important consideration of enterprises in choosing regions. Recently, industrial cluster in China has achieved primary development. Provinces and cities are putting great efforts to creating industrial parks of Industrial Cluster. This plays an important role in promoting regional economic development and regional industrial competitiveness. In this paper, a competitiveness evaluation model of enterprise cluster was established using the GEM model. It was applied to the furniture enterprise cluster in Guangdong Province. Accuracy, time effectiveness, advantages and disadvantages of the model was analyzed through the case study. Advantages and disadvantages of regional enterprise cluster can be identified from competitiveness analysis of the model. Countermeasures to these disadvantages were suggested.

Keywords: GEM model cluster competitiveness

1 Introduction

Industrial cluster is a spatial aggregate of enterprises of an industry with different scales and levels as well as behavioral agents related to their development (e.g. institutions and organizations) within a certain region. They have clear division of labor and are closely connected through a complicated network. Industrial cluster represents a new form of economic organization between enterprise and market. Enterprise cluster development is overwhelming in the world, thus attractting great attentions from economists and managerialists. Facing with economic globalization and regional economic integration, enterprise cluster development has become a new organizational pattern and development trend of regional economy. It enables enterprises to make full use of their advantages and achieve scale economies effect at a small scale, improving competitiveness of both enterprises and enterprise cluster. Currently, research attentions are paid to evaluation of competitiveness, advantages and disadvantages of enterprise cluster as well as analysis of measures to improve competitiveness. Industrial cluster development in China just starts. Scholars focus on qualitative research of enterprise cluster, but reported few model-based researches on the competitiveness of enterprise cluster. This paper established a competitiveness evaluation model of enterprise cluster based on GEM model, which was used to analyze competitiveness of the furniture enterprise cluster in Guangdong Province.

2 Competitiveness evaluation model of enterprise cluster

2.1 MEDEL CONSTRUCTION

GEM model is improved from the diamond model by Tim Padmore and Hervey Gibon, which can be used for quantification. It determines 6 major influencing factors of cluster competitiveness: resources, facilities, suppliers and relative assistance industries, enterprise strategies and competition, local market and external market. They can be divided into three groups: 1) fundamental factors: resources and facilities; 2) enterprise factors: suppliers and relative assistance industries as well as enterprise strategies and competition; 3) market factors: local market and external market. Compared to the diamond model, GEM model is superior for factor quantification. Factor quantification includes three steps:

Step 1: Assignments. These 6 influencing factors are level-1 evaluation indexes. To evaluate them, level-2 indexes will be generated and assigned with value. Suppose these 6 influencing factors are *A, B, C, D, E* and *F*, and their level-2 indexes are $A\left(a_1, a_2, ..., a_i\right)$, $B\left(b_1, b_2, ..., b_j\right)$, $C\left(c_1, c_2, ..., c_l\right)$, $D\left(d_1, d_2, ..., d_n\right)$, $E\left(e_1, e_2, ..., e_m\right)$, $F\left(f_1, f_2, ..., f_u\right)$ and, respectively. Score these level-2 indexes from 0-5. The higher the better and 3 is the mean level or pass line.

Step 2: Calculation. Assignments are often implemented by several experts or senior management of the industry. Mean of their assignments shall be calculated.

[*] ***Corresponding author's*** e-mail: zht@bnu.edu.cn

Moreover, different level-2 indexes have different weights because of their different influences. Scores of level-1 indexes shall be calculated according to weights of corresponding level-2 indexes:

$$\left\{\begin{array}{l} a_1=\dfrac{\sum_{q=1}^{N}a_{1q}}{N} \\ a_2=\dfrac{\sum_{q=1}^{N}a_{2q}}{N} \\ \cdots \\ a_i=\dfrac{\sum_{q=1}^{N}a_{iq}}{N} \end{array}\right\}\left\{\begin{array}{l} b_1=\dfrac{\sum_{q=1}^{N}b_{1q}}{N} \\ b_2=\dfrac{\sum_{q=1}^{N}b_{2q}}{N} \\ \cdots \\ b_j=\dfrac{\sum_{q=1}^{N}b_{iq}}{N} \end{array}\right\}\left\{\begin{array}{l} c_1=\dfrac{\sum_{q=1}^{N}c_{1q}}{N} \\ c_2=\dfrac{\sum_{q=1}^{N}c_{2q}}{N} \\ \cdots \\ c_l=\dfrac{\sum_{q=1}^{N}c_{iq}}{N} \end{array}\right\}\left\{\begin{array}{l} d_1=\dfrac{\sum_{q=1}^{N}d_{1q}}{N} \\ d_2=\dfrac{\sum_{q=1}^{N}d_{2q}}{N} \\ \cdots \\ d_n=\dfrac{\sum_{q=1}^{N}d_{iq}}{N} \end{array}\right\}\left\{\begin{array}{l} e_1=\dfrac{\sum_{q=1}^{N}e_{1q}}{N} \\ e_2=\dfrac{\sum_{q=1}^{N}e_{2q}}{N} \\ \cdots \\ e_m=\dfrac{\sum_{q=1}^{N}e_{iq}}{N} \end{array}\right\}\left\{\begin{array}{l} f_1=\dfrac{\sum_{q=1}^{N}f_{1q}}{N} \\ f_2=\dfrac{\sum_{q=1}^{N}f_{2q}}{N} \\ \cdots \\ f_u=\dfrac{\sum_{q=1}^{N}f_{iq}}{N} \end{array}\right\}. \quad (1)$$

N is number of people participated in the assignments.

$$A=\sum_{p=1}^{p=i}\alpha_p a_p, B=\sum_{p=1}^{p=j}\beta_p b_p$$
$$C=\sum_{p=1}^{p=l}\gamma_p c_p, D=\sum_{p=1}^{p=n}\phi_p d_p \ , \quad (2)$$
$$E=\sum_{p=1}^{p=m}\varphi_p e_p, F=\sum_{p=1}^{p=u}\theta_p f_p$$

where $(\alpha,\beta,\gamma,\phi,\varphi,\theta)$ is weight of level-2 indexes, respectively.

Step 3: Calculate and convert factor groups, and calculate the final results:

$$P_1=\frac{A+B}{2},\ P_2=\frac{C+D}{2},\ P_3=\frac{E+F}{2} \quad (3)$$

$$DEM=2.5\left[\prod\nolimits_{i=1.2.3}P_i\right]^{2/3}. \quad (4)$$

2.2 MODEL APPLICATION

The established model was used to evaluate competitiveness of furniture enterprise cluster in Guangdong Province. Beijing Forestry University investigated furniture enterprise cluster in Guangdong Province in 2012. Level-2 indexes of resources include labor supply, technician supply, raw material supply and fund supply. Level-2 indexes of facilities are facilities of industrial park, transportation conditions, public R&D institutions, support of industry association and business environment. Level-2 indexes of assistance industries include professional level of local suppliers, local device supply and cooperation & communication with local associated enterprises. Level-2 indexes of enterprise strategies and competition are goal orientation of enterprise development strategies, property right structure, competition and ambition of senior executives. Level-2 indexes of local market are domestic market share, influence of domestic market and domestic market prospect. Level-2 indexes of external market include international market share, influence of international market, entry barriers of international market and international market prospect. Assignments of 6 level-1 indexes are listed in Table 1.

TABLE 1 The value of the variable

A	B	C	D	E	F
3.40	3.82	3.96	3.97	3.77	3.59

According to the Table 1, competitiveness of the furniture enterprise cluster in Guangdong Province can be calculated:

$$P_1=\frac{A+B}{2}=3.61$$
$$P_2=\frac{C+D}{2}=3.97,$$
$$P_3=\frac{E+F}{2}=3.68$$

$$DEM=2.5\left[\prod\nolimits_{i=1.2.3}P_i\right]^{2/3}=$$
$$2.5\left[3.61\cdot 3.97\cdot 3.68\right]^{2/3}=35.15.$$

Above analysis data reveal that the furniture enterprise cluster in Guangdong Province has limited competitive edges.

Firstly, resources and external market are inferior. Resources are calculated 3.40, the lowest one and least competitive of 6 influencing factors. The external market is calculated 3.59. Although it is higher than resources, its competitiveness is still unsatisfying and needs further improvement. They are against the overall competitiveness of enterprise cluster. Specifically, labor shortage is a common problem in furniture industry. Some enterprises even suffer collective strikes for low wage level. Although China's furniture industry is facing with raw materials shortage, especially imported timbers, the raw material supply in Guangdong Province is strongly

supported by its geographical advantages and stable import channels built for years. Moreover, Guangdong Province is one of wealthy regions in China, which intensifies support to furniture industrial development. The furniture industry in Guangdong Province depends on external markets significantly. The furniture industry in Guangdong Province developed from processing trade-oriented joint ventures with Hong Kong and Taiwan investment. Later, supported by geographical advantages and the reform and open up, it shows stronger export preference than other regions in China. However, its furniture products enter into the international market as original equipment manufacturer (OEM) products and have no real influence in international market. This determines the high susceptibility of Guangdong furniture export to international market fluctuations. Viewed from international market prospect, with rich export experiences and advantages, furniture industry in Guangdong Province has stronger market penetration than rest regions in China.

Secondly, enterprise strategies and assistance industries have a distinct advantage. In Table 1, enterprise strategies get the highest scores, indicating the strongest competitiveness than other factors. For assistance Industries, the furniture industrial cluster in Guangdong Province is poorer than those in other regions. It is characteristic of "big cluster-small enterprise". According to level-2 indexes, most furniture enterprises in Guangdong haven't made long-term strategic development plan yet. Vicious competition among similar enterprises is universal, such as mutual imitation and race to keep the prices down. Most executives didn't receive high education and have no ambitions and entrepreneur spirit. On the contrary, support of suppliers and relative assistance industries in Guangdong furniture industry ranks the top in China. Since furniture industrial cluster began early in Guangdong Province, it has established a developed division of labor within the industry, leading to the high professional level of suppliers and strong local device supply. However, limited by traditional competition concept between enterprises, cooperation and communication of furniture enterprises in Guangdong Province with local associated enterprises still has a great development space.

Thirdly, facilities and local market are not so competitive. Facilities of industrial park are evaluated less competitive. This is because the furniture enterprise cluster in Guangdong Province is formed spontaneously under the drive of market economy, without government leading and normal planning. In transportation, Guangdong possesses well-developed highways, railways and ports, which improve logistics efficiency of local furniture industry significantly. Furniture enterprise cluster in Guangdong province has cutting edges in public R&D institutions, support of industrial association and business environment. Guangdong Province, the earliest region developing market economy in China, has standard business environment and clear orientation of industry association. These can provide favorable services and supports for industrial development. Furniture enterprise cluster in Guangdong province occupies at least 30% domestic market share, the highest in China. Influenced by traditional regional brands and promoted by furniture exhibition, it takes the dominant role in domestic market. However, its development prospect in domestic market is gloomy. This is caused by its attention on labor-intensive products with low value added, intensifying competition with other domestic regions and higher production cost than Chinese mainland.

2.3 MODEL ANANLYSIS

Accuracy. Model accuracy is mainly determined by the reasonability of level-2 indexes, that is, whether level-2 indexes can reflect conditions of level-1 indexes comprehensively. If yes, the model has high accuracy and the calculated results can reflect the real competitiveness of level-1 indexes; otherwise, the evaluation results will deviates from real performance of level-1 indexes.

Time effectiveness. Given fixed settings, any model can only provide short-term time effectiveness, because any condition change of industry cluster will change the corresponding environment accordingly. As a result, the time effectiveness of mode is limited. However, this can be solved by using same calculation method but changeable indexes (both level-1 and level-2 indexes). Actually, environment keeps stable in a short period. For long-term evaluation, level-2 indexes shall be changed accordingly, followed by level-1 indexes. If there's great environmental change, level-1 and level-2 indexes shall be changed simultaneously.

Advantages and disadvantages. Model advantages can be analyzed quickly from data collected through a simple survey. Since 6 level-1 indexes involve both internal and external environmental changes and the level-2 indexes can be updated as needed, the model can reflect competitiveness of enterprise cluster comprehensively. However, it still has some disadvantages. Firstly, it requires professional surveys. The number of people participated in assignments is proportional to the model accuracy. Secondly, its calculation is simple. Indexes couldn't represent all aspects of enterprise clusters.

3 Suggestions

The model is used to evaluate competitiveness of regional enterprise cluster, but its ultimate goal is to improve competitiveness of regional enterprise cluster. In view of the involved six influencing factors, it is suggested to:

Enhance policy support and perfect soft and hard environments. Policy support includes direct support and indirect support. Direct support refers to financial assistance and preferential policies to enterprises within the cluster. Indirect support refers to governmental supports through optimizing infrastructures and market order. In China hard environmental improvement refers

to perfect infrastructures in places of enterprise cluster, such as transportation, water, electricity, shopping, school, hospital, entertainment, fitness centers, etc. Soft environmental improvement including optimizing market order, enhance market supervision and management, improving unreasonable market provisions, promoting high-efficient market operation, building talent management and training system, as well as establishing channels for communication and cooperation between enterprises.

Increase talent recruitment, strengthen independent innovation, and improve continuous competitiveness of industrial cluster. Entrepreneurs shall bring new ideas to management and system, including reform and innovation of organizational system, management system and human resource system. They shall focus on creating an enterprise cultural atmosphere to train, attract and retain talents. Enterprises shall pay more attentions on vocational education to improve knowledge and quality of workers. Meanwhile, great efforts in implementing talent strategies and introducing advanced talents and management experiences are needed, so that talents can be trained and retained permanently. Furthermore, enterprises can enhance technological innovation through various ways, for example, increasing investment to technological development, establishing technological development institutes with universities and scientific research institutes, etc.

Implement regional information share and reinforce enterprise competition and cooperation. Disordered and vicious competition is a big problem of industrial cluster. Accelerating the construction of resource share and public service platform is the only one solution to this problem. It can overcome transaction difficulties caused by incomplete information and lower information search cost to a certain extent. Nowadays, Chinese enterprises in a region pay more attention to competition rather than harmonious development and copy from each other instead of create new products. Such vicious competition is distressing. Therefore, enterprises focus on mutual cooperation (e.g. Joint-funded R&D center, enterprise cluster or enterprise alliance) when compete with each other. A regional coordination development mechanism shall be established to create a multi-win structure.

References

[1] Xi G, Lei H 2005 The causes of the competitive power of industrial groups and a regression analysis *Nankai Economic Studies* **04** 34-40 (*in Chinese*)

[2] Liu H, Chen J 2004 A review of the research on the competitiveness of industrial clusters *Foreign Economies and Management* **10** 2-9 (*in Chinese*)

[3] Chen L 2009 On the connotation and nature of competitiveness of industrial clusters *Journal of Taiyuan University of Technology(Social Sciences Edition)* **01** 1-6 (*in Chinese*)

[4] Chen L 2009 Industrial cluster competitiveness study *Journal of University of Science and Technology Beijing(Social Sciences Edition)* **02** 15-25 (*in Chinese*)

[5] Zhang C, Huang L, Zhao Z 2013. Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34 (*in Chinese*)

[6] Liu A 2006 Theoretical research and empirical analysis of the competitiveness of industrial clusters *Master Thesis of Fuzhou University* 22-7 (*in Chinese*)

[7] S Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014. PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103 (*in Chinese*)

[8] Jin J 2011 Review of the research on the competitiveness of industrial clusters *HLJ Foreign Economic Relations & Trade*. **13** 188-90 (*in Chinese*)

[9] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81 (*in Chinese*)

[10] Lu J, Huang X, Zhang F 2008 To construct the evaluation model and the theoretical basis of the competitiveness of industrial clusters *Statistics and decision* **19** 67-9 (*in Chinese*)

Authors

Tang Jingxian, born in September 1982, ChangChun, Jilin Province, P.R. China.

Current position, grades: Management MM, ChangChun University of Science and Technology, Chang Chun, Jilin Province, China.
Scientific interest: business administration and higher Education teaching quality.
Publications: 10 papers.
Experience: 7 years teaching experience.

Yu Yan, born in June 1984, ChangChun, JiLin, P.R. China.

Current position, grades: Agronomy MM, ChangChun University of Science and Technology, Chang Chun, Jilin Province, China.
Scientific interest: biotechnology, Crop Genetics and Breeding higher Education teaching quality.
Publications: 8 papers.
Experience: 5 years teaching experience.

Zhou HaiTao, born in December 1972, Beijing, P.R. China.

Current position, grades: Education PhD, Beijing Normal University, Beijing City, China.
Scientific interest: higher education teaching quality.
Publications: 35 papers.
Experience: 15 years teaching experience.

Research on contract management evaluation of construction company based on fuzzy comprehensive evaluation model

Yuejun Liu*, Fang Fang, Yali Zhang

College of Economics and Management, Hebei Institute of Architectural and Civil Engineering, Zhangjiakou, Hebei Province, China, 075000

Received 28 October 2014, www.cmnt.lv

Abstract

Contract management is the core of project management of construction companies. Establishing a perfect contract management evaluation system of construction companies is of important significance to enhance their contract signing and performance as well as competitiveness in international market. Based on the overall contract management of construction companies, this paper discussed how to establish an evaluation system. Combined with fuzzy comprehensive evaluation, the established evaluation system was verified by an empirical analysis.

Keywords: contract management capability, evaluation model, fuzzy comprehensive evaluation, empirical analysis

1 Research significance

Contract management is the core of project management of construction companies and runs through the whole implementation of engineering project. It is of more significance to some large construction companies in China who attempt to step into the international market.

Contract management capability of project management personnel can be evaluated through some way, which enables us to become aware of shortcomings, differences and potential causes of low-efficient project implementation, and give early warnings.

2 Content of contract management of construction companies

Contract management of construction companies is the generic term of a series of legal behaviors on involved contract, including formation, performance, modification, dissolution, assignment, termination, review, supervision and control of contracts. Formation, performance, modification, dissolution, assignment, termination of contracts are the content of contract management, while the rest three are means of contract management [4].

Contract management of construction companies is the most complicated, painstaking and important work, which has significant impact on the whole engineering project.

3 Contract management evaluation model of construction companies

To evaluate contract management capability of China's construction companies, this paper deems that an evaluation model shall be established under the guidance of both domestic and foreign theories of contract management and combining with practical situations of China's construction companies as well as evaluation methods for project management.

3.1 LEVEL-1 EVALUATION INDEXES OF THE MODEL

Three level-1 evaluation indexes for contract management capability were concluded from review of existing associated research results.

1) Support of organizational system. Contract management is implemented through personnel, system and procedure set by the management organization of enterprises. The organizational management is important to contract management. The management organization of enterprises determines division of collaboration and function of departments, providing strong support to contract management.

2) Bidding and contracting management. Offer and commitment are two stages before the conclusion of contracts. Bidding is the offer of construction companies. The bid price not only can influence performance of enterprises directly, but also the key of winning the bidding. After winning the bidding, a "win-win" contract shall be signed by all means. These pave the way for contract performance.

3) Contract performance management. Contract performance is the core of whole contract management. Executives, project department, contract management department and other relevant departments are asked to perform contractual obligations in agreed time, cost and quality by all control means and win agreed rights. Since contract management is very complicated, it has high requirement on capabilities of related personnel.

The above three aspects are the key of the contract

* ***Corresponding author's*** e-mail: 59944615@qq.com

management capability evaluation model of construction companies (hereinafter referred as CMC model, Figure 1. They are level-1 evaluation indexes of the CMC model and the key domain indexes of contract management capability.

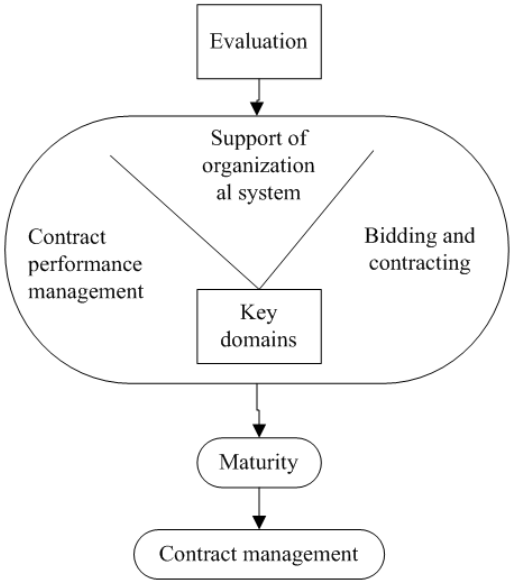


FIGURE 1 CMC model

3.2 LEVEL-2 EVALUATION INDEXES

OF THE MODEL

Due to the inadequate understanding of China's construction staff on contract management and their great differences of education background, opinions of relatively independent experts with rich experiences were used in this paper in order to establish a more scientific and reasonable evaluation system for engineering contract management. The experts mainly include scholars, enterprise managers and project managers in construction field. To get more comprehensive level-2 evaluation indexes, open-ended questionnaire survey was conducted. Firstly, experts were asked to list key indexes of three key domains through face-to-face interview or E-mail. After three questionnaire surveys, a total of 12 level-2 evaluation indexes were concluded under the assistance of experts. These indexes are, in accordance with the design principle of indexes, have certain scientific and independence and can reflect practical capabilities of key domains (Table 1).

3.3 WEIGHT DETEMINATION OF EVALUATION INDEXES

To reflect effect of evaluation indexes in the evaluation system, this paper analyzed significance of evaluation indexes through analytic hierarchy process (AHP). Based on experts' opinions and yaahp0.4.1 AHP software, weights of evaluation indexes were determined (Table 1).

TABLE 1 General situation of contract management evaluation indexes of the construction

goal	Level-1 indexes	Weight	Level-2 indexes	Weight	Maturity grades				
					Chaotic management	*Simple management*	*Standard management*	*Lean management*	*Strategic management*
Contract management capability (C)	Support of Organizational system (C_1.)	0.2972	Completeness of contract management system (C11)	0.2780	0/6	1/6	4/6	1/6	0/6
			Effectiveness of incentives(C_{12})	0.1361	0/6	1/6	3/6	2/6	0/6
			Adaptability of organizational structure (C_{13})	0.0875	0/6	1/6	3/6	2/6	0/6
			Competency of contract management personnel (C_{14})	0.4987	0/6	1/6	4/6	1/6	0/6
	Bidding and Contracting Management (C_2.)	0.1284	Bidding capacity (C_{21})	0.3522	0/6	0/6	3/6	3/6	0/6
			Contract risk management (C_{22})	0.3486	0/6	1/6	3/6	2/6	0/6
			Contract negotiation ability (C_{23})	0.2991	0/6	1/6	4/6	1/6	0/6
	Contract Performance Management (C_3)	0.5744	Contract claim management (C_{31})	0.3202	0/6	2/6	4/6	0/6	0/6
			Contract modification management (C_{32})	0.0584	0/6	0/6	3/6	2/6	1/6
			Contract cost, schedule and quality control (C_{33})	0.3856	0/6	2/6	3/6	1/6	0/6
			Contract analysis and presentation capability (C_{34})	0.0908	0/6	0/6	2/6	3/6	1/6
			Contract information management (C_{35})	0.1450	0/6	1/6	3/6	2/6	0/6

4 An empirical study about fuzzy evaluation on maturity of contract management capability of a construction company

4.1 DETERMINATION OF EVALUTION SET

Similar project management, maturity of contract management was also divided into five grades. Grade 1 represents the poorest contract management, while Grade 5 represents the highest contract management. Therefore, the grade evaluation set of contract management is:

C={chaotic management, simple management, standard management, lean management and strategic management}.

4.2 EVALUATION PREPARATION

4.2.1 Test Projects of the construction company

This paper tried to make a comprehensive evaluation on the contract management of the construction company through its projects under construction, finished projects, domestic and international projects, industrial construction project, civil construction projects and municipal construction projects. Its contract management capability was evaluated based on four domestic and foreign typical projects.

4.2.2 Composition of evaluation personnel

Evaluation personnel includes deputy manager and experts from headquarter, external experts, manager of UAE department, manager of project department, and manager of contract estimation department. The deputy manager from headquarter is the leader of this evaluation group.

4.3 WEIGHT DETERMINNATION OF EVALUATION INDEX

4.3.1 Build AHP structure

Social, economic and scientific management problems were analyzed by AHP. Firstly, problems shall be systematized and layered to build an AHP structure.

The established AHP structure built based on the index system of the CMC model is shown in Figure 2.

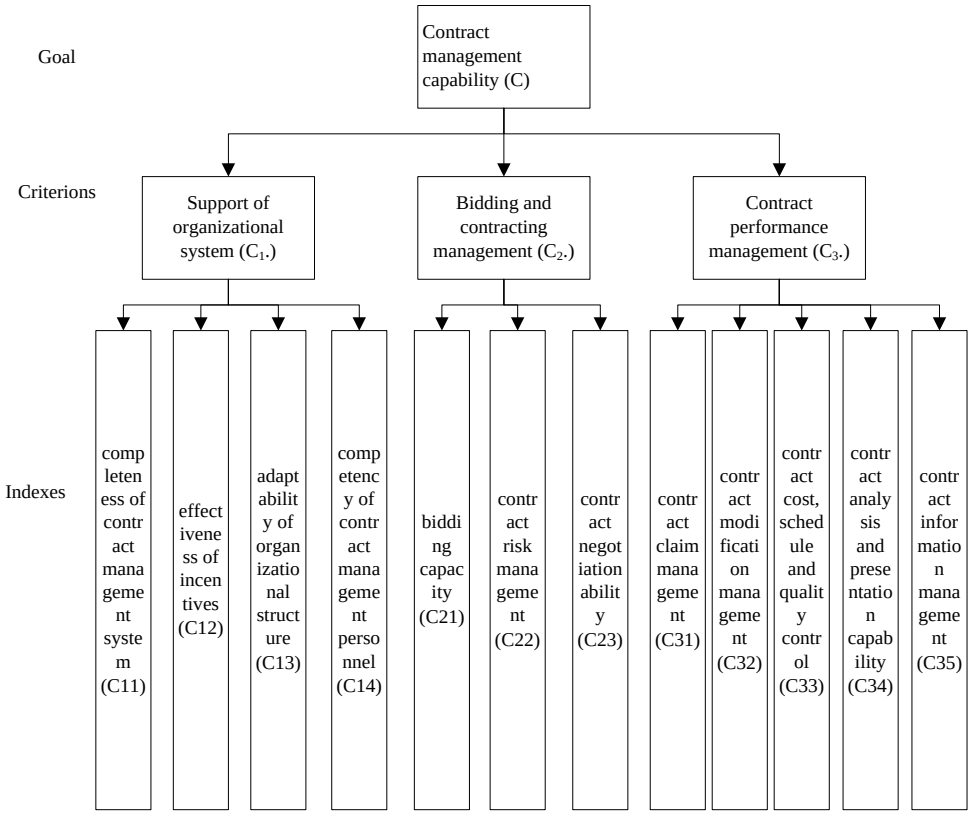


FIGURE 2 The level structure of contract management

4.3.2 Build judgment matrix

Suppose factors of previous layer (*C*) are criteria, which control factors of lower layer $(C_1, C_2, ...C_n)$. Our goal is to give weight of, $(C_1, C_2, ...C_n)$ according to their significance under the criterion Bk. For *n* factors, the judgment matrix of two factors is $C = (C_{ij})n*n$, where C_{ij} is the significance of *i* and *j* to the goal.
The built judgment matrix is:

	C_1	C_2	...	C_n
C_1	C_{11}	C_{12}	...	C_{1n}
C_2	C_{21}	C_{22}	...	C_{2n}
...	...	...	...	...
C_n	C_{n1}	C_{12}	...	C_{nn}

This judgment matrix is characteristic of:

1) $C_{ij} > 0$.

2) $C_{ij} = 1/C_{ji} (i \neq j)$.

3) $C_{ij} = 1 (i, j = 1, 2, ..., n)$.

Next, significance of C_i and C_j in criterion *C* was compared. "Significance" shall be given with certain numerical value, which uses 1-9 scale in this paper [25].

Then, evaluation personnel implemented the pairwise comparison of relevant factors by using their knowledge and experiences, filling in the weight evaluation table of contract management.

Weights of criteria are Criteria include Support of Organizational System, Bidding and Contracting Management as well as Contract Performance Management. Different evaluators got different judgment matrixes of criteria.

The judgment matrix of Evaluator 1 is:

	C_1	C_2	C_3
C_1	1	1/3	1/7
C_2	3	1	1/5
C_3	7	5	1

1) Calculate the product of each row of factors (M_i)

$$M_i = \prod C_{ij},\ i = 1, 2, ..., n,\ j = 1.$$

Therefore, $M = (0.0476, 0.6, 35)^T$.

2) Normalize $M_i (W_i)$

$$W_i = \sqrt[n]{M_i}.$$

Therefore, $W = (0.3625, 0.8434, 3.2711)^T$.

3) Normalize W_i

$$W_i = \frac{\overline{W_i}}{\sum_{j=1}^{n} \overline{W}_j}.$$

The eigenvector is: $W = (0.081, 0.1884, 0.7306)^T$.

4) Calculate the largest eigenvalue of the judgment matrix (λ_{max}). Therefore, $\lambda_{max} = 3.0649$.

$$CI = \frac{\lambda_{max} - n}{n-1} = \frac{3.0649 - 3}{3-1} = 0.0324.$$

5) Calculate the consistency ratio (CR): when $n = 3$, $RI = 0.58$.

$$CR = CI / RI = 0.0324 / 0.58 = 0.0559 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 2 is:

	C_1	C_2	C_3
C_1	1	5	1/5
C_2	1/3	1	1/9
C_3	5	9	1

Similarly, its eigenvectors can be calculated:

$$W = (0.1782, 0.0704, 0.7514)^T$$

$$\lambda_{max} = 3.0291, CI = 0.0145, RI = 0.58, CR = 0.0251 < 0.1$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 3 is:

	C_1	C_2	C_3
C_1	1	7	1/3
C_2	1/7	1	1/9
C_3	3	9	1

Its eigenvectors can be calculated:

$$W = (0.2897, 0.0549, 0.6554)^T$$

$$\lambda_{max} = 3.0803, CI = 0.0401, RI = 0.58, CR = 0.0692 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 4 is:

	C_1	C_2	C_3
C_1	1	3	1
C_2	1/3	1	1/5
C_3	1	5	1

Its eigenvectors can be calculated:

$$W = (0.4054, 0.1140, 0.4806)^T$$

$$\lambda_{max} = 3.0291, CI = 0.0145, RI = 0.58, CR = 0.0251 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 5 is:

	C_1	C_2	C_3
C_1	1	2	1
C_2	1/2	1	1/2
C_3	2	2	1

Its eigenvectors can be calculated:

$$W=(0.4,0.2,0.4)^T$$
$$\lambda_{\max}=3.0, CI=0, RI=0.58, CR=0<0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 6 is:

	C_1	C_2	C_3
C_1	1	3	1
C_2	1/3	1	1/3
C_3	1	3	1

Its eigenvectors can be calculated:

$$W=(0.4286,0.1429,0.4286)^T$$
$$\lambda_{\max}=3.0, CI=0, RI=0.58, CR=0<0.1$$

Therefore, this judgment matrix has satisfying consistency.

As a result, weights of three criteria are:

$$\begin{cases} a_1=(0.081+0.1782+0.2897+0.405+0.4+0.4286)/6=0.2971 \\ a_2=(0.1884+0.0704+0.0549+0.114+0.2+0.1429)/6=0.1284 \\ a_3=(0.7306+0.7514+0.6554+0.4806+0.4+0.4286)/6=0.5744 \end{cases}$$

The weight set of evaluation indexes is $C=(0.2971,0.1284,0.5744)$, representing the weights of Support of Organization System, Bidding and Contracting Management as well as Contract Performance Management, respectively.

Similarly, weight of the goal can be evaluated. Different evaluators got different judgment matrixes of Support of Organization System.

The judgment matrix of Evaluator 1 is:

	C_1	C_2	C_3	C_4
C_1	1	2	5	1/7
C_2	1/2	1	1	1/9
C_3	1/5	1	1	1/8
C_4	7	9	8	1

Then, its eigenvectors can be calculated:

$$W=(0.1628,0.0723,0.0592,0.7057)^T$$
$$\lambda_{\max}=4.2061, CI=0.0687, RI=0.90, CR=0.0763<0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 2 is:

	C_1	C_2	C_3	C_4
C_1	1	3	2	1/3
C_2	1/3	1	1/3	1/9
C_3	1/2	3	1	1/5
C_4	3	7	5	1

Then, its eigenvectors can be calculated:

$$W=(0.2186,0.0647,0.1349,0.5836)^T$$
$$\lambda_{\max}=4.0729, CI=0.0243, RI=0.90, CR=0.0270<0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 3 is:

	C_1	C_2	C_3	C_4
C_1	1	1	3	1/3
C_2	1	1	2	1/9
C_3	1/3	1/2	1	1/9
C_4	5	7	5	1

Then, its eigenvectors can be calculated:

$$W=(0.1421,0.1181,0.0596,0.6803)^T$$
$$\lambda_{\max}=4.0460, CI=0.0153, RI=0.90, CR=0.0171<0.1$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 4 is:

	C_1	C_2	C_3	C_4
C_1	1	5	7	1
C_2	1/5	1	4	1/5
C_3	1/7	1/4	1	1/7
C_4	1	5	7	1

Then, its eigenvectors can be calculated:

$$W=(0.4220,0.1097,0.0464,0.4220)^T$$
$$\lambda_{\max}=4.1398, CI=0.0466, RI=0.90, CR=0.0518<0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 5 is:

	C_1	C_2	C_3	C_4
C_1	1	1	3	1
C_2	1	1	5	1
C_3	1/3	1/5	1	1/7
C_4	1/3	1	3	1

Then, its eigenvectors can be calculated. Therefore, this judgment matrix has satisfying consistency. The judgment matrix of Evaluator 6 is:

	C_1	C_2	C_3	C_4
C_1	1	3	2	1/3
C_2	1/3	1	1	1/3
C_3	1/2	1	1	1/3
C_4	1	3	3	1

Then, its eigenvectors can be calculated:

$$W = (0.3468, 0.1279, 0.1416, 0.3838)^T$$
$$\lambda_{\max} = 4.0206, CI = 0.0069, RI = 0.90, CR = 0.0076 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.
As a result, weights of four indexes of Support of Organizational System are:

$$\begin{cases} a_{11} = (0.1628 + 0.2186 + 0.1421 + 0.4220.0.3754 + 0.3468)/6 = 0.2780 \\ a_{12} = (0.0723 + 0.0647 + 0.1181 + 0.1097 + 0.3241 + 0.1279)/6 = 0.1361 \\ a_{13} = (0.0592 + 0.1349 + 0.0596 + 0.0464 + 0.0834 + 0.1416)/6 = 0.0875 \\ a_{14} = (0.7057 + 0.5836 + 0.6803 + 0.4420 + 0.2168 + 0.3838)/6 = 0.4987 \end{cases}.$$

The weight set of Support of Organizational System is $A_1 = (0.2780, 0.1361, 0.0875, 0.4987)$, representing completeness of contract management system, effectiveness of incentives, adaptability of organizational structure and competency of contract management personnel, respectively.

Different evaluators got different judgment matrixes of Bidding and Contracting Management.

The judgment matrix of Evaluator 1 is:

	C_1	C_2	C_3
C_1	1	3	1/2
C_2	1/3	1	1/3
C_3	2	3	1

Then, its eigenvectors can be calculated:

$$W = (0.3325, 0.1396, 0.5278)^T$$
$$\lambda_{\max} = 3.0536, CI = 0.0268, RI = 0.58, CR = 0.0462 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 2 is:

	C_1	C_2	C_3
C_1	1	2	2
C_2	1/2	1	1
C_3	1/2	1	1

Then, its eigenvectors can be calculated:

$$W = (0.5, 0.25, 0.25)^T$$
$$\lambda_{\max} = 3.0, CI = 0, RI = 0.58, CR = 0 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 4 is:

	C_1	C_2	C_3
C_1	1	1	1
C_2	1	1	1
C_3	1	1	1

Then, its eigenvectors can be calculated:

$$W = (0.3333, 0.3333, 0.3333)^T$$
$$\lambda_{\max} = 3.0, CI = 0, RI = 0.58, CR = 0 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 5 is:

	C_1	C_2	C_3
C_1	1	3	5
C_2	1/3	1	2
C_3	1/5	1/2	1

Then, its eigenvectors can be calculated:

$$W = (0.6483, 0.2297, 0.1220)^T$$
$$\lambda_{\max} = 3.0037, CI = 0.00018, RI = 0.58, CR = 0.0032 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 6 is:

	C_1	C_2	C_3
C_1	1	1/5	1/3
C_2	5	1	1
C_3	1/5	1/2	1

Then, its eigenvectors can be calculated:

$$W = (0.1140, 0.4806, 0.4054)^T$$
$$\lambda_{\max} = 3.0291, CI = 0.0145, RI = 0.58, CR = 0.0251 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.
As a result,

$$\begin{cases} a_{11} = (0.3325 + 0.1852 + 0.5 + 0.3333 + 0.6483 + 0.1140)/6 = 0.3522 \\ a_{22} = (0.1396 + 0.6586 + 0.25 + 0.3333 + 0.2297 + 0.4806)/6 = 0.3486 \\ a_{23} = (0.5278 + 0.1562 + 0.25 + 0.3333 + 0.1220 + 0.4054)/6 = 0.2991 \end{cases}$$

The weight set of Bidding and Contract Management is $A_2 = (0.3522, 0.3486, 0.2991)$, representing bidding capacity, contract risk management and contract negotiation ability, respectively.

Different evaluators got different judgment matrixes of Contract Performance Management.

The judgment matrix of Evaluator 1 is:

	C_1	C_2	C_3	C_4	C_5
C_1	1	7	1/3	3	5
C_2	1/7	1	1/9	1/3	1
C_3	7	9	1	5	9
C_4	1/3	3	1/5	1	3
C_5	1/5	1	1/9	1/3	1

Then, its eigenvectors can be calculated:

$$W = (0.2609, 0.0449, 0.5304, 0.1157, 0.0480)^T$$

$$\lambda_{\max} = 5.0954, CI = 0.0239, RI = 1.12, CR = 0.0213 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 2 is:

	C_1	C_2	C_3	C_4	C_5
C_1	1	7	3	5	7
C_2	1/7	1	1/5	1/3	1/2
C_3	1/3	5	1	3	7
C_4	1/5	3	1/3	1	3
C_5	1/7	1	1/9	1/3	1

Then, its eigenvectors can be calculated:

$$W = (0.5009, 0.0537, 0.2460, 0.1292, 0.0703)^T$$

$$\lambda_{\max} = 5.4323, CI = 0.1081, RI = 1.12, CR = 0.0965 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 3 is:

	C_1	C_2	C_3	C_4	C_5
C_1	1	3	1/3	3	2
C_2	1/3	1	1/5	1/3	1/3
C_3	3	5	1	5	3
C_4	1/3	1	1/5	1	1/3
C_5	1/7	3	1/3	3	1

Then, its eigenvectors can be calculated:

$$W = (0.2235, 0.0729, 0.4613, 0.0729, 0.1694)^T$$

$$\lambda_{\max} = 5.1140, CI = 0.0285, RI = 1.12, CR = 0.0254 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 4 is:

	C_1	C_2	C_3	C_4	C_5
C_1	1	7	1	5	3
C_2	1/7	1	1/5	1	1/5
C_3	1	5	1	5	3
C_4	1/5	1	1/5	1	1/3
C_5	1/3	5	1/3	3	1

Then, its eigenvectors can be calculated:

$$W = (0.3734, 0.0524, 0.3491, 0.0621, 0.1630)^T$$

$$\lambda_{\max} = 5.12, CI = 0.0300, RI = 1.12, CR = 0.0268 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 5 is:

	C_1	C_2	C_3	C_4	C_5
C_1	1	8	1/2	5	2
C_2	1/8	1	1/9	1/3	1/4
C_3	2	9	1	7	3
C_4	1/5	3	1/7	1	1/2
C_5	1/2	4	1/3	3	1

Then, its eigenvectors can be calculated:

$$W = (0.2897, 0.0358, 0.4540, 0.0738, 0.1467)^T$$

$$\lambda_{\max} = 5.0797, CI = 0.0199, RI = 1.12, CR = 0.0178 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

The judgment matrix of Evaluator 6 is:

	C_1	C_2	C_3	C_4	C_5
C_1	1	3	1	3	1
C_2	1/3	1	1/3	1	1/3
C_3	1	3	1	3	1
C_4	1/3	1	1/3	1	1/3
C_5	1	3	1	3	1

Then, its eigenvectors can be calculated:

$$W = (0.2727, 0.0909, 0.2727, 0.0909, 0.2727)^T$$

$$\lambda_{\max} = 5.0, CI = 0, RI = 1.12, CR = 0 < 0.1.$$

Therefore, this judgment matrix has satisfying consistency.

As a result, weights of five indexes of Contract Performance Management are:

$$\begin{cases} a_{31} = (0.2609 + 0.5009 + 0.2235 + 0.3734 + 0.2897 + 0.2727)/6 = 0.3202 \\ a_{32} = (0.0449 + 0.0537 + 0.0729 + 0.0524 + 0.0358 + 0.0909)/6 = 0.0584 \\ a_{33} = (0.5304 + 0.2460 + 0.4613 + 0.3491 + 0.4540 + 0.2727)/6 = 0.3856 \\ a_{34} = (0.1157 + 0.1292 + 0.0729 + 0.0621 + 0.0738 + 0.0909)/6 = 0.0908 \\ a_{35} = (0.0480 + 0.0703 + 0.1694 + 0.1630 + 0.1467 + 0.2727)/6 = 0.0908 \end{cases}.$$

The weight set of Contract Performance Management is $A_3 = (0.3202, 0.0584, 0.3856, 0.0908, 0.1450)$, representing contract claim management, contract modification management, contract cost, schedule and quality control, contract analysis and presentation capability as well as contract information management, respectively.

4.4 FUZZY EVALUATION OF MATURITY

4.4.1 Fuzzy membership

The fuzzy memberships of level-2 indexes were evaluated by evaluators according to project situations (Table 1).

4.4.2 Fuzzy evaluation matrix

1. Fuzzy evaluation matrix of single factor.

According to Table 1, evaluation matrixes of single factor are:

Support of Organizational System:

$$R_1 = \begin{bmatrix} 0/6 & 1/6 & 4/6 & 1/6 & 0/6 \\ 0/6 & 1/6 & 3/6 & 2/6 & 0/6 \\ 0/6 & 1/6 & 3/6 & 2/6 & 0/6 \\ 0/6 & 1/6 & 4/6 & 1/6 & 0/6 \end{bmatrix}.$$

Bidding and Contract Management:

$$R_2 = \begin{bmatrix} 0/6 & 0/6 & 3/6 & 3/6 & 0/6 \\ 0/6 & 1/6 & 3/6 & 2/6 & 0/6 \\ 0/6 & 1/6 & 4/6 & 1/6 & 0/6 \end{bmatrix}.$$

Contract Performance Management:

$$R_3 = \begin{bmatrix} 0/6 & 2/6 & 4/6 & 0/6 & 0/6 \\ 0/6 & 0/6 & 3/6 & 2/6 & 1/6 \\ 0/6 & 2/6 & 3/6 & 1/6 & 0/6 \\ 0/6 & 0/6 & 2/6 & 3/6 & 1/6 \\ 0/6 & 1/6 & 3/6 & 2/6 & 0/6 \end{bmatrix}.$$

Then, the judgment vectors of Support of Organizational System (B_1), Bidding and Contract Management (B_2) as well as Contract Performance Management (B_3) can be calculated from $B = A_i * R_i$:

$$B_1 = (0.2780, 0.1361, 0.0875, 0.4987) * \begin{bmatrix} 0/6 & 1/6 & 4/6 & 1/6 & 0/6 \\ 0/6 & 1/6 & 3/6 & 2/6 & 0/6 \\ 0/6 & 1/6 & 3/6 & 2/6 & 0/6 \\ 0/6 & 1/6 & 4/6 & 1/6 & 0/6 \end{bmatrix} = (0, 0.1667, 0.6296, 0.204, 0),$$

$$B_2 = (0.3522, 0.3486, 0.2991) * \begin{bmatrix} 0/6 & 0/6 & 3/6 & 3/6 & 0/6 \\ 0/6 & 1/6 & 3/6 & 2/6 & 0/6 \\ 0/6 & 1/6 & 4/6 & 1/6 & 0/6 \end{bmatrix} = (1080, 0.5498, 0.3422, 0),$$

$$B_3 = (0.3202, 0.0584, 0.3856, 0.0908, 0.1450) * \begin{bmatrix} 0/6 & 2/6 & 4/6 & 0/6 & 0/6 \\ 0/6 & 0/6 & 3/6 & 2/6 & 1/6 \\ 0/6 & 2/6 & 3/6 & 1/6 & 0/6 \\ 0/6 & 0/6 & 2/6 & 3/6 & 1/6 \\ 0/6 & 1/6 & 3/6 & 2/6 & 0/6 \end{bmatrix} = (0, 0.2594, 0.5382, 0.1775, 0.0249).$$

2. The judgment matrix of multiple factors is:

$$B = (0.2971, 0.1284, 0.5744) * \begin{bmatrix} 0 & 0.1667 & 0.6296 & 0.2040 & 0 \\ 0 & 0.1080 & 0.5498 & 0.3422 & 0 \\ 0 & 0.2594 & 0.5382 & 0.1775 & 0.0249 \end{bmatrix} = (0, 0.2124, 0.5668, 0.2065, 0.0143).$$

4.4.3 Evaluation on the maturity of contract management

Evaluation set for the maturity of contract management is $C = (C_1, C_2, C_3) =$ (Support of Organizational System, Bidding and Contract Management, Contract Performance Management). The evaluation standard is divided into five grades: $U(1, 2, 3, 4, 5)$, representing chaotic management, simple management, standard management, lean management and strategic management, respectively.

1. Single factor evaluation

$$C_1 = B_1 * U^T = (0, 0.1667, 0.6296, 0.204, 0) * \begin{vmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{vmatrix} = 3.0382.$$

This indicates that the maturity of Support of Organizational System is 3.0382, close to standard management. The construction company shall enhance support of organizational system.

$$C_2 = B_2 * U^T = (0, 0.1080, 0.5498, 0.3422, 0) * \begin{vmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{vmatrix} = 3.2342.$$

This reveals that the maturity of Bidding and Contract Management is 3.2342, between standard and lean management, close to standard management. The construction company still shall enhance its bidding and contract management.

$$C_3 = B_3 * U^T = (0, 0.2594, 0.5382, 0.1775, 0.0249) * \begin{vmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{vmatrix} = 2.9679 \cdot$$

This demonstrates that the maturity of Contract Performance Management is 2.9679, between simple and standard management, close to standard management. The construction company shall further standardize implementation of contract management and improve contract performance management continuously.

2. Multi-factor evaluation

$$C = B * U^T = (0, 0.2124, 0.5668, 0.2065, 0.0143) * \begin{vmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{vmatrix} = 3.0227 .$$

This shows that the maturity of contract management capability is 3.0227, close to standard management. The construction company has basically achieved effective contract management and control, but still has to further improve its contract management capability.

5 Conclusions

The evaluation group attempts to make a comprehensive evaluation on the contract management of the construction company through its projects under construction, finished projects, domestic and international projects, industrial construction project, civil construction projects and municipal construction projects. The evaluation reflects that the construction company has standard contract management, far from the lean management. If it wants to get an invincible position in international competition and win more economic benefits, the contract management capability still needs further improvement. Its maturity of Bidding and Contract Management is the highest, while that of Contract Performance Management is the lowest. However, both of them are close to standard management.

Acknowledgements

Project Name: Hebei Construction Technology Research Program in 2012, Project Number: 2012-2031

References

[1] Hu C 2004 Engineering project management *Beijing Higher Education Press* **7** 124-8 (*in Chinese*)
[2] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81 (*in Chinese*)
[3] Chenghu 2005 Engineering project management *China Building Industry Press* **8** 316-7 (*in Chinese*)
[4] He B 1999 International project contract and contract management *China Building Industry Press* **1** 21-9 (*in Chinese*)
[5] Zhang C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34 (*in Chinese*)
[6] Xiong S K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20** (3) 98-103 (*in Chinese*)
[7] Ke C 2006 Research on Application of project management maturity model in construction project in China *Master Thesis of Tongji University* 24-7 (*in Chinese*)
[8] Pan H 2008 Study on Evaluation of construction project management capability *Master Thesis of Tianjin University* 31-3 (*in Chinese*)

Authors

Liu Yuejun, born in August 1982, Zhangjiakou, Hebei Province, P.R. China.

Current position, grades: the lecturer of College of Economics and Management, Hebei Institute of Architectural and Civil Engineering, Hebei, China.
Scientific interest: management science and engineering.
Publications: 10 papers.
Experience: teaching experience of 9 years, 4 scientific research projects.

Fang Fang, born in November 1980, Zhangjiakou, Hebei Province, P.R. China.

Current position, grades: the librarian of College of Economics and Management, Hebei Institute of Architectural and Civil Engineering, Hebei, China.
Scientific interest: management science and engineering.
Publications: 5 papers.
Experience: teaching experience of 9 yeas, 3 research projects.

Zhang Yali, born in May 1987, lvliang, Shanxi Province, P.R. China.

Current position, grades: the assistant of College of Economics and Management, Hebei Institute of Architectural and Civil Engineering, Hebei, China.
Scientific interest: business administration.
Publications: 2 papers.
Experience: teaching experience of 2 years.

Research on the risks of financial derivatives and risk control from the perspective of the financial crisis

Hailei Zhao[1*], Dehuan Jin[2]

[1]*School of Business, Jiangnan University, Wuxi City, Jiangsu province, China, 214122*

[2]*Shanghai University of Finance and Economics, Shanghai City, China, 200433*

Received 28 October 2014, www.cmnt.lv

Abstract

After the economic crisis in Wall Street in 2008, the risks of financial derivatives and risk control have received great attention from countries in the world. The research on the risks of financial derivatives and risk control can effectively prevent the financial crisis. Through the extraction of the characteristics of risks of financial derivatives, the paper expounds the application of sensitivity analysis and in the risk measurement financial derivatives. The result shows that despite certain achievements of China's financial derivatives in the development process, there is still much room for improvement. Based on the result and with reference to the practice of the UK, US and Japan in the risk prevention and control of financial derivatives, the paper proposes some suggestions.

Keywords: financial derivatives, risk prevention and control, VAR model, financial crisis

1 Introduction

The economic crisis in the world financial center Wall Street in 2008 resulted in the global economic depression. After analysis, researchers believed that the lack of control of the investment in financial derivatives was the main factor. At the same time, as shown in Figure 1, some state-owned enterprises in China like China Railway and China Eastern Airlines also suffered serious losses due to the investment in financial derivatives. In order to effectively limit the investment in financial derivatives, two investment channels were added in the Chinese financial market in 2010 and the new two kinds of investment channels, Several Opinions on Further Doing a Good Job in the Utilization of Foreign Investment was released.

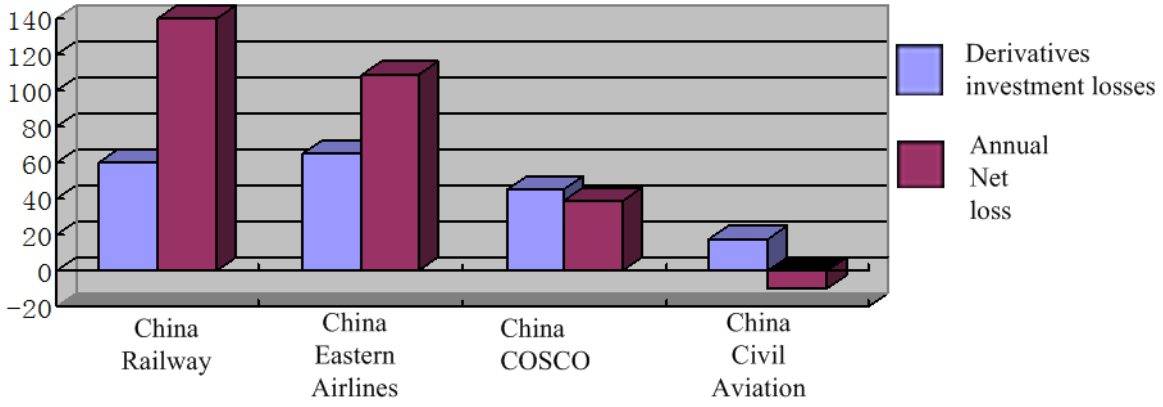


FIGURE 1 The Losses of the investment in financial derivatives in China

The emergence of inflation also let China's economic market face the shock from the development of financial derivatives, which is reflected by the US's economic pressure on China. In addition, financial derivatives are bound to affect China's economy. Many researchers believe that China's market economy will be affected by the development of global financial derivatives and the development orientation of financial derivatives is a common issue concerned by all countries at present. At the same time, whether some derivatives really conforms to our national situation is open to question.

* ***Corresponding author's*** e-mail: 445645889@qq.com

2 Risk types and characteristics of financial derivatives

ISDA explains the industry of financial derivatives as follow: "financial derivative is a financial instrument used for swapping cash flow, transferring risks, etc. After the trade, the debt is determined by the market stock and index futures. It's also a way of securities trading used to reduce risks". This instrument has many characteristics and owns the basic characteristics of some financial products. As a derivative, its income comes from the gains of other financial products. The slight fluctuations in the financial market can also cause the great changes in its gains, so it's very sensitive to the financial market. Futures and risk contract are the typical types.

2.1 RISK TYPE OF FINACIAL DERIVATIVES

The Group of Thirty (G30) is an authority of studying the risk type of financial derivatives which is mainly composed of professional talents in economy, law, banking, financial academics and other fields with the wide and deep research, because the product function determines its risk characteristics. It's found in the process of research that practice and principle are the important reference for risks. In the trading process, the trading market risk is the main source of risk. Then, the performance degree of other risks is different in different periods. There are other standards for classification. For example, the risk analysis by the IOSCO is more specific than the analysis report of G30. It adds the risk analysis of market liquidity, financial operation and transaction settlement. Here, the analysis report of G30 will be discussed.

2.1.1 Market risk

Affected by the price, the trading market risk has great uncertainty. Price change is the source of risk, such as stock, securities, conversion rate, transacting rate and interest rate. According to the trading characteristics, G30 divides the trading market risk into four types such as interest rate risk. Both bank rate and monetary exchange rate can lead to price changes and then result in losses. Equity risk like the change of securities and stock index may cause losses.

2.1.2 Credit risk

Credit risk cite in the paper mainly refers to the risk caused because the other party does not fulfill the stipulated obligations under the contract spirit. In other words, the fiduciary cannot pay principal and interest in the promising time in accordance with the contract conditions, which will lead to the obvious difference between the actual and expected earnings of the accredited party. It's also the main component of the financial risk under the current market economic environment.

2.1.3 Operation risk

Relative to credit risk, operation risk also has great influence on financial derivatives in the market. Operation risk mainly refers to the possibility of being unable to achieve the expected operation goal due to the environmental factors objectively and the insufficient cognition of the influence of environment subjectively. On one hand, operation risk comes from the systematic vulnerabilities caused by poor operation structure, low operation efficiency and unreasonable internal work arrangement; on the other hand, it comes from the external force majeure, such as the unavoidable losses caused by some things which can't be predicted in advance or by natural disasters.

2.1.4 Legal risk

The daily business activities or all kinds of transaction of commercial banks should abide by the relevant business norms and legal principles. In this process, the commercial banks may fail to fulfill the contract because they cannot meet or violate the legal requirements, leading to issues, lawsuits or other disputes and thus causing the risk of economic loss, which is known as the legal risk.

2.2 RISK CHARATERISITICS OF FINACIAL DERIVATIVES

2.2.1 Suddenness

As a product of the development of financial market in a certain period, financial derivative usually refers to the innovation of the financial instruments appeared on the basis of the original assets. It's characterized by completing the trade in full with the corresponding deposit without the problem of principal transfer. Thus, the trading of financial derivatives has the leverage effect. The less the deposit is, the greater the leverage effect is and the greater the risk is. Because of the low threshold, it is difficult to realize the effective control. Therefore, the risks of financial derivatives have certain suddenness.

2.2.2 Strong infection

Due to the characteristic of linkage of financial derivatives, the value of financial instruments is closely linked to basic products or basic variables. Several financial crises in the history were caused due to the spread of the risks of financial derivatives in the market, such as Mexico's financial crisis.

2.2.3 Huge harmfulness

Under the market economy environment, the harmfulness caused by the risks of financial derivatives is huge. It not only brings serious damage to monomer, but also causes the chain harm. For example, due to the bad financial environment, LTcM in the US suffered the crisis of bankruptcy. However, because the financial derivatives of its assets occupy a large share, its influence in the market is huge. Once LTcM sells its assets, the companies buying the financial derivatives of LTcM will also suffer great losses. If LTcM declares bankruptcy, those accreditors will also face dangerous situation. This thing reflects the serious harmfulness of financial derivatives.

3 Methods of risk measurement of financial derivatives

At present, the expansion of financial accessories raises higher requirements on the existing risk prevention system. The specific measures for risk prevention include sensitivity test and VAR. The former is to test the sensitivity of the relevant elements on the basis of multiple financial accessories. It mainly reflects the trading measures through the sensitivity of risks under some simple trading environments. The latter is to show all risks in the numerical value. The value also means the concealed loss of the trade.

3.1 SENSITIVITY ANALYSISI

The maximum advantage of the sensitivity test is that it allows the user to intuitively observe the sensitivity of the accessories and their values in the transaction. Due to its simple operation, it's favored by the users. After getting the change of elements in the market and the correlation of the produced values, the change of values can be obtained through certain calculation. John C Hull proposed the method of the sensitivity test. The values brought by the derivatives are set to be *F*. It's the function of *S*, α and *t* and is expanded with Taylor:

$$\Delta f = \frac{af}{aS}\Delta S + \frac{af}{a\alpha}\Delta\alpha + \frac{af}{at}\Delta t + \frac{1}{2}\frac{af}{aS^2}\Delta S^2 + \ldots,$$

where *f* is the value of the derivatives; *S* is the standard assets; *a* is the changing probability of the assets price; *t* is the time; Δt is the change of time; Δf is the change of *f*; ΔS is the change of *S*; and, $\Delta\alpha$ is the change of *a*.

In fact, Deltas is the first order function and Gamma is the second order function. In the above equation, Delta means the price change of the derivatives in a small scope in a short period of time on the premise of no any explanation. Gamma means the change in a large scope without sensitivity to the assets on the premise of nearly no explanation. These two derivatives symbolize the linear and nonlinear risk test of the derivatives on the change of assets.

Evga represents the effectiveness degree of the instrument value in reflecting the price change in the market environment. Theta and Rho represent the reflection ability of the derivatives in the risk level.

3.2 VAR ANALYSIS

VAR analysis is also known as Value at Risk. It provides an overall guiding value for the existing risks in the market. Value risk is based on the fixed time and the determined proportion, because the market value risk means the possible harm to the enterprises due to the market interest rate change in certain time and at certain probability.

Risk model establishment and risk calculation measure can be included into two aspects: the first is to get the partial calculation result and the second is to get the global calculation result. Those two different aspects in essence are reflected by the different treating methods of the profits brought by different assets.

3.2.1 Delta-the normal method

The following formula shows the profit earned by an investment portfolio in a certain period of time:

$$R_{P^{t+1}} = \sum_{i=1}^{n} W_{i,t} P_{i,t+1},$$

$$VaR = E(\omega) - \omega^* = W_0[1+E(R)] - W_0(1+R^*) = W_0[E(R)-R^*] = -W_0[E(R)-R^*] = W_0 R^*,$$

(setting $E(R)=0$),

where $W_{i,t}$ refers to the ratio of change of different assets based on the change of time. The change of time can show the development form of an investment portfolio. In the actual application, Delta –normal method first acquiesces the extensive existence of the normal distribution in the profit rate. Represented by $R_{i,t+1}$, the profit rate shows the obvious characteristics of the normal distribution variable in the calculation process. Thus, it can be assumed like that. The variance calculation formula of the investment portfolio is:

$$V(R_{p,t+1}) = {}_{\alpha\sigma\rho}\, W = \alpha\sqrt{x\sum x}.$$

3.2.2 Historical data simulation method

As a common method in the current comprehensive valuation, historical data simulation method can provide the valuation result more conveniently and quickly. In the specific application process, the current weight and the

portfolio composed of historical asset profit rates are its core content.

$$R_{p,t+1} = \sum_{i=1}^{n} W_{i,t} R_{i,t+1} .$$

The complete price data is the basis of the successful application of the method. The price in the t[th] period of time can be calculated with the current price and the historical price changes.

$$R^*_{j-\tau} = P_{j-\mu} + \Delta P_{j,\tau} .$$

With the complete virtual price, the key data in the above formula can be solved. In the process, the non-linear relationship can be verified. Then, the virtual profit ratio corresponding to *r* can be obtained:

$$P_{p.t} = \frac{P_{p,t+1} - P_{p,0}}{P_{p,0}} .$$

On this basis, VAR value can be obtained.

3.2.3 Parameter method

It's also known as the variance-covariance method in some literate with the core of estimating the variance-covariance matrix of the return on assets. It sets the current assumption and the normal distribution assumption to make calculation, namely, $X - \mu / \sigma \sim N(0,1)$ and setting $\mu = 0$, then $X \sim N(0, \sigma^2)$. In the process of solving the σ variance, the equal weight formula can be adopted according to the actual situation. It tests the unconditional volatility in the financial derivative market, that is:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n} (R_t - \mu)^2} .$$

The application of the index weight formula can provide us with the conditional volatility, that is:

$$\sigma = \sqrt{(1-\lambda) \sum_{t=1}^{n} \lambda^{n-t} (R_t - \mu)^2} .$$

The equation above can effectively express the volatility of the financial time series in the current market.

4 Empirical Analysis of Risk Measurement with the VAR Model

4.1 NORMAL VAR SINGLE RISK FACTOR δ

The simple statistical processing was done on the trading data in 1308 days with the relevant statistical tools to get the geometric profit rate level in a single period of time. On this basis, the specialized software was adopted to analyze the data statistically and get the daily standard deviation of 0.00142 of the national debt. The data of securities investment funds, stocks and bank deposits were 0.0032, 0.00753, and 0.00097, respectively. In the actual calculation process, we could think that the calculation formula of the maximum level of the negative change of the national debt in a single period of time could be simplified as $0.00142 \times 1.65 \times 1 = 0.00234$ under 95% confidence level. In the same calculation method, the data of securities funds, stocks and band deposits could be calculated to be 0.00528, 0.00753 and 0.0016, respectively.

Thus, the single factor
VAR of the national debt was $27.1 \times 0.00234 = 0.0634$;
VAR of securities was $VaR = 19.4 \times 0.058 = 0.1024$;
VAR of stocks was $VdR = 10.2 \times 0.00753 = 0.1268$
and
VAR of band deposits was $VaR = 43.3 \times 0.0016 = 0.06925$.

After calculation of the total VAR with the factor promoting method, the sum of the absolute values was 0.3618. In the actual application process, the method did not give full consideration to the correlation among four kinds of assets above but made the simple superposition. Thus, from the perspective of historical data, we could get the VAR invested by the insurance company was 0.3618 hundred billion yuan at the confidence level of 95%. In fact, the data proved that in the current market economy environment, more than 95% of people held that the total losses of the insurance company were less than 0.3618 hundred billion yuan in the unit period of time.

4.2 NORMAL VAR OF MULTI-RISK FACTOR δ

Next, on the premise of considering the correlation of assets, the normal VAR of double risk factors was calculated.

$$VaR_{ALL} = \sqrt{VaR_1^2 + VaR_1^2 + 2\rho_{12} VaR_1 VaR_2} .$$

In the equation above, ρ_{12} represented the correlation coefficient of two risk factors participating in the calculation.

The calculation of VAR depending on three risk factors could be expressed as:

$$VaR_{ALL} = \sqrt{VaR_1^2 + VaR_1^2 + VaR_3^2 + 2\rho_{12} VaR_1 VaR_2 + 2\rho_{13} VaR_1 VaR_3} .$$

The calculation formation of VAR of n risk factors could be expressed as:

$$VaR_{ALL} = \sqrt{\left(\sum_i \sum_j \rho_{ij} VaR_i VaR_j \right)^{12}} .$$

The application of matrix sign was the necessary choice in the case of many risk factors. Its basic formula could be simply expressed as:

$$VaR_{ALL} = \sqrt{VCV^T},$$

where V represented the row vector of VAR of n single factors; C was the $n \times n$ correlation matrix between factors; T was the matrix transposition calculation sign.

Through the statistical analysis of four kinds of assets above, the following matrix could be obtained:

$$\begin{bmatrix} 1 & -0.0267 & 0.00716 & 0.54653 \\ -0.0267 & 1 & 0.08848 & 0.05440 \\ 0.00716 & 0.008848 & 1 & 0.00260 \\ 0.54653 & 0.05440 & 0.00260 & 1 \end{bmatrix}.$$

Then, the corresponding normal VAR of δ was:

$$VaR_{ALL} = \sqrt{(0.0634, 0.1024, 0.1268, 0.0693) \begin{bmatrix} 1 & -0.0267 & 0.00716 & 0.54653 \\ -0.0267 & 1 & 0.08848 & 0.05440 \\ 0.00716 & 0.008848 & 1 & 0.00260 \\ 0.54653 & 0.05440 & 0.00260 & 1 \end{bmatrix} (0.0634, 0.1024, 0.1268, 0.0693)^T} \approx 0.19681.$$

The historical market prices of all subjects at the current stage can be known according to the above empirical analysis results. Through the matrix analysis and calculation, the VAR value under 95% confidence level is 0.19681 hundred million yuan, which actually proves that the probability that the company may lose less than 0.19681 hundred million yuan on the current day in the current market economy environment was 95%.

At the same time, the above calculation results also prove that the total VAR of δ is 0.19681 hundred million yuan, which is greater than any single factor VAR obtained above but not greater than the VAR value promoted by total factors. Objectively, the differential investment objects provided in the asset portfolio process are closely correlated. For this reason, the result obtained with the matrix method can ensure the normal VAR of δ is more consistent with the actual situation. Similarly, on this basis, the insurance company can obtain the total VAR and then determine the ultimate loss amount under the extreme situation of 5%, so as to withdraw the reserves in advance and take measures to control the investment risks.

5 Conclusion and Suggestions

Through the above research, despite certain achievements of China's financial derivatives in the development process, there is still much room for improvement. In order to control the risks of financial derivatives in the market economy more effectively, the following aspects must be done:

The first is to strengthen legislation, so as to create a good institutional environment for the development of financial derivatives. The government's laws, regulations and systems on the financial derivative industry are very important for the development of the financial market. Thus, China needs to improve the relevant regulations and systems, make the comprehensive supervision and management of the specific trading processes in the industry, create the reasonable and orderly trading environment for the futures market, protect the interests of both parties to the maximum and realize the good development of the financial derivatives.

The second is to strengthen the external supervision of the finance supervision institutions and ensure the standard development of the financial derivative market. The financial derivative industry in Europe, the US and Japan develops in a good environment, in which the governments play an important role despite their different supervision ways. The strong supervision of the governments ensures the orderly and reasonable development of the market and the bright prospects of the financial derivatives.

The third is to strengthen the supervision on the financial derivative transaction and ensure the fair trading in the exchange. As the important sector managing the financial market and the main place for the financial derivative trading, on one hand, the exchange supervises and manages the orderly development of the financial derivatives; on the other hand, it creates the open and equal environment for the development of the financial market. Considering that the exchange is a sector of investment operation, it should also be fully supervised.

The fourth is to give full play to the role of the self-regulation organization in supervising the financial derivatives. China should draw lessons from the foreign self-regulation organizations of finance to make ours more complete and play the supervision role.

The last is to build the risk measurement model of the financial derivative trading. In order to more accurately predict the potential risks in the financial transactions, the financial enterprises adopt the risk measurement model. The VAR model which is widely used in the foreign countries is an obvious example. VAR model makes precise prediction of the risks in transaction and prevents risks timely or reduce the losses to the minimum when the risks occur in the process of financial derivative transaction, which is conductive to the successful development of the financial derivative market.

References

[1] Han L, Zhang X 2009 Derivatives investment loss by state-owned enterprises: case study and supervision proposals *Management Review* **12** 117-25 (*in Chinese*)

[2] Zhang C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34 (*in Chinese*)

[3] Li R, Wu Y 2009 How the enterprise management of financial derivatives investment risk: based on the Hong Kong dollar CITIC Pacific exchange contract Jukui case *Auditing and Finance* **0**2 5-7 (*in Chinese*)

[4] Li M 2008 Internal control over derivative instruments *Accounting Research* **01** 39-46 (*in Chinese*)

[5] Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103

[6] Ding H 2009 Analysis on Citicpacific's Huge Loss on FX Derivatives Investment and Its Implications *South China Finance* **03** 29-31 Internal control over derivative instruments *Accounting Research* **01** 39-46 (*in Chinese*)

[7] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12** (2) 171-81 (*in Chinese*)

[8] Su Z 2008 Financial risk and internal control of financial derivatives *China Audit* **05** 63-4 *(in Chinese)*

Authors

Hailei Zhao, January 1975, Zhenjiang City, Jiangsu Province, P.R. China.

Current position, grades: **lecturer at the School of Business, Jiangnan University, Wuxi City, Jiangsu Province, China.**
University studies**: PhD student of School of Finance, Shanghai University of Finance and Economics.**
Scientific interests: **systemic risk, credit risk and corporate finance.**
Publications: **more than 10 papers.**
Experience: **teaching experience in finance for more than 15 years.**

Dehuan Jin, April 1953, Shanghai, P.R. China.

Current position, grades: professor and doctoral supervisor of School of Finance, Shanghai University of Finance and Economics, Shanghai City, China.
Scientific interests: financial market and financial engineering.
Publications: more than 20 papers.
Experience: teaching experience in finance for more than 30 years.

A study on the relationship between the developments of China's trading economy and environmental factors

Xueying Zheng*

Economic and Trade College, Zhejiang Business College, Hangzhou, Zhejiang Province, China, 310053

Received 28 October 2014, www.cmnt.lv

Abstract

Trade environment refers to the social and economic environment facing trading economy, and it can directly affect the pace of trade development and trade structure. Establishing a good trade development environment can optimize trade structure, and promote the sustained and healthy economic development. This paper firstly analyzes the current environment of China's trade development with relevant theories, then analyzes the relationship between environmental factors and the development of trading economy based on indexes, builds error correction model, and uses the model to make in-depth analysis on the impacts of environment on the development of trading economy. According to the study results, the ratio of trade dependency over trade surplus exerts side effects on the development of trading economy, while trade openness exerts positive effects on the development of trading economy. In terms of import, the development of China's trading economy shows excessive dependence on foreign energy and other resource-based products, this seriously affects the stable development of economy. In terms of export, some Chinese companies that have overcapacity show over-reliance on international markets, thus reducing the competitiveness of export products, and leading to the situation where China's trading economy is vulnerable to fluctuations of international markets. Based on this, a method is proposed in the paper to improve China's trade environment.

Keywords: error correction model, trading economy, environment; impact

1 Introduction

The economic relations between countries have shifted from simple import and export trading relationship to various forms of economic relations; import and export of goods has evolved into multiple forms, including labor input and output, commodity trade, technology trade and service trade. From commodities that can meet material enjoyment to tourism that can meet the spiritual enjoyment, trade structure and trade patterns have undergone great changes [1,2]. Trade has a very important position in international economic relations. All the world's developed countries are inward-looking, they are interdependent with other countries in terms of economy, so are the United States, Japan, European countries and other developed countries, these countries exert economic impacts on other countries while being under the influence of other countries, and this mutual influence phenomenon is caused by the "transfer" channel named foreign trade [3-5]. Since the 1990s, the rapid development of foreign trade to has made great contributions to China's economic growth. The breadth, depth and speed of China's foreign trading economy have surpassed any other period in history. Dynamic benefits generated from foreign trade enable economic sectors at all levels to have economic growth, thus boosting the overall growth of the national economy [4-6]. Meanwhile, the development of foreign trade has effectively promoted China's diplomatic work, and made significant contributions to the establishment of favorable international environment for China's modernization construction. At present, China has become the world's largest trading nation; China's foreign trade has exerted significant influences on the world trade environment, while China's foreign trade can also be affected by the external environment. Based on the connotation of Kuznets Curve model [7-9], the paper presents an in-depth analysis of the internal and external environmental factors that affect China's foreign trade, and explores the countermeasures for China's trading economy, in order to promote the rapid and healthy development of China's foreign trade [10,11].

2 Analysis of China's current trade environment

First, China has increasing dependence on some foreign resources, especially energy and resource-based commodities. With the continuous development of China's economy, China has become a large energy consumption country. It is estimated that the degree of dependence on foreign energy is more than 80%, while some energy resources are almost 100% imported from other count-

* ***Corresponding author's*** e-mail: ysgao218@163.com

ries. Therefore, with the rapid economic development, China has increasing dependence on soybeans, iron ores, fish meal and other resource-based commodities. With increasing import quantum, China will be one of the world's major importers of energy and resource-based commodities. At present, the degree of China's dependence on imported crude oil, iron ore, aluminum and copper ores has reached 90%, with an increase of 40%. However, China is often forced to accept price increases in international trading of energy and resource-based commodities, and is subject to the external environment. In recent years, the price of various resources including fuel has constantly risen, resulting in the overall increase of China's domestic price level. This situation will be worsened with the development of China's economy.

Secondly, China's economic dependence on international markets is gradually increasing. In recent years, China's labor-intensive goods are basically saturated in the domestic market, this is mainly due to the international market. Therefore, if they are unsalable in the international market, foreign trade enterprises will suffer lethal blow, China's economy will also suffer huge damages, and the damages will increase with the increase of our export value. In addition to transferring excess capacity, China's dependence on the international market environment has also increased significantly in terms of import and export of goods and services, labor and employment. According to statistics, about 100 million people directly work for China's export trade, including 20 million in the textile industry with severe overcapacity of about 20 million, and more than 60 million in electronic product industry. It is estimated that China's output of shoes accounts for 60% of the world's total sales, the output of DVD accounts for 50% of the world's total sales. It can be seen that many China's sectors have over-dependence on the international market.

Thirdly, China still has great dependence on foreign product patents and core technologies. Due to lack of digestion and absorption ability of technologies and lack of innovative capacity, although China is the largest producing country of color TV, computers, mobile phones, stereos and other electronic equipment. China does not have the core production technology of these products, and has to import some key components and equipment. According to statistics, foreign enterprises control 85% of China's integrated circuits, more than 85% of pharmaceutical products, over 80% of chips, more than 70% of CNC machine tools and textile machinery, and more than 90% of automobile patents. This means that China's domestic industrial development and export trade are hollow. As China needs to pay expensive royalties to foreign countries, production costs are increasing, China's technology innovation and development is vulnerable to pressure from foreign companies, forcing many state-owned enterprises and private enterprises to withdraw from the international market competition, thus affecting the core competitiveness of China's foreign trade.

Fourthly, China's foreign trade frictions are increasing, the scope of trade barriers is expanding. China has maintained the trend of trade surplus since the 1990s, and the scale has been increasing; this exerts varying degrees of impacts on related industries of developed countries and developing countries. With China's rapid economic development, China not only seized the international share of developing countries' competitive industries, but also occupied the domestic market, seriously affecting and even marginalizing their competitive industries and pillar industries. Therefore, some countries and regions constantly set various barriers on China's products in order to protect their own industries, and even cause trade disputes. Since 1979, foreign countries have launched a total of more than 1000 cases of trade disputes, involving 4000 kinds of products exported from China, and affected China's exports of nearly $ 20 billion. According to WTO statistics, China's trade disputes have accounted for 15% of WTO anti-dumping cases since 1995, and these trade frictions are mainly caused by Chinese labor-intensive products that have competitive edges.

3 The construction and analysis of error correction model

3.1 MODEL

Based on the above analyses, indexes including the ratio of total export-import volume over gross domestic product (GDP) (JZ), export dependence (YN), the ratio of trade surplus (MY) and the growth rate of total export-import volume (ZZ) are set, in order to analyze the relevance and establish an error correction model. The data are collect from statistical yearbook of 1979 to 2012 on China Statistical Information Network, they are sorted as shown in Table 1.

TABLE 1 Index data

Index	JZ	YN	MY	ZZ	Index	JZ	YN	MY	ZZ
1979	0.1119	0.0521	-0.1284	0.2806	1996	0.3391	0.1767	0.0882	0.0270
1980	0.1254	0.0597	-0.0924	0.2538	1997	0.3415	0.1920	0.2841	0.1174
1981	0.1503	0.0751	-0.0003	0.2900	1998	0.3181	0.1804	0.3094	-0.0044
1982	0.1449	0.0777	0.1575	0.0490	1999	0.3334	0.1802	0.1764	0.1135
1983	0.1442	0.0735	0.0391	0.1151	2000	0.3958	0.2080	0.1071	0.3137
1984	0.1666	0.0805	-0.0645	0.3963	2001	0.3847	0.2009	0.0925	0.0741
1985	0.2292	0.0897	-0.3569	0.7208	2002	0.4270	0.2239	0.1031	0.2180
1986	0.2511	0.1053	-0.2778	0.2486	2003	0.5189	0.2672	0.0612	0.3719
1987	0.2558	0.1219	-0.0893	0.1952	2004	0.5976	0.3071	0.0574	0.3555
1988	0.2541	0.1174	-0.1403	0.2392	2005	0.6322	0.3388	0.1543	0.2238
1989	0.2446	0.1151	-0.1108	0.0874	2006	0.6517	0.3587	0.2244	0.2057
1990	0.2978	0.1599	0.1598	0.3379	2007	0.6278	0.3520	0.2764	0.1836
1991	0.3317	0.1757	0.1260	0.2996	2008	0.5729	0.3197	0.2624	0.0783
1992	0.3387	0.1737	0.0524	0.2621	2009	0.4419	0.2406	0.1954	-0.1627
1993	0.3190	0.1496	-0.1172	0.2359	2010	0.5024	0.2665	0.1301	`0.3390
1994	0.4229	0.2162	0.0464	0.8083	2011	0.4997	0.2605	0.0891	0.1719
1995	0.3866	0.2048	0.1271	0.1530	2012	0.4700	0.2490	0.1268	0.0328

Firstly, unit root test is carried out. ADF test is adopted, it is found that under the condition of no difference, they both have unit root; while after difference of first order, unit root is removed, and the sequences begins to be stable. For details, please see Table 2. Thus, an error correction model is established.

TABLE 2 The ADF test results of variables

Variable	t-Statistic	Prob.*
JZ	-4.6538	0.0007
MY	-5.0222	0.0003
YN	-4.9718	0.0003
ZZ	-6.5761	0.0000
1% level	-3.6537	
5% level	-2.9571	
10% level	-2.6174	

Then, the growth rate of total export-import volume (ZZ) is taken as the dependent variable, while the ratio of total export-import volume over gross domestic product (JZ), export dependence (YN) and the ratio of trade surplus (MY) are taken as independent variables, the least square method is adopted for correlation analysis. Related technologies are -3.4624, -0.5088 and 2.1755, then the error correction term can be established, it is Equation (1).

$$\begin{aligned}&\text{ECM=ZZ(-1)+3.4624*YN(-1)+}\\&\text{0.5088*MY(-1)-2.1755*JZ(-1)}\end{aligned}. \quad (1)$$

Then, the growth rate of total export-import volume (ZZ) is taken as the dependent variable, while the ratio of total export-import volume over gross domestic product (JZ), export dependence (YN) and the ratio of trade surplus (MY) are taken as independent variables for correlation analysis, the results are shown in Table 3. As can be seen from the data in Table 2, the tail probability of all variables is less than 0.01, therefore it meets the requirements at a confidence level of 1%, and the relevant probability is 0.6457, the degree of correlation is very high. Therefore, the fitting degree of dependent variable and independent variables is very good, the model construction is very successful.

TABLE 3 The error correction model parameter estimation

Variable	Coefficient	Std. Error	t-Statistic	Prob.
YN(-1)	-5.6146	3.93E-13	-8.82E+12	0.0000
MY(-1)	-0.2551	2.97E-14	-1.71E+13	0.0000
JZ(-1)	5.8537	2.06E-13	1.05E+13	0.0000
ECM	-0.9841	8.99E-15	1.11E+14	0.0000
C	0.0896	4.52E-15	-2.726609	0.0109
R-squared	0.6457	Mean dependent var		0.2303
Adjusted R-squared	0.6213	S.D. dependent var		0.1842

According to Table 2, the data model can be constructed as follows:

$$\begin{aligned}&D(ZZ)=-5.6146\cdot \mathrm{D(YN)}-0.2551\cdot \mathrm{D(MY)}+\\&5.8537\cdot \mathrm{D(JZ)}-0.9841\cdot \mathrm{ECM}+0.0896+\mu\end{aligned}. \quad (2)$$

In which:

$$\begin{cases}D(ZZ)=ZZ-ZZ(-1)\\D(YN)=YN-YN(-1)\\D(MY)=MY-MY(-1)\\D(JZ)=JZ-JZ(-1)\end{cases}. \quad (3)$$

If Equation (1) and for Equation (3) are put into Equation (2), then:

$$\begin{aligned} &ZZ - ZZ(-1) = \\ &\text{-}5.6146 \cdot [YN - YN(-1)] - \\ &0.2551 \cdot [MY - MY(-1)] + \\ &5.8537 \cdot [JZ - JZ(-1)] - \\ &0.9841 \cdot [\text{ZZ(-1)} + 3.4624 \cdot \text{YN(-1)} + \\ &0.5088 \cdot \text{MY(-1)} - 2.1755 \cdot \text{JZ(-1)}] \end{aligned} \quad . \quad (4)$$

After sorting Equation (4), then:

$$\begin{aligned} &ZZ = 0.0159 \cdot ZZ(-1) - 0.6146 \cdot YN + \\ &2.5140 \cdot YN(-1) - 0.2551 \cdot MY - \\ &0.2456 \cdot MY(-1) + 5.8538 \cdot JZ - \\ &3.7127 \cdot JZ(-1) \end{aligned} \quad . \quad (5)$$

3.2 MODEL ANALYSIS

According to the error correction model, the following points can be seen:

Firstly, the development speed of trading economy has positive correlation with the ratio of total import-export volume over gross domestic product (GDP). The total import-export volume can reflect China's domestic overcapacity. When overcapacity becomes more serious, the development of trading economy can also address capacity issues, thus promoting the growth of trading economy to some extent. However, with overcapacity and excessive development of trading economy, if the ratio of trade economy over the total domestic economy increase, trading economy will exert more impacts on the growth of national economy, China's economy is more vulnerable to the impacts of estimated markets. Meanwhile, the larger the proportion of import-export volume in GDP, the greater the degree of China's opening up, and this can also promote the development of trading economy.

Secondly, the development speed of trading economy has positive correlation with export dependence, but is negatively correlated with the current period export dependence. It is in proportional with export dependence, because current exports are an important part of the current trade economy, therefore the larger the proportion is, the more the volume of trading economy. It is in inversely proportional to previous exports, overall speaking, it is in inversely proportional to export dependence, this indicates that the pace of development speed of trading economy is negatively correlated with trade dependence. This negative correlation indicates that export dependence will exert side effects on the development of trading economy, the greater the export dependence is, the more side effects will be exerted on trading economy. This is mainly because that great export dependence will put China's economy and foreign trade under the influences of fluctuations in international markets, and be short of bargaining power.

Thirdly, the development speed of trading economy has positive correlation with the ratio of trade surplus. This is mainly because that trade surplus will bring more resistances, such as trade friction, deteriorated international trade relations, thus impeding the development of trading economy. This is also the main reason for increasing trade disputes and trade tensions over the years when China maintained trade surplus. Other countries will inevitably take various measures to protect their products, such as preparing various high-specification provisions to limit the entry of Chinese products, or using of a variety of national trade laws to hamper the entry of Chinese products.

4 Measures to improve the trading environment

It can be seen from above analyses that the ratio of trade dependency over trade surplus has side effects on the development of trading economy, and openness degree of trade has positive effects on the development of trading economy. Based on this, in order to promote the growth of China's trading economy and promote economic growth, it is recommended to improve the trade environment from the following perspectives:

Firstly, the construction of international trade environment should be enhanced. This is not only an economic task, but also a political mission. Trade surplus is an inevitable part in the development of China's trading economy. Completing the shift from trade surplus to trade deficit is the basic task of strengthening the construction of international trading environment. Therefore, emphasis should be attached to the maintenance and construction of China's trade environment, promoting the development of foreign trade and construction of international trade environment. A good international trade environment is essential for us to seize strategic opportunities and realize more strategic goals, and it is also a necessary external condition to construct China's harmonious society and achieve peaceful development.

Secondly, to constantly balance the development of China's industries, improve the technical level and reduce some industries' dependence on international trade. Over dependency on international markets, technologies and resources will exert adverse effects on the development of China's trading economy, and affect the domestic economy; this also reflects the imperfections of China's industrial development. The current imbalance in China's industrial development is very serious, the over dependency on imports leads to the increasing price of imported products, and unstable domestic economy. Technology is indeed the vital dilemma in China's economic development, the trade pattern can only be improved when great technological breakthroughs are made.

Thirdly, to actively participate in the competition in international market, and keep the balanced development of foreign and domestic trades. Trading economy is very important, so great attention should be paid to it. However, China, as a major consumer market in the

world, is also China's economic center. Therefore, when promoting the development of trading economy, domestic market should be protected and occupied, the balanced development of domestic and foreign trades should be constantly promoted. In fact, establishing a stable domestic market is the first step of establishing a stable trade environment, and the basic measure to promote the development of foreign trade and trading economy.

References

[1] Wang B 2012 On the influence of international trade environment of our country foreign trade policy C*hina Business & Trade* **12** 219-20
[2] Liu L 2009 The theoretical and empirical analysis of foreign trade of China's environmental effects *Journal of International Trade* **03** 70-7
[3] Jin H 2012 In the new international environment "China trade" required depth adjustment *Economic Information Daily* **08** 31008
[4] Yu H 2006 Analysis of international trade environment effect *Economic Theory and Business Management* **08** 71-5
[5] Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3), 98-103
[6] Wu H Economics analysis for environment problem in international trade *Journal of Hubei Institute of Education* **09** 53-6
[7] Li G, Ma S 2007 The effect of foreign trade on environment: a literature review and evaluations *Ecological Economy* **04** 40-3
[8] Zhang C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34
[9] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81
[10] Liu D, Sun J 2005 The influences of trade environment to adjustments about international production system *Studies in Science of Science* **S1** 113-8
[11] Gong L 2008 With the changes and foreign trade strategy adjustment direction of foreign trade environment in China *Commercial Research* **06** 161-4

Author

Zheng Xueying, born in August 1972, Hangzhou, Zhejiang Province, P.R. China.

Current position, grades: the associate professor of Zhejiang Business College, Hangzhou, Zhejiang Province, China.

Scientific interest: international trade and E-commerce.

Publications: more than 10 papers.

Experience: teaching experience of 14 years, 6 scientific research projects.

Empirical analysis of Chinese cultural products trade based on the gravity model

Pingping Qiao*

Department of Economy and Management, Henan Polytechnic Institute, Nanyang, Henan Province, China, 47300, China

Received 28 October 2014, www.cmnt.lv

Abstract

Based on date of bilateral cultural trades between China and 26 other countries (regions), the paper uses a gravity model to perform an empirical test on the influencing factors and export potential of cultural products. The research shows that Chinese economic size, importing country's economic scale, GDP per capital, trade openness and preferential trade arrangements have a positive effect on the export of Chinese cultural products. Spatial distance and cultural distance from China have a negative effect on exports of Chinese cultural products. The effect of cultural distance is bigger than that of spatial distance. It is very important for promoting the exports of Chinese cultural products to strengthen communication and overcome "cultural discount". In the 26 samples, the trades of cultural products between China and 7 countries (regions) are sufficient, and the trades of cultural products between China and other 19 countries (regions) are insufficient. China has a great potential for cultural products trades, but China should develop it targeted.

Keywords: gravity model, cultural products trade, influencing factors, export potential

1 Introduction

The cultural products have the following features of low consumption, low pollution and high added value, therefore, it becomes a strategic pillar industry which is an important path for many countries to develop and enhance their "soft power". According to international experience, when the ratio of output value in a certain industry and GDP is more than 5%, the industry can be called a pillar industry. In 2010, the cultural industry accounted for 2.75% of GDP in China, and there is still a large gap to reach the target of 5%. As a result, it's imperative to develop the cultural industry in China.

Accelerating the development of cultural trade and expanding the scale of foreign trade of cultural products, which is an important means to promote the cultural industry to be the pillar industry of China. Therefore, we need to research the factors that determine the scale and direction of China's cultural trade, and explore the export potential of cultural products. This paper analyzes the influence factors and export potential of China's cultural products trade based on the gravity model.

The gravity model is used in various social sciences to predict and describe certain behaviors that mimic gravitational interaction as described in Isaac Newton's law of gravity. Gravity model has been widely used in areas of traditional trade. With the continuous expansion of global cultural trade, gravity model has gradually been studied in the field of culture trade. Schulze introduced the gravity model to the field of culture trade first, and he found that the trade flow of artistic products is proportional to the economy scale of the countries, and inversely proportional to the distance [1]. Disdier et al added dummy variable such as common language and colonial relationship to the original gravity model, and he found that, compared with non-cultural products, cultural products trade used to happen in countries that have a short distance [2]. The empirical research from Ferreira & Waldfogel shows that the scale of cultural trade between the close countries that have the same language will be large [3]. They also noted that, with the development of transport and communication technology, the spatial distance between the countries will be shorten, but the spatial distance will be one of the main factors that influence the cultural products trade.

Cultural products that are different from the traditional products will have the phenomenon of "cultural discount" in international trade. "Cultural discount" refers to the value of the cultural products that will be discount in another country because of the cultural differences. The experts usually use the cultural distance to measure the possibility of cultural discount. The research from Disdier et al shows that the more similar the culture of the two countries is, the greater the scale of bilateral trade will be [4]. The paper calculates the export potential of China's cultural products trade with major trading partners on the base of the empirical analysis of the factors affecting Chinese cultural products trade.

2 Application of gravity model

2.1 CONSTRUCTION OF GRAVITY MODEL

The gravity model is a basic law in physics, which pointed out that the gravitational force between two objects is directly proportional to the product of their quality, and inversely proportional to the square of the distance between them.

* ***Corresponding author's*** e-mail: 59103214@qq.com

The gravity model was applied first by Tinbergen (1962) to the study of international trade, introducing GDP of the both sides and geographic distance as explanatory variables to the analysis of bilateral trade volume. The basic form is as follows:

$$T_{ij} = \frac{AY_iY_i}{D_{ij}},$$

T_{ij} represents the bilateral trade volume between country i and country j, A is a constant, Y_i represents GDP of country i, Y_i represents GDP of country j, D_{ij} represents the geographic distance between country i and country j.

According to the characteristics of cultural products trade, the paper added a new explanatory variable in the original gravity model to extend and modify the model, and established the model for the export of Chinese cultural products. Meanwhile, in order to overcome the heteroskedasticity of gravity model, take the natural logarithm on both sides of gravity model, and the final econometric model is as follows:

$$LnEX_{ijt} = A_{ijt} + \beta_1 LnGDP_{it} + \beta_2 LnGDP_{jt} + \beta_3 LnY_{jt} + \beta_4 LnSD_{ij} + \beta_5 LnCD_{ij} + \beta_6 LnOpen_{jt} + \beta_7 FTA_{ijt} + \varepsilon_{ijt}.$$

The subscript i represents China, t represents the time. A_{ijt} represents constant term. EX_{ijt} represents the export of Chinese cultural products to j country(region). GDP_{it} represents China's GDP, and GDP_{jt} represents the GDP of j country (region) as importer. Y_{jt} represents the per capita GDP of j country (region). SD_{ij} represents the spatial distance between China and j country (region). CD_{ij} represents the cultural distance between China and j country (region). $Open_{jt}$ represents the degree of opening in j country (region). FTA_{ijt} is a virtual variable that represents the situation whether China and the importing country (region) j are belong to a trade group. When the two countries are both in the same trade group, the trade flows between them will increase due to preferential trade arrangements. Therefore, FTA_{ijt} was used to investigate the effect of the national or regional economic cooperation on cultural trade. ε_{ijt} represents deviation item.

2.2 SELETION OD DATA

At present, the popular standard of cultural trade statistics is the "Framework for Cultural Statistics" (FCS for short) issued by UNESCO. "2009 Framework for Cultural Statistics of UNESCO" (2009FCS for short) is the new framework for cultural statistics issued by UNESCO in 2011.

Generalized fields of culture are divided into the narrow field of culture (the field of culture) and related fields. The field of culture represents a series of cultural production, activities and practices, which can be classified into the following categories: cultural and natural heritage (coded as A); performances and celebrations (coded as B); visual arts and handicrafts (coded as C); books and newspapers (coded as D); audio-visual and interactive media (coded as E); design and creative Services (coded as F). In addition, it also includes four horizontal fields: non-material cultural heritage, archiving and protection, education and training, equipping and supporting material. The related fields include sports, recreation and tourism. Due to some difficulties in the statistics of horizontal field and related field, the research does not include horizontal field and related field, but only six categories A-F of cultural field (called as culture products by UNESCO), which represent the most core culture field.

UNESCO also encourages countries to collect comparable data at least in these areas. "2009FCS" not only delineate the width of the culture field, but also provides SITC code that is accurate to 5 digits in A-F of cultural field.

According to the SITC code of "2009FCS", the bilateral trade data of cultural products have been obtained from the United Nations' commodity trade database (comrade database). The paper selects 16 developed countries and 10 developing countries (regions) as the sample. In 2010, the export of Chinese cultural products to the 26 countries (regions) accounted for 83.91% of China's total export of cultural products; therefore, the sample is representative. GDP and per capita GDP are from WEO database of the IMF site. The spatial distance between China and the importing country (region), indicated by the distance between the capitals of whom, is from the "distance calculator" of www.indo.com. The trade openness index $Open_{jt}$ is the ratio of total goods trade and GDP of the importing country. The total goods trade is from the official website of WTO. The paper takes the Asia Pacific Economic Cooperation (APEC) as the trade group. If China and the importing country (region) are both in APEC, $FTA_{ijt} = 1$, otherwise, $FTA_{ijt} = 0$.

Cultural distance (CD_{ij}) is used to measure the cultural differences between China and the importing country (region). The paper takes Hofstedes research method of five cultural dimensions for reference, the formula of that is as follows:

$$CD_{ij} = \frac{1}{5}\left[\sum_{k=1}^{5} I_{kj} - I_{ki}\right]^2 / V_k,$$

I_{kj} represents the score of country (region) j in the cultural dimension of k, I_{ki} represents the score of China in the cultural dimension of k, V_k represents the variance of cultural dimension k. The five cultural dimensions are power distance, uncertainty avoidance, individuallism/collectivism, degree of male/female, and the orientation of long/short term. The scores of China and its 26 trading partners in 5 cultural dimensions are all from the Hofstede website. The greater CD_{ij} is, the larger cultural distance will be, which indicates the cultural difference between China and its importing country (region).

The period of sample is 1992-2010. The paper makes an empirical analysis with panel data. The econometric method of panel data combines advantages of time series data and cross-section data, which is more comprehensive to reflect the individual characteristics of the research objects, and makes the regression model more persuasive

than pure time series model and cross section model due to the expansion of the sample capacity.

3 The regression results of the gravity model

In order to test the robustness of the gravity model, the paper uses the method of stepwise regression, and gets the regression results in Table 1.With the gradual adding of independent variables. The goodness of fit (R^2) for the model is improved. The goodness of fit reaches 0.85, and the explanatory power of the independent variables on the equation reaches 85%. The independent variables are stable, and the statistic values (t) of the independent variable pass the significance test, which shows that the model has good robustness.

TABLE 1 regression results of gravity model

	Equation1	Equation2	Equation3	Equation4	Equation5	Equation6	Equation7
constant	8.686 (10.140)	5.020 (5.694)	4.591 (6.385)	5.814 (6.235)	5.672 (8.125)	1.618 (2.587)	1.503 (2.539)
$LnGDP_{it}$	1.369*** (11.164)	1.141*** (9.906)	0.874*** (9.139)	0.332*** (6.541)	0.529*** (13.406)	0.268*** (7.379)	0.303*** (8.744)
$LnGDP_{jt}$		0.546*** (9.166)	0.331*** (6.520)	0.888*** (9.262)	0.773*** (10.738)	0.546*** (9.067)	0.541*** (9.495)
LnY_{jt}			0.698*** (14.585)	0.699*** (14.608)	0.789*** (21.829)	1.215*** (29.954)	1.194*** (31.007)
$LnSD_{ij}$				–0.164** (–2.059)	–0.165*** (–2.779)	–0.117** (–2.417)	–0.120*** (–2.620)
$LnCD_{ij}$					–1.252*** (–18.540)	–0.786*** (–12.547)	–0.718*** (–11.948)
$LnOpen_{jt}$						1.128*** (15.175)	1.037*** (14.511)
FTA_{ijt}							0.234*** (7.292)
R^2	0.220	0.346	0.556	0.560	0.751	0.835	0.852
Adjust–R^2	0.219	0.343	0.553	0.557	0.748	0.833	0.850

Notes: ***, **,* represent 1%, 5%,10% of significant level, The value in brackets is the value of t.

In the gravity model of Chinese cultural products export the elastic coefficient of the independent variable $LnGDP_{it}$ is 0.303, which indicates that when China's GDP increases by 1%, the export of Chinese cultural products will increase by 0.303%. The expanding of China's economy scale will promote the export of cultural products. The elastic coefficient of the independent variable $LnGDP_{jt}$ is 0.541, which indicates that when GDP of the importing country (region) increases by 1%, the export of Chinese cultural products to it will increase by 0.541%. The rising economic might will be conducive to China's cultural products export.

The elastic coefficient of the independent variable LnY_{jt} is 1.194, and its absolute value is the largest in that of all elastic coefficients with the most significant t, which indicates that when per capita GDP of the importing country (region) increases by 1%, the export of Chinese cultural products to it will increase by 1.194%. The per capita GDP of the importing country (region) has great effect on the export of Chinese cultural products.

The elastic coefficient of the independent variable $LnSD_{ij}$ is -0.120, which indicates that when the spatial distance between China and the importing country (region) increases by 1%, the trade flows of the cultural products will reduce by 0.12%. There is a negative correlation between the spatial distance and the export of Chinese cultural products. The spatial distance is still one of the main factors influencing China's cultural products export. The elastic coefficient of the independent variable $LnCD_{ij}$ is -0.718, which indicates that when the cultural distance between China and the importing country (region) increases by 1%, the trade flows of the cultural products will reduce by 0.718%. China tends to export the cultural products to the countries (regions) with the similar culture, which is consistent with the conclusion of Disdier et al (2010) [5,6]. From the absolute value of the elastic coefficient, elastic coefficient of cultural distance is far greater than the elastic coefficient of spatial distance, therefore, the effect of cultural distance on the export of Chinese cultural products is much greater than that of spatial distance.

The elastic coefficient of importing country's (region's) trade openness ($LnOpen_{jt}$) is 1.037, which indicates that when the trade openness of the importing country (region) increases by 1%, the export of Chinese cultural products to it will increase by 1.037%. The liberalization of global cultural trade would promote China's export. The elastic coefficient of FTA_{ijt} is 0.234, which indicates that preferential trade policy would promote China's export of cultural products, if the importing country (region) is the member of APEC.

4 The export potential calculation of Chinese cultural products

The calculation of export potential compares the actual export value with the simulate export value, applying potential export value of product in "theory" or "natural" state of the gravity model.

If actual value / simulate value>1, the actual trade value between the two countries is greater than the simulated values, which indicates "excessive trade"; if actual value / simulate value<1, the actual trade value between the two countries is less than the simulated values,

which indicates "insufficient trade" and the trade potential. Taking 2010 as an example, the paper calculates the export potential of Chinese cultural products. The results are shown in Table 2.

TABLE 2 Export potential of Chinese cultural products in 2010 (Unit: million dollars)

Country (region)	Trade volume (actual value)	Trade volume (simulate value)	Actual value/Simulate value
Australia	408.290	402.352	0.985
Brazil	145.312	321.388	2.212
Britain	1471.281	1081.288	0.735
Canada	683.725	397.160	0.581
Denmark	744.094	50.655	0.068
Philippines	51.854	106.540	2.055
Finland	106.043	65.844	0.621
France	1810.142	418.791	0.231
Germany	3147.851	1526.109	0.485
Hong Kong of China	875.983	7589.268	8.664
India	195.644	297.329	1.520
Italy	549.314	486.593	0.886
Japan	3814.819	1085.347	0.285
South Korea	869.123	346.982	0.399
Malaysia	187.661	196.129	1.045
Mexico	326.426	111.397	0.341
Holland	593.046	718.851	1.212
New Zealand	39.273	33.629	0.856
Portugal	36.466	30.449	0.835
Russia	469.033	319.456	0.681
Singapore	527.868	260.250	0.493
Spain	333.325	275.972	0.828
Sweden	172.977	81.253	0.470
Switzerland	322.573	32.584	0.101
Thailand	167.646	129.896	0.775
America	4591.946	8240.661	1.795

In the calculation of China's export potential to 26 countries, 7countries (regions) who are America, Holland, Hong Kong, Brazil, India, Philippines and Malaysia show "excessive trade", while 19 countries (regions) who are Japan, Italy, France, Germany, South Korea, Mexico, Russia, Singapore, Thailand and so on, show "insufficient trade". Ranking first is Hong Kong of China (8.644), and the last is Denmark (0.068), which shows that the potential of China's export to Denmark is huge. In addition, the ratio of actual value and simulate value in 7 countries, who are France, Germany, Japan, South Korea, Mexico Sweden and Switzerland, is less than 0.5. It indicates that the potential of China's export to the 7 countries above is relatively large. Overall, China's export of cultural products to its main trading partners is in the situation of "insufficient trade".

5 Conclusions and suggestions

The paper selected the panel data of China's cultural products export trade with other 26 countries (regions) in 1992~2010, using the gravity model to empirically analyze the influence factors of Chinese cultural products export trade [7]. On this basis, the paper calculated China's potential of cultural products export. According to the results of empirical analysis, the paper obtained the following conclusions:

First, the traditional factors are still the main factors affecting China's cultural products export. GDP and per capita GDP of China and import country (region) have the positive effect on China's cultural products export, while the spatial distance has the negative correlation with China's cultural products export.

Second, China's cultural products export has its own features. Compared with the space distance, the cultural distance has much more negative influence on China's cultural products export. The trade openness of importing country (region) and whether the trade partners belong to a free trade area, which have a positive effect on China's cultural products export.

Third, Chinese cultural products have great export potential. The relevant research shows that, the export of China to the major trading partners is on the situation of "insufficient trade", the ratio of the actual value to the simulation value is 0.740. In China's 26 major trading partners, only 7 countries' (regions') cultural products trade with China show "excessive trade", and 19 countries (regions) belong to the "insufficient trade".

Based on the research conclusions above, the paper believes that the following measures should be taken to expand the export scale of Chinese cultural product, to promote the development of China's cultural industry.

5.1 TO DEVELOP THE REAL ECONOMY VIGOROUSLY

GDP of the exporting country (region) is one of the important influence factors which influent the cultural products export. The sustained growth of China's GDP is conductive to improve the export capacity of cultural products. The enhancement of cultural products' export capacity, will not only further stimulate China's GDP growth in return, but also promote the upgrading of the national completive structure. However, for a long time, China focused on traditional goods export and ignored the cultural products' export [8]. Therefore, China should not only strive to develop the real economy, but also actively promote the export of cultural industry and make the strategic policy of culture trade:

1) To encourage the establishment of cultural industry organizations (or association), in order to strengthen the development and protection of traditional culture resources;

2) To increase the financial support on the culture industry, for example, supplying export credit subsidy and export credit insurance for the relevant cultural enterprises;

3) To promote the cooperation between cultural enterprises;

4) to promote implementation of the strategy that is "going out" of the domestic cultural enterprises.

5.2 TO STRENGTHEN THE CULTURAL COMMUNICATION AND SHORTEN THE CULTURAL DISTANCE IN ORDER TO AVOID "CULTURAL DISCOUNT"

The cultural distance between import and export country (region) is a very important factor influencing the export of Chinese cultural products. Shortening the cultural dis-

tance to overcome the "cultural discount", which will promote China's export of cultural products. The government should:

1) Strengthen international cultural communication, positive publicize and spread Chinese cultural products;

2) Hold high level cultural communications to promote the understanding of Chinese culture in the world and make Chinese elements melt into the world;

3) Vigorously develop the business of teaching foreigner Chinese through the Kong Zi College;

4) Spread China's language and culture philosophy in the world;

5) Export China's rich cultural goods through the Chinese bridge to all over the world. The export enterprises should make market research fully. The cultural products should not only embody Chinese characteristics, but also meet the foreign consumers' demand and aesthetic habits, in order to avoid "cultural discount" and open the overseas market as soon as possible.

5.3 TO PARTICIPATE IN THE GLOBAL OR REGIONAL ECONOMIC COOPERATION ACTIVELY

Preferential trade arrangement plays an important role in promoting the cultural products export of a country or an area. On one hand, as the member of WTO, China should actively participate in the formulation of world trade rules, as far as possible to reduce trade barriers and unfair treatment which hinder the export of Chinese cultural products; on the other hand, China should actively participate in various organizations of trade cooperation and regional economic integration, and establish close relations of economy and trade with trading partners to create a good international environment for the export of culture enterprise.

5.4 TO DEVELOP TARGETED CULTURAL TRADE

China should actively develop cultural trade with the countries, which have insufficient trade or the export potential, such as Denmark, France, Germany, etc. In addition, GDP, per capita GDP and trade openness of import country (region) have the positive effect on cultural products export. Therefore, China should focus on developing cultural products export with the developed countries which are rich and relatively open. As for the countries "excessive trade" (such as Brazil), China should adjust the export structure of cultural products and improve the grade of export products, in order to avoid the tendency of "excessive trade".

References

[1] Schulze G G 1999 International Trade in Art *Journal of Cultural Economics* **23**(1-2) 109-36

[2] Disdier A C, Tai S H T, Fontagné L, Mayer T 2010 Trade of Cultural Goods *Review of World Economics* **145**(4) 575-95

[3] Ferreira F, Waldfogel J 2010 POP Internationalism: Has a Half Century of World Music Trade Displaced Local Culture? *NBER Working paper* No. 15964

[4] Oh J 2001 International Trade in Film and the Self-Sufficiency Ratio *Journal of Media Economics* **14**(1) 31-44

[5] White R, Tadesse B 2008 Cultural Distance and the US Immigrant-Trade Link *The World Economy* **31**(8) 1078-96

[6] Mattoscio N, Donatella F 2010 A multidimensional model analysis in cultural economics: the Italian case *Tourism Economics* **16**(3) 565-83

[7] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81

[8] Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20** (3) 98-103

Author

Pingping Qiao, born in February 1983, Nanyang, Henan Province, P.R. China.

Current position, grades: the lecturer of department of economy and management, Henan Polytechnic Institute, Henan, China.
Scientific interests: international economy and trade and business management.
Publications: more than 15 papers.
Experience: educational work for 9 years, 3 scientific research projects.

Research on the influential factors of China's logistics demand based on the econometric model

Xuewen Jin*

College of management, Jiangxi University of Technology, Jiangxi Province, China, 330029

Received 28 October 2014, www.cmnt.lv

Abstract

Logistics demand reflects the pace of economic development, and the pace of economic development also affects the change of logistics demand in turn. In fact, logistics demand is affected by many factors. Through the theoretical analysis, the paper points out that the influential factors on logistics demand mainly include four aspects, namely, economic development level, regional industrial structure, macro-economic policy and economic system, and consumption level and concept. Then, the econometric model is established with the data in different periods of time based on those four aspects to make an in-depth analysis of the application of the model. At the same time, in order to promote the development of the logistics industry, improve logistics demand and meet the development demand of the regional economy, it puts forward reasonable suggestions.

Keywords: logistics demand, econometric model, influential factors

1 Introduction

With the continuous development of market economy and information economy, the logistics industry has grown from the end industry to the important one with a great impact on the economy and a role of guiding and promoting the development of other industries, and even the forerunner industry of other industries. Modern logistics industry is a comprehensive industry which integrates systematization, informatization and warehousing modernization together with IT information technology as the support and the transportation industry as the core. The development of the logistics industry plays a huge role in optimizing the industrial structure, enhancing the development of enterprises and improving the quality of economic operation. [1] In recent years, China's logistics industry has been growing rapidly. Various types of logistics enterprises have grown up and distribution center, logistics base, cargo loading center and other logistics related infrastructure facilities have been improved. Both the central and local governments attach great importance to the development of the logistics industry. The State Economic and Trade Commission, the State Planning Commission and the State Council Development Research Center all have made in-depth research and exploration on the problems and developed a series of necessary policies to promote the development of the logistics industry and create a good external environment for the application of logistics by enterprises and for the development of the logistics industry. [2] At present, some cities, especially those with rapid economic development like Shenzhen, Guangzhou and Beijing, provide support to the development of the logistics development from the level of the local government by developing a series of supporting policies. Many cities have listed promoting the development of the logistics industry into the 10th Five-year Plan. Therefore, predictably, the logistics industry will become an important industry and a new economic growth point in China in the 21st century. The development and expansion of the logistics industry is of great significance to realizing the sustainable development of the Chinese economy, improving the quality of economic operation, optimizing the allocation of resources, and promoting the reform and development. [3] This paper makes an in-depth analysis on the influential factors on logistics demand with the empirical method and establishes the econometric model, which can provide the important method and reference for the governments and the enterprises related to the logistics industry to make logistics planning and analyze logistics demand. Thus, this research has important practical significance.

2 Analysis of the influential factors on logistics demand

Logistics demand refers to the demand for space, time and cost generated due to the distribution and circulation of some goods in the social and economic activities in a certain period of time. As an important part of the social and economic activities, logistics demand runs through the whole social and economic activities of production, circulation and consumption and has close relationship with the status and speed of the social and economic development. Thus, the status and speed of the social and economic development is the main factor affecting logistics demand. The factors affecting the status and speed of the social and economic development will inevitably affect the development of the logistics industry and logistics demand. Logistics demand can be reflected

* ***Corresponding author's*** e-mail: 1418758482@qq.com

through the freight turnover rate and freight volume. Figure 1 presents the annual growth trend of China's logistics industry. According to the Figure, China's logistics quantity has been maintaining an average annual growth rate of 8% since 1986, lower than the growth rate of China's economy, which indicates the imbalance between supply and demand of logistics in China. Generally speaking, the growth rate of logistics is positively correlated with that of economy. At the same time, it can be known from the figure that the annual growth rates of freight volume and turnover are very unstable. The figures fell down very sharply particularly in recent years: the growth rate of freight volume was 14% in 2011, 11% in 2012 and still about 11% in 2013; the growth rate of freight turnover was 12% in 2011, 9% in 2012 and -3% in 2013. The change of logistics demand can reflect the development status of China's economy. Meanwhile, it's affected by many factors as follows:

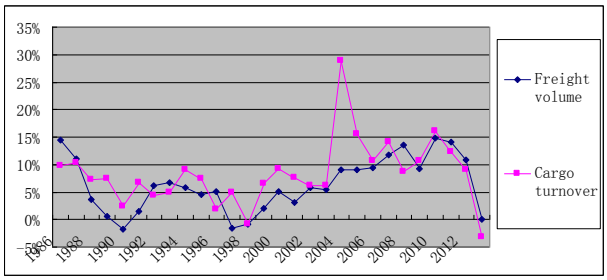


FIGURE 1 The changing trend of China's logistics demand

Note: the data are from www.tjcn.org.com

The first is the level of economic development. The level of economic development determines the level of output and consumption, that is, the regional economic aggregate. In the short term, it's the determinant and basic drive of logistics demand. This is because the object of logistics is always the basic output and consumption of regional economy. If the regional output and consumption level is high, the economic aggregate is big and the logistics demand must be great. If the output and consumption level is low, the economic aggregate is small and the object of logistics will not be virtually high. [4,5]. Thus, the high economic development level leads to the high logistics demand. For example, after comparison between the central and west China and the east China, the relative per capita logistics demand in the central and west China is low and that in the east China is high.

The second is the industrial structure. Different industrial structures have a significant impact on the logistics demand function, logistics level and logistics demand structure. The logistics demand of agriculture, forestry and animal husbandry in the primary industry is extensive with large quantity but low value. Manufacturing and mining in the secondary industry provide the products with physical substance which are reliant on logistics from production and consumption, so they have great logistics demand. The tertiary industry is given priority to the service industry which is less dependent on the logistics activity but needs more the modern logistics service based on the information technology [6]. Thus, different industries have obviously different degrees of logistics demand. In the regional industrial structure, if the primary and secondary industries occupy a large proportion, under given conditions, its logistics demand is large; but if the proportion of the tertiary industry is high, its logistics demand is small.

The third is the macro economic policy and management system. Due to the derivation of logistics demand, the macro economic policy directly affects the logistics quantity. First of all, if the policy can promote the economic development, the economic aggregate will increase, which will promote the development of the logistics industry and increase the logistics demand. If the macro economic policy fails to keep up with the development requirements of regional economy, leading to the economic stagnation or impossibility of recovery, it will affect the increase of the logistics demand [7]. Since the logistics service is strictly dependent on the infrastructure, the management system exerts a direct impact on the construction of the logistics infrastructure. For example, the increase of the investment in the logistics industry by the governments at all levels provides good conditions for the rapid development of the logistics industry. The investment scale in railways, highways and waterways directly affects the density and level of the transport network and further exerts the corresponding impact on logistics demand.

The fourth is the consumption level and concept. The ultimate purpose of circulation is consumption, so like the economic aggregate the consumption level can also reflect the regional logistics demand. If the regional consumption level increases, the demand for the products will be large. In other words, the local products cannot meet the consumption. Then, the products produced by other places will enter through logistics. The consumption concept will also affect the demand of products [8]. The circulation of different products needs different logistics demands. This is because the weight, volume and basic requirement for logistics which are important factors affecting logistics demand are different. In fact, the consumption concept directly affects the enterprises' management decision, production, sales and further the logistics size. New products and service always need the high-level logistics service for support.

The fifth is other factors. Market environment and technological advance also affect logistics demand. The logistics activity serves production and life. The change of the market environment affects the development speed of the regional economy, the residents' consumption expectations and thus the regional logistics quantity [9]. At the same time, different macro environmental conditions in different regions have different influence on the regional economy, which will also affect the flow direction and service way of logistics goods as well as the quality and quantity of service. The technological advance including the logistics technology will also affect the demand. For example, the Internet technology will increase the logistics demand. In fact, the domestic and foreign trading ways, economic globalization and other market environment all affect logistics demand.

3 Empirical analysis

3.1 MODEL ESTABLISHMENT AND PREDCTION

According to the theoretical analysis, the factors affecting the model include economic development level, regional industrial structure, macro economic policy and economic system, and consumption level and concept. The following indicators can be set: the added value of the primary industry (DY), the added value of the secondary industry (DE), the added value of the tertiary industry (DS) and the consumer level (XF). The development among industries can not only reflect the economic development level, but also the proportion change of the tertiary industry. At the same time, the logistics industry is also affected by its own development. Thus, the model can be established as follow:

$$WL = a_1 \cdot DY + a_2 \cdot DE + a_3 \cdot DS + a_4 \cdot XF + a_5 \cdot WL(-1) + u_t \quad (1)$$

The data are from China Statistical Yearbook from 1980 to 2013. To verify the model, it's necessary to test the unit root. It's found through the ADF test that the sequence is in the unstable state under both no difference and the first order difference, but it's in the stable state after the second order difference. The specific results are shown in Table 1.

TABLE 1 Augmented dickey-fuller test statistic

Variable	t-Statistic	Prob.*	Variable	t-Statistic	Prob.*
DE	-8.6250	0.0000	XF	-4.6302	0.0134
DS	-8.1326	0.0017	1% level	-3.6702	
DY	-8.2647	0.0000	5% level	-2.9640	
WL	-3.2338	0.0285	10% level	-2.6210	

Thus, it can be seen that the model can be used for the correlation analysis. When WL(-1) is listed in the influential factors for analysis according to the expectation, the result is not significant, so it should be excluded. Here, the analysis is made from four directions. The sequence coefficients of the correlation analysis are shown in Table 2 and the model is as shown in Equation (2):

$$\begin{cases} WL_{1980-2013} = 205.9716 \cdot XF + 11.8296 \cdot DY + 9.5704 \cdot DS + 1.0517 \cdot DE + u_t \\ WL_{1980-1990} = 520.4910 \cdot XF + 4.4747 \cdot DY + 8.6536 \cdot DS + 13.2296 \cdot DE + u_t \\ WL_{1990-2000} = 201.7310 \cdot XF + 9.2186 \cdot DY + 2.0680 \cdot DS + 9.2289 \cdot DE + u_t \\ WL_{2000-2014} = 273.3543 \cdot XF + 2.5861 \cdot DY + 2.2714 \cdot DS + 9.9887 \cdot DE + u_t \end{cases} \quad (2)$$

TABLE 2 Technical Drawing of the Econometric Model

Time	XF	DY	DS	DE
1980-2013	205.9761	11.8296	9.5704	1.0517
1980-1990	520.4910	4.4747	8.6536	13.2296
1990-2000	201.7310	9.2186	2.0680	9.2289
2000-2014	273.3543	2.5861	2.2714	9.9887

3.2 THE PREDICTION RESULT

Figure 2 makes prediction with the data from 1980 to 2013. According to the figure, the difference between the budget line and the balance error line is small, which indicates that the prediction value is about 42 billion tons in 2014 according to the prediction. Figure 3 makes prediction with the data from 1980 to 1990. The prediction value is greatly different from the actual one, but the trends are consistent. The prediction value of the latter is greater than that of the former. Figure 4 makes prediction with the data from 1990 and 2000. It can be seen from the figure that the difference between the error and the actual value is the largest and the prediction value is small, which indicates that the data growth area during 1990 and 2000 is significantly different from that in other years and its growth rate is smaller than that of others. Figure 5 makes prediction with the data from 2000 to 2013. According to the figure, the prediction value is closer to the actual one and the difference between the balance error line and the prediction line is better than that of Figure 1, indicating that the model is more accurate.

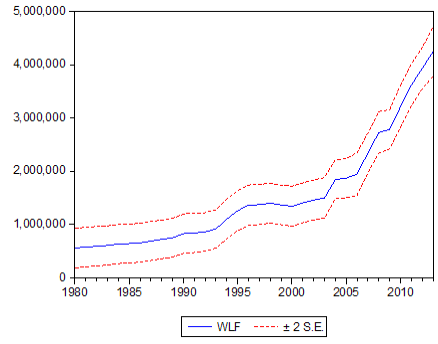


FIGURE 2 The prediction diagram based on the data from 1980 to 2013

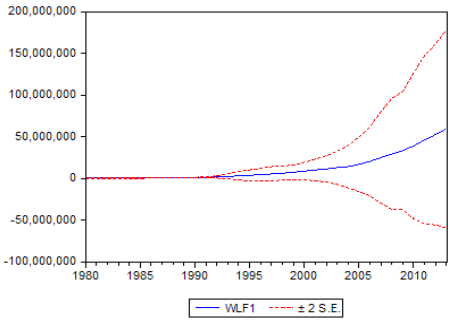


FIGURE 3 The prediction diagram based on the data from 1980 to 1990

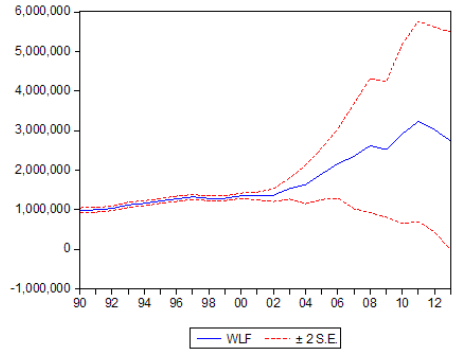


FIGURE 4 The prediction diagram based on the data from 1990 to 2000

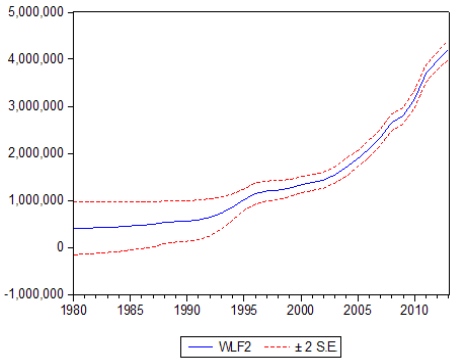


FIGURE 5 The prediction diagram based on the data from 2000 to 2013

3.3 APPLICATION ANALYSIS OF THE MODEL

First of all, the accuracy of the model can be verified through the model. The comparison among the parameters of four prediction models and four prediction diagrams indicates the accuracy of the model. It can be found that more cited data does not mean the more accurate model. The estimation should be based on the data in recent years. Thus, Figure 5 with the data in the past 13 years is more accurate than Figure 1 with the data in the past 34 years. This is mainly because the environment is always changing, so is the parameters of the model.

Secondly, the coefficients of the model change greatly, which indicates the great change of the model. Generally speaking, the greater the coefficient is, the greater the influence is. The great change of the coefficient is mainly because it's dependent on the environment. For example, the coefficient of the secondary industry changes greatly, indicating that its influence on logistics demand is small due to the development environments of different regional economies. When the trade volume of the regional goods is large, the demand for logistics will be huge; when the regions tend to satisfy the current scale of the industry, the trade volume is small and the demand for logistics will be small.

Thirdly, attention should be paid to the timeliness in the application of the model. In the current period, the model is affected by four influential factors. However, with the changes of the environment, the influential factors and the degree of influence, the model will also change, which is the timeliness of the model. To ensure the long-term practical application of the model, it's necessary to update the influential factors according to the change of the environment, get rid of those with little influence or no influence and add some new ones. Thus, the empirical analysis should be made often to estimate the model with the data in the past dozen years to predict the future data.

4 Conclusion and suggestion

Analysis of the influential factors on logistics demand and prediction of the demand is to analyze the development of the logistics industry according to the changing trend of the influential factors and better satisfy the development requirements. This is because if the supply of the logistics industry can't meet the future development of logistics demand, it will limit the development of the regional industry and even the whole regional economy and affect the economic efficiency; if the supply exceeds the demand, it will waste the logistics resources and affect the economic efficiency, too. Therefore, in order to balance the relationship between the logistics demand and supply, ensure the moderate development of the logistics industry, meet the demand of economic development and promote the upgrading and transformation of the logistics industry at the same time, the following aspects should be improved.

The first is to cultivate the concept of modern logistics, grasp its development trend and develop the development plan of modern logistics based on the full consideration of regional development, economic structure, transportation information, city planning and logistics demand. The government should avoid the excessive interference in the logistics market and establish the socialized and specialized modern logistics system with the market as the orientation and the logistics enterprises as the main players.

The second is to integrate the logistics resources and cultivate the main players in the logistics market. In the large area of China, it should break the regional block, promote advantages and abolish disadvantages of each region. In addition, it should guide the manufacturing and trade enterprises to promote the recombination of enterprise logistics management and enterprise logistics orderly based on their own development requirements and establish the logistics model to meet their special requirements.

The third is to accelerate the construction of the informatized and standard logistics enterprises and comprehensively improve the logistics service and technical level. Informatization is the soul of modern logistics. Modern information technology realizes the information communication and sharing in different economic sectors and enterprises and thus achieves the purpose of effecttively coordinating, managing and integrating the logistics elements and functions. It should promote the application of modern information management technology like ERP and MRP in logistics enterprises and logistics management of business enterprises, make great efforts to promote the construction of the public information platform, establish the sound e-commerce certification system, on-line payment system and logistics distribution management system, so as to create conditions for the smooth and efficient communication of logistics information.

References

[1] He G 2008 Forecast of regional logistics requirements and application of Grey Prediction Model *Journal of Beijing Jiaotong University(Social Sciences Edition)* **01** 33-7

[2] Zhang C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34

[3] Hu H 2008 Research on the model of regional logistics demand forecast *Statistics and Decision* **17** 62-4

[4] Luo Y, Liu G 2011 Study on the relationship between logistics service innovation and logistics demand -- Based on the Symbiosis Theory *Contemporary Finance & Economics* **02** 61-8

[5] Wan L, Li Y, Wu J 2011 Application research on prediction of regional logistics demand *Microelectronics & Computer* **09** 160-4

[6] S Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103

[7] Huang M, Feng Y 2009 The Grey Prediction Model applies to regional logistics demand forecasting *Logistics Sci-Tech* **03** 17-20

[8] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81

[9] Liu Z, Li C, Chen B 2012 Regional Logistics Demand Forecast Based on Factor Analysis and Neural Network *Computer Simulation* **06** 359-62

Author

Jin Xuewen, born in November 1977, Nanchang, Jiangxi Province, P.R. China.

Current position, grades: the lecturer of College of management, Jiangxi University of Technology, Jiangxi, China.

Scientific interest: business management.

Publications: more than 6 papers.

Experience: teaching experience of 14 years, 4 scientific research projects.

The review and research on agile oriented method n the pilot industry system

Hong Wang*

School of Art and Design, Wuhan University of Technology, Wuhan, Hubei, 430070, China

School of Art, Hubei University, Wuhan, Hubei, 430062, China

Received 1 March 2014, www.cmnt.lv

Abstract

Agile methodology has getting wide recognition within the software industry due to its flexibility and ability to cope with rapid changes in the software development environments. It is however comes with a number of demands that must be complied with. This paper presents a pilot systematic literature review (SLR) study on the limitations of Agile methods in the industry based on primary research. In this study, conference and journal papers in the IEEE, published between 2007 and 2012 were investigated. 29 papers were found as the most relevant. While the SLR findings have brought to both limitations in the implementation and in the Agile methods, the former becomes the most addressed issues. The result revealed that high dependency on people/personnel, organizational dependency, as well as high impact on organizational structure and culture as the three most repeatedly addressed factors. While these three factors are mutually related, people factor especially upper level management strong involvement and support can be regarded as a primary necessity in the Agile implementation. In spite of apparent emphasis on people critical function stated in the Agile principles, plus the excellent rules in the gist of the principles, the problem still arose when it comes to the implementation part. This indicates the need for future work on proper guidelines for management, given that existing guidelines for Agile adoptions and implementations are general and less focus given to upper level managers.

Keywords: Agile, limitation, human factor, pilot management, systematic review

1 Introduction

Agile methods are now becoming the mainstream and extensively being adopted outside of its initial intended scope of small and collocated project teams. They are now being implemented in all project sizes (small, medium and large), in both distributed (locally and globally) and non-distributed project environments, as well as in various project domains such as engineering, manufacturing, banking and medical. As a result, new versions of Agile methodology have been developed, Agile methods have been enhanced or being integrated with other models to support the increasing demands of different project environments. The initial intention was to integrate Agile method (AM) with an existing framework originated from non-software industry to enhance the AM adoption coverage. To further explore the limitations of Agile practices before deciding their suitability for integration, a pilot systematic literature review (SLR) study has been conducted.

While the limitations of AM in the industry have been aimed at, the limitations in adopting and implementing the AM were identified more during the reviewing. The Agile Manifesto that was addressed as too informal [5] and embraced "abstract principles" [1] raised a lot of issues in the implementation and adaptation of the Agile practices. Regardless of abundance suggestions and recommendations from existing experiences, the flexibility and generality of AM open for various interpretations and therefore, inviting issues in the way to effectively perform AM in practice notably for early adopters.

Agile methodology is an iterative and evolutionary approach for software project development. Initiated by seventeen experts of organizational anarchists in February 2001, it operates under four core values and twelve principles, namely Agile Manifesto [7]. This widely known people-centric process model is inspired among which by sentiments, working with people who shared compatible goals and values based on mutual trust and respect, promoting collaborative, people-focused organizational models and building the types of professional communities in which we would want to work [7]. There are different types of AM. For example, Dynamic Systems Development Method.

(DSDM), Extreme programming (XP), Scrum, Crystal Clear, Feature-Driven Development and Lean Software Development. Each of the different types of Agile methods such as Scrum and XP have their own practices which are based on different concentrations. While XP concentrates on the project level activities of software deployment, Scrum "concentrates on the management aspects of software development" [8].

In this paper, we firstly highlight the background of this pilot SLR study which encompasses a brief introduction on Agile methodology, previous related studies as well as motivation behind the study. Then, the

* ***Corresponding author's*** e-mail: 408569488@qq.com

review method or the steps taken in performing the SLR is elucidated. Subsequently, the result consisting of the overview of the study and the compiled limitations is presented. Last but not least, we conclude our SLR study together with our future work.

2 Background

Agile methods are getting wider attention and widely used at present. However, early investigation (end of 2011) finds their usage are limited (i.e. more suitable for experienced and skilled software engineers, and less appropriate for large and complex projects) despite of their ambitious purpose. AM is regarded as an ambitious practice because it allows for changes within rapid software development environment but the anticipated outcome is yet to be stably attained. At the same time, no paper is found to discuss the drawbacks of AM inclusively.

As a result, pilot SLR was conducted to dig out the limitations of Agile practices and prioritized them based on frequencies. The SLR review process was very helpful as the reviewer could clearly understand the situations and issues surrounded the AM practice. This study also brings the reviewer to see different angles in the research area and help her to firmly decide on what direction that should be focused.

First systematic review study on AM done by Dingsøyr and Dyba [6] intends to tackle the issue of anecdotal evidence in the agile adoption success story. The SLR specifically reviews empirical studies of agile software development since beginning up to 2005 with earlier search result of 1996 papers. The benefits and limitations of AM from their finalized 36 empirical studies as well as the strength of evidences were discussed. Their inclusive analysis provides guidance and comparison for industrial readership based on situational applicability. To conclude, the study mainly advocated for further similar study with better quality in the similar area.

This next systematic review of Agile methodology was specific for global software engineering (GSE) environment [9]. The SLR includes papers from five electronic databases between the year 1999 and 2009 where 77 most relevant papers were analysed. The study found that Agile practices in most cases were customized to fit the project environments and requirements. Thus, future work in incorporating existing experiences was pointed as required to aid agile adopters in distributed project settings.

Subsequently, Causevic et al. [4] have conducted a systematic review on factors limiting industrial adoption of test-driven development (TDD), one of the Agile practices. This SLR study examines both mainly focus and non-specific TDD empirical studies covering papers of industrial and academic studies from 2005 to 2009. The study was based on earlier research findings that the TDD practice "is not followed to the extent preferred by industry". Based on the result, the paper addressed the need for guidelines "to overcome these limiting factors for successful industrial adoption of TDD" [4]. While papers from seven electronic databases were investigated, the SLR merely focuses on one basic agile software development practice.

The only paper of similar focus with this SLR study was by Livermore [12]. Nevertheless, it was a worldwide online survey of general focus involving various industrial areas. With a low response rate of 5.76% or 112 survey responses, factors that were found to significantly impact the Agile implementation include; training, management involvement, access to external resources, as well as corporate size. Through this smaller scope of findings (compared to this SLR study), the management was identified as having main influence over most of the other factors.

Lastly, Neves et al. [14] have presented a paper on a thorough analysis and evaluation (empirically) of the benefits and limitations of Agile practices in relation to knowledge formation and transition experienced by Agile teams. In the last phase of evaluation, a SWOT (Strength Weakness Opportunity Threat) analysis was utilized to evaluate Agile contributions toward the productivity of the software development teams. Our SLR findings show a number of similar limitation items with this paper results of weaknesses and threats of Agile processes in knowledge management.

3 Review Method

First and foremost, an informal review is conducted to gain some understanding on the issues surrounded the AM. In the process, the author review articles, journals and conference papers by searching through the "google" and several well-known electronic databases such as ACM portal, IEEE Xplore, ScienceDirect and SpringerLink that pointed out the limitations or constraints of AM. The information obtained from this preliminary search and review is used as a basis to plan the systematic review.

To form an appropriate systematic review, separate search on papers discussing systematic review method is performed. There are two combination of search strings used; "systematic review AND software" and "systematic review AND Agile". In the process, several papers are retrieved based on title where final selection is made based on abstract and recent published date. Here, two systematic review papers on different AM research focus are chosen and referred; Empirical studies in Agile software development [6] and Agile practices in Global Software Engineering [9]. Their review techniques are carefully assessed and combined in building this systematic review plan. In actual, their SLR methods which focus to the AM study based on several SLR guidelines (i.e. [10]) provide a very helpful and clear-cut guidance for the reviewer to work on the SLR.

Basically, the design of this SLR review protocol was closely influenced by the review method presented by Dingsøyr and Dyba [6] and Jalali and Wohlin [9]. The combination of their review techniques resulted to the following construct. In a systematic literature review, research question that is required to be initially built defines the scope of the review study for the paper. In this

review study, the following research question has been formulated with the objective of finding and gathering the limitations of AM.

A lot of previous research reports pointed out the issue of the validity of Agile methods (AM) study, saying most of the reported results were anecdotal. Based on the situation, research community asks for more concrete study on AM where more and more AM study focusing on tangible results were published afterwards. Therefore this SLR study started with searching AM related papers that were based on strong foundation such as empirical or case study as well as survey and industrial experience report with concrete evidences or strong justifications. In the process, a reviewer went through the research method used in each paper to identify their validity.

Since this review study is about finding the limitations of Agile practices, the primary keyword used in the searching process is "agile" meanwhile "limitation" becomes the secondary keyword. For the first keyword, there are different types of AM such as extreme programming (XP), scrum, crystal clear, dynamic systems development method (DSDM), feature-driven development and lean software development. Separate search is conducted to identify the limitations of each of these AM types. Since the second keyword "limitation" is a common noun, similar meaning words are listed. The words are spotted from papers in the earlier informal review process. The formulation of several search strings from the keywords ensures that only related papers are extracted. The searching activity is conducted using IEEE Advanced Search, the sole electronic database used in this pilot SLR. IEEE is one of the general sponsors for Agile Conference annual events since 2003 and the main publisher for papers in the conferences.

Altogether, there were 7 search strings formed and used in the searching process. This is to ensure that all the related papers, according to the research question could be extracted. Basically, there were 2 types of search string as indicated below:

First search string: to extract all Agile practices in software project related papers that having Agile keyword in their abstract. Second to seventh search string: to extract papers that discuss specific Agile practices such as Scrum which is not included in the first searching, where these papers having no Agile keyword in their abstract.

The reviewer went through four stages of searching and filtering to identify relevant papers In the first stage, a total of 306 search results were returned after 7 individual searches in the IEEE Advanced Search. Next stages were a three phase filtering process. While the first filter is mostly quite straightforward (scanning the titles or abstract), there were cases where the reviewer need to go beyond the abstract to further understand and decide. During the last two phases, the reviewer carefully went through each paper where in most of the cases, need to go through the whole paper or study the full-text to identify its relevancy. In all the three phases, intended data as well as potential points were extracted and stored in an excel file. At the final stage, 29 papers are found to be most relevant to the SLR topic. Subsequent process, the data extraction was not straightforward as the process was delayed by the way of the studies/experiences were reported. The limitation elements are most of the time generally pointed or non-transparent (not in the surface). Thus, the intended points could not be easily extracted and require critical reading and understanding. Furthermore, the points to be extracted are not fix or unique whereby keyword search is not possible. To put in a nutshell, the limitations data are extracted based on the reviewer's analysis and conclusion by reading the whole paper normally more than once to understand the cases/experiences reported.

4 Results and Discussions

4.1 OVERVIEW OF THE RESULT

29 studies that addressed limitations in the Agile method implementations have been identified. The constraints of AM are constantly addressed through the year 2007 to 2012 (until 8th of August 2012), though the papers were extracted merely from an electronic database.

14 of them were case studies of either single or multiple case studies, 11 of them were experience reports, 3 were empirical studies, and the remaining 1 paper is a survey. 38% or 11 papers were affirmed as having large projects or large scale Agile adoptions.

Out of the 29 relevant papers, majority of them or 90% were published in conferences while only three papers (10%) were journals. In this review study, Scrum is found as the most adopted Agile techniques (38%), followed by Agile practices in general (28%). The rest are extreme programming (XP) (2 studies), Lean (1 study), Lean and Agile (1 study), Evo (1 study) and a mixture of several AM (1 study). Having single Scrum adoption as the majority, there are also hybrid implementation of Scrum and XP (1 study), Scrum and User Stories as well as Scrum and Lean (1 study). A new approach in practicing Scrum namely Enterprise Scrum was also developed and exercised (1 study). Our result aligns with several earlier reported studies where at present, Scrum is noted as the most favoured approach among other AM [2,3,17,S2,].

More than half or 55% of the reviewed papers implement Agile in collocated or non-distributed project setting, 21% applied Agile in distributed project environment, 14% of the papers discussed globally distributed Agile projects and the remaining three studies (10%) reported Agile implementation in both environments.

There are five Agile transition cases reported where four of them involve large projects and/or large scale Agile adoptions and four of the five cases as well adopted Scrum [S11,S16,S22,S25]. Besides, an experience report also discussed how the Scrum method has been extended to the executive level to effectively handle large scale Scrum implementation [S10]. The final result of this review compilation exhibits a variety of industry have involved in the Agile adoption other than software such as medical, system engineering, and embedded systems.

4.2 LIMITATION FACTORS

The limitations in the Agile implementation are compiled into several categories. High dependency on people/personnel, organizational dependency, as well as high impact on organizational structure and culture are found as the three factors with highest priority that limiting the Agile implementations in the industry. Here, we are going to highlight the issues identified from the related papers on these three elements.

The most addressed element, people or personnel dependency is divided into four groups. The first one that is found to enforce most barriers to the AM implementation is strong reliance on management. Without management support or having no full support, the agile implementations were either reported as facing greater pain and challenges [S1,S1s] or less effective due to so many problems and difficulties arose [S23], and in the worst case, the implementation was terminated [S11]. In one case, the developers felt insecure and threatened with their Agile adoption status due to suspiciousness after there was a change in key management personnel, lead to work impairment [S9]. While a case study reported successful in the Scrum implementation, strong management support has been gained from the beginning and above that, the company has been practicing Scrum for five years [S13].

On the other hand, a company was clearly reported as failed with lacking in management support as the main cause [S11]. In the case, management approval was raised as fundamental towards AM realization. Developer's with experienced and highly skilled was the second most concentrated factor within the people dependency constraint. Adopting AM, using merely product backlog in place of traditional requirements engineering (RE) practices requires for vast expertise to handle huge gap between the user requirements and coding [S22]. Besides, the lack of documentation [S29], the emphasis on working autonomously (both individually and as a team) through the self-organizing approach [S19], having no design work prior to applying user stories technique for requirements elicitation [S3] and doing things in a speedy way without adequate experience lead to technical debt. Technical debt is a term utilized by Kaiser and Royse [S14] that made use of financial debt as parable to refer to code issues where "when a developer cuts a corner (whether they are implicitly asked to do so or not) it is potentially something that the company will pay for down the road. That is because it will generally cost more (sometimes much more) to resolve it later". All of these conditions were addressed as factors contributing to the favorable of highly skilled and experienced developers or project team in exercising AM. The fact that Agile approaches favors skillful personnel has been acknowledged based on Smith and King [S24] experience report on a project adopted XP method stated that, "a clear benefit was the high level of expertise of candidates that applied for positions throughout the project" [S24].

Subsequent dependency concern was on other personnel such as agile/scrum master and product owner (PO). To start, clear specification of each and everyone's role and tasks for Agile project team is a must according to Jakobsen and Poppendieck [S13] based on their multiple case study findings. In one case, clear blueprint of PO's roles to the extent "how the team collaborated with the PO" has been recognized as a key factor for successful project sprints. Nevertheless, no proper guidelines available for either PO's role and job description, agile/scrum master or other main roles in an Agile project until to-date. At the same time few literatures discussed about PO's role regardless of its clear need [S16]. To conclude, the applications of PO or Agile and Scrum master position in the studied literatures are inconsistent and reported in disappointed ways [S11, S14, S16, S20, S26]. They were either inexperienced, inactive, rarely available, unavailable, or inefficient. So to say, those who want to adopt Agile needs to place extra initiative and effort to carefully design main roles job description based on their project requirements and environment.

Last but not least, Agile practices depend highly on intensive communication with customer as customer collaboration is one of the four core values of Agile methodologies. As a result, when there is a lack of trust on the customer side or the customer is unavailable, relevant information and feedback will be in scarce [S7]. The information shortage issue has been identified also in Korkala and Abrahamsson [S15] report that had caused serious problems due to lack of well-defined customer. In their case of distributed agile project, the customer became intermediate between two remotely located teams but inactively held the responsibility and failed to deliver important information between the two teams. While according to Jakobsen and Poppendieck [S13], "many projects struggling with clarifying features in collaboration with the customer. Clarifications from the customer were late, leading to a decrease in flow, which we know causes schedule and cost overrun". In the case, too much expectation was placed on the customer that exceeds the customer's ability and readiness.

Srinivasan and Lundqvist [S26] reported that inappropriate tool has been used to support Agile processes as it was mandated by the company, given a lot of investment has been made on the tool. Meanwhile organizational restructures have obstructed the communication and collaboration between developers and testers as well as with the QA team (the team felt the lack of support from the QA organization) when they are placed separately from each other [S27]. According to Thomas [S27], organizational boundaries such as enterprise policies that exercise separate team for testing and release activity from the development team made agile technique like Test-Driven Development (TDD) inapplicable [S4]. This kind of policies disallow the adoption of testing process like TDD besides impractical to be incorporated in a short release cycle due to longer time taken for testing activity. "Organizational resistance may be the main barrier to other organizations trying it, because top executives and engineers must be willing to give it a serious try" [S10]. It refers to a Scrum practice extension initiative at the executive level. To truly gain the commitment from the people on the changes brought about by the Agile through its values, the effort must come from the highest organi-

zational level [S1]. It has been concluded by Lalsing et al. [11] where it is imperative that before any organization decides to adopt an AM, it needs to assess whether the company culture, operating structure, business processes and projects are suited for the use of an Agile Project Management Methodology.

To adopt AM, there will be a need for changes not only in the working style but also in the organizational structure and culture whenever necessary. For example, Thomas and Baker [S28] stated that there was "an inherent conflict" between the AM and organization operated under legacy processes, mindsets and cultures. While Middleton and Joyce [S17] pointed out the drawback of Lean methods as might be conflicting with current corporate standards. Among the obstacles specified; "Lean does not work well with targets, milestones, Gantt charts and traffic light reporting methods" (the artifacts/requirements demanded in organization with heavy plan driven process), besides the need for changes in the job roles and responsibility such as the manager being a facilitator [S17] instead of a planner and controller [13]. To conclude, organizational changes are necessary in the Agile adoption process [S11,S20], especially for traditional corporate governance [S28]. Nevertheless, culture and mindset among the people within organization were recognized as the most difficult aspects to change [S25].

4.3 SUMMARY

Firstly, recognized AM limitations from the final 29 relevant papers are compiled and grouped through profound analysis made throughout the SLR study. Meanwhile discoveries and inputs from other literature readings (e.g. [14]) contribute to the refinement of the limitation items. Nonetheless, instead of classifying the limitations based on project sizes (small, medium and big) and/or project types (non-distributed and distributed), they are generally itemized.

Secondly, the two factors; organizational dependency as well as high impact on organizational structure and culture are separated from each other. While organizational dependency refers to constraints posed by the decision at the organizational level, the latter factor signifies the required organizational changes that had imposed strong opposition and challenges. To conclude, these two factors are mutually related and can be associated back to the people factor.

To compare, this pilot SLR finding aligns with several earlier research results on the importance of upper level management high commitment and support for better implementations of AM. For instance, Young and Jordan [18] result on multiple case studies on projects "ranged from complete failure to complete success" shows that top management support (TMS) is the most critical determinant for project success or failure. While Livermore [12] through his worldwide online survey found that, "a number of the factors that impact the implementation of an agile development methodology are completely under the control of management". Lastly, recent study on agile deployment in three companies [15] identified management clear vision and support as very significant and thus provided recommendations for management on effective agile deployment plan.

4.4 LIMITATIONS OF THIS REVIEW

The limitation in the SLR is on the electronic database (e-database) used where only papers in IEEE is reviewed (one e-database only). Therefore, this review study can be regarded as pilot SLR to identify and prioritize the factors limiting Agile implementations in the industry. Since the SLR inclusion criteria is broad; includes case study, empirical study, experience report, survey, and expert opinion of all project types (small, medium and large) in both distributed and non-distributed Agile project management and development, using only one e-database has taken considerable time. This is because the whole SLR process is done by a single reviewer through a profound analysis.

This SLR study as well does not specifically include the obstacles of implementing AM in the distributed projects. The reviewer found that other than technology restrictions, the relevant papers do not distinctly stress the limitations concerning distributed environments. In spite of that, the shortcoming of implementing AM in such project setting is an important and a huge area to be focused on its own.

The equation shown below was used to calculate the error rate. It was required to calculate two measures; the number of error patterns and the total number of patterns, which was used to find the error rate in the improved K-means clustering algorithm and the K-means clustering algorithm in all data sets of this study. In next section, it will be seen that the RER-K-means algorithm reduced the error rate and iteration. In this algorithm, additional operations that have a negative effect on the calculation must be avoided. In all the data sets, the K-means clustering and the RER-K-means algorithms implementation were similar and only the data set name and data set coordinates were changed by the algorithms.

6 Conclusion

From 29 most relevant papers identified in the IEEE database, high dependency on people/personnel was found as the most mentioned factor imposing barriers to the Agile implementations in the industry. People dependency encompasses management, developers, customers, and others such as agile master and product owner. This is followed by organizational dependency as well as high impact on the organizational structure and culture as the second and third limitation factor. The SLR demonstrates that the first three limitations with highest frequencies are mainly originated from the people factor where management plays the most important role. While the rest of the limitations exhibit that most of them can be strongly associated back to the people factor (management), they are not highlighted in this paper.

If compared to the Agile principles illustrated in the Agile Manifesto [7], elements such as people dependency and heavy reliance on communication are clearly mentio-

ned and therefore should be mostly expected. To conclude, clear relations between the SLR findings and the Agile principles stated in the Agile Manifesto could be identified. High dependency on people that is supposed to be vastly anticipated during the Agile adoption became the most addressed factor that bring limitations to the Agile adoption and implementation in the industry.

In spite of apparent emphasis on people critical function stated in the Agile principles, plus the excellent rules in the gist of the principles, the problem still arose when it comes to the implementation part. Therefore, future research is to construct a set of guidelines to tackle the identified issue of "lacking in specific guidelines in the practical application of AM" as an approach in preparing the top/senior management with proper direction (on their actual role) in managing people in the practical implementation of AM (aligns with Agile principles).

Appendix

The 29 relevant papers:

[S1] Asnawi A L, Gravell A M, Wills G B 2012 Emergence of Agile Methods: Perceptions from Software Practitioners in Malaysia Agile India IEEE Computer Society 30-9

[S2] Azizyan G M, Magarian K Kajko-Matsson 2 2011 Survey of Agile Tool Usage and Needs Proceedings of the 2011 Agile Conference IEEE Computer Society 29-38

[S3] Babar M A 2009 An exploratory study of architectural practices and challenges in using agile software development approaches. Proceedings of the Joint Working IEEE/IFIP Conference on Software Architecture 81-90

[S4] Bass J M 2012 Influences on Agile Practice Tailoring in Enterprise Software Development AGILE India IEEE Computer Society 1-9

[S5] Beckhaus A, Karg L M, Hanselmann G 2009 Applicability of Software Reliability Growth Modeling in the Quality Assurance Phase of a Large Business Software Vendor The 33rd Annual IEEE International Computer Software and Applications Conference 2009 209-15

[S6] Cagle R 2012 Enterprise Architecture facilitates adopting Agile development methodologies into a DoD acquisition IEEE International Systems Conference 2012 1-5

[S7] Cao L, Ramesh B 2008. Agile Requirements Engineering Practices: An Empirical Study IEEE Software 25(1) 60-7

[S8] Chookittikul W, Kourik J L, Maher P E 2011 Reducing the Gap between Academia and Industry: The Case for Agile Methods in Thailand The Eighth International Conference on Information Technology New Generations 239-44

[S9] Evans M 2008 The FrAgile Organisation Agile Conference 181-5

[S10] Greening D R 2010 Enterprise Scrum: Scaling Scrum to the Executive Level *The 43rd Hawaii International Conference on System Sciences* 1-10

[S11] Hajjdiab H, Taleb A S 2011 Agile adoption experience: A case study in the UAE IEEE 2nd International Conference on Software Engineering and Service Science 31-4

[S12] Hanssen G K, Yamashita A F, Conradi R, Moonen L 2009 Maintenance and agile development: Challenges, opportunities and future directions IEEE International Conference on Software Maintenance 487-90

[S13] Jakobsen C R, Poppendieck T 2011 Lean as a Scrum Troubleshooter Agile Conference 168-74

[S14] Kaiser M, Royse G 2011 Selling the Investment to Pay Down Technical Debt The Code Christmas Tree Agile Conference 175-80

[S15] Korkala M, Abrahamsson P 2007 Communication in Distributed Agile Development: A Case Study The 33rd EUROMICRO Conference on Software Engineering and Advanced Applications 203-10

[S16] Lehto I, Rautiainen K 2009 Software development governance challenges of a middle-sized company in agile transition ICSE Workshop on Software Development Governance 36-9

[S17] Middleton P, Joyce D 2012 IEEE Transactions on Engineering Management 59(1) 20-32

[S18] Miller J R, Haddad H M 2012 Challenges Faced While Simultaneously Implementing CMMI and Scrum A Case Study in the Tax Preparation Software Industry The Ninth International Conference on Information Technology 314-8

[S19] Moe N B, Dingsoyr T, Kvangardsnes O 2009 Understanding Shared Leadership in Agile Development A Case Study The 42nd Hawaii International Conference on System Sciences 1-10

[S20] Moe N B, Dingsoyr T, Dyba T 2008 Understanding Self-Organizing Teams in Agile Software Development The 19th Australian Conference on Software Engineering 76-85

[S21] Rottier P A, Rodrigues V 2008 Agile Development in a Medical Device Company Agile Conference 218-23

[S22] Savolainen J, Kuusela J, Vilavaara A 2010 Transition to Agile Development – Rediscovery of Important Requirements Engineering Practices The 18th IEEE International Requirements Engineering Conference 289-94

[S23] Shatil A, Hazzan O, Dubinsky Y 2010 Agility in a Large-Scale System Engineering Project: A Case-Study of an Advanced Communication System Project IEEE International Conference on Software Science 47-54

[S24] Smith C, King P 2008 Agile Project Experiences The Story of Three Little Pigs Agile Conference 378-83

[S25] Smits H, Rilliet H 2011 Agile Experience Report: Transition and Complexity at Cisco Voice Technology Group Agile Conference 274-8

[S26] Srinivasan J, Lundqvist K 2009 Using Agile Methods in Software Product Development: A Case Study The Sixth International Conference on Information Technology: New Generations 1415-20

[S27] Thomas J 2008 Introducing Agile Development Practices from the Middle The 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems 401-7

[S28] Thomas J C, Baker S W 2008 Establishing an Agile Portfolio to Align IT Investments with Business Needs Agile Conference 252-8

[S29] Xu J, Lippert D 2007 Lesson Learned in Managing IT Departments Portland International Center for Management of Engineering and Technology 2107-15

References

[1] Abrahamsson P, Warsta J, Siponen M T, Ronkainen J 2003 New Directions on Agile Methods A Comparative Analysis *Proceedings of the 25th International Conference on Software Engineering* 244-4

[2] Asnawi A L, Gravell A M, Wills G B 2011 Empirical Investigation on Agile Methods Usage Issues Identified from Early Adopters in Malaysia *XP 2011 Lecture Notes in Business Information Processing* Springer 192-207

[3] Bustard D 2012 Beyond Mainstream Adoption: From Agile Software Development to Agile Organizational Change *The 19th International Conference and Workshops on Engineering of Computer Based Systems* 90-7

[4] Causevic A, Sundmark D, Punnekkat S 2011 Factors Limiting Industrial Adoption of Test Driven Development: A Systematic Review *IEEE Fourth International Conference on Software Testing, Verification and Validation* 337-46

[5] Conboy K, Fitzgerald B 2004 Toward a conceptual framework for agile methods: a study of agility in different disciplines *Proceedings of the 2004 ACM Workshop on Interdisciplinary Software Engineering Research* 37-44

[6] Dingsøyr T, Dyba T 2008 Empirical Studies of Agile Software Development: A systematic Review *Information and Software Technology* **50**(9-10) 833-59

[7] Fowler M, Highsmith J 2001 The Agile Manifesto. www.pmp-projects.org/Agile-Manifesto.pdf (Accessed on September 17 2013)

[8] Fowler M 2005 The New Methodology http://www.martinfowler.com/articles/newMethodology.html (Accessed on October 3, 2013)

[9] Jalali S, Wohlin C 2010 Agile Practices in Global Software Engineering – A Systematic Map *The 5th IEEE International Conference on Global Software Engineering* 45-54

[10] Kitchenham B A, Charters S 2007 Guidelines for performing Systematic Literature Reviews in Software Engineering *EBSE Technical Report* Keele University and University of Durham UK
[11] Lalsing V, Kishnah S, Pudaruth S 2012 People Factors in Agile Software Development and Project Management *International Journal of Software Engineering and Applications* 3(1) 117-37
[12] Livermore J A 2008 Factors that Significantly Impact the Implementations of an Agile Software Deployment Methodology *Journal of Software* **3**(4) 31-6
[13] Nerur S, Mahapatra R K, Mangalaraj G 2005 Challenges of migrating to agile methodologies *Communications of the ACM* **48**(5) 73-8
[14] Neves F T, Correia A M R, Rosa V N, de Castro Neto M 2011 Knowledge creation and sharing in software development teams using Agile methodologies: Key insights affecting their adoption *The 6th Iberian Conference on Information Systems and Technologies* 1-6
[15] Pikkarainen M, Salo O, Kuusela R, Abrahamsson P 2012 Strengths and barriers behind the successful agile deployment-insights from the three software intensive companies in Finland *Empirical Software Engineering* **17**(6) 675-702
[16] United States Government Accountability Office (GAO) 2012 Software Development: Effective Practices and Federal Challenges in Applying Agile Methods *GAO* Washington DC USA
[17] West D, Grant T 2010 Agile Development: Mainstream Adoption Has Changed Agility. Forrester Research, Inc. http://www.ca.com/~/media/Files/IndustryResearch/forrster-agile-development-mainstream-adoption.pdf (Accessed on May 22, 2013)
[18] Young R, Jordon R 2008 Top management support: Mantra or necessity? *International Journal of Project Management* **26**(7) 713-25

Authors

Hong Wang, born in February 1977, Wuhan, Hubei Province, China.

Current position, grades: PhD candidate at Wuhan University of Technology. Lecturer at Hubei University.
Scientific interests: include information interactive visual design and computer simulations.
Publications: 10 papers.

Study on GIS based intelligence transportation

Jianda Zhu*

School of Architecture & Urban Planning, Suzhou University of Science and Technology, 215009, China

Received 1 March 2014, www.cmnt.lv

Abstract

Recently a vast amount of geographic information system (GIS) data make road networks around the world as polylines with attributes. In this form, the data are insufficient for applications such as simulation and 3D visualization—tools which will grow in power and demand as sensor data become more pervasive and as governments try to optimize their existing physical infrastructure. With the increased dissemination and computing power of mobile devices, it is now possible to execute distributed artificial intelligence applications for various situations: intelligent routing using algorithms, planning, is tribute optimization of traffic lights. This paper reports on our development of a GIS-based traffic network analysis system, named GIS-based Transport Decision Support System, which provides a graphical analysis platform to transportation planners and researchers for transportation network analysis. The system has the functions of designing traffic networks on digital maps and doing traffic equilibrium analysis, as well as a novel function to integrate local detailed structures of intersections into global networks. The latter is particularly useful for the analysis of large traffic network where the detailed local network structures of some intersections have to be taken into account. The system links great volumes of traffic data and geography information data accumulated for visualization traffic analysis. We added information data on the following: road structure data, zone geography information data and node geography information data. The system also enabled us to extract traffic data by road section and by specific condition.

Keywords: K-means, clustering algorithm, error rate, iteration, reduction, stable

1 Introduction

Traffic simulation describes large numbers of vehicles on a traffic network by taking advantage of the reduced dimensionality typically found on road networks: vehicles follow roads and their motion can be described with few degrees of freedom. Research on techniques for traffic simulation has been carried out since the 1950s; see the survey of Helbing for a good overview of the field. Traffic is an integral component of any virtual environment that attempts to realistically portray the contemporary world, be it a video game, movie, or virtual globe. Traffic is also a global challenge with a direct impact on the economy, energy consumption, and the environment in today's society. Traffic simulation is a key tool to address both the challenges of traffic and its visualization. However, traffic simulation takes place on a complex domain and realistic road networks. The main objective of this work is to create road network representations from polyline data that can be used directly for real-time traffic simulation and visualization in a virtual world.

Traffic simulation presents unique challenges in the acquisition and representation of the underlying simulation domain, namely, the road network. Digital representations of real-world road networks are commonly available, but the level of detail of these data is not immediately usable for many queries related to traffic simulation. Traffic simulations take place on a network of lanes. This network needs to be represented with all its details, including the number of lanes on a road, intersections, merging zones, and ramps.

Optimizing vehicle routes in the context of current road congestions can reduce fuel consumption and transportation time. Maximizing the load of each vehicle in a transport company depending on route, will reduce the amount of trips per goods. Giving the driver information about Point of Interests (POI), such as safe parking spots or resting places, reduces the down time having to search for them manually. Online simulation of trafic can assist existing route guidance systems by predicting problems such as congestion. Accurate predictions require accurate status information about vehicles – the fact that vehicles are distributed over large-scale road infrastructure makes this particularly challenging.

A fundamental problem in transportation network planning and traffic engineering is the evaluation of the impacts of a design plan or some management policy on the distribution of traffic flow on the network, which is solved by the equilibrium network analysis method. The essence of equilibrium network analysis is the consideration of the influence of traffic volumes on travel times, and consequently on route choices of travellers, which then reversely change traffic volumes. There have been developed several software packages in which the equilibrium network analysis algorithms are implemented.

This work entails numerous scientific challenges. First, constructing the intersection, ramp, and road geometries presents numerous special and degenerate cases, typical of geometric computation. Our method is speci-

* ***Corresponding author's*** e-mail: jiandazhusz@163.com

fically designed to automatically handle as many of these cases as possible. Second, GIS data of road networks are not intended to be used for simulation. We reformulate these networks in order to extrapolate a network on which simulation can be done. Third, the data as available require filtering in order to be processed; while this is not the main focus of our work, it is a challenge that we have addressed in this paper. Fourth, these networks are large in scale, and so efficient algorithms and implementations are required. Fifth, the scale of the implementation itself is a challenge as this project is a combination of multiple systems, a road network importer, a road network representation, a simulation system, and a visualization system. Finally, there are algorithmic challenges in capturing details such as over-passes and in defining arc roads, which further address the needs of traffic simulation.

The idea in this study is to consider active vehicles in the traffic as agents, which send traffic reports to a centralized server via wireless internet. A centralized server behaves as an agent which collects data from vehicle agents and estimates average traffic flow speed for each road piece on the map. The vehicles are able to retrieve real-time traffic flow speed of a specific set of road pieces when they wish to plan a route between two points. The purpose of the study is to investigate if route planning based on real-time data is more effective and efficient than route planning based on static and statistical traffic data.

2 Related works

While digital road networks are widely available, the amount of detail varies widely across sources. Data for North America and Europe are freely available from the US Census Bureau's TIGER/Line database [4] and "crowd-sourced" community projects like OpenStreetMaps [5], but these data sets contain polyline roads with minimal attributes–information about lanes and intersection structure is wholly missing. Commercially available data sets, such as those provided by NAVTEQ [6], often contain some further attributes, such as the lane arrangements at intersections, but they are expensive to obtain, the techniques used are not known, and they do not capture all of the desired detail.

Numerous methods have been proposed for automatic and semiautomatic GIS road extraction from aerial and satellite images. Extensive surveys include [7-9]. These methods are complementary to our work: the GIS network we assume as input could be the product of a satellite image extraction method.

Procedural modelling of cities and roads has been an active area of research interest in computer graphics. For example, recent work in [10,11], among a notable body of investigation, has enabled the generation of detailed, realistic urban layouts, and roads for visualization. Commercial procedural city modelling software is also available. For example, consider the intersection geometry generated by CityEngine. Here, the intersection is modelled as a square connected to neighbouring rectangles with narrow triangles. In our work, we construct the geometry for every lane, not just the roads; the lane connections are C1 continuous, and the geometry defines all the needed parameters for vehicle animation, including orientation and steering angle.

Numerous spatial representations of curves have been developed over the years – see the comprehensive books by Farin [13] and Cohen et al. [14]. However, road networks and traffic behaviour have specific requirements: existing curve representations are not the best suited for modelling road networks to support real-time traffic simulations. For example, the popular NURBS formulation [15], despite of its generality of representations, is costly in space and efficiency. In particular, many splines do not readily admit arc-length parameterizations: those must be obtained using relatively expensive numerical integration techniques for establishing vehicle positions and for describing quantities of vehicles on each lane in traffic simulators. Willemsen et al. [16] describe ribbon networks, specifically discussing the need for "fattened" splines to describe road shapes, and our technique is potentially complimentary to the modelling technique for road networks they present. How-ever, they use the representation of Wang et al. [17], which is only approximately arc length parameterized and requires iterative techniques for evaluation. In contrast, our method only needs a simpler and much cheaper direct evaluation. Van den Berg and Overmars [18] proposed a model of road maps for robot motion planning using connected clothed curves. However, their choice of representation is based solely on the need to generate vehicle motion. For both traffic visualization and simulation, the representation must also be suitable for the generation of road surfaces, which are not necessarily clothed curves. Additionally, clothed curves are expensive to compute–requiring the evaluation of Fresnel integrals–whereas our method relies solely on coordinate frames, sins, and cosines. Nieuwenhuisen et al. [19] use circular arcs, as we do, to represent curves, but these arcs are used to smooth the corners of road maps for motion planning as in [18]. Furthermore, neither of these techniques have been investigated for the case of extracting ribbon-like surfaces, as we do, nor is there an established technique for fitting them to multisegment, no planar polylines.

3 The proposed method

3.1 PRELIMINARIES

The common formulations for traffic simulation are lane-based. These lanes are treated as queues of cars, represented either as discrete agents or by continuous density values. For traffic simulation, lane geometry is irrelevant as long as speed limits and distances are available. However, geometry matters for visualization and for localizing data, such as cell phone or GPS transmissions sent to inform about traffic conditions. These lanes are connected in various ways to form a road network, and cars traverse these connected lanes by crossing intersections and merging between adjacent lanes. The principle requirement for simulation is the creation of this network of lanes. This includes the division of roads into lanes, but also the creation of transient "virtual" lanes within

intersections: these virtual lanes exist only during specific states of a traffic signal. The creation of the network of lanes also entails determining the topological relationships be-tween lanes (so that vehicles can change lanes and take on-and off-ramps) and making geometric modifications to the road network to allow the construction of 2D or 3D road geometry.

To efficiently support traffic simulation, there are a number of queries the network needs to be able to answer in a computationally efficient manner. The nature of these queries depends on the simulation technique (i.e., whether the technique is continuum-based or discrete). Additionally, it is desirable that the road network representation abstracts away the details of the queries on the road network to maintain clear separation and software modularity between the traffic simulation and the road network.

3.2 GIS DATA FILTERING

We filter the GIS data we use to remove the most commonly occurring errors. These changes are not meant to change the underlying geometry or topology of the network, only to correct sloppy data creation. The first filter removes points that are coincident, where is a distance argument that is kept on the order of feet. This is done prior to the splitting and joining algorithms, while the remaining filters are applied afterward. The second filter removes collinear points within roads. The third filter ensures that no point added to a road causes it to turn too sharply or double back on itself. This filter calculates the offset, as in Figure 1, that would be required for a circle of minimum turning radius to be inscribed within the polyline segments. If this offset is greater than half the length of either segment, the node is not added. This ensures that when a point is added to the road, the road still satisfies the kinematic constraints of a typical car. Further filtering includes ensuring that one way roads are defined in the correct direction and that roads have been assigned the correct classification.

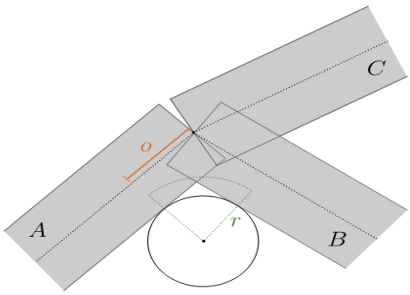


FIGURE 1 A simple intersection with three roads

3.3 THE PROPOSED COMPONENTS

We have implemented the Traffic Map Editor (TME), as a traffic simulator in Java in which infrastructures can be edited easily by using the mouse, and different levels of road usage can be simulated. A large number of fixed and dynamic traffic light controllers have already been tested in the simulator and the resulting average waiting times of cars have been plotted and compared. We will speak about dynamic traffic lights system as the adaptability component of the solution.

1) Application Architecture: Traffic Map Editor (TME) has been divided in several main components that describe the features that come together with the application. The architecture allows developers to not only build on top of TME but also allows them to enrich core features by adding new (or even replace) plug-ins. To the core of the application have been attached several important components as: db, editor, exporter, importer, network view, visualizer, simulation statistic or TmeEventServer.

2) Traffic Data: Traffic data that is aimed to be shared between agents is the speed of the vehicles moving along road segments and average speed of unique road segments. Vehicle agents request traffic data from the server agent by sending a list of road identifiers. Server agent's response to these requests contains the calculated average flow speed of the requested road segments. The reports that vehicle agents send to the server agent contain the momentary movement speed of the vehicle and the unique id of the road segment on which the vehicle is moving. These reports also contain the reporting time and the exact coordinates of the vehicle at that moment. These reports are received by the server agent and the average flow speeds of road segments are calculated.

In our development of GIS-TDSS for traffic road network mapping applications, objectives on geographic processing providing operations for processing or transforming data in a manner determined by user specified data. Through the system, we are able to integrate with other new traffic analysis model, this including road information, data management, and traffic model execution and management. Figure 2 shows the structure of our development of GIS-TDSS system.

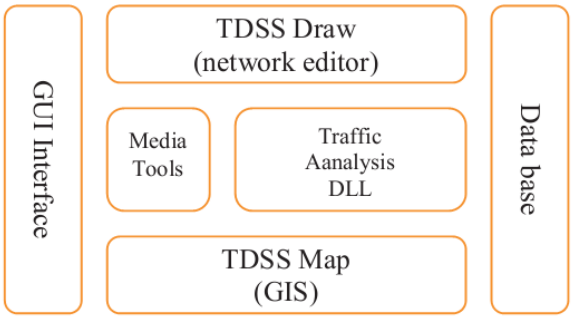


FIGURE 2 Structure of GIS-TDSS system

Similar to other GIS software, geographic information in our development is organized into 6 groups as follows: GUI system, database system, TDSS Map, TDSS Draw, media information tool and traffic network analysis algorithm DLL (Dynamic link library) files. GUI is the basic function group in a graphic interface built using Microsoft GDI (Graphics Device Interface). TDSS Map is the core module for processing digital map.

TDSS Draw is the module for network drawing and editing. The media tools module provides tools for sto-

ring various kind of regional multimedia information, e.g., traffic video at important intersections. The traffic network analysis algorithm DLL is the dynamic link library of network analysis programs. GIS-TDSS uses conventional layer models for storing and editing information. The topological relation between a global network and a local network is described in the data table for the local network data, detailed structure of this is described in the following section.

3.4 DATA STRUCTURE

The characteristics of the geospatial data set are changing. First and foremost, in order to meet users demands effectively, the capacity for the real-time collection, synthesis and access must exist; data import and export is essential. The data should be seamless, without artificial boundaries, and linked to attributes table that has become critical to many applications, for example, traffic flow management, road of disaster, and flood and earthquake traffic.

The GIS-TDSS system project file consists of a workspaces file, multiple geometric files, multiple relational data tables and multiple attribute data tables. The map file is a direct access, variable-record-length file in which each record describes a shape with a list of its vertices. In the index file, each record contains the offset of the corresponding main file record from the beginning of the main file. The data table contains feature attributes with one record per feature. The one-to-one relationship between geometry and attributes is based on record number. Attribute records in the data file must be in the same order as records in the main file. Figure 3 shows the data file structure of GIS-TDSS system and data file flow.

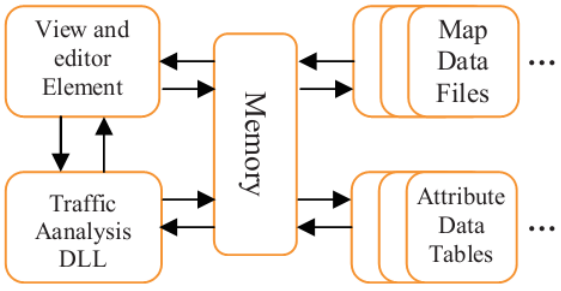


FIGURE 3 Data file structure of GIS-TDSS system

We propose new concepts for the GIS-DSS system: seamless operation of a detailed local network into a global network. As mentioned in the Introduction section, the integration of a detailed local network for an intersection into a global network is necessary to improve traffic analysis accuracy without involving too large a computational burden.

In the general structure of GIS (9), spatial data can be divided into three groups: 1) geometric data – data for describing space characteristics of spatial data, also known as location data, positioning data; 2) attribute data – data for describing attributes of spatial data such as type, grade, name, status, and so on; 3) relational data – data for describing topological relationship between spatial data. These three data groups are stored in separate tables in conventional GIS. In this study, the geometric data and relational data are placed in a unified data management table. In the following an example is given to illustrate our method.

The module for implementing the above data modification in the GIS-TDSS is shown in Figure 4.

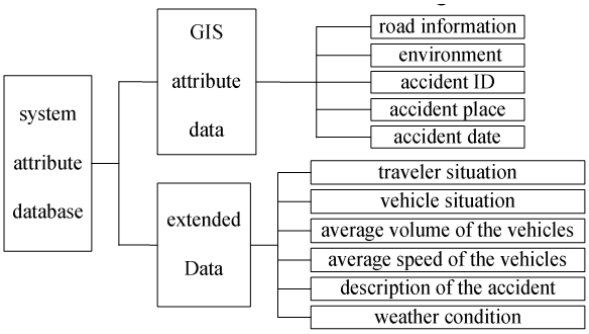


FIGURE 4 Attribute data organization chart of the traffic

The layer manager extracts data from data file in accordance with the requirements of the system, forms a graphic layer with the data of the geometric table, and then loads the required layer attribute data, and loads them into memory to output the required layer. The data manager controls the definition and value assignment of the structure and metadata in the layer-refining table, defines the geometric shapes in network, extracts data, etc.

3.5 DATA PROCESSING

Geographic entities are organized by layers in spatial database, mainly including point, line, and area entity. Geographic entities of map are called graphic element which are display as Feature object. Each geographic entity has its own attribute and is stored in attribute list in the form of one record. Entities with similar attributes or functions are stored in the same layer when needed in practical application. Each layer corresponds with one attribute list, to realize operation and inquiry of attributes of graphic element.

Actually, in the traffic information management and promulgating system based on GIS, the spatial database which storing and managing various spatial information and attribute data of graphic element is a map file of MapInfo in the system, including spatial database of non-related data structure with spatial information and attribute database with related data structure. The files or updated ones can be downloaded when running the GIS viewer. Spatial data mainly include point, line, area entity such as urban road network, urban terrain, traffic facility map, administrative regionalization, road transect and vertical section. Spatial data on the road map are collected and input by electronic map or engineering drawing of AutoCAD format. Then the data are edited and processed by MapInfo. At last, basic layers are integrated by Geoset Manager, in order to link and operate.

4 Experiments

Traffic attribute data in GIS attribute list are applied to inquire and analyse basic information about road traffic by users, not including parameters of traffic flow and all the information of traffic accidents, while they are stored in extended database and defined according to data list field. Entity attribute list of traffic flow and accident is formed by part of extended data list field. Traffic reference point data are from writing and electronic data related to traffic police department, which are processed by manual inputting, scanning, and format changing and so on. At last, the data are inputted into SQL Server 2000 database. The data collected from section detector through the analysis and decision-making module processing, the necessary flow, with an average speed of traffic flow parameters and other important information are archived, stored in SQL Server2000 database, which provide basic data for the future transport planning and road design and transportation research.

Spatial data are the basic loaders to indicate, query, analyse space and construct road network. Since the range of urban road network is large, traffic information is refreshed at any time. The spatial data of different traffic information in road network are on adjusting, which need to maintain and refresh spatial data frequently, in order to ensure integrity, accuracy, and validity of spatial data. Attribute data are stored in a list in the form of database, usually extended database. There are many alternatives here and relation database is often chosen, such as SQL Server, Oracle, Sybase, Access, dBase, Paradox, and so on. Generally, SQL Server and Oracle are used to develop large-scale database, Access and dBase are used to develop medium and small-scale database. The attribute and spatial data are linked by data index mechanism in GIS. Graphs and characters can be checked quickly with each other too. Data index mechanism is a method associating spatial objects with attribute data.

1) When we query spatial information from the attribute information, in the first the corresponding database records should be found in the attribute data file, such as the number N record. You can find the Nth pointer in the cross-index file, the pointer is pointing to a map of an object which is the spatial objects correspond with the database records.

2) When we query attribute information from the spatial information, if you have found a space object from the map, the system read out its spatial information and the corresponding database record number from the spatial data file, according to a database record number, the attribute information of the map object could be query in the attribute database. History data records of traffic information are stored in extended database, which should also be connected with processing module of extended detectors. Besides, there should be index linking attribute data of graphic elements. Related database can be realized on this point.

Suppose that we are interested in the traffic flow crossing a special intersection indicated in the map. For this the detailed network for the intersection is integrated into the global city road network. Due to space restriction, the detailed data for the parameters of the network are omitted here.

Step 1: The Conventional equilibrium analysis result of the common transport network, and the part surrounded by the dotted line. We can find out the traffic volume of the four links (two-way) connected to node intersection.

Step 2: The node intersection is set to be a common intersection (local detailed network), so that new twelve links at the intersection can be inserted into data table and the start point or the end point of four links connected to the node intersection can be revised.

Step 3: A further calculation on the new network obtained in Step 2 is taken to get the new equilibrium analysis results. The traffic volume change on links connected to node happened.

For detailed analysis of the intersection denoted as node Intersection in the network, Intersection is expanded as a local detailed network with twelve new links. It can be seen that some of the traffic volumes have turned onto other links due to the delay impact of the intersection.

Two alternative evaluation procedures were used in testing the accuracy of the road traffic noise prediction model developed in this study. First, grouped data were collected from 85 locations near road segments where sound propagation was not interrupted by barriers. Second, an additional set of grouped data were collected from 148 reception points within housing estates in Beijing, where the noise environment was believed to be much closer to real conditions. Results of these two data collection procedures are presented in Figures 5 and 6.

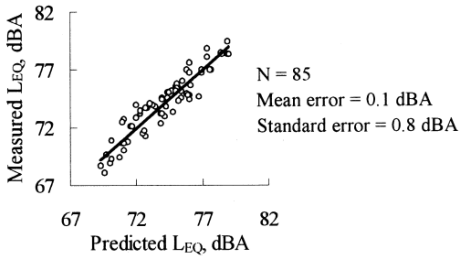


FIGURE 5 Evaluation of accuracy of model by field data

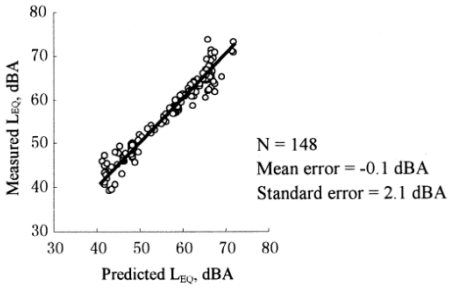


FIGURE 6 Evaluation of accuracy of model.

Evaluation of the modelled results indicates that noise prediction is more accurate at locations closer to the road carriage way where the environment of sound propagation is less complex. The model has an accuracy of 0.8

dBA for traffic noise prediction at locations nearer the road, and was accurate to 2.1 dBA for locations within the housing estate. Further, there were no systematic errors in the prediction results observed for either set of conditions. The accuracy in predictive results of the adjusted model for China is comparable to those of the FHWA model, whose predictive accuracy is 2.0 dBA.

5 Conclusions

We have presented a method for transforming GIS data into a topological and geometric representation suitable for use in traffic simulation. Our geometric representation of roads is visually smooth, including at ramps and intersections. Our method preserves the topological relationships of the GIS road network. We have shown examples of GIS data that have been processed by our method and composed with satellite images. These figures illustrate features of the road networks generated by our method, such as intersection handling and highway ramps, as well as the extensive scale of models that our method can process within a matter of few seconds.

References

[1] Helbing d 2001 Traffic and Related Self-Driven Many-Particle Systems *Rev of Modern Physics* **73**(4) 1067-141

[2] Sewall J, Wilkie D, Merrell R, Lin M 2010 Continuum Traffic Simulation *Computer Graphics Forum* **29**(2) 439-48

[3] Treiber M, Hennecke A, Helbing D 2000 Congested Traffic States in Empirical Observations and Microscopic Simulations *Physical Rev E* **62**(2) 1805-24

[4] U.S. Census Bureau "TIGER/Line," http://www.census.gov/geo/www/tiger/ 2010

[5] OpenStreetMap Community "OpenStreetMap," http://www.openstreetmap.org/ 2010

[6] NAVTEQ, "NAVTEQ" http://www.navteq.com/ 2010

[7] Mena J 2003 State of the Art on Automatic Road Extraction for GIS Update: A Novel Classification *Pattern Recognition Letters* 24 3037-58

[8] Park J, Saleh R, Yeu Y 200 Comprehensive Survey of Extraction Techniques of Linear Features from Remote Sensing Imagery for Updating Road Spatial Databases *Proc ASPRS-ACSM Ann Conf and FIG XXII Congress*

[9] Fortier A, Ziou D, Armenakis C, Wang S 1999 Survey of Work on Road Extraction in Aerial and Satellite Images *Technical report Center for Topographic Information Geomatics*

[10] Galin E, Peytavie A, Marechal N, Guerin E 2010 Procedural Generation of Roads *Computer Graphics Forum* **29**(2) 429-38

[11] Chen G, Esch G, Wonka P, Mueller P, Zhang E 2008 Interactive Procedural Street Modelling *Proc ACM SIGGRAPH*

[12] Prodecural Inc "CityEngine Manual http://www.procedural.com:8080/help/topic/com.procedural.cityengine.help/html/manual/is/create/streetshapes.html Feb 2011

[13] Farin G 1996 Curves and Surfaces for Computer-Aided Geometric Design: A Practical Code *Academic Press Inc*

[14] Cohen E, Riesenfeld R, Elber G Geometric Modeling with Splines: An Introduction *AK Peters Ltd*

[15] Piegl L, Tiller W 1997 The NURBS Book *Springer Verlag*

[16] Willemsen P, Kearney J, Wang H 2006 *IEEE Trans Visualization and Computer Graphics* **12**(3) 331-42

[17] Wang H, Kearney J, Atkinson K 2002 Arc-Length Parameterized Spline Curves for Real-Time Simulation *Proc Fifth Int'l Conf Curves and Surfaces* 387-96

[18] Van den Berg J, Overmars M 2007 Kinodynamic Motion Planning on Roadmaps in Dynamic Environments *Proc. IEEE/RSJ Int'l Conf Intelligent Robots and Systems* 4253-8

[19] Nieuwenhuisen D, Kamphuis A, Mooijekind M, Over-mars M 2004 Automatic Construction of Roadmaps for Path Planning in Games *Proc. Int'l Conf. Computer Games Artificial Intelligence Design and Education* 285-92

[20] Balaban I 1995 An Optimal Algorithm for Finding Segments Intersections *Proc 11th Ann Symp Computational Geometry* 211-9

Author

Jianda Zhu, born in January 1964, Shanghai, China.

Current position, grades: an associate professor in School of Architecture & Urban Planning, Suzhou University of Science and Technology, China.
University studies: Urban Planning, Design and Theory.
Scientific interest: computer simulation, urban planning and design.
Publications: 20 papers.

Financial management based decision making in the big data era

Ting Hao[1,2*], Xi Zhao[1]

[1]*Tianjin University, Tianjin, 300072, China*

[2]*Inner Mongolia University of Science & Technology, Inner Mongolia, 014010, China*

Received 1 March 2014, www.cmnt.lv

Abstract

Financial management are derived for the profitability analyses of demand-side management (DSM) alternatives. The present value of cost and equivalent uniform annual cost models are selected to determine the least-cost solution, while the net present value, pay back year and benefit/cost ratio models are proposed for the execution of cost–benefit analyses. In a market economy, Market orientation is the only correct choice for economic development strategy, so it is also significant for financial management innovation. While most studies in the past decade focused on the consequences of fund in financial management, few have investigated antecedents to market orientation concept. This paper derives two fuzzy financial profitability models, namely, a least cost solution model and a cost-benefit analysis model, to evaluate the fuzzy financial profitability of load management alternatives. A straightforward vertex parameters' fuzzy mathematics operation using the function principle is de-rived as an alternative to the traditional extension principle and is applied to evaluate a number of different financial decision indexes. The developed models represent readily implemented possibility analysis tools for use in the arena of uncertain financial decision-making.

Keywords: financial management, decision making, finance decision making, big data

1 Introduction

Financial management includes the functions of activities for solution to enterprise funds, costs, capital structure, cash flow, financial risk control, etc. As the innovation of financial in management level, Financial management innovation is the re-combination of elements of financial management to adapt to environmental change, then thereby remodelling and the establishment of their own financial capability. Market orientation has received much attention for its apparent positive effect on organizational performance. In a market economy, Market orientation is the only correct choice for economic development strategy, and the poor or rich of natural resources is the second. Because financial activities are the lifeline of business operations and financial management innovation is an important part of business innovation, so Market orientation is also significant for financial management innovation.

With the advances of information communication technologies, it is critical to improve the efficiency and accuracy of modern data processing techniques. The past decade has witnessed the tremendous technical advances in Sensor Networks, Internet/Web of Things, Cloud Computing, Mobile/Embedded Computing, Spatial/Temporal Data Processing, and Big Data, and these technologies have provided new opportunities and solutions to data processing techniques. Big data is an emerging paradigm applied to datasets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Such datasets are often from various sources (Variety) yet unstructured such as social media, sensors, scientific applications, surveillance, video and image archives, Internet texts and documents, Internet search indexing, medical records, business transactions and web logs; and are of large size (Volume) with fast data in/out (Velocity). More importantly, big data has to be of high value (Value) and establish trust in it for business decision making (Veracity). Various technologies are being discussed to support the handling of big data such as massively parallel processing databases, scalable storage systems, cloud computing platforms, and MapReduce. Big data is more than simply a matter of size; it is an opportunity to find insights in new and emerging types of data and content, to make business more agile, and to answer questions that were previously considered beyond our reach.

Strategic financial planning, or strategic planning for short, is a common way for organizations to plan to achieve their long-term financial objectives by focusing on where best to invest and allocate resources. Typically, strategic planning starts by establishing a baseline of performance and setting up goals over the planning horizon, including financial objectives. A strategic plan is then developed to map out the path to achieving these goals, along with indicators to track and measure per-

[*] ***Corresponding author's*** e-mail: tinghao_tju@126.com

formance with respect to the plan. While there is extensive agreement among business leaders that strategic planning is essential to enabling the continued success of a company, there are differences in the approaches used. Most of these, at least implicitly, acknowledge the uncertainties associated with the future as well as the existence of risks that may interfere with achieving the planned objectives. However, these approaches have proven not to be effective in reflecting and managing strategic risks. In fact, a recent study finds that over 80% of all cases of significant value loss among Top 1000 companies over the last 10 years are attributable to failure to adequately recognize and manage strategic risks.

Conventional fuzzy mathematical operations using the extension principle are applicable only to normalized fuzzy numbers. However, generalized fuzzy numbers (i.e., normalized and non-normalized fuzzy numbers) have the advantage that the degrees of confidence of a decision-makers' opinions can be rep-resented by their heights. Moreover, fuzzy mathematical operations using the extension principle change the membership function type of the fuzzy number following mathematical manipulation and involve complex and laborious mathematical operations. Accordingly, Chen proposed the function principle, which can be used to perform fuzzy mathematical operations on generalized fuzzy numbers. The authors pointed out that the fuzzy mathematical operations preserve the membership function type of the fuzzy number following mathematical manipulation and reduce the complexity and tediousness of the mathematical operations. Consequently, the present paper develops an easily implemented and conceptually straightforward vertexes operation using the function principle for application to fuzzy mathematics. The developed fuzzy mathematics operations are then applied to evaluate fuzzy financial indexes as part of a decision-making process. The proposed financial decision analysis method is more flexible and more intelligent than other methods since it takes the degree of confidence of the decision-makers' opinions into account.

Following the manipulation of fuzzy financial functions by fuzzy mathematics, the task of comparing or ranking the resultant fuzzy numbers can invoke a further problem since fuzzy numbers do not always yield a totally ordered set in the same way that crisp numbers do. Many authors have investigated the use of different fuzzy set ranking methods. These methods have been reviewed and compared. However, the majority of previous studies focused on the ranking of normalized fuzzy numbers, while relatively few considered the case of non-normalized fuzzy numbers. In this paper, a geometric moment model is derived to rank generalized fuzzy numbers based on the probability measure of fuzzy events. The geometric moments of a fuzzy number comprise the domain moment sand the grade moments.

2 Related works

Over the years, scholars has different opinions and views of what is Market orientation. After years of debates, at present, there are two points of view of Market orientation was generally accepted by marketing strategy literatures: Market orientation concept of culture and behavior. In the two views, formed different definitions and division of content of the Market orientation. Several scholars describe the marketing concept as a form of organizational culture. Deshpande&Webster [1-3] put forward a Market orientation is a kind of organizational culture, which could have the efficiency and effectiveness of the creation of the necessary acts and create outstanding customer value. Narver and Slater [4-6] further defined Market orientation in terms of culture and related it to the fundamental characteristics of the organization, although they asserted. That it is the organization culture that most effectively creates the necessary behaviors for creating superior value for buyers, and thus, continuous superior performance for the business. And it is characterized by:

l) Aim to profit and excellent customer value, whereas taking into account other interests of stakeholders;

2) To provide the behavior of organizational development and reflecting market information.

MO in their structure included elements of the three acts: customer orientation, competitor orientation and intersectoral collaboration, and there have two decision-making criteria: long-term perspective and attention to profits. As two of the components of Market orientation, customer orientation focuses on "the sufficient understanding of one's target buyers," whereas competitor orientation emphasizes the understanding of "the short-term strengths and weaknesses and long-term capabilities and strategies of both the key current and potential competitors" [7-9]. Further, customer and competitive orientations are two primary means that firms employ to interact with environments [10]. Behavior perspective of Market orientation Kohli and Jaworski [11] use the term Market orientation to describe organizational behaviors and activities that manifest the adoption of the marketing concept philosophy. Different from Culture Perspective of MO, Behavior Perspective is more focused on the study of specific behaviors related of MO [12]. Kohli and Jaworski [13-15] define market-oriented behaviors (MOBs) is in terms of three behavioral processes, namely the generation and dissemination of, and responsiveness to, market intelligence [16]. Concepted in the conduct of information infrastructure point of view, evolved several other acts of behavior perspective of MO.

The cash-flow models applied in many financial decision-making problems often involve some degree of uncertainty. In the case of deficient data, most decision-makers tend to rely on an expert's knowledge of financial information when carrying out their financial modelling

activities. Since the nature of this knowledge often tends to be vague rather than random, Dr. Zadehi introduced fuzzy set theory, which aimed to rationalise the uncertainty caused by vagueness or imprecision. However, practical applications of fuzzy set theory in the profitability arena require two laborious tasks, namely fuzzy mathematical operations and the comparison or ranking of the resultant complex fuzzy numbers. Fuzzy mathematics is based on the extended principles presented in [17], in which the traditional addition, subtraction, multiplication, division, power, logarithmic and exponent mathematical operations are applied to fuzzy numbers. Dubois and Prade [18] demonstrated that, when performing the binary manipulation of fuzzy numbers, the resultant increasing (decreasing) part arose from binary operations on the non-decreasing (non-increasing) parts of the two fuzzy numbers. The extended operations ensured that the resultant fuzzy number continuously maintained its fuzzy properties during the arithmetic operating procedure. It is found that fuzzy mathematics tends to be cumbersome for even the m or e straightforward operations such as addition and subtraction. Unfortunately, financial and engineering applications involving fuzzy sets typically require the more complex nonlinear mathematical operations such as product, division, power and logarithmic manipulations [19]. In some cases, fuzzy operations of this type may require an insurmountable computational effort. Consequently, it has been proposed that approximated triangular fuzzy numbers be used to examine the resultant fuzzy profitability indexes [20].

3 Methodology and analysis

3.1 MARKET ORIENTED FINANCIAL MANAGEMENT

The goal of financial management is to maximize stakeholder's value. In broader terms, business stakeholders include shareholders, creditors, and employees, suppliers, senders, the Government and the public. Because multiple targets will draw away financial manager's attention and not been as an action standard, selection the part of stakeholders in many stakeholders and maximizing their value is essential to the target of financial management. According with the foregoing summary, Culture Perspective of MO consider that Market orientation is an organizational culture containing behavior elements. Its characteristic is that let create profit and superior customer value as the supreme principle, while taking into account the interests of other stakeholders. This orientation point out the new standards to ensure financial management objectives.

In accordance with the time the methods of financial management including the operating methods of asset management, cost management, investment and financial management, income distribution methods, can be divided into financial decision-making and budgeting, financial control, financial analysis. As the traditional financial sector usually directly faces to the capital market but not to the product market, therefore, in terms of cost control, investment and financing management and distribution of income, the financial sector obtain corresponding data from the enterprise but not the market to carry out the management activities.

Market orientation has a greater impact on the use of financial management methods in enterprises. Under Market orientation, it requires the financial sector to face directly the product market and capital market at the same time, and access to product market information and capital market information related with financial management, to analysis customer and competitor, let financial activities directly linked to customer value.

Actions of decision-making, budgeting, control and analysis of financial management need to obtain a variety of financial information, including internal and external financial information, and the macro-and micro-financial information. To deal with the financial information cannot be separated from the market. Behavior Perspective of Market orientation hold that MO is the implementation of marketing concepts, as generation, dissemination of and response to market intelligence to the main line. Behavior Perspective of Market orientation can conduct the behavior criterion for financial work and market information response. This behavior criterion requires not only dealing with information on the capital market, but also carrying out customer orientation and competitor orientation to collect information on customers and competitors. In order to fully understand the needs of current and future customers, and competitors strengths, weaknesses and trends, coordinate financial sector functions with other internal departments in the financial management of enterprises, so as to give best to create more customer value than competitors offer, and then come true the purpose of maximizing customer value.

3.2 FINANCIAL MANAGEMENT BASED DECISION MAKING

The value creation mechanism and management of logistics financial management of supply chain system as shown in Figure 1. The external influence factors of the value creation of logistics financial management of supply chain system are macroeconomic environment, market environment, supply chain length and related parties factors, etc. The logistics financial management value of supply chain system mainly influenced by the supply chain cost and profit. The supply chain system value chain contains suppliers, manufacturers, distributors, final customers, reverse logistics and their respective logistics financial management elements such as technology infrastructure investment, financing costs, warehousing costs, supply chain optimization cost, customer service cost, reverse logistics cost, logistics transport cost, etc., respectively constitutes the subsystem's logistics financial management

structure of a supply chain system. Logistics financial management of supply chain system through use of logistics financial planning, organizing, control and other methods. Effect on the role of capital flow, logistics, information flow, and obtain a corresponding logistics financial management performance, optimize suppliers, manufacturers and distributors of their respective logistics financial management activities, so as to create the overall value of logistics financial management of supply chain system.

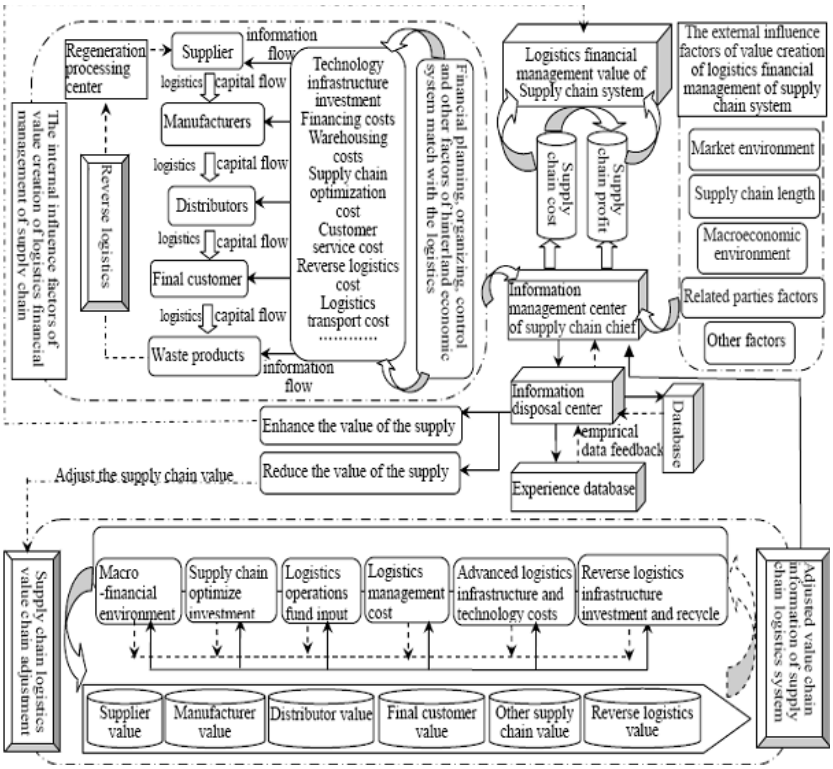


FIGURE 1 Logistics financial management

The status and data of the subsystems of the logistics financial management of supply chain system are changing constantly, and it can collect, dispose and feedback data depending on the information management center, information disposal center, data base and experience data base that built by the chief of supply chain. The advance or reduction of the value of logistics financial management of supply chain system can be measured through making the best of the data proposal function of information management center, receiving the input of information data from every process of supply chain and internal and external environment. The value chain of logistics financial management of supply chain will make adjustment according to the feedback received from information disposal center. Feed back to information management center of the chief, and hammer at the adjustment and improvement of logistics financial management during every part of value chain of logistics financial management of supply chain, so that the systematic value creation level of financial management is improved. As a result every aspect that affects the value improvement of logistics financial management of supply chain shapes the feedback loop of value creation mechanism and management.

Value chain is the process of value increase through the production of final products or services inside and between manufacturers. It covers all phases experienced by the commodities or services in the creation process from raw materials to final consumer goods. Logistics financial management of supply chain can achieve value maximum of the subsystem of every parts through value chain management, increasing combined running efficiency, in order to get competitive advantages through the improvement of market reaction speed, and the value creation efficiency of systematic logistics financial management will be improved finally. During the value chain of logistics financial management of supply chain, every logistics point of suppliers, manufacturers, distributors, final customers, reverse logistics and other stakeholders will be affected by macroscopic financial environment. Supply chain optimize investment, logistic operation fund input, logistic operation management cost, logistic technical investment, and reverse logistic infrastructure investment and recycle, and so forth, so that the efficiency of financial management and even the value creation efficiency of logistics financial management of supply chain will be affected.

3.3 LOGISTICS FINANCIAL MANAGEMENT MECHANISM

Macroscopic financial influencing factors: value creation of logistics financial management of supply chain can be influenced by macroscopic financial environment, and the main influencing factors include unemployment rate, market rate of interest, inflation rate, exchange rate and

international balance of payment, and so on. It is positive correlation with target variable.

Supply chain optimization: the value variable of supply chain optimization is mainly the increase or decrease of logistics points, making the currency of logistics reduced in supply chain system, logistics cost decreased, logistics points more reasonable, and supply chain more smooth. At present, the important problems of supply chain optimization are selecting suppliers, transfer warehouse and distributions, building proper inventory, establishing supplies related to the uncertainty of market demands, and fully displaying storage worth of supply chain. It is also possible to use logistics storage financing of supply chain to avoid financial deposit. Logistics storage financing is supply chain financing points using the document of title to get loan from business bank. Storage logistics finance can be third-party warrantor, avoiding financial risks efficiently, and also can accommodate logistics capital flows to blow up systematic logistics quantity and efficiently use long-standing inventory. Logistics finance/storage financing is one of the approaches of supply chain optimization, positive correlated to target variable.

Logistics operation fund input: under the influence of demand and supply of logistics finance, supply chain forms comparative condition of supply and demand of logistics fund and cost of supply chain. The comparative condition may be that the supply of logistics fund fulfils the demand of logistics fund, reaching the balance, so that the logistics of supply chain runs smoothly. Therefore, this variable is correlated positively to target variable in system.

Logistics operation management cost: relating to the manufacturing cost and quantity during logistics system of supply chain in ports-hinterland, including material cost, transportation, load and unload, transit, inventory, package, process, dispatching, wages and tax and interest and so on. It is correlated negatively to target variable. Backward status of logistics infrastructure and technology: it is necessary to increase investment to keep the advancement of logistics technical infrastructure of supply chain, and it is also good to improve logistics finance management value of supply chain through enhancing technology and consummating infrastructure.

While backward logistics infrastructure and technology block the logistics finance management value of supply chain, mainly incarnating at the aspects of transportation technology, load and unload technology, inventory technology, package technology, containerization technology and logistics information technology and so forth. It is correlated negatively to target variable. Reverse logistics infrastructure investment and recycle value: as a result of the pressure of living environment and law and regulations, and the drive of economic benefit, the reverse logistics is getting more and more attention. It contains reverse logistics in manufacture, reverse logistics in distribution, reverse logistics after distribution, customer B2B business reverse logistics, reverse logistics of customer service, and reverse logistics after use. The investment and recycle value of reverse logistics can improve supply chain value. Almost every parts of supply chain will come down to reverse logistics. So it is correlated positively to target variable

4 Case study

The fuzzy financial decision-making procedures are described briefly. First, the estimated input parameters, such as interest rate, inflation rate, investment, and operating revenue and/or cost, which are needed in financial index calculation, should be provided by the expert in form of fuzzy numbers. The fuzzy financial decision indexes are then calculated according to the models developed in Section III. The fuzzy financial decision is made finally according to the relative ranking of the resultant fuzzy financial indexes, which is performed following the process described in Figure 2.

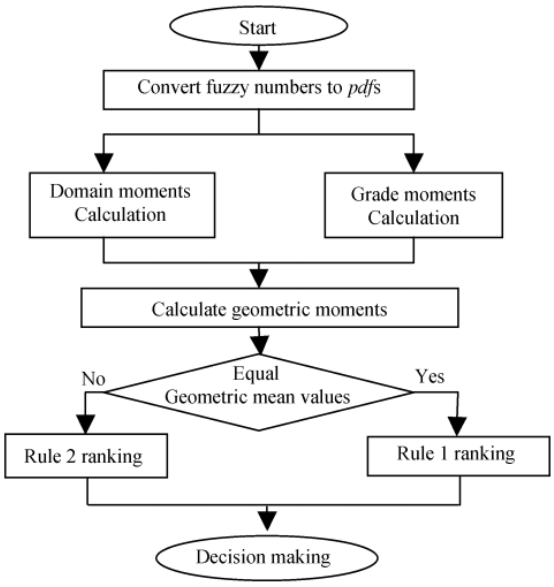


FIGURE 2 Flow chart of fuzzy number ranking process

A Cool Energy Storage air conditioning system (CES-system) functions by removing heat from a thermal storage medium during periods of low cooling demand and then releasing the stored cooling at a later time to meet an air-conditioning or process' cooling load. A CES-system has the advantage that it can meet the same total cooling load in a given period as a non-storage system but with a lower instantaneous cooling capacity. From the perspective of the utilities, a CES-system not only generates a reduction in the electrical demand but also shifts the consumer's energy consumption from peak time to off-peak hours. By decoupling the chiller operation from the instantaneous loads, CES-systems enable a more stable loading on the refrigeration equipment and increase the chiller efficiency due to the lower condensing temperatures associated with night time operation. It has been demonstrated that the use of cheaper off-peak energy and the reduction or elimination of peak-time charges enables CES-systems to pro-vide substantial operating cost savings. Various forms of cool storage

media may be used, including chilled water, ice, or a eutectic salt phase-change material. Depending upon the desired mode of operation, various combinations of these coolant paths can be achieved through the appropriate adjustment of the CES-system control mode.

5 Conclusions

This study contributes to the existing literature in three ways. First, most financial management innovation studies have been conducted at the fund orientation. The present study focuses on the financial management innovation at the MO. Second, this paper takes a first step on researching the impact of culture of Market orientation on financial management. The next Market orientation's three core activities (market information generation, dissemination and responsiveness). Finally, we research financial management innovation across four different aspects of the financial management, and thereby show the consequences of financial management innovation more than previous studies. While most studies in the past decade focused on the consequences of fund in financial management, few have investigated antecedents to Market orientation concept. To our knowledge, no study has investigated several of the antecedents included here. We hope to fill a part of this gap in extant knowledge.

Acknowledgement

Doctoral Research Fund of Ministry of Education, Grant No.: 2011002110035; Issue of Philosophy and Social Science of Baotou, Grant No.: 2014-sk30

References

[1] Vine E 2003 Public policy analysis of energy efficiency and load management in changing electricity business *Energy Policy* 31 1307–14
[2] Talukdar S, Gellings C W 1986 Load Management *New York IEEE Press*
[3] Nahmais S 1977 Fuzzy variables *Fuzzy Sets Syst* **1**(2) 97–110
[4] Dubois H, Prade D 1978 Operations on fuzzy numbers *Int. J. Syst. Sci* **9**(1) 613-26
[5] Zadeh L A 1975 The concepts of a linguistic variable and its application to approximate reasoning *Part 1, 2 and 3" Inf Sci* 8 199–249
[6] Dubois D, Prade H 1980 Fuzzy Sets and Systems: Theory and Applications *New York Academic*
[7] Kaufmann A, Gupta M M 1985 Introduction to Fuzzy Arithmetic: Theory and Applications *New York: Van Nostrand Reinhold*
[8] Chen S H 1985 Operations on fuzzy numbers with function principle *Tamkang J Manage Sci* **6**(10) 13-25
[9] Chen S J, Chen S M 2003 *IEEE Trans Fuzzy Syst* **11**(1) 45-56
[10] Chen S H 1998 Operations of fuzzy numbers with step form membership function using function principle *Inf Sci* 108 149-55
[11] Hsieh C H 1999 A model and algorithm of fuzzy product positioning *Inf. Sci* **121**(1 –2) 61 –82
[12] Chang W 1981 Ranking of fuzzy utilities with triangular membership function *in Proc Int Conf Policy Analysis Information Systems* 263-72
[13] Jain R 1976 Decision-marking in the presence of fuzzy variables *IEEE Trans. Syst., Man, Cybern* **6**(10) 698-703
[14] Dubois D, Prade H 1983 Ranking fuzzy numbers in the setting of possibility theory *Inf Sci* 30 183-224
[15] Chen S J, Hwang C L 1992 Fuzzy multiple attribute decision making methods and applications *in Lecture Notes in Economics and Mathematical Systems New York Springer-Verlag*
[16] Klir G J,Folger T A 1988 Fuzzy Sets, Uncertainty, and Information. Englewood Cliffs *NJ: Prentice-Hall*
[17] Delgadoet M, et al 1988 A procedure for ranking fuzzy numbers using fuzzy relations *Fuzzy Sets Syst* **26**(1) 49-62
[18] Lee E S, Li R J 1988 Comparison of fuzzy numbers based on the probability measure of fuzzy events *Comput Math Appl* **15**(10) 887-896
[19] Yoon K P 1996 A probabilistic approach to rank complex fuzzy numbers *Fuzzy Sets Syst* 80 167-76

Authors

Ting Hao, born in January 1987, Baotou, Inner Mongolia, China.

Current position, grades: a PhD Candidate in Tianjin University, a lecturer in Inner Mongolia University of Science & Technology, China.
University studies: management science and engineering.
Scientific interest: cost control, financial management, management science and engineering.
Publications: 15 papers.

Xi Zhao, born in June 1955, Tianjin, China.

Current position, grades: a professor in Tianjin University, China.
University studies: Management Science and Engineering
Scientific interest: corporate cost control and budget management, financial management.
Publications: 50 papers.

Extending opinion dynamics model for collective online behaviours analysis

Shixiong Wang*, Yi Jiang

Department of Management Science and Engineering, Zhejiang Sci-Tech University, 310018, Hangzhou, China

Received 10 September 2014, www.cmnt.lv

Abstract

In online social networks, opinion dynamics generally lead to different types of collective online behaviour such as consensus, polarization and fragment. Then an open problem arises: how are different typical collective online behaviours emerged from the behavioural decisions of individual and interactions among individuals during the process of opinion dynamics? This work examines the process of opinion dynamic in online social networks and different types of interactions among individuals on this process. An opinion-driven dynamics model, which combines a social network-based opinion dynamics model with generative individual behaviour, is proposed by adding antagonistic responses to the DW model. The proposed model integrates three types of interactions and setting up two thresholds to characterize individual behaviour. The behavioural component utilizes an initiation threshold such that if an individual's opinion exceeds this threshold, the individual will initiate the behaviour. In order to verify the effectiveness of the model, simulations are presented to examine how different typical collective behaviours emerge. As a result, we find that opinion dynamics with different threshold lead to different types of collective online behaviours. The openness of individuals to a differing opinion is the key factors to consensus or fragment.

Keywords: opinion dynamics, consensus, antagonism, fragment, online social network

1 Introduction

Online social networking sites such as Weibo are popular platforms for social interaction which may lead to collective behaviour emerged from local interactions among individuals [1,2]. The interaction can include a broad range of individual decision such as behavioural choices and transitions [3]. Therefore, collective online behaviour is driven by behaviour decisions of individuals in a social network environment, but it is not simply the aggregation of individual behaviours [4-6]. In online social networks, opinion and behaviour are woven into the fabric of individuals' daily life. This naturally leads to opinion and behaviour correlation between connected individuals. Behavioural decisions of individuals are triggered by their opinion and attitudes on a certain topic or events and influenced by the opinion and behaviour of friends or neighbours. Studies have demonstrated that individuals register an immediate and automatic reaction of "good" or "bad" towards everything they encounter in less than a second, even before they are aware of having formed an opinion [7]. Advertising, educational campaigns, and other persuasive media messages are all built on the premise that behaviour follows opinion, and opinion can be influenced with the right message delivered during the process of exchanging [8]. Therefore, there is an unprecedented opportunity to analyze collective online behaviour based on opinion-driven behavioural dynamics model, which assumes that collective online behaviour dynamics combines a social network-based opinion dynamics with generative individual behaviour.

In online environments, there are three typical collective behaviours [9-11]. The first is characterized by the emergence of a global consensus, in which all individuals reach the same state in the long run. Consensus behaviours are driven by opinion agreement and can lead to a movement towards uniformity such as collective condemning. In the consensus behaviours, all individuals have to interact to achieve the same state through changing opinion and behaviour. The second is characterized by the emergency of a bipartite consensus, in which all individuals achieve a double extreme convergence with identical magnitude but opposite sign. Bipartite consensus behaviours are driven by antagonistic opinion and can lead to a polarization phenomenon that often happens in a two-coalition community such that opposite opinions are held by two fractions. The third is characterized by the emergency of a global fragment, in which all individuals hold disagreement in the long run. Fragment is driven by diversity of opinions and lead to an anarchy. In examining three typical collective behaviours, we shall draw heavily on the interactions among individuals and opinions change of individuals. We also need to work on two different levels: the microscopic level, where the

* ***Corresponding author's*** e-mail: xim_wang@163.com

behavioural decisions of the individual occur, and the macroscopic level where collective behaviour can be observed. The greatest promise lies in analysis of linking microscopic opinion change and behavioural decision to macroscopic behaviours [12,13]. Then two open problems arise: how can we describe the behavioural decisions of individual and analyze different typical collective behaviours emerging from both individual interactions and opinion dynamics? Can we change individuals' opinions threshold to influence the behavioural decisions of individual and collective behaviours?

Analysis of social network-based interactions has proven effective and efficient to the understanding of collective behaviour. Recently opinion dynamics modelling has been provided for the analysis of social influences on individual opinions and the emergency of resulting collective behaviour. Several formal mathematical models have been proposed to simulate opinion dynamic, in which interactions between individuals differ from model to model. However, all opinion dynamic models are built on common theoretical roots, and all individuals adjust opinions based on local rules governing the interaction range, which is so-called neighbour-based method. In these models, a set of individuals are used to populate a community, and are seeded with initial opinion value. Each individual adjusts his opinion in the light of interactions with his neighbours. In order to answer those questions aforementioned, a proper choice of the dynamics modelling is the neighbour-based method. Therefore, an opinion-driven behavioural dynamics model is presented in this paper. In the proposed model, behavioural dynamics combines a neighbour-based opinion dynamics with the generative behavioural decision of individual.

Our purpose in this paper is achieved according to the following three aspects:

(i) As collective online behaviours are driven in part by opinions that individuals hold regarding a certain topic, the opinion-driven mechanism is described by opinion-behaviour mapping based on the interaction among individuals, which is the first step to analyze collective online behaviour.

(ii) As collective online behaviours emerged from individuals' opinions and behaviours which are influenced by their personal social network and updated based on neighbour-based method, an opinion-driven behavioural dynamics model is established to analyze three typical collective online behaviours, namely, consensus, polarization or bipartite consensus and fragmentation.

(iii) This study will examine how individuals' opinion dynamics influence the behavioural decisions of individuals and collective behaviours? Interventions to influence collective online behaviour are presented by applying the proposed model.

The following parts of this paper are organized as follows: In Section 2, we review classical opinion dynamic models and explain why an opinion-driven behavioural dynamics can be used to analyze collective online behavioural. In Section 3, opinion-behaviour mapping is established, and an opinion-driven behavioural dynamics model is presented. In Section 4, the proposed model is applied to analyze collective online behaviours using the method of computer simulating. Finally, the paper is concluded in Section 5.

2 Classical opinion dynamics

Consider that online social network is a graph $G=(V,E)$ with $V=\{1,2,\ldots,n\}$ and $E \subset V \times V$, where a node represents an individual and the edges represent the interactions between two individuals. In the graph, the neighbour set of the vertex i is defined by $N_i=\{ j \in V \mid (j, i) \in E \}$.The graph may be directed or undirected. A directed graph is used to model networks where relationships are not symmetric, for example, the follower relationship in Weibo. An undirected graph models a network with symmetric relationships where $(i, j) \in E$ implies $(j, i) \in E$ such as the friend relationship in Wechat. Two individuals i and j are called adjacent if there is an edge connecting them, i.e., $(i, j) \in E$, and individual i is a neighbour of individual j. Although the original work concentrated on mutual exchanges of opinion, interaction relationship in online social network are often regarded as directed edges. In this study, we focus our analysis on directed social networks.

Early works to study opinion dynamics were focused on exploring the patterns of interactions and consensus problem that can explain what kind of interactions will lead to agreement of opinions. There are two classical models including "binary opinions dynamics", where opinions are represented by binary value, and "continuous opinion dynamics", where opinions are represented by real positive numbers. In contrast, the latter deals with the problem of what happens to the probability of choosing one decision over another.

In recent years bounded confidence (BC) models have received significant attention. BC models are genuinely models of continuous opinion dynamics in which individuals have bounded confidence in others opinions. The first version of BC models was developed and investigated by Hegselmann and Krause [14,15], called HK model where agents synchronously update their opinions by averaging all opinions in their confidence bound. The other version was presented by Deffuant and Weisbuch [16,17], called DW model, where a pairwise-sequential updating procedure is employed. Both HK and DW are very similar, and assume that individuals have a continuous opinion and tolerance threshold. The principle of them is that an individual takes into account opinions from others in a limited zone which is defined by the tolerance threshold, around its own opinion. They differ in their update rule. In the DW model agents meet in random pair-wise encounters after which they compromise

or not. In the HK model, each agent moves to the average opinion of all agents which lie in his area of confidence.

This study considers the model of group opinion dynamics within an online social network that was DW model. DW model was initially constructed through randomized interactions among individuals who are assigned a random opinion, resided as a value on the interval [0,1] drawn from a uniform distribution and a threshold which limits the number of interactions that will result in an opinion change. The threshold can also be taken as a measure of uncertainty about a given issue, in which the individual opinion moves closer to his neighbours. If the difference between his opinion and that of his neighbours is considered too far apart (that is, exceeds their threshold), he is not willing to receive his neighbours' opinion, and no adjustment takes place. Each individual i has an opinion $x_i(t) \in [0,1]$ in the round t, which is the representation of the individual's support for a certain topic. Let $w_{ij} \geq 0$ be the weight that individual i places on the opinion of individual j with the normalization requirement that $\sum_{j=1}^{n} w_{ij} = 1$. Let $\varepsilon_i \in (0,1)$ be the threshold that represents the effect tolerance of individual i. In each round, each individual adjusts his opinion based on based on opinions of neighbouring agents and their own tolerance, taking a weighted average of his own opinion and the opinions of his neighbours and ignoring neighbours whose opinion is outside individual's tolerance. Collective behaviours take places in the discrete rounds. Specifically, the individual i updates his opinion as follows:

$$x_i(t+1) = w_{i1}x_1(t) + w_{i2}x_2(t) + w_{i3}x_3(t) + \ldots + w_{in}x_n(t), \quad (1)$$

where $w_{ij} > 0$ only if $(i, j) \in E$, and $|x_i(t) - x_j(t)| < \varepsilon_i$, otherwise, $w_{ij} = 0$.

3 Opinion-driven behavioural dynamics model

3.1 OPINION-BEHAVIOUR MAPPING

Collective online behaviours are emerged from individuals' behaviour decision and interactions among individuals. According to cognitive-behaviour theory, individuals' behaviour decision can be categorized as a behaviour change to the interaction which is triggered by opinion dynamics. In the original definition of DW opinion dynamics, there are two types of potential interactions between two individuals who are neighbours: positive interaction, in which the individuals' opinions and behaviours move closer to one another, and neutral interactions, in which the opinions are considered too far apart so that no adjustment takes place. Although these two possibilities capture a wide range of potential interactions, there is a third possibility in online social network: a negative interaction that drives the opinions of the individuals further apart. The proposed model extends it by adding antagonistic responses in order to enable us to capture a more complete range of interactions between different types of individuals. Besides, we extend opinion dynamics by adding behaviour as illustrated in Figure 1.

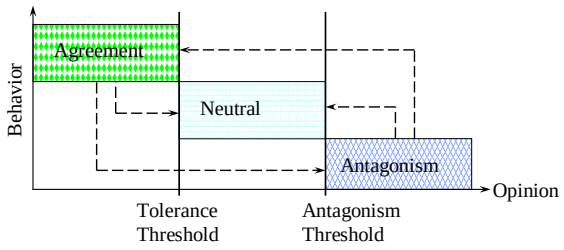


FIGURE 1 Graph illustration of opinion-behaviour mapping in online social network

For purposes of simplicity, a simple step function with the value of the behaviour being either true or false is proposed. The extended model describes the interactions between individuals by postulating that individuals' actions are driven by the objective of maximizing their own expected utilities, which depend on the state of collective behaviours in their own world. Individuals observe the actions of their neighbours, and update their opinions, optimally amalgamating public opinions obtained by observing their neighbours' behaviours. There are two initiation thresholds for every individual, including tolerance threshold and antagonism threshold, especially, the tolerance threshold is less than antagonism threshold. When an individual whose opinion differs from his neighbours by an amount exceeds the antagonism threshold value, he initiates the behaviour to give antagonistic responses to group opinions and collective behaviours. When an individual whose opinion differs from his neighbours by an amount is less than the tolerance threshold value, his opinion and behaviour move closer to group opinions and collective behaviour. When an individual whose opinion differs from his neighbours by an amount is between two thresholds, his opinion and behaviour stay the place. Therefore, addition of antagonistic response and mapping opinions to behaviour to classical DW model enables us to simulate a more complete range of opinion-driven behavioural dynamics among individuals in online social network as follows:

1) Agreement with individuals adjusting new opinions and behaviours closer to their friends whose opinions are similar.
2) Polarization with individuals adopting widely divergent opinions and behaviours when opinion difference exceeds the antagonism threshold.
3) Disagreement with individuals not affected by opinions of their friends where opinion differences are greater than the tolerance threshold.

3.2 THE PROPOSED MODEL

We extend the DW model to directed online social network and interpret the opinion-driven behavioural dynamics to represent overall social influences from all individuals rather than discrete pair-wise interactions. That is, the proposed model considers continuous interactions between individuals rather than the discrete exchanges.

There is a population with N individuals. In the round t each individual i has an opinion $x_i(t) \in [0,1]$ which is the representation of the individual's support for a certain topic, a tolerance threshold determining the latitude of reception $t_i \in (0,1)$ and an antagonism threshold determining the latitude of rejection $a_i \in (0,1)$ with $a_i > t_i$. Let $w_{ij}>0$ be the weight that individual i places on the opinion of individual j with the normalization requirement that $\sum_{j=1}^{n} w_{ij} = 1$. We refer to w_{ij} as the weight of edge (i, j), and $w_i = 1 - \sum_{j \neq i} w_{ij}$. In each round, each individual adjusts his opinion based on based on opinions of neighbouring agents and their own tolerance, taking a weighted average of his own opinion and the opinions of his neighbours and ignoring neighbours whose opinion is outside individual's tolerance. Besides, on account of individuals' different influence in online social network, a parameter is introduced to control for the strength of influence, which is determined by individuals' edges. Collective behaviour takes places in the discrete rounds. More specifically, individual i updates his opinion as follows:

$$x_i(t+1) = w_i x_i(t) + k \sum_{j \in T_i} w_{ij} x_j(t), \tag{2}$$

where T_i is the set of all out-degree neighbours of x_i (t) whose opinions fall out the bounds of antagonism threshold; $w_{ij}>0$ only if $(i, j) \in E$, otherwise, $w_{ij}=0$; k is a coefficient and $k \in \{-1,0,1\}$. If $|x_i(t) - x_j(t)| < t_i$, then $k = 1$, and T_i is the set of all neighbours of $x_i(t)$ whose connecting edge points from $x_i(t)$ and whose opinions fall within the tolerance threshold. The process can be viewed as individuals seeking to gain consensus with their friends in online social network. If $|x_i(t) - x_j(t)| \geq a_i$, then $k = -1$, and T_i is the set of all out-degree neighbours of $x_i(t)$ whose opinions fall out the bounds of antagonism threshold. The process can be viewed as some agents seeking to antagonistic response to polarization. If $t_i \leq |x_i(t) - x_j(t)| \leq a_i$, then $k = 0$, $|T_i|=0$ and $w_i = 1$.

4 Collective online behaviour analysis

The proposed model is applied to analyze three typical collective behaviours, and explore how three typical collective behaviours are driven by opinion dynamics and what will lead to different collective online behaviours. First we examine how adjusting tolerances influences collective online behaviour, which modify individual tolerances at strategic locations in the network. Next we examine how adjusting antagonisms influences collective online behaviour, which modify individual antagonisms at strategic locations in the network. Finally we examine how adjusting both tolerance and antagonisms influences collective online behaviour, which modify individual antagonisms at strategic locations in the network

According to complex network theory, the ability for a node to influence the network via opinion propagation is primarily determined by an individual's centrality. Online social network members who regularly exchange information with many others in the network are represented by nodes having greater centrality. In online social network, there are some opinion leaders with high centrality, who might have greater influence on making other individuals update opinions toward their own opinion. To effect behavioural change across the network, the proposed model allows us to examine the effectiveness of different nodes importance metric such as the betweenness centrality. Therefore, we adjust the tolerance threshold or the antagonism threshold for the 10 percent nodes with highest betweenness ranking and explore how it influences on collective online behaviour.

We simulated collective online behaviour using the proposed model based on *NetLogo*, which is a programmable modelling environment for simulating natural and social phenomena. The conditions in the simulations are the following: the population N indicates that there are N individuals in online social network, and $[w_{ij}]$ is a $n \times n$ matrix obtained from the adjacent matrix of the network graph. In the experiment, the population is setup $N = 1000$. Every agent has an opinion $x_i(t) \in [0,1]$ at time t and has two thresholds, tolerance T and antagonism A. Given a random network, interactions between two agents have been abstracted as equation (2).

4.1 ADJUST TOLERANCE TO INFLUENCE COLLECTIVE ONLINE BEHAVIOUR

Tolerance in the proposed model indicates the openness of an individual to a differing opinion, and is also termed uncertainty about one's own opinion. In the first experiment, we assume that the tolerance threshold for the ten percent nodes with the highest centrality ranking have the same initial opinion, the same tolerance $T \in (0, 1)$, and a constant antagonism $A = 0.8$. In Figure 2, the diagram shows that adjusting tolerance T results in the emergence of different collective online behaviours in case of the same initial opinion for central nodes. In the second experiment, we assume that the tolerance threshold for the ten percent nodes with the highest centrality ranking have opposite initial opinions, the same tolerance $T \in (0, 1)$ and a constant antagonism $A = 0.8$. One half of the ten percent nodes with the highest centrality ranking have initial opinion $x_i(t) \geq 0.9$, the other half of those nodes have initial opinion $x_i(t) \leq 0.1$. In Figure 3, the diagram shows

that adjusting tolerance T results in the emergence of different collective online behaviours in case of having opposite initial opinion for central nodes.

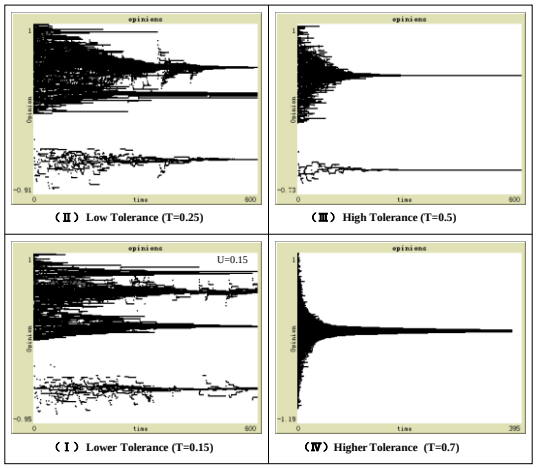


FIGURE 2 The emergency of collective online behaviour in case of different tolerance for central nodes having the same initial opinion and a constant antagonism threshold

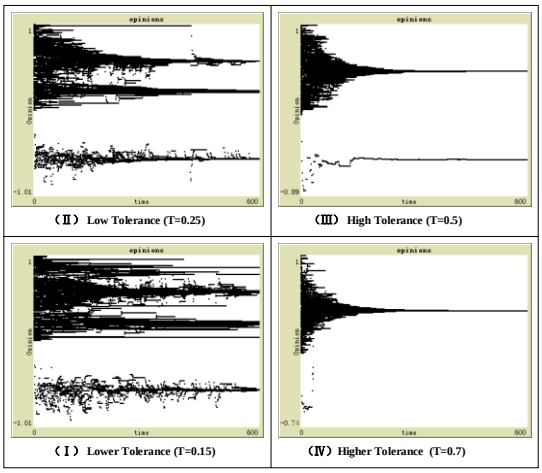


FIGURE 3 The emergency of collective online behaviour in case of different tolerance for central nodes having the opposite initial opinion and a constant antagonism threshold

From the result of simulations, we find that a network composed of individuals with a low tolerance threshold to their neighbours lead to fragment, and a network composed of individuals with a high tolerance typically results in one or bipartite consensus. Our simulations indicate that adjusting the tolerance threshold of the ten percent nodes with the highest centrality can have a dramatic influence on collective online behaviours. Raising the tolerance value of those ten percent nodes above 0.5 resulted in a large increase the possibility of clustering and consensus. Lowering the tolerance value of those 10 percent nodes below 0.5 strongly mitigated the possibility of clustering and consensus, and fragment emerges more easily. Especially, we find that initial opinions of the individuals have no significant impact on collective online behaviour by comparing the results of two experiments. In a word, an online social network composed of nodes with lower tolerance threshold leads to fragment more easily, those networks composed of nodes with lower tolerance threshold result in consensus even if there are three typical interactions including agreement, antagonism and neutrality.

4.2 ADJUST ANTAGONISM TO INFLUENCE COLLECTIVE ONLINE BEHAVIOUR

Antagonism in the proposed model determines the latitude of rejection of an individual to a differing opinion, and is also termed tenacity about one's own opinion. In the third experiment, we assume that the antagonism threshold for the ten percent nodes with the highest centrality ranking have the same initial opinion, the same antagonism $A \in (0,1)$ and $A > T$, and a constant tolerance $T = 0.5$. In Figure 4, the diagram shows that adjusting antagonism A results in the emergence of different collective online behaviours in case of the same initial opinion for central nodes.

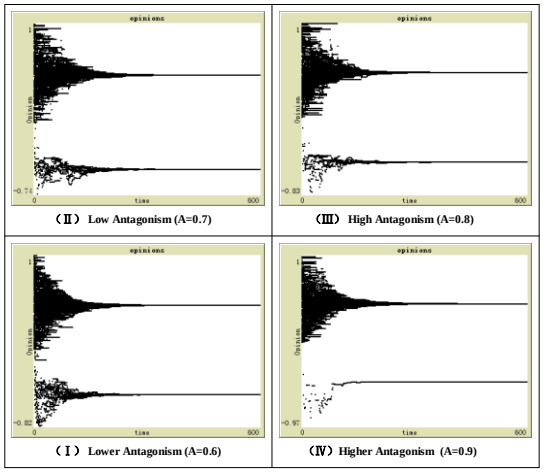


FIGURE 4 The emergency of collective online behaviour in case of different antagonism for central nodes having the same initial opinion and a constant tolerance threshold

In the forth experiment, we assume that the antagonism threshold for the ten percent nodes with the highest centrality ranking have opposite initial opinions, the same antagonism $A \in (0, 1)$, and constant tolerance $T = 0.3$. One half of the ten percent nodes with the highest centrality ranking have initial opinion $x_i(t) \geq 0.9$, the other half of those nodes have initial opinion $x_i(t) \leq 0.1$. In Figure 5, the diagram shows that adjusting antagonism threshold A results in the emergence of different collective online behaviours in case of having opposite initial opinion for central nodes.

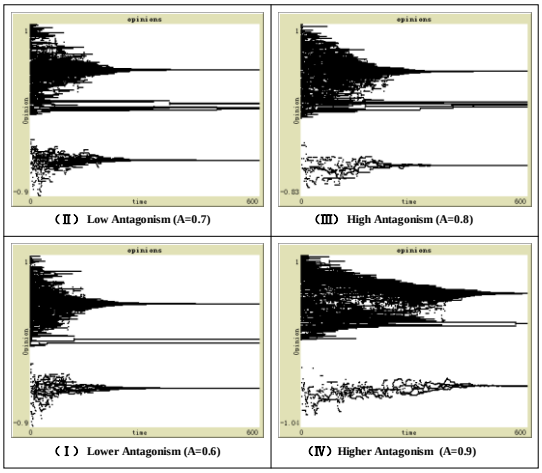


FIGURE 5 The emergency of collective online behaviour in case of different antagonism for central nodes having opposite initial opinion and a constant tolerance threshold

From the result of simulations, we find that a network composed of individuals with both a low and a high antagonism threshold to their neighbours and lead to the same type of collective behaviour, and the difference is only the speed of clustering. A network composed of individuals with a lower antagonism threshold has a higher speed of clustering. Our simulations indicate that adjusting the antagonism threshold of the ten percent nodes with the highest centrality can have a tiny influence on collective online behaviour. With the constant tolerance threshold both raising and lowering the tolerance value of those 10 percent nodes does not change the type of collective online behaviour. Especially, we find that initial opinions of the individuals have no significant impact on collective online behaviour by comparing the results of two experiments.

4.2 ADJUST BOTH TOLERANCE AND ANTAGONISM TO INFLUENCE COLLECTIVE ONLINE BEHAVIOUR

The difference between tolerance and antagonism decides whether an individual takes a neutral interaction with his neighbours. In the last experiment, we assume that individuals are initialized with the same initial opinion, the same antagonism $A \in (0, 1)$ and the same tolerance $T \in (0, 1)$. The values of A and T is varied between 0.1 and the maximum of 1.0 with the constraint of $A > T$.

In Figure 6, the diagram shows that different conditions for the values of T and A result in the emergence of consensus and anarchy. According to the result of simulations, we find that consensus emerges only in case of high uncertainty and high antagonism threshold as (III) shown in Figure 6. In other cases, polarization and fragment emerge as (I), (II) and (IV) shown in Figure 6. Fragment is more notable in case of low uncertainty and high antagonism, while polarization is more notable in case of low antagonism, and bipartite consensus is more notable in case of high uncertainty and low antagonism. As a result, difference between tolerance and antagonism is the key indicator to the emerging of consensus and fragment, and when the difference is smaller, consensus is attained more easily

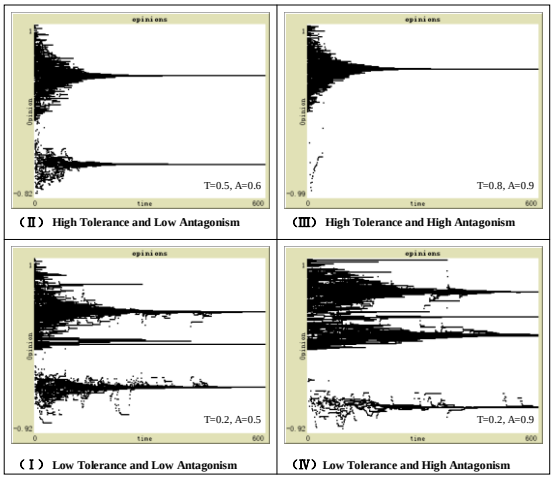


FIGURE 6 The emergency of collective online behaviour in case of a typical opinion trajectory for individuals with different antagonism A and different tolerance threshold T

5 Conclusions

In this paper, we study the process of collective online behaviour triggered by opinion dynamic in online social networks and the influences of different types of interactions among individuals on this process. Our work extended the classical DW model and investigated a possible source of consensus, polarization and fragment by adding antagonistic responses in order to enable us to capture more generative individual behaviours. We consider online social networks as directed graph, and there are three interactions among individuals such as agreement, antagonism and neutrality. The proposed model maps opinion dynamics to collective online behaviour and considers the continuous interactions among individuals. All individuals have a continuous opinion and two thresholds including tolerance and antagonism. We simulated collective online behaviour for the proposed model considered three types of interactions into online social networks based on *NetLogo* to observe the emerging of three typical collective behaviours. As a result, we find that opinion dynamics with different threshold lead to different types of collective online behaviours. The openness of individuals to a differing opinion is the key factors to consensus or fragment. Certainly, our study only focuses on the static social network and simulating for a constructed network. It is a next step to study real and dynamic social networks of collective online behaviour in our future research.

Acknowledgments

This research was supported by the Humanity and Social Science Youth foundation of Ministry of Education of China under Grant No.13YJCZH183 and Zhejiang Province Philosophy and Social Science Planning Projects of China under Grant No. 14NDJC204YB.

References

[1] Vispute A, Jadhav P, Kharat P V 2014 Collective Behavior of social Networking Sites *Journal of Computer Engineering* **16**(2) 75-9
[2] Ioanăs E, Stoica I 2013 Social Media and its Impact on Consumers Behavior *International Journal of Economic Practices and Theories* **4**(2) 295-303
[3] Patterson S, Bamieh B 2010 Interaction-driven opinion dynamics in online social networks *Proceedings of the First Workshop on Social Media Analytics* ACM New York NY USA 98-105
[4] Azeem A, Sujata D 2013 Collective behavior of online user based on social media *International Journal of Research in Advent Technology* **1**(5) 572-580.
[5] Bellomo N, Herrero M A, Tosin A 2013 On the dynamics of social conflicts looking for the Black Swan *Kinetic & Related Models* **6**(3) 459-79
[6] Gleesona J P, Cellaia D, Onnelab J P, Porter M A, Reed-Tsochas F 2014 A simple generative model of collective online behaviour *Proceedings of the National Academy of Sciences of the USA* **111**(29) 10411-5
[7] Moore T W, Finley P D, Linebarger J M, Outkin A V, Verzi S J, Brodsky N S, Cannon D C, Zagonel A A, Glass R J 2011 Extending Opinion Dynamics to Model Public Health Problems and Analyze Public Policy Interventions *Proceedings of 29th International Conference of the System Dynamics Society*, Washington DC 1-15
[8] Sobkowicz P, Kaschesky M，Bouchard G 2012 Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web *Government Information Quarterly* **29**(4) 470-9
[9] Li L, Scaglione A, Swami A, Zhao Q 2013 *IEEE Journal on Selected Areas in Communications* **31**(6) 1072-83
[10] Butail S, Bollt E M, Porfiri M 2013 Analysis and classification of collective behavior using generative modeling and nonlinear manifold learning *Journal of Theoretical Biology* **336** 185-99
[11] Bortolussi L, Hillston J,Latellac D, Massink M 2013 Continuous approximation of collective system behaviour A tutorial *Performance Evaluation* **70**(5) 317-49
[12] Tang L, Wang X F, Liu H 2012 Scalable Learning of Collective Behavior *IEEE Transactions on Knowledge and Data Engineering* **24**(6) 1080-91
[13] Hu J P, Zheng W X 2014 Emergent collective behaviors on coopetition network *Physics Letters A* **378**(26-27) 1787-96
[14] Deffuant G, Neau D, Amblard F, Weisbuch G 2000 Mixing beliefs among interacting agents *Advances in Complex Systems* **3**(1) 87-98
[15] Zhang J B 2014 Convergence analysis for asymmetric Deffuant-Weisbuch model *Kybernetika* **50**(1) 32-45
[16] Hegselmann R, Krause U 2002 Opinion dynamics and bounded confidence: models, analysis and simulation *Journal of Artificial Societies and Social Simulation* **5**(3) 2
[17] Yang Y, Dimarogonas D, Hu X 2014 Opinion consensus of modified Hegselmann-Krause models *Automatica* **50**(2) 622-7

Authors

Shixiong Wang, born in November 1976, Huanggang, China.

Current position, grades: associate professor.
University studies: PhD degree in Management Science and Engineering at the Donghua University.
Scientific interests: complex network, online user behavior analysis, information management, and complex system.
Publications: over 15 papers.
Experiences: employed as a full-time teacher since July 2004.

Yi Jiang, born in November 1979, Yichang, China.

Current position, grades: lecturer.
University studies: PhD degree in information science from the Central China Normal University.
Scientific interests: personalized information service, tech mining, complex network and online user behavior analysis.
Publications: over 10 papers.
Experiences: employed as a full-time teacher since 2002.

A student profile model based online English learning

Aiyun Guo*

School of Foreign Studies, Xi'an University of Finance and Economics, Xi'an Shaanxi, 710061, China

Received 1 March 2014, www.cmnt.lv

Abstract

Many Chinese universities have begun reforms to enhance educational competitiveness in our globalizing economy. This study aims to ascertain the status of English education and English-medium instruction at a Chinese engineering school and to offer workable suggestions for English communication training for Chinese graduate engineering students. Colleges and universities across China are adopting bilingual education to meet the need for well-rounded personnel with sound knowledge in specialized areas and competency in foreign languages. The development and difficulties of current bilingual education in China are discussed. First, the short-term English word context is generated to identify related concepts of the word. Second, the user context is generated based on the click through data of users. Finally, a forgetting factor is introduced to merge the independent user context in a user session, which maintains the evolution of user preferences. It is significant to the reform from teacher-centered to student-centered teaching mode. It is helpful to the cultivation of the students' collaborative ability and spirit, and has important theoretical significance and practical value.

Keywords: English education, online learning, student profile

1 Introduction

Under the increasing influence of globalization, many Asian governments have begun university reforms to enhance educational competitiveness and national competence in a globalizing economy. Subsequently, internationalization has impelled the restructuring of higher education institutions in a number of Asian nations, including China, Hong Kong, Japan, Singapore, South Korea, and Taiwan [1]. Individual governments have taken different approaches and measures to enhance the competitiveness of their institutions of higher education. The Chinese government has taken quantitative and qualitative measures such as the massification of higher education and the development of world-class universities since the mid-1990s [2]. With financial support from the government, Chinese universities have actively recruited high-quality academics from around the world to improve their educational quality [2]. The Japanese government has had concerns about the diminishing positions of their universities and has supported international collaborations and exchanges to facilitate the repositioning of Japanese higher education institutions [3]. In the case of Singapore, the government has carried out extensive university reforms with the intention of establishing the island country as a center of education in the Asia-Pacific area. To achieve this goal, the government has promoted the establishment of branch campuses of reputable foreign universities in addition to the reformation of university curricula [4].

The English-Chinese bilingual education in China has two main objectives. The first one is to empower China's elite youths to get advanced sciences and technologies directly from English literatures. Second, it helps students develop bilingual skills for efficient and effective worldwide communication. For these purposes, research has been conducted nationwide to help promote English-Chinese bilingual education, which includes evaluations of the training of bilingual teachers, and revisions of college curriculums for bilingual requirement. However, little research is done on the network aided learning environment for bilingual education. Current e-learning systems in China are mostly focused on continuing education and adult education. They are incapable of providing an original English environment and are often not suitable for bilingual courses.

The extensive use of digital technologies in China has greatly changed the college English teaching methodologies. College English Curriculum Requirements was issued to replace the Syllabus, and from then on it has served as an outline in the transformation effort. Accordingly, The Testing Syllabus for College English Test-Band 4 states, the objective of College English is to develop students ability to use English in a well-rounded way, especially in listening and speaking, so that in their future studies and careers as well as social interactions they will be able to communicate effectively in both oral and written forms. In this sense, CET-4 and CET-6 are aimed at measuring precisely college students' comprehensive employment of English and thus play an active role in realizing the objective of college English teaching." With the new orientation of the test objectives in The Testing Syllabus, college English teaching is expe-

* ***Corresponding author's*** e-mail: missguo168@163.com

riencing a change from the test whether the students have attained the various purposes established in the Teaching Syllabus after they have finished the four-year college learning to test how well the students can use English, which means developing students' all-round English abilities and The Requirements symbolizes that an era demanding a uniform teaching requirement is over, and a new era stressing the individualized and diversified college English teaching starts. After working under the guidance of the Syllabus and its refined version for about a decade and half, CE was found to be less and less up to date because of challenges from globalization, student demands, enrollment expansion, and changes in primary and middle school curricula. Therefore, in the new e-learning environment college English teaching must set up a new teaching system according to the societies' needs, the teachers' strong points and students' interests. In order to meet the students' requirements, some universities already started to offer some enrichment courses for undergraduates who have passed CET Band 4 & 6. College English teaching is changing from fostering English language knowledge to cultivating practical abilities. The main features of the transforming period is that the terms and the hours per week for college English teaching is shortening, while the number of the enrichment courses (the elective or optional courses) which emphasize developing ability is increasing.

In this paper, first, the short-term English word context is generated to identify related concepts of the word. Second, the user context is generated based on the click through data of users. Finally, a forgetting factor is introduced to merge the independent user context in a user session, which maintains the evolution of user preferences. It is significant to the reform from teacher-centered to student-centered teaching mode. It is helpful to the cultivation of the students' collaborative ability and spirit, and has important theoretical significance and practical value.

2 Related Works

As higher education rapidly accepts the notions and practices of e-Learning, the learning and teaching processes are experiencing ever-increasing changes [5-7]. Several universities, including Iranian universities, are currently providing online learning courses and facilities for students' use [8], which have significantly increased the quality of each field of study especially language education including language learning, language pedagogy, and language use [9]. However, the integration of leT tools in language classroom setting offers a broader concept of e-Learning which is defined variously in many studies. The most general educational demotion of e-learning is provided by Urdan and Weggen [l0] as a term for all types of electronic-supported learning and teaching which deliver the knowledge through electronic media such as the Internet, intranets, extranets, satellite broadcast, audio/video tape, interactive TV, CD-ROM. Moreover, sveral studies have compared the traditional language learning to e-learning of language and revealed a number of e-Learning benefits. Fageeh [11], for instance, indicated that e-Learning has effectively reduced the shortcomings of traditional learning including inadequate time devoted in the classroom, limited opportunities for students to discuss freely in the classroom as a result of their shyness or time restraints, lack of well-organized and better interaction with their peers and teachers synchronously or asynchronously. To confirm his comparison, Fageeh [11] carried out a descriptive research, including a survey and in-depth interview, to make out the perceptions of EFL students and their teachers towards e-Learning. The participants declared that e-Learning environment not only decreases the affective obstacles or inhibitors which motivate students to be more active and less silent and shy but also enhance inputs, outputs, and the processes of educational practices. Similarly, Mohammadi, Ghorbani and Hamidi [12] asserted that learning language via ICT resources is a learner centered, self-paced, cooperative, comprehensive, communicative learning which is dynamic and convenient any time and any place. Moreover, e-Learning provides opportunities for language learners to get familiar and interact with target language culture via online videos, chat rooms and videoconferencing [13].

Nevertheless, despite the growing numbers of e-Learning benefits, some studies have focused on the inhibitors in the implementation of e-Learning. One of the major obstacles seems to be related to the characteristics of ICT key users in e-learning settings, namely, language learners. Several research studies indicated that user-related inhibitors include lack of ICT literacy and computer anxiety [2], age and gender [14] as well as lack of readiness to use ICT resources [7].

Since the emergence of ICT tools in education in the 70s, language programs have enjoyed the implementation of these beneficial tools in language classes. It is suggested that ICT with its different forms from educational software to using networking has created a revolution in language teaching and learning [16]. ICT has not only been increasingly used in the practice of the skills of reading, writing, speaking, and listening as well as into corpus linguistics and testing but also provided lots of opportunities for language learners to enhance their communicative abilities [16]. Also, Beatty [17] characterized ICT-based language education as a process in which learners use computers and as a result of that they improve their language proficiency.

In sum, in line with the vital need of e-learning readiness assessment, this study investigated e-learning readiness among EFL students for the implementation of the e-learning language education in Shiraz Azad University. Moreover, the relationship between EFL students' English proficiency and their readiness was investigated.

3 The Proposed User Profile based Method

3.1 THE BASIC FRAMEWORK

An investigation of e-learning products leads to the fact that teaching materials are usually prepared in advance

and then lose flexibility in customer's interests as well as up-to-dateness in contents. On the other hand, abundant online resources (dictionaries, thesaurus, lexicons or encyclopedia etc.) are separated from contexts and becomes hence to some extent to be "big white elephant". The conceptual guideline of the framework here is thus allocated on the pedagogical motivation of "The customers decide what to learn and we help it to be well learned". The concrete technical measures to support this idea consist of three parts:

1) The texts to be processed are supplied or selected by customers;

2) related analytical measures based on web service discovery are introduced in the process of text analysis and teaching explanation generating, including language point determination, looking-up online dictionaries for new words;

3) the output of analytical results can be individually represented according to the specific need of customers both in forms and contents. The schematic framework is illustrated in Figure 1.

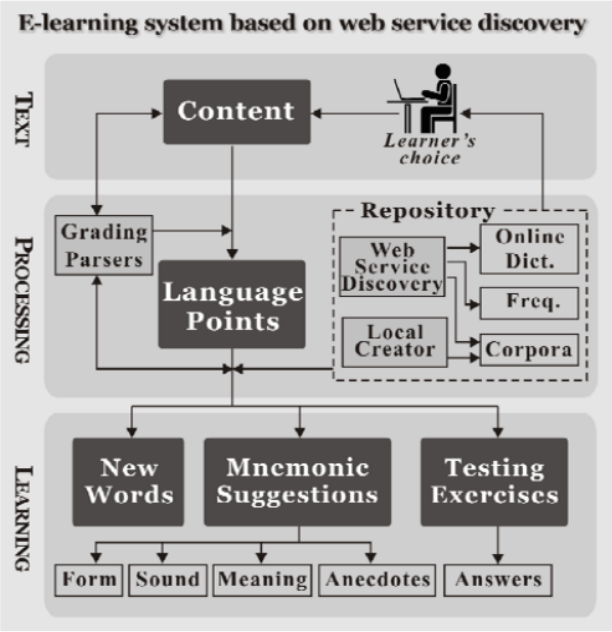


Figure 1 Flowchart of E-learning system

3.2 BUILDING THE WORD CONTEXT

Context, in its general form, refers to any additional information associated with the query [14]. In this paper, we narrow the context to represent a piece of text (e.g., a few words, a sentence, a paragraph) that has been authored by the users. Generally speaking, a word context can be represented as a concept vector. An obvious choice for extracting concepts of the query q is mining the Web pages returned by Web search engines, such as Google which provides the URLof each search result. However, the above choice is impractical. The reasons are given as follows:

1) Time consuming. Though Google provides the URL of each search result, it is time consuming to download these Web pages;
2) Parsing infeasible. Due to the huge number of Web pages and the high growth rate of the Web, it is impractical to analyze each Web search result pages directly and separately. Meanwhile, different Web sites have different HTML format, it is infeasible to parse the different Web sites at the Web scale.

Therefore, the web-snippets of the word are used for extracting concepts instead of the Web pages. Snippets are useful information resources provided by Web search engines, which are brief summaries of Web pages along with the search results. Generally speaking, a snippet contains a brief window of text selected by a Web search-engine around the query *q* in a Web page. Since many stopwords such as preposition, pronoun may occur on the snippets, it is necessary to do some preprocessing steps to reduce noise from the snippets before extracting concepts. Given the real time of building query context, we do not use some time consuming language dependent preprocessing steps such as part-of-speech tagging. Instead, we only remove the stopwords using standard SMART stopword list. After preprocessing the snippets of the query *q*, it is time to extract concepts of the query q. Our concepts extraction method is inspired by the famous problem of finding frequent patterns in data mining. When the query q is submitted to the Web search engine, a set of snippets are obtained for identifying the concepts. According to the theory of cognitive science, if a concept appears frequently in the snippets of the query q, it represents an important concept related to the query q since it coexists in close proximity with the query q in the top Web search results.

Besides the weight of concept ci, the relations between concepts can also be mined from the snippets. A famous formula named Pointwise Mutual Information (PMI) from information theory [20] is used to compute the relation between concepts.

3.3 BUILDING STUDENT PROFILE

The query context reflects the related concept of the query *q* in the Web search result pages, which can be derived without any user click through data. Different from query context which is static, the user context is dynamic and based on the click through data of the users. In other words, the user context is user-oriented. Given the query *q*, the problem of building user context can be viewed as a three-stage process.

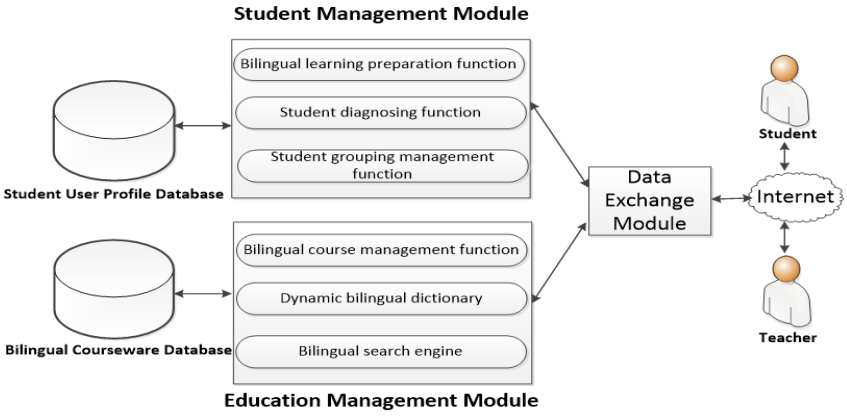


Figure 2 System organization

1) Obtain the explicit concepts of user context. The explicit concepts mean the concepts appear in the click snippets of the users. For example, when the user searches the query "apple," then he/she clicks the snippet which contains the concept "iPhone," thus the concept "iPhone" is the explicit concept of the user.
2) Obtain the implicit concepts of user context. The implicit concepts mean the concepts do not appear in the click snippets but may be interested by the users. For example, if the user is interested in the concept "iPhone," the concepts which are related to the "iPhone" such as "iPod" may be the implicit concepts of the users.
3) Process the sequential click snippets of user context. Each click snippet of users can be used to generate a context, respectively. The merging of these user contexts in a user session should be considered.

Intuitively, the concepts which appear in the user click snippets can be thought as the explicit concepts of the user context. For instance, if a user submits the query "apple," and he/she is interested in the concept "iPhone," then he/she may click the snippet containing the concept "iPhone". In other words, the concepts which appear in the snippets of the Web search result may be interested by the user who clicks it. Besides the concepts which appear in the snippets clicked by the users, other concepts may also be interested by the users. The implicit concepts mean the concepts do not appear in the click snippets but may be interested by the user which is shown in Figure 2.

The concept relationship graph of query context gives us a chance to find the implicit concepts of the user context. If a user is interested in the concept ci, the concepts which are the neighborhood of the concept ci in the concept relationship graph are the implicit concepts of the user context, which means these concepts are more likely interested by the users. For instance, if a user submits the query "apple," and he/she is interested in the concept "iPhone," then he/she may click the snippets containing the concept "iPhone". Obviously, the concepts which are neighborhood of the concept "iPhone" in the concept relationship graph of query "apple" such as "stock" and "store" may be interested by the users. Therefore, we compute not only the weight of the explicit concepts appearing in the click snippet, but also the weight of the implicit concepts which are the neighborhood of the explicit concepts in the concept relationship graph. An intuitive method for computing the weight of the implicit concepts ci is using the strength of the link between the implicit concept ci and the explicit concept cj. Unfortunately, this method is impractical since one implicit concept may relate to many explicit concepts. For example, suppose "iPhone" and "iPod" are two explicit concepts appearing in the click snippet, but the implicit concept "Mac" may link to both "iPhone" and "iPod." In that case, it is difficult to select which link should be the weight of the implicit concept "Mac" to the user context. To address this problem, three strategies are proposed to compute the weight of the implicit concepts of the user context.

3.4 RESEARCH INSTRUMENTS

For the purpose of this study two major instruments were used including e-learning readiness questionnaire and a retired version of Test of English as a Foreign Language (TOEFL). First, EFL students' e-learning readiness was assessed by using e-Learning readiness questionnaire adapted from Watkins, Leigh, and Triner. The privilege of using this scale is to provide practitioners with tools for improving both individual and organizational performance through useful e-learning experiences. The questionnaire has two major sections. The first section includes demographic characteristics such as age, gender, computer use, computer ownership, internet use for academic activities. The second part consists of 27 closed-ended items categorized under six factors in order of:

1) technology access (three items),
2) online skills and relationships (nine items),
3) motivation (three items),
4) online audio/video (three items),
5) internet discussion (four items),
6) importance to one's success (five items).

The participants were asked to rate their readiness in e-learning environment on a 5-point Likert scale based on Strongly Disagree (1 point), Disagree (2 points), Unsure (3 points), Agree (4 points), Strongly Agree(5 points).

The overall reliability for this questionnaire was found by estimating Cronbach's alpha (0.75) making it acceptable in terms of internal consistency. The results of reliability analysis for six components of the scale indicate acceptable levels of internal reliability for all components ranging from 0.63 to 0.87. Secondly, a retired version of TOEFL test was used to investigate the proficiency score of the participants. This test consists of 30 questions in three parts (reading comprehension, structure and written expression).

4 Analysis

Digital technology makes it possible to construct a system of teaching contents, which can combine words, pictures, sounds and visual clips together, and thus content-based and multi dimensional approach of college English enrichment course teaching arises. This teaching mode is based on the digital technology and multimedia teaching environment, so it can embody the principle of combining practicality, interest and knowledge together. It can be realized both in the classroom activities and after-class Internet-based autonomous learning. During the classroom activities, the teachers will concentrate on some role play activities, so as to develop the students' autonomous and initiative learning spirits.

Then how to adopt the content-based teaching mode of CEEC in the e-learning environment? Before class activities: This technology provides a very good environment for both the students and the teachers to finish their tasks before each lesson. With online accessibility, they can read up-to-date English materials on the websites of the BBC or The New York Times, visit English learning websites to complete various after-class assignments online. They can also easily find English-speaking partners through the Internet or participate in online forums in English. After the students collect the materials closely related to the certain theme, they can discuss and communicate in various online ways and make some PPT documents so that they can present group work in class. This process must involve students' autonomous and cooperative learning. Compared with the confinement of the traditional classroom, an e-learning environment gives them many more opportunities to improve their English ability dramatically. Incorporating digital technology to college English enrichment courses is not only effective in improving students' English proficiency, but also offers teachers many more possibilities. At our university, we can teach every multimedia college English class, which means we can incorporate video and other visual aids into classroom work.

The students are enthusiastic about using video clips of English movies for a lesson with the particular theme. In class activities: In the e-learning environment, tasks are more easily assigned to cater to students' needs and interests. The teacher design some pair work, group work role play and imitating activities to improve students' oral expressing abilities. The e-learning environment will surely make it feasible for the students to be immersed in authentic English learning environments, so as to enhance their sense of language and cultivate their culturally communicative abilities. The e-learning environment also makes it possible to combine after-class work with the in-class work, combine the online work with the offline work, and combine the multimedia teaching with the traditional teaching. Thus, the maximum amount of authentic communicative activities make students strongly stimulated by audio video clips, enrich their intellectual functions and train their cooperative and initiative consciousness.

A sectional interface of the online teaching and learning system is illustrated. It is composed of three sections: (1) text display; (2) new word annotation and mnemonic suggestion; and (3) new word association (e.g. synonyms and antonyms). The Gettysburg Address by U.S. President Abraham Lincoln is selected as a demonstrative built-in text in our created repository to illustrate the functionalities of our system. The entire speech consisted of 272 words as Encyclopedia tells. Our system gives exactly the same result and uncovers more facts about it as illustrated in Table II. Of the total 272 words, 138 are distinct. From the results, readers could find that easy texts for primary learners (like B in the table) are with a lower coefficient and a condensed reading text for advance learners (like C) has a higher value. A speech, which is usually composed of colloquial expressions, is often a medium degree of difficulty and medium coverage of new words.

The facts of texts B, C and D in Table II reveal that text B belongs to essential course with easy words and low degree of difficulty; text C advocates an intensive expanding of vocabulary; and text D in the last volume of the series does not have the largest vocabulary coverage but rather tackle with cultural issues as the author designed, as can see from the series names: First Things First, Practice And Progress, Developing Skills, and Fluency in English. Such an arrangement is also introduced in author's pedagogical instructions. The analytical results of the example texts demonstrate the important function of text evaluation for teachers to select a proper text aiming a specific group of students. After the reading text is scanned through above algorithm, new words come out from the distinct words that are equal or higher than the level the customer has assigned. They are marked by bold and blue letters. Once the list of new words is determined, the list is sent to the repository system, where the standing MAS serves for vocabulary data collection and administration in advance and real-time if necessary. It sends back all word information on the list, including vocabulary notes, mnemonic suggestions, learning points, and test questions. Under the category of mnemonic suggestions and learning points, useful learning materials are provided there. Distinction notes on similar words in form, for example "repository" and depository, are compared. Similarly, "sun" and "son" are pronounced the same; both "attain" and "obtain" have a meaning of "to gain or get" etc. They are all served for improving the impressive memory of learners to new words. Some anecdotes about certain words are also recorded in the mnemonic corpus of the repository, for example, "arithmetic" as the acronym of "a rat in Tom's house may eat Tom's ice cream". The webpage frame of test questions shows the test question of corresponding

new word to deliver practice for learners to master new words. After all new words are learned and tested, a score is then fed back to the learners to reminder them the degree of their understanding and digestion.

5 Conclusion

This paper first brings about three problems in the traditional teaching mode of college English interpretation class. They are disconnection between teachers' instruction and students' learning, shortage of interaction between teachers and students, the out-of-date content and the single teaching methods, static instruction and shortage of authentic practice. Then, the author points out that applying E-Learning in college interpretation instruction may help solve the three problems. e-learning, as a new way of education, has many characteristics, such as individuality of the subject, interaction of the learning process, integration of learning content and the opportunity for life-long learning. In the e-learning environment, college English interpretation teaching shows a new development trend. The roles of teachers and students will be changed. Students become the subject of teaching and they have more autonomy. While teachers play the role of monitoring and directing and they become the designer of the learning, the organizer and the attendant of teaching activities, and also the guidance and the consultant of the students' learning. In order to meet the requirements of the modem college English interpretation course, teachers have to learn the computer and network techniques well, or they will encounter some difficulties and problems. Altogether, nowadays e-learning and network learning have become the new trend of education and their advantages must greatly facilitate college English interpretation teaching and learning in the future.

References

[1] Hogo M A 2010 Evaluation of E-Learning systems based on fuzzy clustering models and statistical tools *Expert Systems with Applications* **37** (10) 6891-6903
[2] Huang T 2005 Qualitative Enhancement and Quantitative Growth: Changes and Trends of China's Higher Education *Higher Education Policy* 18, 117–130
[3] Sun Y, Todo N, Inoue N 2014 Exploration of Higher Education Indicators and Universities' Characteristics in Japan *IIAI 3rd International Conference on Advanced Applied Informatics*
[4] Comerchero M 2006 E-Learning concepts and techniques
[5] Kau K, Abas Z W 2004 An assessment of e-Learning readiness at Open University Malaysia *International Conference on Computers in Education (ICCE)* Melbourne
[6] Rahimi M, Yadollahi S 2009 Computer anxiety: A comparison between campus-based and distance learning EFL students *The First International Conference on E-Learning and Teaching*
[7] Hung M, Chou C, Chen C, Own Z 2010 Learner readiness for online learning: Scale development and student perceptions *Computers & Education* **55**(3) 1080-90
[8] Darab B, Montazer G H 2011 An eclectic model for assessing e-learning readiness in the Iranian universities *Computers & Education* **56**(3) 900-10
[9] Tanveer M 2011 Integrating E-Learning in Classroom-based Language Teaching: Perceptions, Challenges and Strategies *International Conference iCT for language learning*
[10] Urdan T, Weggen C 2000 Corporate E-learning: Exploring a New Frontier
[11] Fageeh A 2011 EFL students' readiness for e-Learning: Factors influencing e-learners' acceptance of the Blackboard in a Saudi university *JALT CALL Journal* **7**(1) 19-42 http://www.jaltcall.orgljournallarticles17_1 Jageeh.pdf.
[12] Mohammadi N, Ghorbani V, Hamidi F 2006 Effects of e-Learning on language learning Procedia *Computer Science* 3 464-8
[13] Chen Y 2008 A mixed-method study of EFL teachers' Internet use in language instruction *Teaching and Teacher Education* 24(4) 1015-28
[14] Jones A 2004 British Educational Communications and Technology Agency 'http://dera.ioe.ac.uk/id/eprintI1603
[15] Warner D, Christie G, Choy S 1998 The readiness of the VET sector for flexible delivery including on-line learning: a guide for developers of curriculum and training packages *Australian National Training Authority*
[16] Roohi E 2009 Exploring the Effects of Using CALL as a Discriminator of L2 Performance of Iranian EFL Learners *The 7th International TELLS Conference* Yazd lran
[17] Beatty K 2004 Teaching and researching computer assisted language learning: Applied Linguistics in Action *Harlow Longman*

Author

Aiyun Guo, born in June 1970, Xi'an Shaanxi Province, China.

Current position, grades: lecturer at the School of Foreign Studies.
University studies: Master's degree in Language and Literature.
Scientific interest: language and literature, semantic web.
Publications: 15 papers published in the international or national journals.

Industrial design based on computer aided simulation

Jianxun Wang*, Qingyun Zhou

Zhengzhou Technical College, Zhengzhou, 450121, China

Received 1 March 2014, www.cmnt.lv

Abstract

Recently, new technologies have emerged in industrial automation platforms. A rapid modelling and simulation environment is required to integrate these new technologies with existing devices and platforms to reduce the design effort and time to market. System-level modelling is a popular design technique that provides early simulation, verification, and architectural exploration. However, integration of real devices with system models is quite challenging due to synchronization and hard real-time constraints in industrial automation. Simulations are software tools approximating and predicting the behaviour of real industrial plants. Unlike real plants, the utilization of simulations cannot cause dam-ages and it saves time and costs during series of experiments. A shortcoming of current simulation models is the complicated runtime integration into legacy industrial systems and platforms, as well as ad-hoc design phase, introducing manual and error-prone work. This paper contributes to improve the efficiency of simulation model design and integration. It utilizes a semantic knowledge base, implemented by ontologies and their mappings. The integration uses the Automation Service Bus and the paper explains how to configure the runtime integration level semantically.

Keywords: assembly systems, estimation, part feeding, pose statistics, simulation

1 Introduction

Industrial automation systems are experiencing a paradigm shift due to incorporation of new technologies, such as embedded real-time devices and communication networks. It has been shown that the demand for plug and play mechatronic solutions increases significantly and the investments on automation systems that utilize hardware–software and communication infrastructures cannot be ignored in today's factory systems [1-3]. As the complexity of these systems increase, a rapid modelling and simulation environment is required to reduce the design and verification time. System-level modelling is a very effective way of reducing the development cycle while providing early prototypes and enabling architectural exploration and verification.

The role of simulation models is becoming more essential as industrial systems are being more complex and their structure more sophisticated. Simulation models are necessary for the advanced process control and they are important for optimization or as test-beds to analyze the behavior of both a control system and the whole plant under normal, extreme or testing conditions, when an engineer compares different parameters or control algorithms under various operation scenarios.

A lot of research effort has been invested into problems dealing with mathematical modelling, some effort has been in-vested into the description of models. But only few works have been concerned with the description of the whole automation system including knowledge about the real plant, control system, software tools, and communication interfaces not only on the hardware level, but also on the software level. Such pieces of information should be explicitly conceptualized, interfaces for the integration of sub-systems should be described in a machine-understandable form and the integrated solution has to be consistent and easy to maintain.

System-level modelling also allows reuse of different forms of intellectual properties (IPs), which is a common practice in the industry. Since industrial automation systems are increasingly connected with other IPs or industrial components, they should also benefit from system-level modelling techniques to reduce the development costs. In traditional system-level modelling, all components need to be modelled. However, modelling has no added value for components that are already physically implemented because modelling is an abstraction mechanism and requires human effort. Therefore, techniques have to be developed for incorporating real devices with virtual models. Traditionally, there are communication mechanisms between virtual models. These mechanisms are defined through the constructs of modelling languages. Similarly, real implemented devices communicate with each other through physical mediums with predefined exchange data formats. However, there is no established mechanism that provides communication between real devices and virtual models, and this paper fills this gap. Specifically, we present synchronization mechanisms between virtual models and real devices so as to achieve real-time communication.

The goal of this paper is to introduce an ontology-based approach for the plant description that is focused on both

* ***Corresponding author's*** e-mail: wangjianxuanzz@126.com

complex and flexible automation systems. Although this paper deals especially with simulation models, the approach provides additional support for integration of diverse engineering tools that are involved in the automation system engineering as well. All phases of the automation system life-cycle are based on the description of the real industrial plant. In the early design stage, such a description is needed for optimization of the process structure and parameters, afterwards, it is useful to design a simulation model as a virtual test-bed for the control system design and fine-tuning. Since every industrial process is modified and maintained, the plant description has to reflect the current state. Particularly, modifications of the industrial process bring fundamental benefits of the explicit plant description that is independent on other engineering tools, such as simulators. For example, in approaches used nowadays a simulation expert has to modify the simulation model according to the modified plant, which is usually time-consuming, because he or she has to study principles of the simulation model, meaning of required parameters and often empirically check the results. Using the explicit plant description a process specialist modifies this description according to the real plant and all necessary changes are propagated into the simulation model and other systems in the automated way. Traceability of changes is one of the important features of modern system engineering.

This paper addresses problems related to the integration of automation systems including simulations, human-machine interface (HMI) as a part of SCADA systems (standing for supervisory control and data acquisition), and interaction with real systems solved via OPC UA or OPC classic. Some tools can be legacy, i.e., they had been designed before the need for interoperability emerged. One of the automation system tools are simulations, being the software approximations of the real plant behavior. Since their design phase, reuse and access to historical data pose problems which have not been satisfactorily solved by now, this paper addresses these issues. The presented simulation model design is oriented on structural issues dealing with simulation blocks, their interconnections, and parameters; whereas mathematical equations are not handled.

The proposed approach utilizes ontologies comprising knowledge about design and integration of a real plant, a simulation model, as well as other industrial automation tools. The knowledge base provides the information in a machine-understandable form, hence the computer is able to process and query the knowledge required for the integration and the simulation model design. The presented approach is an application of the semantic integration [4-7]. In contrast to classic integration approaches, the semantic integration is based on utilizing mappings between adequate entities. For example, semantic map-pings relate really measured variables and their simulated approximations, real devices and their equivalents in a simulation model or local tag names used in a particular tool with the global representation of tag names.

2 Related Works

In the last decade, there have been significant studies to develop design platforms for intelligent industrial automation systems [8-10]. These automation systems rely heavily on a distributed computer-based infrastructure, where smart sensors and actuators, intelligent machines, robots, and other automation devices can interact using industrial protocols and take decisions in real time. In these systems, system-level communication, device synchronization, and the integration of new devices to the system are extremely challenging issues. Hence, there is a demand for sophisticated tools for the design of intelligent complex industrial automation systems. In SIMOO-RT [11-13], an object-oriented framework is pro-posed for modelling a real-time industrial automation system. A system model with real devices cannot be developed with SIMOO-RT. OOONEIDA [14-16] complies with IEC 61449 [17], which is a standard to design distributed control systems with intelligent devices. It proposes encapsulation of different types of IPs into reusable portable software modules called as function blocks. To achieve this, interfaces are created between various kinds of automation IPs such as devices, RTOSes, machines, systems, and industrial enterprises. However, in IEC 61499, the real-time properties of applications (e.g., reaction time) and resources (e.g., polling of data by function blocks, communication) are unspecified [18]. In OOONEIDA, real-time properties have to be handled via embedded controllers that are introduced to the system by encapsulation. There has also been a proposal on handling real-time issues in function blocks [19], but it has not been implemented yet. The encapsulation modules are developed in Java for the Eclipse and NetBeans integrated development environments. In RIMACS [20] project, a service-oriented architecture with real-time capabilities is proposed. Temporal behavior of each activity is isolated as much as possible by using dedicated hardware and software [4]. Linux kernel is modified to provide temporal isolation.

On the technical interoperability level, there can be found approaches using general-purpose distribution techniques such as DCOM, CORBA, J233, as well as architectures or frameworks related to the simulation area, such as DIS, SEDRIS, HLA. Especially HLA is used in many approaches, but its shortcomings are the lack of a semantic layer and data source description. The ex-tension of HLA with semantics is proposed in [12]. Ref. [114] proposes the Ontology Driven Simulation Design Tool (ODS). It is based on two ontologies that are mapped: the domain ontology categorizing knowledge including a problem vocabulary, and the modelling ontology used for the simulation model description.

3 Motivation and Formulation

The major challenges for the automation system integration in runtime are expressed in Figure 1. It depicts the typical architecture used in the industrial practice now.

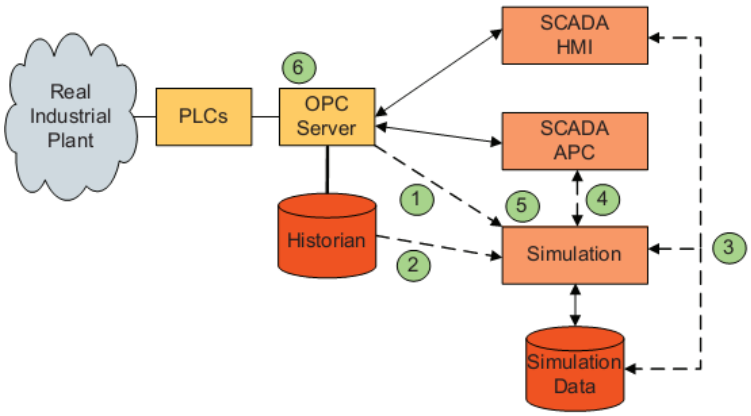


FIGURE 1 Current approach in industrial automation practice and its challenges

The meaning of circled labels refers to the issues as follows:

1. Import of runtime data into simulations.
2. Import of historical data into simulations.
3. Visualization of simulation data in standard HMIs and testing dispatchers' commands entered via HMI on simulation models.
4. Testing of SCADA advanced process control (APC) control actions on simulations.
5. Semi-automated design of simulation models.
6. Semi-automated configuration of OPC tags and other integration interfaces.

Analyzing these major challenges, our approach is based on two integration levels:

1) The technical integration level dealing with the technical transmission of data between particular tools.
2) The semantic integration level, which captures the meaning of the relevant knowledge, mapping common pieces of knowledge, and configuring the technical integration level. The paper addresses especially the semantic level of integration, whose research issues are summarized in the following overview.

The problems, which are investigated in this paper, were summarized into following research issues. Simulation models are hard to design and integrate manually. The goal of this research issue is to design an approach, which is based on explicitly represented engineering knowledge and to support the design of simulations semi-automatically. Since the runtime integration requires the proper structure of data in design time, the knowledge base must respect the integration request and involve it into the design time as well.

Current technical integration approaches require repeating manual work. This research issue is aimed at supporting runtime integration of industrial automation tools with a strong focus on simulations, which is flexible and easy to configure. Since the Simulation Framework is used as the technical integration level in the presented approach, one of the required output of the knowledge base tool is a generated set of configuration files for the Simulation Framework.

4 The Design of simulation based model

Although simulations are becoming widely used in industrial practice, their integration within the remainder of the automation system has not been satisfactorily solved. In addition, the design phase of simulations is usually done ad-hoc, which causes problems when either the real system is modified or some parts of the simulation should be reused. The goal of the presented work is to improve the design phase of simulation models in order to integrate them within the remainder of the automation system easily. We propose to capture the knowledge required for the simulation model design in the knowledge base.

4.1 SEMANTIC DESIGN OF SIMULATION MODELS

The typical engineering workflow [16] starts with the under-standing and summarization of a terminology. In other words, a domain specific language (DSL) is adopted for a particular class of plants, such as for hydraulic systems. DSL includes typical components and possible kinds of their connections, typical parameters and their types, relevant physical variables, definitions of relevant mathematical models, etc. Such pieces of information have to be structured and stored in a real plant ontology.

Plant device classes are represented as ontology classes, their parameters and relationships as ontology proper-ties. The creation of a general plant ontology structure enables to accelerate further project stages and it is done only once for every type of plant.

Having the ontology structure for the real plant, the description of the particular physical plant is created, i.e., there are created individuals of ontology classes. The individuals are labeled with unique names and relations to other individuals are entered. Parameters and their values can be entered, too. Based on the real plant ontology, simulation specialists are able to implement a simulation library, comprising generic simulation blocks. In a basic case, every plant ontology concept corresponds to a simulation block; in a more general case, a real plant element can be modeled by several blocks or vice versa. In the further subsection, this problem is discussed in more details.

The methodology for simulation model design and integration strongly depends on the available knowledge at the initial state of the automation project. In simulation practice, it is usual to utilize libraries containing generic blocks that are parameterized to approximate a specific device. In most cases, engineers are not able to create the library from scratch and afterwards assemble the simulation using these blocks. Usually, they must gain experiences on early versions of the simulations, perform a set of tests and comparisons of measured and simulated data and finally create or finalize the simulation library. Therefore, the first project in a specific area, such as a water distribution network simulation, requires iterative work on the simulation model and on the library. Other simulation projects of such networks reuse the library and even if some blocks must be added or modified, it is a simple task for an experienced simulation expert.

The simulation design phase is strongly influenced by the availability of the simulation library. If the simulation library exists, the semantic knowledge base tool is able to generate the simulation model semi-automatically. If the library is not available at the beginning, the knowledge base tool must be used in a similar way as an expert system, driving the user throughout a workflow collecting relevant pieces of knowledge required for the simulation model design.

In both cases, the general knowledge about the type of a real plant must be captured in the automation ontology. For example, the general knowledge about water distribution networks contains information that plants can have pumps, pipes, tanks, water wells, consumers or some disturbances; pumps have flow and pressure on their input and flow and pressure on their output, real parameters can be length, diameter and elevation, etc. The general knowledge is a kind of knowledge skeleton, which can be filled with real values when describing the specific plant. If needed, it can be extended with other parameters which are device-specific.

If the simulation library is available, the library blocks are annotated in the knowledge base, including their inputs, outputs and parameters. Consequently, the plant description must be formalized, including real plant topology and parameters. Technically, this step means populating the automation ontology with individuals denoting the real devices, their interconnections, tags, etc. Based on those pieces of knowledge, the semantic knowledge base tool is able to generate the simulation model structure automatically. Finally, the simulation expert is required to insert values of simulation block parameters, such as diameters or lengths of pipes.

In the case that the simulation library does not exist, the general knowledge is extended with a particular plant description in a similar way, but afterwards, the simulation model structure cannot be obtained automatically. The process of gathering the expert knowledge starts up with a specification which devices will be simulated. The step includes both grouping devices into complex subsystems (such as the whole plant can be decomposed to several simulation modules or a group of devices can be modeled as one block) and specification of block interfaces (such as the utilization of just one of the variables flow and pressure or both of them, an extension with further signals such as control values for devices). Afterwards, for each simulation block, there must be declared simulation parameters.

4.2 BLOCK-BASED DESIGN OF SIMULATION MODELS

System analysis and synthesis is typically based on the decomposition of problems into sub-problems. Engineering disciplines usually utilize components or blocks for the system design. For example, software engineers use methods as well as classes to encapsulate specific functionalities. In the simulation design area, the behavior of devices is often encapsulated by simulation blocks, which represent for example pumps, pipes, etc. Such blocks are comprised in a (universal) simulation library and the implementation of these blocks is generic in terms of their parameterization for the specific use.

This section addresses the research issue RI-2, i.e., the integration of whole simulations and other tools into the automation system environment. On the semantic level, the proposed solution is based on the representation of knowledge about local tag data models and mapping them to the global tag representation in the ontology. On the technical level, the approach is based on the enterprise service bus (ESB) approach. Although this paper is focused on the semantic level, the following paragraphs provide an overview of the technical layer, which is crucial for under-standing.

The proposed integration architecture, implemented by the Simulation Framework, is depicted in Figure 2. Tools are interconnected via connectors (having domain

specific and tool specific parts) to the ASB. The ASB transfers data according to the defined service bus workflow. This architecture poses a solution to the technical integration, nevertheless, it must be configured via a set of XML files to work properly. The preparation of such configuration files can be solved either manually, bringing short-comings in repeating error-prone manual work, or semi-automatically. We propose the semi-automated approach, which guarantees a consistent solution being flexible and capable to adopt modifications of real plant or software automation tools easily. The proposed solution is based on the structural tag description based on mappings between global tag representations and their local tool names.

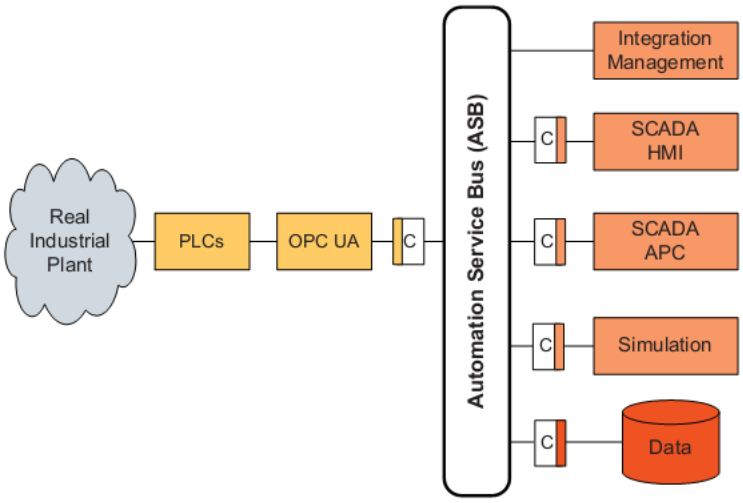


FIGURE 2 Industrial integration based on automation service

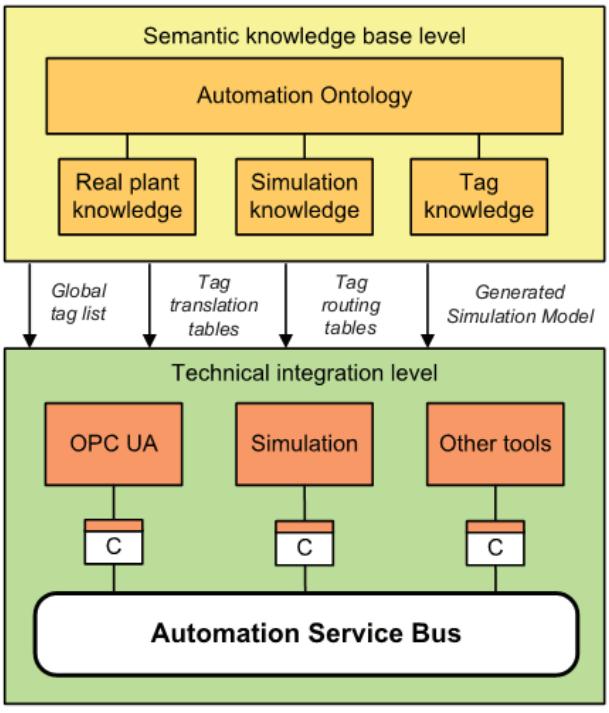


FIGURE 3 Technical layers including their interface

On the top of the Simulation Framework, we utilize the semantic knowledge base tool, which encapsulates the included ontologies, and provides interfaces to access it easily. The conceptual diagram of the proposed layered architecture is depicted in Figure 3. The knowledge base provides configuration XML files for the Simulation Framework, which are responsible for the definition of runtime behavior and furthermore, the knowledge base is able to support the design of particular tool models (such as simulation models in the current version).

4.3 PARAMETER MANAGEMENT OF THE SIMULATION SYSTEM

Since simulation parameters, i.e., parameters required for the parameterization of simulation blocks, can differ from parameters of real plant devices in both count and scale, the descriptive paradigm has to reflect this fact and manage them efficiently.

The presented approach distinguishes 3 classes of parameters. Real system parameters are parameters of devices included in the real system. Simulation model parameters are parameters of simulation modules. Simulation solver parameters are parameters of the simulation solver, such as simulation start time or stop time, required precision, etc.

In order to guarantee the consistency and avoiding errors while the real plant is changed, the presented approach involves the parameters in ontologies: the simulation ontology comprises definitions of required parameters to run the simulation for each block and plant ontology involves especially real plant elements description. A process specialist, for example, is not interested in the model parameters of a pump, which is modeled and simulated as the first-order system. Although he knows

almost everything about the real pump, he does neither know how simulation specialists would model pumps, nor the expected value, that can be calculated even by some algorithm. On the other hand, to reach the consistency in large-scale projects, changes should be automatically propagated between engineering domains and tools, hence these parameters are inter-linked. Mapping corresponding entities, such as real system parameters and simulation model parameters, which depend on them, is very easy to handle in ontologies that is one of the advantages of using them in this area.

The configuration of the Simulation Framework is based on the following XML files: Global tag list, tag translation tables, and tag routing tables. The global tag list file is aimed at setting all existing tags in the ASB environment and their properties (such as type, minimal and maximal values or others, which are useful for a control system). The tag translation tables are related to ASB connectors, which translate tag names between local names (occurring in the particular tool data model scope) and global tag representations, being uniquely available in the ASB. The tag routing tables define how tag values should be distributed between the tools (such as simulation results must be transmitted to HMI, simulation inputs must be entered with an operator-training data-set). Note that the global tag list and tag translation tables are project-specific, whereas routing tables are task-specific. Every project can have several tasks, such as simulation can be used as a soft-sensor to estimate unmeasured states, or as a test-bed to analyse control system behavior and train operators.

5 The real simulation cases

We selected a sample project to test the presented approach and to evaluate its efficiency. Since the complexity of typical industrial systems is very high and too complicated for a conference paper, we selected the simulation of the passive house. The goal is to describe a particular passive house in the unified ontology way and to generate its simulation model automatically. The simulation library with generic blocks is available. It is called Bldsimlib library [22], and it is implemented in MATLAB-Simulink. To create a simulation model of a particular house, the appropriate generic blocks from the library bldsimlib.mdl have to be inserted into a blank simulation model file simulation.mdl. Their names have to be set uniquely, and the blocks have to be interconnected according to the passive house floor plan. As well, the required parameters of the blocks have to be entered into the simulator workspace, in our case MATLAB Workspace.

For the first experiments, a simplified passive house was defined. It offers enough testing possibilities and the obtained results can be easily checked by people. To describe the testing house, a passive house ontology was designed. Its concepts define general elements of passive houses, such as rooms, walls, windows, etc. Afterwards, a description of the testing house was created by making ontology individuals representing the real passive house equipment.

The task for engineers is to design a simulation model for this plant and to integrate it within the remainder of the automation system. According to the workflow presented in Figure 2, the simulation model design requires general knowledge about hydraulic systems at the beginning. The second step is to populate the knowledge base (i.e. the automation ontology) with individuals representing devices in this particular industrial plant. In case of the universal simulation library being available, the simulation model structure can be generated automatically; otherwise, the simulation expert is driven through the process of gathering relevant simulation data. In this case, the simulation expert selects which devices create a simulation module and which blocks will be included in the modules. Since the system does not include repeating segments such as pump-stations, every device will be approximated by one simulation block. Interfaces of the blocks will be input and output flow and pressure; furthermore, tanks will have level sensor variables as outputs. Pumps and valves will have input control signals to distinguish their states. As tags are a subset of variables, there are expected tank level tags and pump flows to be considered as tags. The subsequent step in the methodology is to define simulation parameters. One possible solution is to accord to every device an elevation to the reference point and to define for every tank its height, diameter, and diameters of inlets and outlets. Pipes are expected to have resistances and pumps maximal volumetric flows. In a further step of the design methodology, the simulation library will be derived automatically. Finally, the simulation parameter values must be entered.

6 Conclusion

Industrial plants and processes are becoming very complex and sophisticated. Appropriate control algorithms, being necessary for their efficient operation, can be tested and fine-tuned on simulation models; for advanced process control the models are even necessary. Therefore, the fundamental precondition for control algorithm design is to implement and fine-tune a simulation model. Since plant description methods used nowadays are insufficient for describing modern flexible plants, this paper presents an ontology-based approach to the plant description. The proposed paradigm enables querying and inferring new pieces of knowledge. The solution supports the efficient integration of diverse engineering tools, such as simulators, SCADA systems, or even proprietary software used in particular projects. The proposed approach supports reasoning techniques and efficient utilization of domain specific languages for each engineer or expert. Ontologies comprise explicitly specified knowledge,

enable to interchange it between automation tools and can guarantee consistency of the whole integrated solution. The presented approach can be used for both continuous and discrete systems, especially the large-scale ones. A practical example, describing a passive house with ontology individuals and generating a structure of a simulation model automatically by the implemented tool, is involved in this paper.

References

[1] Moser T, Mordinyi R, Sunindyo W, Biffl S 2010 Semantic service matchmaking in the atm domain considering infrastructure capability constraints *Canadian Semantic Web* 133-57

[2] Gruber T 1993 A translation approach to portable ontology specifications *Knowledge Acquisition* **5**(2) 199-220

[3] Gómez-Pérez A, Fernandez-Lopez M, Corcho O 2004 Ontological Engineering with examples from the areas of Knowledge Management e-*Commerce and the Semantic Web second printing ed London* Springer-Verlag

[4] Obitko M 2007 Translations between ontologies in multi-agent systems *PhD dissertation Czech Technical University Faculty of Electrical Engineering* Prague

[5] Lacy L, Gerber W 2004 Potential modelling and simulation applications of the web ontology language – OWL *Proceedings of the 2004 Winter Simulation Conference* **1**

[6] Silver G, Hassan O H, Miller J 2007 From domain ontologies to modelling ontologies to executable simulation models *Proceedings of the 2007 Winter Simulation Conference* 1108–17

[7] Hu J, Zhang H 2009 Ontology based collaborative simulation frame-work using HLA and Web Services *World Congress on Computer Science and Information Engineering* **5** 702-6

[8] High Level Architecture (HLA) 2003 US Defence Modelling And Simulation Office

[9] Sarjoughian H, Huang D 2005 A multi-formalism modelling compos-ability framework: agent and discrete-event models *Proceedings of Ninth IEEE International Symposium on Distributed Simulation and Real-Time Applications* 249-56

[10] Benjamin P, Akella K Towards ontology-driven interoperability for simulation-based applications *Proceedings of the 2009 Winter Simulation Conference* 1375-86

[11] Durak U, Güler S, Guztüzün H, Ider S K 2007 An exercise in ontology driven trajectory simulation with MATLAB SIMULINK (R) *Proceedings of the 21th European Conference on Modelling and Simulation* Prague

[12] Barth M, Fay A 2013 Automated generation of simulation models for control code tests *Control Engineering Practice* **21**(2) 218-30

[13] Silver G, Bellipady K, Miller J, Kochut K, York W 2009 Supporting interoperability using the discrete-event modelling ontology (DeMO) *Proceedings of the 2009 Winter Simulation Conference* 1399-410

[14] Silver G, Lacy L, Miller J 2006 Ontology based representations of simulation models following the process interaction world view *Proceedings of the 2006 Winter Simulation Conference* 1168-76

[15] Vrba P, Tichý P, Mařík V, Hall K, Staron R, Maturana F, Kadera P 2011 Rockwell Automation's holonic and multiagent control systems compendium *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews* **41**(1) 14-30

[16] Novak P, Sindel R 2012 Semantic Design and Integration of Simulation Models in the Industrial Automation Area *2012 IEEE 17th Conference on Emerging Technologies & Factory Automation (ETFA)* 1-8

[17] Melik-Merkumians M, Zoitl A, Moser T 2010 Ontology-based fault diagnosis for industrial control applications *2010 IEEE 15th Conference on Emerging Technologies & Factory Automation (ETFA)* 1-4

[18] Novak P, Sindelar R2011 Applications of ontologies for assembling simulation models of industrial systems *On the Move to Meaningful Internet Systems: OTM 2011 Workshops* Hersonissos Springer Dordrecht 148-57.

[19] Gawthrop P, Bevan G 2007 Bond-graph modelling *IEEE Control Systems Magazine* **27**(2) 24-45

[20] Sindelar R, Novak P 2012 Simulation integration framework *The 10th IEEE International Conference on Industrial Informatics (INDIN 2012)* 80-5

Authors

Jianxun Wang, born in May 1972, Zhengzhou, Henan Province, China.

Current position, grades: associate professor at Zhengzhou Technical College, China.
University studies: Master's degree in Mechanical Design and Simulation.
Scientific interests: mechanical design and simulation.
Publications: 15 papers.

Qingyun Zhou, born in September 1972, Zhengzhou, Henan Province, China.

Current position: associate professor at Zhengzhou Technical College, China.
University studies: Master's degree in computer science and technology.
Scientific interests: computer software, system simulation and automatic control.
Publications: 12 papers.

Intelligent human resources systems in the information technology era

Jie Guo*

School of Philosophy and Public Administration, Henan University, Kaifeng, Henan, 475001, China

Received 1 March 2014, www.cmnt.lv

Abstract

Effective human resource management facilitates the success of an organization and the progress of a society. We describe an evolutionary computer model that simulates different modes of interaction between people and their environment. A two-level genotype-phenotype structure is used to represent the characteristics of an individual. The environment is modelled as a two-dimensional array of regions in which each region is characterized by a set of regional features and organizational culture (e.g., leadership strategies). Human resource decisions are subject to limitations, because they always depend on human knowledge, judgment and preference. Decision support applications can be used to provide fair and consistent decisions, besides to improve the effectiveness of decision making processes. This study consists of three parts; the first part is to understand the IDSS concepts, applications and related research in human resources decision making application known as HR DSS. The second part is to identify the potential intelligent techniques that can be used in HR DSS application, and the third part is to suggest the HR DSS framework that is related to human resource decisions. Finally, the paper proposed the HR DSS framework and the potential intelligent techniques that can be used to develop the IDSS application in any phases of decision making processes.

Keywords: human resource decision support, human resource management systems, human resource strategy

1 Introduction

At this point, information technology (IT) can play an essential role, providing the tools for the modeling of an integrated system. However, IT is not enough to convert information into knowledge. Appropriate qualitative and quantitative models must be added to analyze the information and produce outputs that support the decision procedure [1]. Operational research (OR) can provide these models. In the rapidly changing knowledge–based economy, organizations that manage knowledge effectively survive and develop through the years. Knowledge is acquired, retrieved and created by people [2]. Thus, high quality human resources (HR) is the determinant factor that can lead to a competitive and inimitable advantage. In that respect, both HR academics and HR professionals have turned their focus to the strategic role of human resource management (HRM) [3]. In order HRM to perform this role, individual HR practices should be engaged in a way to build a coherent HRM system that supports decision making at corporate level.

The vast majority of human beings are genetically distinctive from each other. Furthermore, as we grow physically and develop mentally, each of us becomes unique in the world. There is no single mechanism that can underlie all human behaviour [4]. In some cases, people want to conform, while in others they want to be different. The immediate decisions made by individuals are sometimes unpredictable. The diversity of human beings makes human resource management more of an art than a science. The common approach for studying human resource management problems has involved statistical measurements. Individual behavior has been totally neglected. However, under certain circumstances, a small number of people can cause a system to change drastically. That is, a system might respond unexpectedly when slight turbulence is introduced. In the real world, human resource management problems are extremely complex [5-8]. A fair measure of fitness for people is a very subjective and controversial notion. As we know, an individual might respond differently to the same things at different times or places, or with different people. Schelling wrote, "People are responding to an environment that consists of other people responding to their environment, which consists of people responding to an environment of yet other people's responses."

It is difficult to predict what behavior will emerge from such intensive human interactions. As a consequence, it is not feasible to conduct experiments on people in various environments and with different management strategies, and to study the effect of altering environments and management strategies on human satisfaction. It is becoming increasingly clear that the HR system is one important component that can help an organization become more

* ***Corresponding author's*** e-mail: guoguo@henu.edu.cn

effective and achieve a competitive advantage. However, a larger question remains unanswered: How does HRM contribute to firm performance? Disappointedly, this question has not been answered until now, thus, the relationship between HRM and organizational performance is called a 'black box' problem. Much of the debate over the relationships between HRM and organizational performance has been based on the distinction between two views typically referred to as 'best practice' and 'best fit' [9-11]. The best practice perspective identifies a set of HR practices that is associated with improved performance in all types of organization and, by implication, for all kinds of employees. The best fit approach argues that performance is maximized when the HR policies adopted are consistent with the business strategy. Both of these perspectives assume that the HR policies adopted will be implemented as intended and have the same effect on all employees who work for the organization. Various authors have questioned these assumptions because of the differences between intended HR policies and employees' experience, because complex organizations have different types of employees who may be managed successfully through diverse sets of HR policies.

This paper presents a new modeling platform for human resource management using computer simulation and evolutionary computational techniques. Our objective was to construct a computer model that served as a tool for investigating the phenomena that could occur in a human resource management system. In this model, individual behavior was taken into account. Through computer simulation, we were allowed to perform a variety of experiments in a feasible manner. The application of evolutionary computational techniques to solve real-world problems has received more attention in recent years. Our model, an artificial world simulation system, abstracts the above essential features of human resource management. Likewise, the goal of our model is not to draw every possible aspect of human resource management into the system and then try to optimize it. Instead, it should be pictured as a large relational model better thought of, like an evolutionary ecosystem, as an existential game. Self-organizing dynamics are the essence of this model. The criterion for the success of this model is reproduction of the essential features of human resource management. If we generate a new phenomenon that is unknown, then either something new has been discovered that could come into effect under appropriate circumstances, or there is some universal constraint that is operative that quenches this effect. However, we cannot describe all of the human resource management features in great detail. Simplifying assumptions and compromises are thus unavoidable.

2 Related work

Human resources refers to a production capacity in the human body, it is reflected in workers who, as measured by the quantity and quality of resources of laborers, it play a productive role in economic, the national income growth. It is the most active labor factor of production is the most active, create and accumulate material capital, development and utilization of natural resources, and promote the development of national economy, and promote the social transformation of infinite power [12]. Rural human resources refers to the total rural population has a range of physical and mental, it includes two aspects of quality and quantity, the total is the product of the two. The number of rural human resources, refers to the number of the population of the rural labor range; the quality of rural human resources, rural human resources refers to has the physique, intelligence, knowledge, skills, generally reflected in the rural labor population fitness level and culture level [13,14].

According to the resource – based view of the firm [15], the basis for competitive advantage lies primarily in the resources it possesses. Even though this point of view has received much criticism, it is widely accepted that the way an organization manages its own resources can contribute to a superior or inferior performance. Considering also the fact that HR are an organization's most valuable asset, HRM can indeed plays a significant role to the excellence of a firm. In that respect, the concept of strategic HRM (SHRM) has been developed. SHRM focuses on organizational performance rather than individual performance. It also emphasizes the role of HRM systems rather than individual HR practices in isolation as solutions to business problems [16]. Before proceeding with the concept of this study, it is important to remind the definition of a "system". It is a set of interacting or interdependent entities, each one being described by a number of characteristics, forming an integrated whole that is in constant communication with the external environment. One of the main features of a system is the relationships among the entities. That differentiates a system from a sum / group of entities. So, in particular, we cannot consider a HRM system only as a sum of unrelated HR practices. The interactions among practices lead to what Ichniowski et al. [17] declared: The basic assumption is that the effectiveness of any practice depends on the other practices in place. If all of the practices fit into a coherent system, the effect of that system on performance should be greater than the sum of the individual effects from each practice alone.

According to Tannenbaum [18], HRIS is a systematic procedure to acquire, store, manipulate, analyze, retrieve, and distribute information about an organization's HR. Usually, HRIS is restricted to a transaction processing or record–keeping role. Even though the above mentioned tasks are included in the responsibilities of a HRIS, the important strategic role is neglected. HRIS support an organization's operations in two ways, by facilitating the daily tasks of the personnel and by providing useful outcomes to decision makers.

3 The proposed artificial human resource system

With our understanding of the human resource structure, we built an artificial human resource system that uses an agent-based simulation method to model and simulate human resource dynamics in PM. Using the basic elements that comprise the human resource system, employees are modeled as the main agents. Each agent has its own properties, such as age, gender, working experience, grade, and so on. Accordingly, each agent has five types of behaviors: recruitment, retirement, grade promotion, position promotion, and leaving, which is shown in Figure 1.

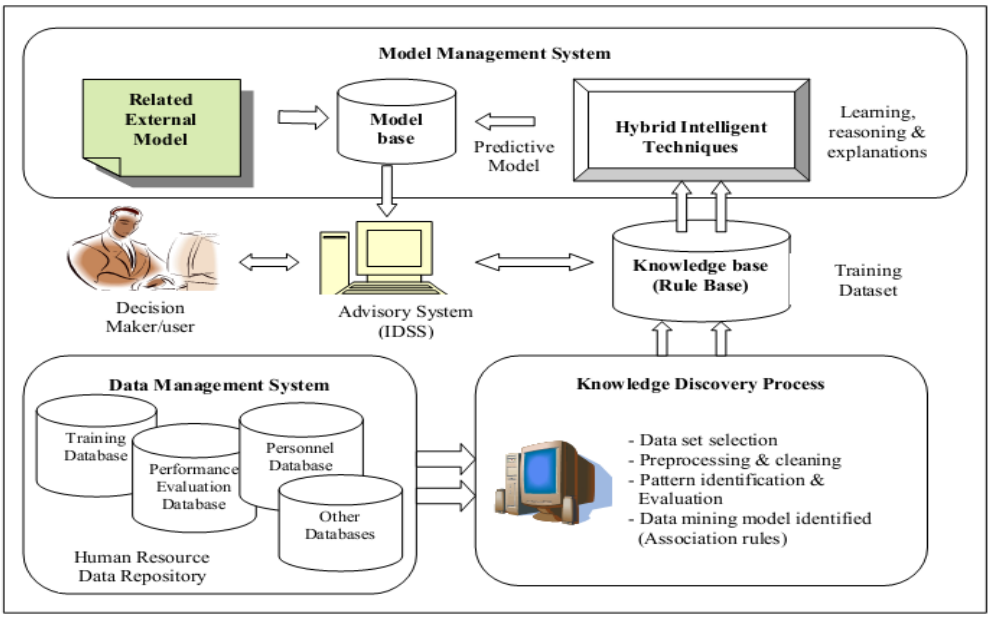


FIGURE 1 The proposed HR system

3.1 SYSTEM LEVEL

Personnel selection is one of the most critical functions of HRM. Apart from the straightforward fact that an organization should be staffed with the right people, a wrong hiring decision can lead to serious consequences. This is more obvious at high level managerial positions (e.g. director of department) where the decision may drive the future overall performance. Extensive literature can be found on selection methods, their validity and reliability [26]. Recently, endeavors from operations research and decision support fields strived to highlight the complexity of the selection function [8,9]. In general, the problem of hiring or promoting the right person should be considered in its multidimensional and incorporate all these parameters that influence the final decision.

An important factor that is linked to long term success and growth of an organization is its capacity to evaluate employees' performance and in addition the capacity to use the relevant information in order to improve employees' as well as organizational performance. Employee appraisal should not be conducted only against the outputs / results but also against behaviors and attitudes come into practice. Employee appraisal is the backbone of the concept called performance management. The latter links the main functions of HRM, aiming at improving the performance of a firm as a whole.

Compensation system has become more and more a really complicated issue for employers. It is not only to determine the base salary, considering the years of experience. Many other rewards do motivate and retain employees and might determine even the organizational performance. Compensation strategy can be used as a motivation and satisfaction incentive and as a recruitment and retain tool. In parallel, it is the mirror of the values and beliefs that form the organizational culture.

HR development (HRD) is a wider concept than the one of HR training. Training is mainly applied to improve the technical skills that cover short term needs of an organization. It is essential for workers in industry production business environments. Development includes training but is considered as a more strategic aspect. It refers basically to the key personnel of the organization. HRD facilitates the development of core capabilities that are critical in developing and maintaining sustained competitive advantage [14]. Core capabilities include positive behavior and ability to adapt to a rapidly changing environment and rational decision making, creativity development, motivation, human behavior understanding skills. The above mentioned are some examples of what is called management development [12], i.e., the development of specific capabilities of the future leaders of an organization.

3.2 APPLICATION LEVEL

Related enterprises and organizations should adjust their roles, to change service providers play by the traditional role of the manager's role, but also to change the location of mechanical, routine of their own, to play the initiative for the rural human resources to provide services. Management Department of rural human resource in its basic functions of the case, to make the management and practice to coordinate, and vigorously in practice, of course, said the enthusiasm to help the enterprises to mobilize human resources from the fundamental, to serve the development of enterprises.

Operation of the system, the rural human resources system maintenance work can take the relevant departments within the system of mutual cooperation; reduce the manpower, financial expenditure. Of course, with the overall quality of rural human resource enhancement, grasp the computer level, also can let them finish related work, so as to mobilize the enthusiasm of the migrant workers, the enterprise management personnel, also can liberate part of the labor force, to spend more time and energy into the development strategy of the enterprise. Migrant workers, give full play to the spirit of master, which is shown in Figure 2.

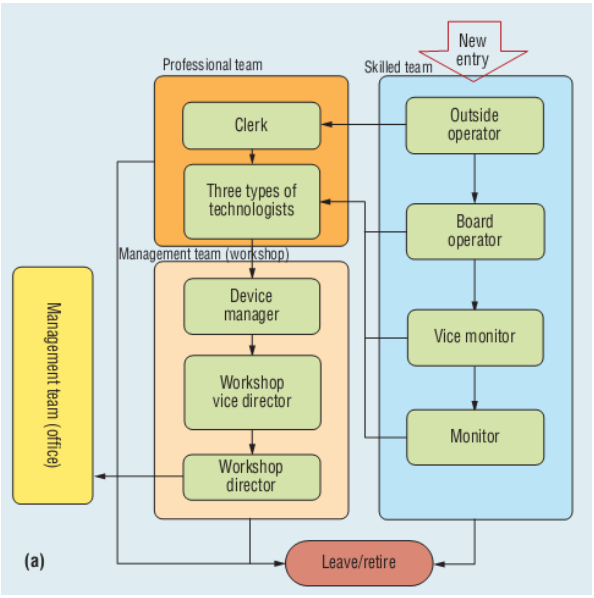


FIGURE 2 Potential employee career paths.

Enterprise organization should according to the characteristics of the rural human resources present and characteristic of the enterprise, make the rural human resources development management system continuous improvement system function, make the platform more open, more flexibility and innovation ability value, increasing rural human resource management information system and the outside resources embedded together, so as to enhance the efficiency, reduce the cost of.

Of course, with the maturity and application of the system of rural human resources, a bridge of communication with management system to complete between managers and workers, this time facing managers rely too much on the computer and the opportunity to meet the workers reduced risk, as a communication tool, we don't want is highly dependent on the human resource management system on how to improve the quality and reduce labor, enhance enterprise mission. At the same time, with the development of network technology, as well as to the rural human resources system to recognize and its operation process standardization, in particular with the advanced idea promotion, how to make the rural human resources information system more flexible, excellent price, more humane and the function is more perfect, for the design of the HRIS producers is grim the challenge, for enterprises using the HRIS system is a very real problem.

3.3 RELATION BETWEEN COMPONENTS

During personnel selection process, information on technical skills, abilities, professional experience and even personality is gathered through either self–report tests or interviews, methods that does not picture the future performance. For instance, successful management of a project team in a previous workplace does not ensure the success in managing a similar team in this organization. Selection process can only be a predictor of performance. Thus, the objective is to design as better a predictive system as possible. Employee appraisal is the phase that confirms the application of particular competences, based on performance standards and demands of the specific position. Employee appraisal can comprise a critical tool monitoring the relevance of the selection process. For example, following a specific methodology for the selection of an IT manager, the organization leads to the choice of a candidate with many years of experience in similar position, with strong academic background and with developed interpersonal, team management and creativity skills, among others. In case employee appraisal reveals behaviors and performance results that do not correspond to the initial evaluation (during selection process), there are several issues to be considered. Either the wrong person for the organization and the position was selected or external factors (e.g. the current situation of the market and the economy) affected the performance negatively or the standards and the criteria of the appraisal had not been set correctly.

The first one of the above mentioned potential reasons is closely related to the selection process. Wrong selection methods, wrong selection criteria or bad choice of criteria weights can result in wrong decisions. Moreover, inadequate evaluators/interviewers usually affect the whole process. Thus, a well-designed employee appraisal, can give feedback, through learning mechanisms, to the

selection process in order that the right people are hired, placed and promoted in the right positions.

There are two dominant compensation strategies. The first one uses academic titles and professional experience as the only decision criteria in order to set the level and type of salary. The other follows the concept of linking each job position with a respective wage. These two practices, bearing in mind that high level positions are occupied by experienced personnel, tend to become identical. As an employee acquires more experience, he occupies higher positions and receives higher wage. With these approaches, two employees, just having the same years of experience and the same studies, get equal pay. Similarly, two employees, just being at the same level of hierarchy, get equal pay. The fact that one may have achieved the targets set in comparison to the other is not taken into consideration in the compensation process. In addition, one may add value to the organization in financial (e.g. contribution to sales increase, cost reduction) or social terms (contribution to colleagues motivation, creativity behavior) and not be rewarded for this. The above mentioned situations can lead to injustice feelings and negative attitudes. For these reasons, compensation should be associated to a well-designed, objective employee appraisal process so that a good organizational environment to be secured. In parallel, compensation strategy can play a motivational role to performance improvement. That means a two way relationship between employee appraisal and compensation.

3.4 CONFIGURATION OF HRM SYSTEM

As mentioned in a previous section, HRD plays a dual, both tactical and strategic role, targeting on the one hand the coverage of specific short term technical needs of an organization (mainly through technical training) and on the other hand the management and development of potential future leaders and their careers. The decisions that should be made refer to when, on what and who to be trained. The answer can be supported by the employee appraisal process. Appraising employees, it is possible to sort them into categories, each one to be linked to particular training and development programs or even to create individual programs for each one of them. Giving an example, the appraisal of a programmer/developer can reveal weaknesses on specific features of performance (e.g. inefficiency on programming techniques). A personal training program can be designed using material relevant to these weaknesses. Similarly, a positive output of an appraisal of a programmer can be an indication for further development of the technical skills and in parallel a guide for future evolution into higher hierarchical levels (e.g. IT manager). Thus, an appropriate development program must be designed, with all necessary technical and managerial knowledge to be provided.

In general, there is a two way relationship between training/development and employee appraisal process. Theoretically, training output advances an improved future performance for the trainee. This indeed must be proved in practice. An effective training can be confirmed through the next employee appraisal. If this appraisal is positive, it could be at a certain degree due to the applied training program. If not, it is possible that the training practice should be redesigned. In other words, a learning mechanism is essential, which allows the respective HR practices to be adjusted dynamically.

Training can be closely linked to personnel selection practice. It is common, due to limited time or budget or because of the inadequacy of the recruitment phase to attract high quality candidates, the final selection decision not to be the ideal one. In that case, a gap analysis between the required and the available competences of the employee should be conducted so that the latter to be trained accordingly. Thus, training and development practice should receive feedback from selection process and provide the necessary tools to close a potential competence shortage.

4 Discussions and suggestions

Intelligent system and soft computing technologies are new technological platforms, whereby intelligent logic is now usually inherent in the processing of all decision support tools. HR DSS as a part of Intelligent System applications play the same roles to assist decision making process. In addition, applications and intelligent techniques of HR DSS need a lot of attention and efforts, from both academicians and practitioners. From this study, we can see the potential of HR DSS applications for future works. Firstly, there are many problem domains in HRM that can be explored by intelligent system researchers. In this case, the researchers should have the effort to identify problem domains where tools are needed to transform uncertain and incomplete data into useful knowledge. For that reason, we are trying to explore HR DSS applications for human resource decision. Secondly, researchers agree that hybrid intelligent techniques are the best approach to support decision making especially in reasoning and learning. We have embedded the HR DSS framework using hybrid techniques i.e., Knowledge-based system and Artificial Neural Network (ANN) approaches. Thirdly, the academicians and practitioners should continuously improve the core knowledge of effective HR DSS. This process can be enhanced by continuous development in web-enable tools, wireless protocol and group decision support system, which can expand the interactivities and perverseness decision support technologies. In our system development, we plan to use this technology to expand the capabilities of the application.

5 Conclusions

Effective human resource management facilitates the success of an organization and the progress of a society. We describe an evolutionary computer model that simu-

lates different modes of interaction between people and their environment. A two-level genotype-phenotype structure is used to represent the characteristics of an individual. The environment is modelled as a two-dimensional array of regions in which each region is characterized by a set of regional features and organizational culture (e.g., leadership strategies). Human resource decisions are subject to limitations, because they always depend on human knowledge, judgement and preference. Decision support applications can be used to provide fair and consistent decisions, besides to improve the effectiveness of decision making processes. This study consists of three parts; the first part is to understand the IDSS concepts, applications and related research in human resources decision making application known as HR DSS. The second part is to identify the potential intelligent techniques that can be used in HR DSS application, and the third part is to suggest the HR DSS framework that is related to human resource decisions. Finally, the paper proposed the HR DSS framework and the potential intelligent techniques that can be used to develop the IDSS application in any phases of decision making processes.

References

[1] Qian Z, Huang G H, Chan C W 2004 Development of an intelligent decision support system for air pollution control at coal-fired power plants *Expert System with Applications* **26**(3) 335-56

[2] Shim J P 2002 Past, present, and future of decision support technology *Decision Support System* **33**(2) 111-26

[3] Faye R M, et al.1998 An Intelligent Decision Support System for Irrigation System Management *in IEEE International Conference*

[4] DeNisi A S, Griffin R W 2005 Human Resource Management *New York Houghton Mifflin Company*

[5] DeCenZo D A, Robbins S P 2005 Fundamentals of Human Resource Management *8th Ed. ed 2005 New York John Wiley & Son Inc*

[6] Palma-dos-Reis A, Zahedi F M 1999 Designing personalized intelligent financial support systems *Decision Support System* **26**(1) 31-47

[7] Turban E, et al.2007 Decision Support and Business Intelligence Systems *Eighth ed 2007 New Jersey Pearson Education International*

[8] Quintero A, Konare D, Pierre S 2005 Prototyping an Intelligent Decision Support System for improving urban infrastructures management *European Journal of Operational Research* **162**(3) 654-72

[9] Malhotra P, et al. 2003 Brest Cancer Knowledge On-Line Portal: An Intelligent Decision Support System Perspective *In 14th Australasian Conference on Information System Perth*

[10] Viademonte, S, Burstein F 2006 From Knowledge Discovery to computational Intelligent A Framework for Intelligent Decision Support System *London Springer London*

[11] Baba N, Suto H 2000 Utilization of artificial neural networks and the TD-learning method for constructing intelligent decision support system *European Journal of Operational Research* **122**(2) 501-8

[12] Kuo R J, Chen C H, Hwang Y C 2001 An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network *Fuzzy Sets and Systems* **118**(2) 21-45

[13] Gorzalczany M B, Piasta Z 1999 Neuro-fuzzy approach versus rough-set inspired methodology for intelligent decision support *Information Sciences* **120**(1-4) 45-68

[14] Linger H, Burstein F 1998 Learning in Organisational memory Systems: An Intelligent Decision Support Perspective *In Proceedings of the Thirty-First Hawaii International Conference on System Sciences*

[15] Seder I, Weinkauf R, Neumann T 2000 Knowledge-based databases and intelligent decision support for environment management in urban systems *Computers, Environment and Urban Systems* **24**(3) 233-50

[16] Liqiang, G 2001 An intelligent decision support system for management of petroleum-contaminated sites *Expert System with Applications* **20**(3) 251-60

[17] Sajjad A, Slobodan P S 2006 An Intelligent Decision Support System for Management of Floods *Water Resources Management* 20 391-410

[18] Adla A, Zarate P 2006 A Cooperative Intelligent Decision Support System *in International Conference on Service Systems and Service Management Troyes*

Author

Jie Guo, born in March 1977, Kaifeng, Henan Province, China.

Current position, grades: Master, an associate professor in School of Philosophy and Public Administration, Henan University, China.
University studies: Public Management.
Scientific interest: public sector human resource management.
Publications: 20 papers.

PCA-based analysis on factors of English translation ability

Ting Hong*

Zijin College, Nanjing University of Science and Technology, Nanjing, 210000, China

Received 1 March 2014, www.cmnt.lv

Abstract

This work studied the relationship between English proficiency and translation ability as well as that between English and translation teaching among English majors. PCA is utilized to quantify and analyze the relationship between translation ability and English skills such as listening, reading, error correction and writing. Then we obtain the quantitative relation between translation ability and its factors, providing decision basis for the improvement of English translation skills and teaching.

Keywords: PCA analysis, translation ability, English teaching, quantitative analysis

1 Introduction

With the deepening of economic globalization, there is a growing demand for talents of foreign language translation. Therefore, the improvement of students' English translation ability has become an important task for college English major and common English teaching.

For a long time, translation ability has been regarded as a language skill due to unclear concepts about translation and English teaching. Currently, the mainstream of translation teaching still focuses on language skills. While strengthening the teaching of language knowledge, it neglects the variability of texts for translation, differences in translators' cognition and some non-linguistic factors (social, historical and cultural). However, translation behavior is a dynamically-generated process of translation and also a reflection of translators' comprehensive abilities. As many scholars interpreting the nature of translation capabilities, they made it clear that comprehensive ability is not equal to language application ability.

2 Principal component analysis (PCA)

2.1 BASIC IDEAS OF PCA

Principal component analysis utilizes the method of mathematical dimension reduction to substitute comprehensive variables for original ones. These unrelated, comprehensive variables contain the information of the original ones as much as possible. The statistical analysis method converting several variables into few unrelated, comprehensive ones is called as principal component analysis (PCA).

2.2 MATHEMATICAL MODELS OF PCA ANALYSIS

For a data sample, p variables $(x_1, x_2, \cdots, x_p)$ are observed to form the data matrix of n samples as follows.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \left(x_1, x_2, \cdots, x_p\right),$$

where

$$x_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}, \qquad j = 1, 2, \cdots, p.$$

PCA can integrate p observed variables into p new comprehensive ones.

$$\begin{cases} F_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ F_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \cdots \\ F_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p \end{cases}.$$

Namely

$$F_j = \alpha_{j1}x_1 + \alpha_{j2}x_2 + \cdots + \alpha_{jp}x_p,$$
$$j = 1, 2, \cdots, p$$

The model should meet the following criteria.

1) F_i, F_j $(i \neq j,\ i, j = 1, 2, \cdots, p)$ are unrelated to each other.

2) F_1 has a larger variance than F_2, and F_2 has a larger variance than F_3, etc.

3) $a_{k1}{}^2 + a_{k2}{}^2 + \cdots + a_{kp}{}^2 = 1, \qquad k = 1, 2, \cdots, p.$

There are p principal components. F_1 is called as the first principal component; F_2 as the second principal

* ***Corresponding author's*** e-mail: zj_ht123@163.com

component, and so on; a_{ij} as the principal component factor.

The model can be represented by a matrix. A is the principal component coefficient matrix.

$$F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{pmatrix}, \qquad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}.$$

$F = AX$, where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}.$$

3 Data sources

As psychological characteristics, English translation and language skills cannot be measured only by one's external manifestations or characteristics.

As the study objective of this work, TEM-8 score of a college includes a total score and individual score. TEM-8, with high reliability and validity, is the most authoritative, standardized test for measuring students' English proficiency.

For data processing and analysis, listening and dictation are merged into listening comprehension, measuring students' listening ability; Reading One and Two are merged into reading comprehension to measure English reading ability; translation from English-Chinese and Chinese-English translation are also combined together.

In terms of differences in English translation ability of candidates, translation scores are divided into two groups – one with scores not less than 48 points (qualifying criterion) and the other less than 48 points.

4 Results analysis

4.1 MEAN CONTRAST OF CHINESE-ENGLISH TRANSLATION SCORES

Mean data contrast of the two groups and T-test results of independent samples are shown in Table 1, Table 2, respectively.

According to Table 1 and Table 2, the results of Levene variance homogeneity test are F = 0.772 and P = 0.381> 0.05, so the two groups of students have equal variance in Chinese-English translation score; t = 3.303, and P = 0.001 <0.05. Thus, the difference between two groups in Chinese-English translation has statistical significance.

TABLE 1 Mean contrast for two groups of Chinese-English translation scores

Translation Score	N	Mean	Std.Deviation	Std.Error Mean
>=48.00	40	6.35	.796	.140
<48.00	94	5.71	1.031	.112

TABLE 2 Independent samples T-test for two groups of Chinese-English translation scores

Levene's Test for Equality of Variances		T-test for Equality of Means							
	F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variance assumed Equal variance not assumed	.772	.381	3.303	122	.001	.64	.194	.257	1.024
			3.693	80.182	.000	.64	.173	.295	.985

4.2 MEAN CONTRAST OF CHINESE-ENGLISH TRANSLATION

Table 3 and Table 4 show the mean contrast, independent sample T-test of two groups of data, respectively.

According to Table 3 and Table 4, the results of Levene variance homogeneity test are F = 2.534 and P = 0.115> 0.05, so variances of two groups of Chinese-English translation scores are equal: t = 3.158, and P = 0.002 <0.05. Thus, the difference between the two groups in Chinese-English translation has statistical significance.

TABLE 3 Mean contrast of two groups of Chinese-English translation scores

Translation Score	N	Mean	Std.Deviation	Std.Error Mean
>=48.00	40	6.74	1.067	.183
<48.00	94	5.89	1.457	.161

TABLE 4 Independent samples T-test of two groups of Chinese-English translation scores

Levene's Test for Equality of Variances		T-test for Equality of Means							
	F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variance assumed	2.534	0.115	3.158	122	.002	.86	.275	.320	1.391
Equal variance not assumed			3.613	85.121	.002	.86	.237	.385	1.330

The statistical contrast in Chinese-English and English-Chinese translation shows that students with stronger English ability are doing better in English translation.

4.3 PEARSON CORRELATION ANALYSIS

For data of all the samples, Table 5 shows pearson correlation coefficients among Chinese-English and English-Chinese translation, listening comprehension, reading comprehension, error correction, writing and translation score.

TABLE 5 Pearson correlation analysis results

		Listening Comprehension	Reading Comprehension	Error Correction	Writing	Translation Score	Chinese to English	English to Chinese
Chinese to English	Pearson Correlation	.273(**)	.214(*)	.332(**)	.124	.333(**)	1	.454(**)
	Sig.(2-tailed)	.002	.017	.000	.171	.000	.000	.000
English to Chinese	Pearson Correlation	.403(**)	.348(**)	.271(**)	.041	.431(**)	.454(**)	1
	Sig.(2-tailed)	.000	.000	.002	.650	.000	.000	.000

** Correlation is significant at level 0.01 (2-tailed)
* Correlation is significant at level 0.05 (2-tailed)

According to Table 5, Chinese-English translation has the biggest correlation coefficient (0.454) with English-Chinese translation. It indicates that Chinese-English and English-Chinese ability are correlative, so teaching for the two kinds of translations should be complementary; Chinese-English translation has strong correlation (0.333) with translation score, so it is closely related to listening, writing and reading; error correction and Chinese-English score have the strongest correlation (0.332) among all individual scores. With language knowledge including grammar, rhetoric and language structure, error correction can greatly influence translation ability by identifying faulty wording or formulation and proposing corrective approach.

4.4 PCA ANALYSIS

4.4.1 Statistical test

After setting the significance level as 0.05, KMO and Bartlett's Test of Sphericity are conducted on the selected data. Results are shown in Figure 1.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.604
Bartlett's Test of Sphericity	Approx. Chi-Square	56.843
	df	28
	Sig.	.001

FIGURE 1 KMO and Bartlett's Test of Sphericity

In Figure 1, KMO statistic is 0.604 > 0.5; Bartlett spherical statistic 56.843; significant probability 0.001 <0.05. Then the null hypothesis in Bartlett Test of Sphericity is rejected, and PCA analysis is more suitable for the index data of translation ability.

4.4.2 Determination of common factors

In this work, eigenvalues is greater than one, and the cumulative contribution rate greater than 80%. Figure 2 shows the PCA extraction results.

In Figure 2, the characteristic curve shows a turning point at factor 5, so common factors can be extracted from the first five factors. In Table 6, among the five factors with eigenvalue greater than one, the first four factors contribute to 80.571% (> 80%) of cumulative rate. Therefore, the four factors can reflect the influencing factors of English translation ability.

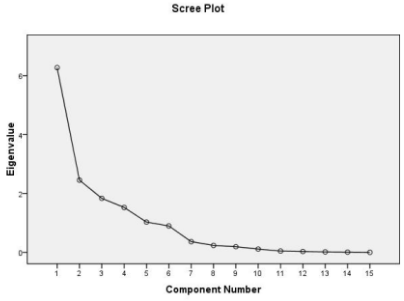


FIGURE 2 Scree plot

TABLE 6 Eigenvalues of correlation matrix R

	Initial Eigenvalue			Eigenvalue after extracting four factors and orthogonal rotation		
	Eigenvalue	**Proportion**	**Cumulation**	**Eigenvalue**	**Proportion**	**Cumulation**
1	6.274	41.829	41.829	5.278	35.189	35.189
2	2.453	16.352	58.181	2.477	16.511	51.700
3	1.834	12.227	70.408	2.462	16.413	68.112
4	1.525	10.164	80.571	1.869	12.459	80.571
5	1.028	6.853	87.424			
6	0.897	5.977	93.402			
7	0.362	4.415	97.816			
8	0.235	2.184	100			

4.4.3 Establishment of factor loading matrix

Varimax rotation method is used to calculate factor loading matrix. Results are shown in Table 7.

TABLE 7 Factor loading matrix

	Component			
	1	2	3	4
Chinese to English x1	.926	-.084	.058	-.183
English to Chinese x2	.914	-.197	.187	.204
Correction x3	.906	.142	.227	.059
Writing x4	.901	-.129	.220	.270
Reading 1x5	.881	.359	.171	.002
Reading 2x6	.215	.888	.184	.060
Listening x7	-.437	.758	-.060	.375
Dictation x8	.358	.562	-.277	-.210

4.4.4 Scores of calculated factors

Score linear calculation model can be constructed for each factor based on the factor score matrix in Table 8, and the specific steps are as follows.

$$F_1 = 0.170x_1 + 0.047x_2 + ... - 0.101x_8,$$

$$F_2 = -0.039x_1 - 0.052x_2 + ... + 0.298x_8,$$

$$F_3 = -0.018x_1 - 0.470x_2 + ... + 0.063x_8,$$

$$F_4 = 0.120x_1 + 0.087x_2 + ... + 0.216x_8,$$

$$F = 0.35189F_1 + 0.16511F_2 + 0.16413F_3 + 0.12459F_4.$$

Then mathematical evaluation model of English translation ability can be concluded based on variance contribution rate of common factors in Table 6.

TABLE 8 Factor score matrix

	Component			
	1	2	3	4
Chinese to English x1	.170	-.039	-.018	.120
English to Chinese x2	.047	-.052	-.470	.087
Correction x3	.180	.150	-.021	-.019
Writing x4	.212	-.031	-.114	-.121
Reading 1x5	.090	-.031	.094	-.327
Reading 2x6	-.017	-.023	.020	.252
Listening x7	.145	-.118	-.295	.356
Dictation x8	-.101	.298	.063	.216

5 Conclusions

In this work PCA analysis reveals the relationship between English proficiency and translation ability, as well as the quantitative relationship between English and translation teaching among English major students. Then the quantitative relationship between translation ability and English abilities – listening and reading comprehension, Chinese-English and English-Chinese translation – is obtained through extracting factors and calculating factor scores. In the end, quantitative calculation and evaluation for English translation are achieved, thus providing decision basis for teaching English translation.

References

[1] Lao, She. Camel Xiangzi. trans. Shi Xiaoqing. Beijing: Foreign Languages Press. 1981 Leech G N Style in Fiction *New York Longman* 2010 11-16

[2] Li C N, Thompson S A 2011 Subject and topic: a new typology of language *In Charles N Li (ed) Subject and Topic London/New York Academic Press* 457-90

[3] Newmark P 1988 A Textbook of Translation Hertfordshire: Prentice-Hall Newmark Peter Approaches to Translation *Oxford Pergamon* 200 95-101

[4] Nida E 2011 A Language, Culture and Translating *Shanghai: Shanghai Foreign Language Education Press* 312-5

[5] Nord C 2007 Translating as a Purposeful Activity: Functionalist Approaches Explained *Manchester St Jerome* 67-74

[6] Quirk R 2011 A Grammar of Contemporary English *Longman* 34-8

[7] Robinson D 2002 Western Translation Theory: from Herodotus to Nietzsche *Cornwall: St. Jerome Publishing* 112-20

[8] Snell-Hornby M 2011 Translation Studies An Integrated Approach *Revised ed Amsterdam John Benjamins* 321-30

Author

Ting Hong, born in Oktober 1980, Nanjing, Jiangsu Province, China.

Current position, grades: Master, a lecturer of Zijin College, Nanjing University of Science and Technology, China.
University studies: American literature, English teaching and Research.
Scientific interest: English teaching and research, computer aided Instruction.
Publications: 15 papers.

Collaborative mechanism of project management based on complex system theory

Jiale Tian*

School of Civil Engineering and Architecture, Wuhan University of Technology, Hubei, 430070, China

Received 1 March 2014, www.cmnt.lv

Abstract

Although it is arguable that humans have been studying complex systems for thousands of years, the modern scientific study of complex systems is relatively young in comparison to conventional fields of science with simple system assumptions, such as physics and chemistry. The history of the scientific study of these systems follows several different research trends. The project management community is actively demonstrating substantial interest in the development of viable methods to assess and improve project management maturity. There is little empirical evidence on the benefits of deploying a project management office (PMO) and/or conducting project reviews. The increasing complexity of exploratory activities in pharmaceutical innovation makes less likely that a project can stand alone. Project managers not only resort to in-house innovation but also external sources to propel a central project. This paper introduces the notion of a quality function for individual tasks and uses the functional form of the bivariate normal, to model quality at the task level. Using real data from two case studies, a translation agency and a software development company, the quality function is specified and incorporated into a mathematical programming model that allows quality to be explicitly considered in project planning and scheduling. An alternative model formulation leads to the creation of quality level curves that enable managers to evaluate the nonlinear tradeoffs between quality, time, and cost for each of the example projects. The results of these analyses lead to specific decisions about the planned values for these three fundamental dimensions at the task level and provide insights for project planning and scheduling that can be gained through improved understanding of the choices and tradeoffs.

Keywords: project management, project planning, project scheduling, quality management

1 Introduction

A complex system is a damped, driven system (for example, a harmonic oscillator) whose total energy exceeds the threshold for it to perform according to classical mechanics but does not reach the threshold for the system to exhibit properties according to chaos theory. We have witnessed a dramatic and steady increase in the extent to which the modern enterprise adopts and relies upon project management to secure a competitive advantage. As project management becomes the dominant way that work is accomplished, organizations strive to become good at delivering projects successfully. The predictable consequence is widespread commitment to improvement initiatives that may include the establishment of an enterprise project management process, the development of a career path for project managers, the implementation of project management education and training programs, and investment in project management tools and information systems. But the modern enterprise cannot afford to improve recklessly or randomly. The modern enterprise must approach improvement purposefully. Committing an organization to a significant improvement effort requires a thorough understanding of where the organization is and, perhaps more importantly, where the organization needs to grow. This is the need that is addressed by the recent interest and attention devoted to the development of project management maturity models.

It is widely agreed that the choice of management structures used to implement innovative, temporary, cross-functional and complex project endeavours has important implications for project success. The discussion of alternative project management structures dates back to Galbraith's conceptual introduction of the matrix organization and its differentiation from functional and product organizations. He systematically compared the advantages and dis-advantages of alternative matrix organization structures. Based on Galbraith's typology some authors favoured matrix project organization structures for their flexibility, their economical use of resources, and the clear differentiation between project authority and functional authority. Others criticized matrix project organization structures due to their complexity and lack of unity of command. On balance, these conceptual discussions lack agreement, thus providing little conclusive theoretical direction concerning the relationship of specific project structures to project success. These conceptual disagreements are also reflected in the empirical research. With the exception of and empirical studies have not generally revealed significant associations between project organization types and project success. However, an alternative stream of empirical research suggests the importance of project management structures for project success. Several studies concordantly identified significant and strong associations between project

* ***Corresponding author's*** e-mail: 714823@qq.com

managers' (PM's) decision authority and project success. The differences between these two empirical research streams accentuate the inconclusive discussion of project management structures and their association with project success.

Project management addresses cost, schedule, and performance targets while providing an outcome that satisfies the client. A measure of the value of the project to the client is the level of quality associated with the completed project. It follows then that important managerial decisions relate to the level of quality achieved for each of the project's tasks, since the quality of the tasks defines the quality of the project. The emphasis in project planning and scheduling has been on managing the relationship between time and cost, with an implicit assumption of a fixed level of quality that is seldom explicitly examined. However, in many situations there are alternative approaches for completing each task, each having its own time, cost, and quality considerations. Differences in quality can arise due to bids offered by competing subcontractors to complete specific tasks. Even different bids by the same subcontractor could imply different quality levels. For example, subcontractors might have some flexibility with time and cost that would result in different quality levels for the same task. This can also be true for alternative work plans offered in-house. For example, in completing a foundation for a building there are choices related to the depth of the excavation and the compressive strength of the concrete used. Each of the possible alternatives will achieve different levels of time, cost, and quality associated with this task.

A complex system is a damped, driven system (for example, a harmonic oscillator) whose total energy exceeds the threshold for it to perform according to classical mechanics but does not reach the threshold for the system to exhibit properties according to chaos theory. Although it is arguable that humans have been studying complex systems for thousands of years, the modern scientific study of complex systems is relatively young in comparison to conventional fields of science with simple system assumptions, such as physics and chemistry. The history of the scientific study of these systems follows several different research trends.

This paper introduces the notion of a quality function for individual tasks and uses the functional form of the bivariate normal, to model quality at the task level. Using real data from two case studies, a translation agency and a software development company, the quality function is specified and incorporated into a mathematical programming model that allows quality to be explicitly considered in project planning and scheduling. An alternative model formulation leads to the creation of quality level curves that enable managers to evaluate the nonlinear tradeoffs between quality, time, and cost for each of the example projects. The results of these analyses lead to specific decisions about the planned values for these three fundamental dimensions at the task level and provide insights for project planning and scheduling that can be gained through improved understanding of the choices and tradeoffs.

2 Related works

In the field of project management, "Quality management has equal priority with cost and schedule management" [1]. This statement makes inherent sense, since project management is concerned with not only managing cost and schedule, but also the actual work completed in order to achieve the project goal. The quality of the work completed then is an important project outcome, since it directly relates to the value of the project deliverables. The Project Management Body of Knowledge (PMBOK) [2] has adopted the ISO 9000, clause 3.11 definition of quality as "the degree to which a set of inherent characteristics fulfils requirements" [3]. Quality issues must be addressed in both the management of the project and the product of the project [2]. Specifically, the PMBOK suggests that quality must be addressed throughout the project life cycle, beginning in the project planning phase and continuing through quality assurance and quality control [2]. Unfortunately, no guidance is provided in terms of how quality can be measured in a project context. The construction industry has been concerned about quality for a long period of time and has conducted research to address this issue. A study by the Quality Performance Measurements Task Force of the Construction Industry Institute resulted in the quality measurement matrix [4] and led to the development of an approach to measure quality performance of engineer-procure-construct (EPC) projects [5]. The measures were tied to four total quality management components: customer focus, leadership, delivery, and employee empowerment. Under delivery, for example, the subcategories include cost, time, safety/health/environment, and product deliverables. The focus of this approach is on identifying and then tracking critical quality measures for each project phase. A project quality performance model based on empirical study of project control variables was developed for Hong Kong construction projects [6]. These variables are grouped under the headings of client, project, project environment, project team leader, project management action, and project procedure. The literature on quality issues and problems in the construction industry is summarized in [7].

Several authors have attempted to develop methods to measure project quality. Paquin et al.[8] assess project quality by decomposing client satisfaction into a hierarchical structure of quality dimensions that are measured and aggregated using a multi-criterion approach such as the analytical hierarchy process (AHP) [9] or multi-attribute utility theory [10]. A limitation of the multi-criterion approach for quality measurement is the necessity of identifying and evaluating a possibly unique set of quality dimensions for each task. Pollack-Johnson and Liberatore [11] illustrate how the quality of a task option can be determined using the AHP. They extend the discrete time–cost problem by developing a mathematical programming model that determines optimal discrete options defined in terms of time, cost, and quality combinations for specific tasks to maximize overall project quality, whether determined by AHP or by some other method, subject to time, and cost constraints. Icmeli-Tukel and

Rom [12] present two models that integrate quality considerations into resource constrained project scheduling. In their study, quality is measured by the amount of time and money spent on reworking activities that do not satisfy specifications. Several methods of integrating rework times and costs into the models are proposed. An alternative approach for measuring quality at the task level is to make a direct subjective assessment of quality [13,14].

3 Method

3.1 THE MODEL

This research was conducted using the PM Solutions Project Management Maturity Model. The PM Solutions model was developed by a team of highly experienced project managers and has been used successfully to conduct maturity assessments in several organizations. This model adopts a two-dimensional (2-D) framework. Both of the dimensions are based on accepted industry standards. The first dimension reflects the level of maturity. It is based on the structure of the SEI Capability Maturity Model. The SEI model has received widespread acceptance as a standard for process modeling and assessment of organizational maturity in several process areas. Additionally, research evidence from several studies including case studies and survey research have demonstrated important organizational performance improvements are associated with improved process maturity. The second dimension depicts the key areas of project management addressed. This dimension adopts the structure of the PMI's nine knowledge areas. Each of the nine knowledge areas were further decomposed into key components that pro-vide for a more rigorous and specific determination of project management maturity based on specific capabilities. There were a total of 42 components included in this study, which is shown in Figures 1 and 2.

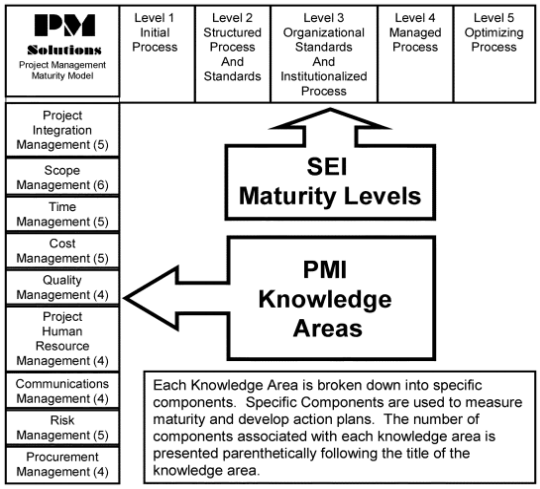


FIGURE 1 PM solutions project management maturity model

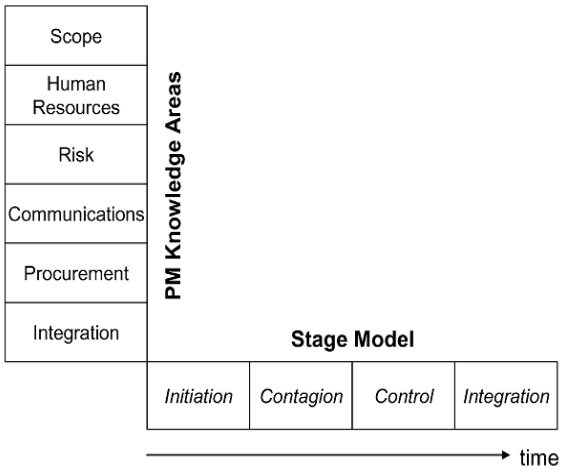


FIGURE 2 Project management knowledge areas and stage model

3.2 THE FORMULATION

In this study, we establish a triple-stage stochastic decision process, which illustrates a typical pharmaceutical R&D project in which the committee periodically reviews a project and decides whether to continue with in-house research or buy existing technology from outside.

a) Once the R&D project is initiated, it has two process lines: the monitoring process and the stochastic innovation process. The monitoring process is periodically scheduled (e.g., twice a year), whereas the innovation process is a random time duration.
b) The focal R&D project has a fixed total time span stipulated by the monitoring process, meaning, practically, that the project is not allowed to last forever.
c) The R&D project's innovation process can be divided into three phases as mentioned previously: the toxicity phase; the efficacy phase; and the clinical test phase. The duration of each phase is random.
d) Scientists develop a lead compound by exploring molecules (e.g., via HTS) and optimize the compound by adding more molecules to its basic structure or replacing old molecules with new ones. Scientists face uncertainties in the sense that their in-house developed compound may fail or their in-house research lasts too long and thus they run out of time without having a positive result.
e) The R&D project can buy existing molecular entities or novel entry technology developed by external sources as substitutes and combines them with its own in-house compound.
f) The committee decides at scheduled monitoring times whether to buy substitute molecular entities or continue a drug's in-house development. However, by using externally

procured molecules, we assume that the probability of producing a successful lead compound ultimately decreases. This condition reflects Pisano's argument that bought-in molecular entities may increase development uncertainty and thus reduce the probability of success. It also reflects the make-or-buy dilemma the committee has to face in science-based pharmaceutical R&D.

g) By "project success," we mean that the compound turns out to have positive effect within the project's allotted total time. By "failure," we mean that the compound either turns out to have negative effect or the research runs out of time. The following details our model's formulation.

3.3 THE DEVELOPMENT

Our overall goal is to find a rich method for modeling the relationship between quality, time, and cost at the individual task level within a project, as well as at the overall project level. Our approach begins by formulating a model of the quality of each task as a function of the time and cost allocated to it. We assume that there is an entity (an individual or group within an organization, a subcontractor, etc.) that could complete a given task with different allocations of time and cost. We further assume that there is a quality function that assigns to each combination of time and cost a corresponding quality value. We limit time and cost c to reasonable values for the task at hand and assume that, within that domain, the quality function for a task has two basic properties.

1) Holding time constant, quality q is an increasing function of cost. Thus, if time is fixed, we assume that allocating more money to the task will increase quality.
2) Holding cost constant, quality is an increasing function of time. Thus, if cost is fixed, we assume that allocating more time to the task will increase quality.

If we normalize quality to be on a 0–100 scale, based on the two non-decreasing assumptions earlier, we would expect the graph to show quality being lowest at the corner of the domain with the smallest values of time and cost and highest in the opposite corner (the highest values of time and cost). For a fixed quality, we would expect a standard time/cost trade off curve that is decreasing and convex. Thus, to maintain the same level of quality, to reduce the time, one has to pay increasingly more money per unit, such as in standard project activity crashing. This suggests a basic hill shape rising out of a plain, although we would only be interested in a one-quarter wedge of the hill.

3.4 THE MEASUREMENT

The questionnaires used in this study include 199 single items and some quantitative measures of project-specific characteristics. Out of these, 67 items were directly taken from questionnaire, with permission of the author. The remaining items were developed for the purposes of this study. Each item was assessed on seven-point rating scales with a range from strongly agree" to "strongly disagree." The original questionnaire was developed in the German language and the adopted Pinto's items were translated from German to English. Inorder to collect data on the U.S. projects the German questionnaire was translated into English. To demonstrate consistency and accuracy in translations the documents were back-translated using two experts who discussed the translations and together corrected any inconsistencies. Prior to final data collection in the U.S. the instrument was tested in a pilot of 20 projects.

Following our conceptual discussion we use a set of five variables measuring project management structure attributes and four variables measuring project success. All constructs consisting of multiple items were tested for composite validity using Cronbach's alpha and factor analysis. Some of the initial scales had to be modified to achieve satisfactory composite validity.

1) Project Management Structure Variables: The scales used to measure the project management structure attributes could be portrayed as follows. The PM project authority scale describes the level of authority delegated to the PM and covers different important decisions typically made or influenced by PMs. The PM functional responsibilities scale measures the position of the PM within the functional organization. The PM personnel authority scale measures whether the PM had the authority to reward project team members. The PM project responsibility scale is used to differentiate the project responsibility of the PM from the project responsibility of other constituents within the organization. The project influence of steering committees was measured by the level of top management involvement in project decisions.
2) Unidimensional Project Management Structure Scale: We used the scale developed by Larson and Gobeli. This scale is a blend of the scales suggested by Might and Fischer and Murphyet al. It differentiates between five project management structure types and describes the spectrum from a functional organization to a pure project organization.
3) Project Success Variables: Pinto and Mantel identified three distinct aspects of project performance: 1) the implementation process; 2) the perceived value of the project; and client satisfaction with the delivered project outcome. Shenhar et al. have suggested four different criteria to assess project success: 1) meeting design goals; 2) benefits to customers; 3) commercial success; and 4) future potential. In this study, we use four different project success measures: efficiency, effectiveness, customer satisfaction, and business results.

4 Analysis

The results of this study complement previously published results concerning overall project management maturity in which a clear majority of respondents indi-

cated that their organizations are relatively immature in terms of the project management maturity model. In the referenced study, nearly 67% of respondents indicated their organizations were operating at level 1 – initial processes (13.7%) or at level 2 – structured process and standards (53.2%). While a notable portion of respondents rated their organizations had reached level 3 – organizational standards and institutionalized process (19.4%), a mere 7.3% indicated their organizations were operating at level 4 – managed process and only 6.5% assessed their organizations to have achieved level 5 – optimizing process.

This instant research reveals very comparable results at the more detailed component level. Reviewing the results across all component areas, the median project management maturity level was level 2 out of 5 in 36 of the 42 areas. In fact, there was only one component area in which the median level of project management maturity was above level 2. In the change control component of the project integration management knowledge area, the median project management maturity was level 3. It is also instructive to note that reported practices in the project risk management knowledge area are the least mature of all knowledge areas. In fact, this area accounts for four of the five cases where the median level of project management maturity is level 1. In particular, the respondents indicated level 1 practice with respect to risk quantification, risk response development, risk control, and risk documentation. Risk management practices appear to deserve the attention of those charged with leading their organizations toward improved project management maturity.

The most significant difference between industries identified in this study occurs with respect to the schedule development component area. The median maturity reported by respondents in the professional services industry was level 3. This value is above the median results reported for the information and finance industries (level 2) and well above the median maturity level reported by those respondents engaged in the manufacturing industry (level 1). These results indicate that in the professional services industry, a majority of the organizations that participated in this study have adopted the practices that characterize level 3 maturity in schedule development. These practices include the definition of the project schedule at the appropriate level of detail, in line with the project scope and work breakdown structure. Additionally, the project schedules are established as baselines that are managed. Finally, the use of project management scheduling tools is standard across all projects. In the remaining three industries, a preponderance of respondent organizations are currently operating at level 2 or below. At level 2, schedule development is based on a repeatable process that relies on expert knowledge, access to industry methods, and access to commercial databases. Project teams develop staffing plans to support the project schedule and work with management to secure the resources required.

Finally, at level 2, project management software tools are standard for large, visible projects. We should also note that over 70% of respondents in the manufacturing industry indicated their organizations were operating at level 1 maturity. At level 1, scheduling occurs in an ad hoc fashion. The organizations lack a formal process for developing project schedules. The results for cost resource planning reveal that once again the respondents in the professional services industry report the highest level of project management maturity with over 38% reporting at level 3 or higher, and the respondents from the manufacturing industry report the lowest level of project management maturity with over 70% reporting practices at level 1. Among industries, project management maturity levels reported for cost resource planning are less than those reported for schedule development. In the case of professional services and finance and insurance companies, the majority of respondents report practices consistent with level 2 or above. These organizations are identifying resource requirements for all labour categories, equipment, and material. They are also employing planning processes that develop and document the resources required, as well as the methods for determining resource requirements. The planning process in level 2 organizations is supported by management and is gaining acceptance through the organization. In the manufacturing and information industries, level 1 practice is considerably more prevalent. The respondents indicated that project managers in their organizations employ methods to identify resources that vary from one project to the next. At times, the functional support areas are overlooked. Most indicated the cost resource planning process was undocumented.

5 Conclusions

The value of the project to the client can be measured in part by the level of quality associated with the completed project. Quality is acknowledged to be an important component of project management, but previously has received limited consideration in planning and scheduling. The implicit assumption behind standard time/cost tradeoffs is that some unspecified level of quality is maintained for each task. However, in many situations project managers must evaluate alternative options for accomplishing project activities, and these involve differing levels of time, cost, and quality. In such situations, it makes sense to analyze the relationship between cost, time, and quality, and decide on their levels for each project task that best achieves the project's objectives. We have introduced the concept of a quality function that represents the relationships between quality, time, and cost for each task. Using two case studies with real data, a translation agency and a software development company, the quality function is specified for each task and incorporated into a nonlinear programming model that allows quality to be explicitly considered in project planning and scheduling. An alternative formulation minimizes cost with bounds on project quality and completion time and leads to the creation of quality level curves. Both formulations can be very useful tools in making final project planning and scheduling decisions that explicitly model and incorporate quality.

References

[1] Kerzner H 2003 Project Management: A Systems Approach to Planning, Scheduling and Controlling *8th ed New York Wiley*
[2] Project Management Institute A Guide to the Project Management Body of Knowledge 3rd ed Newtown Square PA Project Management Institute 2004
[3] International Organization for Standards ISO 9000:2000 2000.
[4] Stevens J, Glagota C, Ledbetter W 1994 Quality-measurement matrix *Manage Eng* **10**(6) 30-35
[5] Stevens J 1996 Blueprint for measuring project quality *Manage Eng* **12**(2) 34-9
[6] Chan A 2001 A quest for better construction quality in Hong Kong *CIOBConstruction Informat Quarterly* **3**(2) 1200-13
[7] Heravitorbati A, Coffey V, Trigunarsyah B 2011 Assessment of requirements for establishment of a framework to enhance implementation of quality practices in building projects *Int J Innov Manage Technol* **2**(6) 465-70
[8] Paquin J P, Couillard J, Ferrand D J 2000 *IEEE Trans Eng Manage* **47**(10) 88-97
[9] Saaty T L 1996 The Analytic Hierarchy Process *Pittsburgh RWS Publications*
[10] Keeney R, Raiffa H 1976 Decision with Multiple Objectives: Preferences and Value Tradeoffs *New York Wiley*
[11] Pollack-Johnson B, Liberatore M 2006 *IEEE Trans Eng Manage* **53**(4) 534-42
[12] Icmeli-Tukel O, Rom W O 1997 Ensuring quality in resource constrained project scheduling *Eur J Operational Res* **103**(3) 483-96
[13] Babu A J G, Suresh N 1996 Project management with time cost and quality considerations *Eur J Operational Res* **88**(2) 320-7
[14] Khang D B, Myint Y M 1999 Time cost and quality trade-off in project management A case study *Int J Project Manage* **17**(4) 249-56
[15] Morgan LO, Morgan R M, Moore W 2001 Quality and time-to-market tradeoffs when there are multiple product generations *Manuf. Serv. Oper.Manage* **3**(2) 89-104
[16] Brucker P, Drexl A, Mohring R, Neumann K, Pesch E 1999 Resource-constrained project scheduling Notation, classification, models, and methods *Eur J Operational Res* **112**(1) 3-41
[17] Interagency Language Roundtable (16 Sep 2011) ILR skill level descriptions for translation performances [Online] Available http://www.govtilr.org/Skills/AdoptedILRTranslationGuidelines.htm
[18] American Translators Association, ATA Certification Program Rubric for Grading, Version 2009 Available from the American Translators Association Alexandria VA
[19] Lindo Systems Inc Extended Lingo/Win 32 Version 13.0 Chicago 2011

Author

Jiale Tian, born in May 1981, Yiyang, Hunan Province, China.

Current position, grades: senior engineer, PhD candidate in School of Civil Engineering and Architecture, Wuhan University of Technology, China.
University studies: project management.
Scientific interest: project management and information management.
Publications: 15 papers.

Soft sensor system of coke oven flue temperature based on CBR and PCA-RBFNN

Wentao Xiao[1], Gongfa Li[1,2], Honghai Liu[2*], Guozhang Jiang[1], Ze Liu[1], Disi Chen[1], Weiliang Ding[1], Wei Miao[1], Zhe Li[1]

[1]*College of Machinery and Automation, Wuhan University of Science and Technology, Wuhan 430081, China*

[2]*Intelligent Systems & Robotics Group, School of Computing, University of Portsmouth, Portsmouth, PO1 3HE, United Kingdom*

Received 1 October 2014, www.cmnt.lv

Abstract

The key process indicator – coke oven flue temperature – is difficult to detect online with instruments in the coke oven heating process, thus an intelligent forecasting model is developed which is composed of four parts: the data gathering and handling unit, the optional forecasting unit, the online amendment unit and the effect evaluation unit. The optional intelligent forecasting model and its corresponding algorithm are established for different categories of practice operating conditions. In normal operating condition, the nearest neighbor clustering algorithm based on the principal component analysis and neural network with the radial basis function is selected. In unconventional operating condition, the case-based reasoning technology is selected. The models of different conditions are validated and applied according to the actual data in a steel enterprise coke production, the results show that the established forecasting model can reflect different practice conditions and meet the real-time control requirements.

Keywords: intelligent prediction, neural network, principal component analysis, neighbor clustering algorithm

1 Introduction

In the complex process of industrial production, some of the important variables closely related to the quality of products cannot be measured online directly due to various reasons by measurement instrumentation so that it has become a key factor to restrict the product quality and production efficiency. Coke oven flue temperature is a very important process parameter and refers to the measurement flue temperature average of the whole furnace chamber, which can reflect the whole oven temperature level. At present, the existing flue temperature soft measurement models are mostly linear regression models with the establishment of the regenerator top temperature and the coke oven flue temperature or creating the double parabolic model between them [1]. These models just regard the relationship of the regenerator top temperature and the flue temperature seen as a linear relationship or double parabolic relationship which exists certain limitations and low measurement accuracy. To meet the need of how to achieve the real-time control for coke oven flue temperature has become a constraint key issue.

Soft sensor technique is developed to solve such problems: According to some auxiliary variables being directly measured and the corresponding mathematical model in the process of industrial production to estimate the dominant variables not being measured online [2,3]. At present, soft sensor technique has more and more applications as the effective way to estimate some unpredictable variables in the industrial process. Main research content of soft sensor technology is: Mathematical model is established for a certain process variable. This mathematical model can be called the prediction model based on its characteristic and role modeling. Currently, the neural network soft measurement technology and the case-based reasoning soft measurement technology have been applied more and more widely.

Neural network has some itself characteristics: approximated to any complex nonlinear continuous function, the parallel handling of information and self-learning, adaptive, etc. Among them, the neural network of Radial Basis Function belongs to the local approximation network, and it has some significant features in the approximation capability, the classification ability and the learning speed, etc. [4,5]. Principal Component analysis method is a statistical analysis method to transform multiple related variables into a few independent variables. Soft sensor methods based on PCA and soft RBFNN is able to take advantage of neural network modeling at reduced dimensionality situation [6,7].

Case-Based Reasoning directly uses the past knowledge and experience, the model is easy to implement and maintain, and its training is simple and effective. Case-Based Reasoning adopts incremental learning method and has a strong self-learning ability, being good at using the knowledge and experience to learn [8,9]. The new cases are added increasingly in the learning process to modify the old cases for improving the judge reasoning ability.

The intelligent forecasting model is developed for the problems and needs above which is composed of four

[*]***Corresponding author's*** e-mail: honghai.liu@port.ac.uk

parts: the data gathering and handling unit, the optional forecasting unit, the online amendment unit and the effect evaluation unit [10,11]. The respective model and the corresponding algorithm are established with neural networks and case-based reasoning combined for different categories of actual operating conditions. In normal operating condition, the nearest neighbor clustering algorithm based on the principal component analysis and neural network with the radial basis function is selected. In unconventional operating condition, the case-based reasoning technology is selected. These two techniques are simultaneously applied to the coke oven flue temperature soft measurement of a coking plant, the practical application results show that the established forecasting model can reflect different actual conditions and meet the real-time control requirements.

2 The entire system of optional intelligent forecasting model

The optional intelligent forecasting model structure which can predict key variables of complex industrial process is displayed in Figure 1. $\hat{X}$ is the output of the optional prediction unit. $\bar{X}$ is the output after amendment; the distributed control system (DCS) generates the process data set Σ. Θ is the manual measurement data set from the online monitoring model. e is the amendment input parameters from the online amendment unit. u is the input of control object; y is the output of control object.

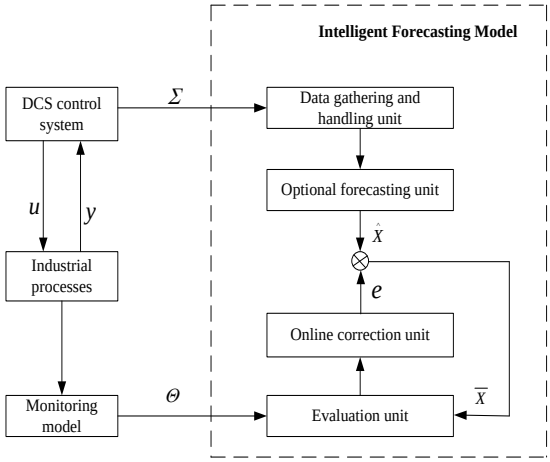


FIGURE 1 The entire system of optional intelligent forecasting model

a) Data gathering and handling unit: Data collected from the scene usually contains various interfering noise, such as the zero drift for the instrument, etc. In order to minimize these noise interference when modeling, these input data must be converted and difference handling, namely data pre-handling. Under certain circumstances, the output of prediction model should be appropriately processed, the reason is that the model is obtained in the context of a series of assumptions, it is impossible fully consistent with the practice situation and there is a model discrepancy.

b) Optional forecasting unit provides different categories forecasting model, depending on the actual condition, a prediction model and algorithm is selected. In normal condition, the neural network forecasting model is selected. Through learning and training, the neural network can obtain the knowledge of the key variables for the forecasting. In unconventional conditions, the case-based reasoning model is selected, the similar cases in the case base are retrieved according to the characteristics working conditions and the retrieved cases are matched and reused based on resemblance threshold to obtain the case solution of current condition which is the soft measurements of estimated dominant variable to predict.

c) Online amendment unit: the output of prediction model will drift with the passage of time after the forecasting model being put into use, therefore, some appropriate corrective measures must be taken to increase the accuracy of forecast values. Yet general, the online amended parameters is taken:

$$e = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^* - \hat{X}_i\right), \tag{1}$$

$\hat{X}_i$ is the output of prediction model, X_i^* is the practical measured value.

The amended output is:

$$\bar{X} = \hat{X} + e\,. \tag{2}$$

This amendment pattern is easy to implement, the output of neural network prediction model should be calibrated to compensate for drift according to Equations (1) and (2) to increase the accuracy of forecast.

d) Evaluation unit will make corresponding evaluation results for forecasting the results of different categories, the output for neural network prediction model is compared with the manual measurement data obtained with the monitoring model, if the difference exceeds the set of upper and lower limit, the training and learning of neural network is conducted. The output of forecasting model for case-based reasoning is compared with manual measurement data obtained with monitoring model to evaluate the prediction accuracy.

3 Two different categories of the model structure in the optional prediction unit

3.1 THE FORECAST MODEL STRUCTURE FOR NORMAL OPERATING CONDITIONS

When the prediction model of flue temperature established, after considering the actual values of these factors: the statistics of regular testing T^*. The load pressure of furnace n, the flow of heating gas u, the pressure of heating gas p,

the energy value of heating gas *h*, the flow of heating air *v*, the flue temperature of combustor is analyzed by principal component analysis and neural network technology in order to achieve the intelligent prediction of flue temperature. The structural principle is displayed in Figure 2.

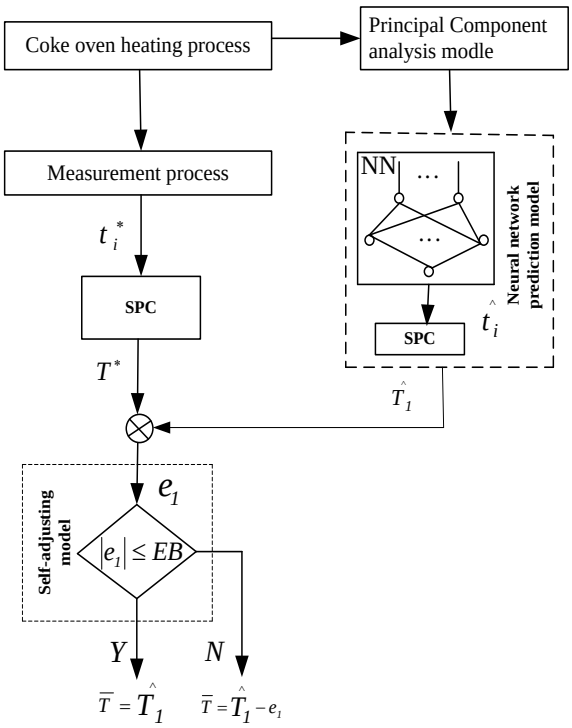


FIGURE 2 Neural network forecasting model of coke oven flue temperature

Intelligent forecasting model for flue temperature is composed of several parts which mainly are the principal component analysis model, neural network prediction model, self- amendment model. $\hat{T}_1$ is the output of neural network prediction model; $\hat{t}_i\,(i=1,2,...,n)$ is the output value of neural networks within a certain period of time. $t_i^*\,(i=1,2,...,n)$ is the actual measured value of flue temperature within a certain period of time. T^* is the statistics of manual measurement. e_1 is the difference between output value of neural network prediction model and manual measurement result; *TB* (>0) is the preset difference limit. $\overline{T}$ is the amended output of flue temperature.

The model structure of PCA and RBFNN is displayed in Figure 3. PCA extracts PCA components about some preliminary selection input variables to reduce the number of input variables and simplify RBFNN model under the premise of avoiding the loss of multi-variable information. RBF neural network is composed by three layers, the input layer nodes only pass the input signal to the hidden layer, the hidden layer nodes is composed by the effect function of radial structure like Gaussian functions, and the output layer node is a simple linear function. The basis function of hidden layer nodes will be generated in the local response for the input signal, that is, when the input signal is close to the central range of the base function, the hidden layer node will have a great output.

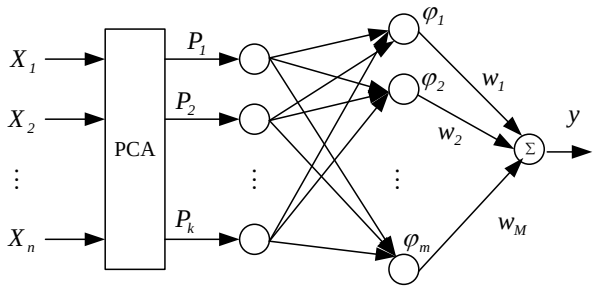


FIGURE 3 The model structure of PCA and RBFNN

The tasks of the hidden layer and output layer are different in the RBF neural Networks, therefore their learning strategies are also different. The hidden layer adopts nonlinear optimization strategy to adjust the mapping function parameters, thus it has the relative slower learning rate. The output layer adopts linear optimization strategies to adjust linear weights, thus the learning rate is faster.

There are many forms of radial basis functions, this paper selects the Gaussian function:

$$\varphi\left(\left\|P-P_i\right\|\right)=\exp\left(-\frac{\left\|P-P_i\right\|^2}{2\delta^2}\right), \tag{3}$$

In the type, $\varphi(\cdot)$ is the radial basis function; $\|\cdot\|$ is the European norm; $P_i \in R^{n_i}$ is the center of RBF; δ is the standard deviation of the neurons P_i determining the shape of the Gaussian function:

$$\delta=\frac{d_m}{\sqrt{2M}}, \tag{4}$$

where d_m is the maximum distance among the selected centers; *M* is the center number (the number of hidden layer units).

The output is:

$$F(x)=\sum_{i=1}^{M} w_i\varphi\left(\left\|P-P_i\right\|\right)+w_0, \tag{5}$$

where w_i is the weight of hidden unit to output unit; w_0 is the bias item.

PCA model needs a data set of the normal working conditions as the modeling data. Suppose the test data vectors matrix process $X \in R^{n\times n}$ is composed by *m* process variable and *n* data vector samples. In order to avoid different dimensions impact on the results and ease to handle mathematical, it is necessary to normalize the data. μ is the mean vector of set *X*, σ is the standard deviation vector, the normalized process variable is:

$$\tilde{x}_{ij}=\frac{\left(x_{ij}-\mu_j\right)}{\sigma_j}, i=1,2,...,n;\ j=1,2,...,m, \tag{6}$$

$\tilde{X}$ is the normalized vector of process variables, *V* is the covariance matrix for $\tilde{X}$. Based on principal component

analysis for $\tilde{X}$, the process of principal components selected as the input vector of RBF network based on Principal Component Analysis is as the followings.

a) The each eigenvalue λ_j and corresponding unit orthogonal eigenvector P_j are calculated for the covariance matrix;

b) The principal component t_j of the order j is calculated:

$$t_j = \tilde{X}P_j, \tag{7}$$

c) The model of principal component is calculated:

$$\hat{X} = t_1P_1^T + t_2P_2^T + ... + t_jP_j^T, \tag{8}$$

where $t_j \in R^n$ is called as the score vector; $P_j \in R^m$ is called as the load vector; P_j is the corresponding eigenvector matched with the order j eigenvalue in descending order for the matrix V, including various interrelated information among the variables. Each pair t_j, P_j is sorted by the descending of the eigenvalue λ_j for the corresponding eigenvector P_j. Where in the first pair intercepts the maximum information volume among all decomposition of the load vector and principal component vector. So on, $T = [t_1, t_2, ..., t_m]$ is the score matrix; $P = [P_1, P_2, ..., P_m]$ is the load matrix, the Equation (9) can be written as:

$$\hat{X} = TP^T. \tag{9}$$

d) The contribution rate is calculated before the principal component $k(1 \le j \le m)$:

$$\eta_k = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{m}\lambda_i}. \tag{10}$$

Next step, the number of principal component for PCA model should be determined. Since there is high correlation between the process variables based on experience, generally, the cumulative contribution rate of all principal components before the principal component k more than 85% is considered to reflect the body information of process variables. The k vectors for the corresponding principal components are selected main element vectors in the matrix as the input vector of RBF neural network training.

3.2 THE FORECAST MODEL STRUCTURE FOR UNCONVENTIONAL OPERATING CONDITION

The intelligent forecasting model of case-based reasoning is established for flue temperature after considering the actual values of these factors: the statistics of regular testing T^*, the load pressure of furnace n, the flow of heating gas u, the pressure of heating gas p, the energy value of heating gas h, the flow of heating air v. The model structure is displayed in Figure 4. Ξ is the set of measurement data obtained from the coke oven heating process. $\hat{T}_1$ is the output of case-based reasoning prediction model; $t_i^*(i = 1, 2, ..., n)$ is the actual measured value of flue temperature within a certain period of time. T^* is the statistics of manual measurement. e_1 is the difference between output value of case-based reasoning prediction model and manual measurement result. TB (>0) is the preset difference limit. $\bar{T}$ is the amended output of flue temperature. Model selector gets the measurement data from the coke oven heating process, the forecast values are completed by case-based reasoning model, and the self-correction model based on artificial measured temperature statistics to determine how to deal with the results of prediction, the target of forecast is to get at last.

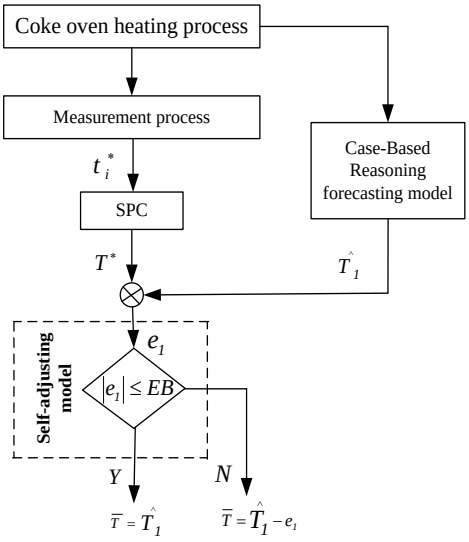


FIGURE 4 Case-based reasoning forecasting model of coke oven flue temperature

4 Model algorithms under different conditions

4.1 THE SELF- CALIBRATION ALGORITHM

The initial forecast results should be amended online to ensure the forecast accuracy, while the self-calibration algorithms for the prediction models of two conditions are the same, the same self-calibration algorithm is available [12,13].

Generally, the flue temperature of combustion is measured through the Infrared Thermometer held by the operator. Setting that the set for the manual measurement data is $\{t_i^*, i = 1, 2, ..., k\}$, statistical process control (SPC) mechanism is used to deal with these data.

$$T^* = \frac{\sum_{i=1}^{k} t_i^*}{k}, \tag{11}$$

$\hat{T}_1$ is the original forecast value of flue temperature from the forecasting model, the effect of flue temperature forecast can be assessed by the following equation:

$$e_1 = \hat{T}_1 - T^*, \tag{12}$$

where e_1 is the difference between the original forecast value of flue temperature and the manual measurement

data. If the absolute value of the difference e_1 is beyond a preset difference limit *TB*, thus $|e_1| > TB$, then the output of prediction model should be amended, the amended parameter is e_1. Otherwise, the output needn't be amended, the value of amended parameter e_1 is 0. The amended output is as follows:

$$\bar{T} = \hat{T}_1 - e_1. \tag{13}$$

4.2 THE NEAREST NEIGHBOR CLUSTERING LEARNING ALGORITHM OF RBF NEURAL NETWORK

K-mean clustering algorithm is a classical clustering algorithm to determine the centric value of RBF network, its main problem is that the number of network hidden layer nodes must match the number of input patterns, or it may seriously affect network performance. The number of centers is fixed in advance and depends on the initial position of the cluster center, there is likely to fall into local extreme point. This paper uses the nearest neighbor clustering algorithm with a variable number of hidden layer center to select the center of the hidden layer, the algorithm is as follows:

Select the appropriate width of the Gaussian function;

Begin from the first learning sample (P^1, Y^1), establishing the first cluster center for P^1, taking $C^1 = P^1$;

Consider a second study sample (P^2, Y^2); calculate the distance $|P^2 - C^1|$ of the cluster centers between P^2 and C^1, if $|P^2 - C^1| < r$, then C^1 is the nearest neighbor clustering for P^2, if $|P^2 - C^1| > r$ then P^2 is as the new neighbor clustering center, taking $C^2 = P^2$; then add a hidden unit in the establishment of the RBF network.

Suppose there are *M* cluster centers before consider the order *k* learning sample (P^k, Y^k). The central points are $C^1, C^2, ..., C^M$. The distance of these points to *M* cluster centers is respectively calculated, $|P^k - C^i|, i = 1, 2, ..., M$. If $|P^k - C^i|$ is the minimum distance among them, then C^i is the nearest neighbor clustering for P^k; if $|P^k - C^i| < r$, then C^i is the nearest neighbor clustering for P^k. If $|P^k - C^i| > r$, then P^k is as the new neighbor clustering, taking $C^{M+1} = P^k$, $N = M + 1$.

For each pair of the input – output data is likely to generate a new clustering, the self-adjust of two processes for parameters and structure is carried out simultaneously; therefore, the value of determines the complexity of the dynamic adaptive RBF network, reflecting one kind of generalization ability of the RBF network.

After determining the network center, the weights of the hidden layer of the network to the output layer can be adjusted, this paper adopts the recursive least squares algorithm with a weighted forgetting factor to adjust the weights of the output layer, the specific algorithm is as followings:

a) The initial values of the weights $w_i (i = 1, 2, ..., M)$ are given, the initial value $Q(0)$ of the inverse correlation matrix is the positive definite diagonal matrix for the matrix $M \times M$ (*M* is the hidden layer nodes).

b) Taking a loop variable $k = 1$.

c) The output hidden layer node output is calculated:

$$\varphi_k = \varphi\left(\left\|P^k - C^i\right\|\right) = \exp\left(-\frac{\left\|P^k - C^i\right\|}{2\delta^2}\right), \tag{14}$$

$$J(k) = Q(k-1)\varphi_k \left[\rho + \varphi_k^T Q(k-1)\varphi_k\right]^{-1}, \tag{15}$$

$$Q(k) = \frac{1}{\rho}\left[Q(k-1) - J(k)\varphi_k^T Q(k-1)\right], \tag{16}$$

$$\hat{w}(k) = \hat{w}(k-1) + J(k)y_d(k) - \varphi_k^T \hat{w}(k-1), \tag{17}$$

where $\rho(0 < \rho < 1)$ is the weighted forgetting factor; this article takes: $\rho = 0$. $\hat{w}(k)$ is the amendment result of weight matrix after the sample in the order *k* is input. φ_k is the output vector of hidden node after the sample in the order *k* is input. $y_d(k)$ is the expected output of the sample in the order *k*.

d) The target cumulative difference is calculated:

$$E(k) = \rho E(k-1) + \frac{1}{2}\left[y_d(k) - \varphi_k^T \hat{w}(k-1)\right]^2, \tag{18}$$

$E(k)$ is the cumulative difference between the network output and the actual output after the sample in the order is the input.

$E(k) < E$ is judged, if it do, then $k < N$ is judged, if this relationship $k < N$ is true, taking $k = k + 1$, going to the step 3, else going to the step 1. *N* is the number of samples, *E* is the preset target difference.

4.3 LEARNING ALGORITHM OF CASE-BASED REASONING

The specific algorithm of case-based reasoning is as the followings:

The case is expressed, the specific method is as the following: The following parameters are selected as the dominant variable for the flue temperature soft measurement model: the flow of heating gas, the suction of flue, the energy value of heating gas, the pressure of heating gas, and the load pressure of furnace. The cases of flue

temperature are stored in the computer in the form of database. The case database are composed of several of the case data, they is displayed in Table 1.

TABLE 1 The case for flue temperature of machine side and coke side

	Case for flue temperature of machine side	**Case for flue temperature of coke side**
Condition Expression	Machine-side gas flow	Coke side gas flow
	Machine-side flue suction	Coke side flue suction
	Gas heat value	Gas heat value
	Machine-side gas pressure	Coke-side gas pressure
	Furnace load pressure	Furnace load pressure
Case Solution	Machine side forecast value	Coke side forecast value

The resemblance of case is identified: The cases in the case base are defined as:

$$F_k = \left(f_{1,k}, f_{2,k}, ..., f_{n,k} \right), k = 1, 2, ..., m, \tag{19}$$

m is the number of cases in the case base.

Assuming the current operating condition is C_r. The resemblance function is defined as the following between the current condition C_r and the described case C_k:

$$Sim\left(C_r, C_k\right) = \sum_{i=1}^{n} w_i Sim\left(f, f_{i,k} \right), \tag{20}$$

where $Sim\left(f, f_{i,k} \right)$ is the resemblance between the feature $f_i \left(1 \le i \le n\right)$ of operating condition C_r and the described feature $f_{i,k}$ of the case, which is defined in the real interval $\left[0,1\right]$ and meets the symmetry and reflexivity. w_i is the weight coefficient of case described feature and $\sum_{i=1}^{n} w_i = 1$.

The resemblance threshold is identified. The threshold value of resemblance is determined before the case is matched, Sim_t is the threshold value of resemblance.

$$Sim_t = \begin{cases} X_t, Sim_{\max} \ge X_t \\ Sim_{\max}, Sim_{\max} < X_t \end{cases}, \tag{21}$$

where the threshold value X_t is decided by specific process or experience according to the situation of object, generally taking $X_t = 0.9$. The $Sim_{\max}$ is marked as $\max\limits_{k=1,2,...,m} \left(Sim\left(C_r, C_k\right) \right)$.

The case is retrieved and matched: $\tilde{C}$ is the case with the greatest resemblance $Sim_{\max}$. $\tilde{J}$ is marked as the solution after the case reused.

$$J_u = \begin{cases} \tilde{J}, \left(Sim_{\max} = 1\right) \vee \left(Sim_t = Sim_{\max}\right) \\ \sum_{k=1}^{h} Sim_k \times J_k \,/ \sum_{k=1}^{h} Sim_k, others \end{cases}, \tag{22}$$

In the type, $\vee$ is the "OR" operator, expressing the "or" relationship between before and after. The result is output after the case reused, and this time, the description working condition and its solution of solving this case are stored in real-time database.

5 Application example of intelligent forecasting model

The intelligent prediction model of coke oven flue temperature proposed in this paper has been successfully applied to the heating intelligent control system of a steel company's coke oven [14]. The forecast results of the machine side and coke side are shown respectively in Figures 5 and 6; According to the statistics, the model prediction difference of the machine side and coke side which is within ±7°C respectively reaches 92.5% and 90.8%, meeting the requirements of industrial production.

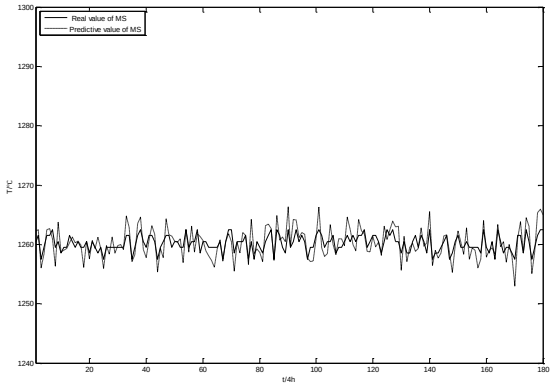


FIGURE 5 The forecast result for the machine side

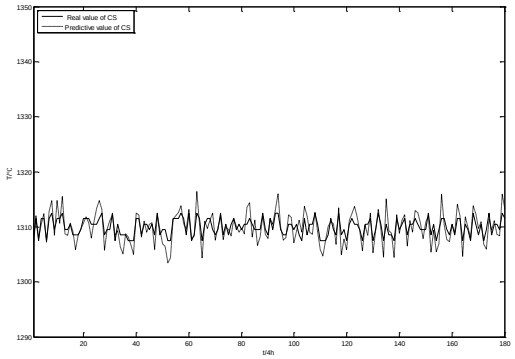


FIGURE 6 The forecast result for the machine side

6 Conclusions

The key process variables are difficult to detect online with instruments for the complex industrial processes, an intelligent forecasting model is developed which is composed of four parts: the data gathering and handling unit, the optional forecasting unit, the online amendment unit and the

effect evaluation unit. The optional forecasting model and its corresponding algorithm are established for different categories of practice operating conditions. In normal operating condition, the nearest neighbor clustering algorithm based on the principal component analysis and neural network with the radial basis function is selected. In unconventional operating condition, the case-based reasoning technology is selected. The two techniques are simultaneously applied to the intelligent forecast of coke oven flue temperature which can give full play to their strengths in different conditions. The model has a certain self-learning ability with higher intelligence. At the same time, the application results of the prediction model in the optimazation control of coke oven heating process show that the model has high accuracy. This method is effective and can be applied in real-time control of coke oven flue heating process.

Acknowledgments

This research was supported by Hubei Provincial Department of Education (Q20141107).

References

[1] Kong J-Y, Li G-F, Xiong H-G 2007 Research on Soft-sensing Modeling Methods and its Application in Industrial Production *Machine Tool & Hydraulics* **35**(6) 149-51
[2] Yoo C K, Lee I B 2004 Soft Sensor and Adaptive Model-based Dissolved Oxygen Control for Biological Wastewater Treatment Processes *Environment Engineering Science* **21**(3) 331-40
[3] Fortuna L, Graziani S, Xibulia M G 2005 Soft Sensors for Product Quality Monitoring in Debutanizer Distillation Columns Control *Engineering Practice* **13**(2) 2499-508
[4] Rank E 2003 Application of Bayesian Trained RBF Networks to Nonlinear Time-series Modeling *Signal Processing* **83**(7) 1393-1410
[5] Robert J S, Jamsex J C 2001 *IEEE Transactions on Neural Networks* **19**(4) 1-14
[6] Waheed A, Adeli H 2005 Case-based reasoning in steel bridge engineering *Knowledge Based Systems* **18**(1) 37-46
[7] Arshadim N, Jurisica I 2005 *IEEE Transactions on Knowledge and Data Engineering* **17**(8) 1127-37
[8] Yang B, Zhu Z Y 2005 A Nonlinear PCA Algorithm Based on RBF Neural Networks *Journal of Harbin Institute of Technology* **12**(1) 101-4 *(in Chinese)*
[9] Zamprogna E, Barolo M, Seborg E 2005 Optimal Selection of Soft Sensor Inputs for Bath Distillation Columns Using Principal Component Analysis *Journal of Process Control* **15**(1) 39-52
[10] Yan A-J, Yue H, Zhao D-Y 2005 Intelligent Soft-sensor model for a Class of Complex Industrial Process and Its Application *Control and Decision* **20**(7) 794-7 *(in Chinese)*
[11] Ning F, Zhang S, Yan W 2004 Generalized Predicative Control in Flue Temperature *Journal of UEST of China* **33**(1) 53-5 *(in Chinese)*
[12] Chen K, Zhang K, Qiao W 2003 Application of Low Cost Intelligent Heating Control System on Coke Oven Control and Instruments in *Chemical Industry* **30**(1) 29-30 *(in Chinese)*
[13] Hu Q-Y, Huang S-T, Li H-Z 2007 Application of Adaptive Neuro-fuzzy Inference System in Inverted Pendulum System *Journal of Mechanical & Electrical Engineering* **24**(1) 35-8 *(in Chinese)*
[14] Gu M-H, Zhu H-P, Cheng Y-F 2013 Adaptive relaying algorithm for protection of distribution networks integrated with wind farm *Journal of Mechanical & Electrical Engineering* **30**(6) 649-53 *(in Chinese)*

Authors

Wentao Xiao, born in 1989, Tianmen, China.

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at City College of Wuhan University of Science and Technology, Wuhan, China, 2013.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Gongfa Li, born in October 1979, Hubei province, P.R. China.

Current position, grades: associate professor at Wuhan University of Science and Technology, China.
University studies: PhD degree in mechanical design and theory at Wuhan University of Science and Technology in China.
Scientific interest: intelligent control, modeling and optimal control of complex industrial process.
Publications: 110.

Honghai Liu, born in 1973, P.R.China.

Current position, grades: professor in Intelligent Systems, Head of Intelligent Systems and Biomedical Robotics, University of Portsmouth.
University studies: PhD in Intelligent Robotics in 2003 at Kings College, University of London, UK.
Scientific interest: approximate computation, pattern recognition, multi-sensor based information fusion and analytics, human machine systems, advanced control, intelligent robotics and their practical applications.
Publications: 320.

Guozhang Jiang, born in December 1965, Tianmen, P.R. China.

Current position, grades: professor of Industrial Engineering, and the Assistant Dean of the college of machinery and automation, Wuhan University of Science and Technology.
University studies: PhD degree in mechanical design and theory at Wuhan University of Science and Technology, China, 2007.
Scientific interest: computer aided engineering, mechanical CAD/CAE and industrial engineering and management system.
Publications: 130.

Ze Liu, born in 1989, Hubei province, P.R.China.

Current position, grades: M.S. degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: B.S. degree in mechanical engineering and automation at Wuhan Institute of Bioengineering, Wuhan, China, 2013.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Disi Chen, born in 1992, Hubei province, P.R. China.

Current position, grades: M.S. degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: B.S. degree in mechanical engineering and automation at Wuhan Textile University, Wuhan, China, 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Weiliang Ding, born in 1991, Hubei province, P.R.China.

Current position, grades: Currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in measurement and control technology and instrumentation program from Changzhou Institute Of Technology, Changzhou, China, 2013.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Wei Miao, born in 1993, Henan province, P.R. China.

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Zhengzhou Huaxin College, Zhengzhou, China, 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Zhe Li, born in 1991, Hubei province, P.R. China.

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Wuhan University of Technology Huaxia College, Wuhan, China, 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Intelligent transportation systems based on computer aided simulation method

Hongli Xiao*

Baicheng Vocational and Technical College, Baicheng, Jilin, 137000, China

Received 1 March 2014, www.cmnt.lv

Abstract

Road transportation is one of the main sources of greenhouse gas emissions, which lead to global warming and climate change. Promoting the decarbonisation of this sector through more efficient and greener mobility is a challenging task that can be achieved by intelligent transportation systems (ITS) enabled by vehicular communications. Intelligent Transportation Systems (ITS) have been developed for more than ten years in China. Furthermore, a new generation Intelligent Transportation Systems should be launched to meet the requirement of rapid development of transportation in China. For the last two decades, intelligent transportation systems (ITS) have emerged as an efficient way of improving the performance of transportation systems, enhancing travel security, and providing more choices to travellers. A significant change in ITS in recent years is that much more data are collected from a variety of sources and can be processed into various forms for different stakeholders. This paper presents an overview of the background, concepts, basic methods, major issues, and current applications of Parallel transportation Management Systems (PtMS). In essence, parallel control and management is a data-driven approach for modelling, analysis, and decision-making that considers both the engineering and social complexity in its processes. The developments and applications described here clearly indicate that PtMS is effective for use in networked complex traffic systems and is closely related to emerging technologies in cloud computing, social computing, and cyber physical social systems.

Keywords: agents, intelligent transportation systems (ITS), mobile agent systems, multiagent systems

1 Introduction

In the European Union (EU), the majority of the population (72%) and economic development [85% of the gross domestic product (GDP)] concentrates in urban areas according to [1]. However, mobility in these areas is strongly constrained, i.e., by traffic and inefficient (private and/or public) transportation. These inefficiencies lead to congestion, pollution, noise, increased energy consumption, and the associated economic losses. For instance, in the USA, congestion wastes a massive amount of time, fuel, and money. i.e., 1.9 billion gallons of wasted fuel, 4.8 billion hours of extra time, and 101 billion dollars of delay and fuel cost [2]. In the EU, congestion costs 1% of the total GDP annually [3].

Recently, there has been significant interest and progress in the field of intelligent transportation systems (ITSs) from both industry and academia. The main contribution of ITSs is to significantly increase road safety. However, the role of infotainment has rapidly taken an important place [4]. The new emerging infotainment applications or safety-related mobile applications call for vehicular communication networks to support seamless wireless Internet services in fast-moving vehicles [5]. Internet access in automotive scenarios is a particularly relevant case, particularly because people in modern cities spend much time in transportation and vehicles. Public transportation systems, such as subways, suburb trains, and city buses, represent one relevant scenario because of a group of users and the time spent by these users in transportation and vehicles [6].

There are many challenges in conducting full-scale ITS research, particularly when vehicles are involved. For example, modifying conventional vehicles to autonomous vehicles is usually costly in terms of money and time. Testing in real traffic environments is usually dangerous. Such challenges have posed significant obstacles to many researchers in ITSs and forced them to use pure computer simulation [7], in which it is extremely hard to mimic certain aspects such as communication, vehicle dynamics, and driving experience. For many ITS research problems, a scale-down platform may be very useful for preliminary study and feasibility tests. However, there is very limited previous work in developing such a multipurpose ITS research platform. First, congestion has become an increasingly important issue worldwide as the number of vehicles on the road increases. For example, Beijing, China, had a total of 4 million vehicles at the beginning of 2010 and added another 800,000 in that year. Congestion can lead to an increase in fuel consumption, air pollution, and difficulties in implementing plans for public transportation [8]. It can also increase the risk of heart attack, as indicated by a medical report [9]. Second, accident risks increase with the expansion of transportation systems, particularly in several developing countries. Zheng et al. [10] showed that in China, there were 104,373 fatalities in 2003 and 67 759 fatalities in 2009. It was pointed out by Malta et al. [11] that almost three fourths of all traffic accidents can be attributed to human error. The reports published by the

* ***Corresponding author's*** e-mail: honglixiao_bc@126.com

U.S. Federal Highway Administration indicated that traffic accidents that happened in cities account for about 50%–60% of all congestion delays [12]. Undoubtedly, there is a need to reduce traffic accidents and to detect accidents once they have occurred to minimize their impact. Third, land resources are often limited in several countries. It is thus difficult to build new infrastructure such as highways and freeways. After the terrorist attacks in New York City on September 11,2001, the effectiveness of transportation systems is increasingly tied to a country's capability to handle emergency situations (e.g., mass evacuation and security enhancement) [13]. The competitiveness of a country, its economic strength, and productivity heavily depend on the performance of its transportation systems [14].

The domain of traffic and transportation systems is well suited to an agent-based approach because of its geographically distributed nature and its alternating busy-idle operating characteristics [15]. From the traffic and transportation management perspective, the most appealing characteristics of agents are autonomy, collaboration, and reactivity. Agents can operate without the direct intervention of humans or others. This feature helps to implement automated traffic control and management systems. Agents are collaborative. In a multiagent system (MAS), agents communicate with other agents in a system to achieve a global goal. Agents can also perceive their environment and respond in a timely fashion to environmental changes. Agent-based transportation systems allow distributed subsystems collaborating with each other to perform traffic control and management based on real-time traffic conditions. A distributed vehicle monitoring test bed presented in [16] is an early example of the distributed problem-solving network. Recently, more and more agent-based traffic and transportation applications have been reported. Our literature survey shows that the techniques and methods resulting from the field of agent system and MAS have been applied to many aspects of traffic and transportation systems, including modelling and simulation, intelligent traffic control and management, dynamic routing and congestion management, driver-infrastructure collaboration, and decision support. This paper reviews agent applications in traffic and transportation systems. These applications are classified into five categories: 1) agent-based traffic control and management sys-tem architecture and platforms; 2) agent-based systems for roadway transportation; 3) agent-based systems for air-traffic control and management; 4) agent-based systems for railway transportation; and 5) multiagent traffic modelling and simulation. The selected projects in each category are listed in a tabular format with information of project name, research group, application domain, and key features.

2 Related work

ITS aim to streamline the operation of vehicles, manage vehicle traffic, and assist drivers with safety and other information, along with provisioning of convenience applications for passengers [17]. In [18], these applications have been classified into three main categories: 1) safety-related applications (e.g., collision warning); 2) traffic efficiency and management (e.g., speed management); and 3) infotainment and entertainment (e.g., cooperative local services). In this paper, we will rather focus on traffic efficiency and management applications since these have a direct impact on traffic flow, vehicle coordination, and information dissemination. This last application type can be further subdivided into (as partly done in [19]): 1) speed management (provide regulatory or recommended contextual speed information to drivers to improve road efficiency); 2) navigation (information and services provided to drivers to assist in-vehicle routing); 3) road systems (e.g., traffic management systems); 4) vehicle control [e.g., adaptive cruise control (ACC)]; and 5) driver-related services (e.g., eco-driving).

Over the past decade, ITS technologies have greatly improved traditional transportation conditions, and improved the traffic capacity of the road network and transport security in China. In 1999, the National Engineering Technology Research Center of Intelligent Transportation System (ITS) was established. Many universities and research institutions also set up ITS Research Center to research ITS theories and technologies. The Tenth Five-Year period (2001-2005) first took the fields of intelligent transportation systems as content of the national planning. In 2001, China had selected ten cities as model cities for ITS field testing and evaluation. Those cities included Beijing, Shanghai, Guangzhou and so on [20]. During this period, many important aspects and key issues in ITS research and development were addressed on a high scientific and engineering level: agent-based and vision-based technologies; traffic modelling, control, and simulation; communication and location-based services; and driving safety and assistance, etc. For instance, the digital bus station systems has come into use in many cities, such as Beijing, Guangzhou, Chongqing, Shanghai, Hangzhou, Shenzhen, Nanjing, Shenyang, etc. Many information technologies of this system, such as computer control technology, wireless network communication technology and LED display control had been completely developed. During the Eleventh Five-Year period (2006-2010), ITS had more opportunity to develop, especially in transport services for major international events. Olympic traffic management command and control system was established in Beijing in 2008. Which include bus operations management system, floating vehicle dynamic traffic information collection processing and publishing systems. These systems have four major functions: command, signal control, integrated monitoring and regional traffic optimization. The systems comprehensively support Beijing public security traffic management work in 2008 Olympic.

3 Simulation based ITS

The operation of agents is supported and managed by distributed software platforms known as agent systems. The name of MASs usually refers to systems that support stationary agents, and mobile agent systems support mobile agents. An agent system provides mechanisms for agent management, agent communication, and agent directory maintenance. A mobile agent system provides additional mechanisms to support the migration and execution

of mobile agents. In an agent system, agencies are the major building blocks and are installed in each node of a networked system, in which agents reside and execute. To facilitate the interoperation of agents and agent systems across heterogeneous agent platforms, agencies designed to comply with agent standards are highly desired.

The integration of a driving simulation and traffic simulation environment may be achieved by implementing a communication structure that can effectively exchange data between the two applications (see Figure 1).

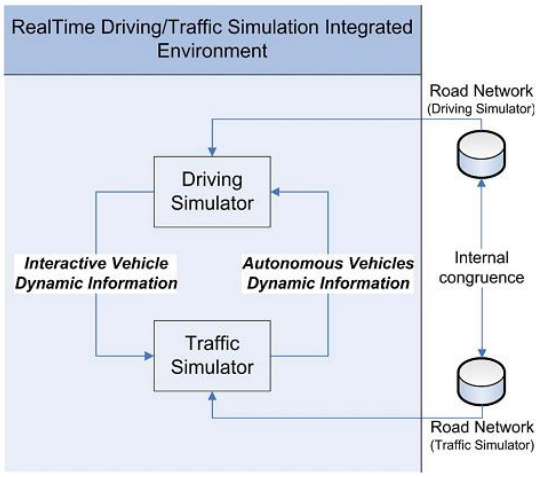


FIGURE 1 Communication framework

The integrated environment has to ensure that, during the driving simulation, the autonomous vehicles are consistently moved by the traffic microsimulation model with the movements of the interactive vehicle. Furthermore, during the simulation, the software that manages the driving simulator needs to send the kinematic characteristics of the interactive vehicle (e.g., position, angle, and speed) to the traffic-simulation model. Once such information has been received, the traffic-simulation model can calculate kinematics of the surrounding traffic for the step ahead and send them back to the driving simulation environment, which has the task of updating this information in the driving scenario.

To implement the communication framework, it is necessary to tackle several issues, which are summarized as follows. Accurate Road Matching Between Traffic and Driving Modules: The first issue for integration is the consistency between the road networks: Position related data to be exchanged have to be framed within a well-known reference system to allow appropriate conversion between the two environments during communication. The two road networks thus have to exactly match to have strict correspondence between the two simulation environments.

With regard to the first aspect, there is a simple way of overcoming the problem. Network creation usually starts in both software applications with importing the road alignment. Both types of applications should import drawing files (.dwg) or shape files (.shp). Thus, it is only important to avoid any transformation of geographical information during the import process to ensure the consistency of the two reference systems. This way, data retrieved by the two environments are ready to be exchanged. As for the road matching problem, this case is mainly because a driving simulator requires very detailed road geometry to present the road as realistically as possible for the driver, which is obviously far from the purposes of a microscopic traffic simulator, where, in most cases, for instance, the road alignment curvature, even when represented in detail, does not affect vehicle behavior.

Synchronization of Traffic and Driving Modules with Real Time: As aforementioned, the communication framework has to allow real-time data exchange required to perform a driving simulation. Unlike the driving simulation engine, traffic microsimulators are not designed to work in real time, but according to the hardware performance and scenario complexity, they try to perform all the needed calculations in the shortest time. This condition means that the time needed for a simulation is usually shorter than the real time and generally involves the need to slow down the computing speed of the traffic-simulation package. However, if the amount of data to be exchanged is large, the traffic module may slow down to less than real time. This case would naturally rule out any chance of successfully integrating the modules.

Consistency of the Updating Calculation Frequency: Each simulation model performs calculations at a given frequency. The finer the desired outputs are, the higher the updating frequency of calculations becomes. Both the microscopic traffic-simulation model and driving simulation software work this way. The difference is that the minimum simulation step in traffic simulation is usually 0.1 s, whereas realistic visualization in a driving simulation environment needs information to be updated at a higher frequency (generally 60 Hz). Therefore, if higher simulation frequency is not allowed by the microsimulation software (or it would slow down the simulation speed to below the real-time speed), outputs have to be augmented, e.g., through interpolation. Management of Autonomous Vehicle Visualization: This issue is twofold. First, a microscopic traffic simulation can manage thousands of vehicles at the same time. The management of such a large number of vehicles in a driving simulation environment is unfeasible for the aforementioned real-time issue and useless for driving simulation purposes. Hence, only the information of vehicles that surround the interactive one is exchanged during the simulation, which can be done by defining a "bubble" around the interactive vehicle. It is thus necessary to create, on the one hand, a vehicle in the traffic model to reproduce the movements of the interactive vehicle (maneuverer by the driving simulator) in the traffic simulation environment.

On the other hand, in the driving-simulation environment, a certain number of "ghost" vehicles have to be created, which come into action when some vehicles enter the bubble in the traffic simulation. When a vehicle enters the driving simulation, a clear one-to-one correspondence has to be created with the corresponding one in the traffic simulation. More specifically, as long as a vehicle in the traffic simulation stays within the surroundings of the interactive vehicle, it always has to be represented by the same vehicle in the driving simulation. Both aspects of this issue have to specifically be dealt with and, depending on

how the solutions are implemented, can seriously affect the possibility of having a real-time data exchange. The aforementioned issues do not constitute the exhaustive list of problems to deal with when implementing integration. They are, however, the most basic issues to obtain an effective framework. Additional issues may arise, depending on the particular driving and traffic simulation models used, as well as on the particular strategy adopted. The framework implemented in a particular case study, and the way in which such issues have been taken into account are shown as follows.

4 System Architectures and Operation Process

4.1 SYSTEM ARCHITECTURE AND OPERATION PROCESSES

Figure 2 presents the system architecture of ACP-based PtMS. It should be pointed out that normally, there is more than one ATS used in parallel traffic control and management. For example, different ATS can be created, respectively, for the purposes of historical traffic situations, normal and average performance, optimal and ideal operations, or worst-case scenarios for disasters and emergency management. Through interaction and parallel operations between an actual transportation system and its corresponding multiple ATS, the effectiveness of different traffic strategies under various conditions and expectations can be evaluated and analysed, both offline and online, and useful information can be obtained timely and combined to generate and select decisions for control and management.

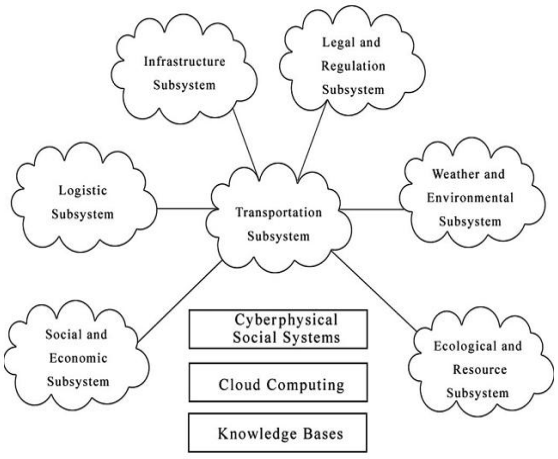


FIGURE 2 Scope, framework, and facilities of ATS

Typically, PtMS consist of five major components: 1) actual transportation systems; 2) ATS; 3) traffic operator and administrator training systems (OTSt); 4) decision evaluation and validation systems (DynaCAS); and 5) traffic sensing, control and management systems (aDAPTS). The operation process of PtMS is based on the co-evolution of real systems and the corresponding ATS. The operation process can be divided into three modes: 1) training; 2) testing; and 3) operating. General operational procedures can be derived from the functional specification of operating systems OTSt, Dyna CAS.

4.2 TRAINING AND LEARNING

Operator Training Systems for transportation (OTS) is developed for learning and training mode operations for traffic operators and administrators. The use of OTS was partially inspired by the applications and success of operator training simulation in many other advanced and complex industrial operations, such as in petrochemical production processes. Task requirements and procedures for both regular traffic operations and emergency situations are incorporated into OTS in order to make its functionality more useful and closer to reality. Sessions by OTS can be generated manually by human operators or automatically by agent programs. Manual sessions are also used to collect behavioral data from trainees and learners. Using agent-based behavioral modelling, automatic sessions can be employed when conducting accelerated testing and evaluation on the reliability and effectiveness of traffic operational procedures and regulations.

4.3 TESTING AND EVALUATION

Dynamic network assignment based on Complex Adaptive Systems (DynaCAS) is constructed to design, conduct, evaluate, and verify computational transportation experiments, detect existing and emerging traffic patterns, and support the use of advanced traveler information systems, advanced traffic management systems, and other ITS modules. DynaCAS facilitates the estimation and predication of traffic network conditions, performance testing and evaluation of different traffic control and management measures and information dissemination strategies, and decision support to traffic operators and individual drivers. Using ATS, DynaCAS is able to pay special attention to rule-based computational modelling of the social and behavioral aspects of people, vehicles, roads, and environments involved in transportation activities. In addition to conventional microscopic, mesoscopic, and macroscopic specification, a level of logic representation has been introduced in DynaCAS to represent the transportation networks so that factors in social and economic, ecological and resource, construction and infrastructure, logistical, legal, and regulatory aspects can easily be incorporated. On the logic level, transportation modelling extensively employs qualitative information in linguistic forms, and computing with words and methods in linguistic dynamic systems is used to achieve quantitative analysis. Data-mining techniques are also used to discover useful patterns from simulation results and computational experiments on all levels. Compared with other traffic estimation and predication systems (TrEPS), such as the well-known DynaSMART and DynaMIT, methods in AI and complex systems are much more extensively used in DynaCAS to provide additional flexibility and efficiency in traffic-condition analysis and decision evaluation.

4.4 CONTROL AND MANAGEMENT

Agent-based Distributed and Adaptive Platforms for Transportation Systems (aDAPTS) is built to provide supporting and operating environments to design, cons-

truct, manage, and maintain autonomous agent programs for various traffic tasks and functions. Those agents are delivered to traffic-control centers, roadside controllers, sensing devices, and information systems via communication networks to make the right decisions and collect the right information at the right time. We have designed and manufactured special intersection light controllers and sensing systems that are capable of hosting and processing traffic-control agents for different functions. Generally, an agent can autonomously move in networked environments, identify its tasks, and actively improve its performance. Transitioning from traffic-control and management algorithms to traffic-control and management agents is a natural step forward in this age of networks and connectivity. This step enables an intelligent and proactive mechanism that will significantly improve the performance and reliability of traffic operation control and management, yet it will have low cost for networked transportation systems.

Traffic task agents can be distributed at different operating centers and information sites. To ensure a coherent control and communication mechanism among those agents, we must integrate and coordinate their objectives and activities. A hierarchical intelligent control architecture consisting of organization, coordination, and execution levels has been used to facilitate the activities and operations of traffic agents, which is shown in Figure 3.

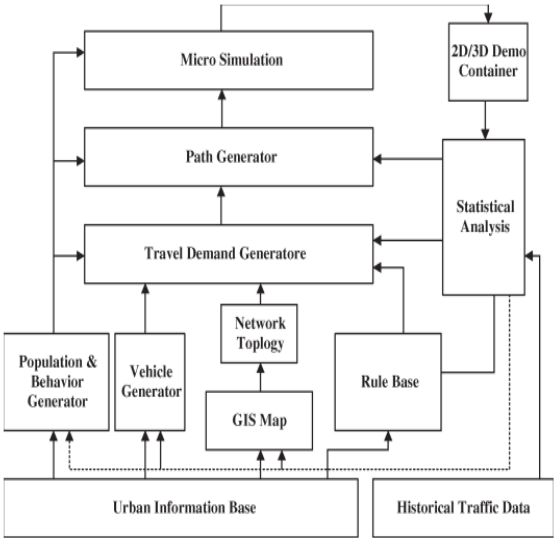


FIGURE 3 System architecture and components

InaDAPTS, a global traffic operating center (GTOC) (virtual or real) is designed to construct, organize, and maintain different types of agents for various traffic tasks, such as intersection light control, traffic-incident detection, and route guidance. WANs link the GTOC to several regional traffic-operating centers (RTOC), which host various agent repositories for regional demands and requirements of traffic control and management. RTOC coordinate and dispatch traffic agents to thousands of intersection light controllers, ad hoc networks, display devices, and guidance equipment at hundreds of locations. At each location, a gateway downloads task agents from WANs to traffic controllers, sensors, displayers, etc. through LANs and uploads information and requests from traffic devices to RTOC. Each location also has an agent coordinator to conduct cooperative control and management among traffic devices and within an individual device. Clearly, there is a natural link between PtMS and CPS, CPSS, and cloud computing. For example, CPS can be used to implement smart intersection traffic-control systems via aDAPTS, CPSS can be employed to build ATS in cyberspace and construct intelligent web-based DynaCAS and OTS t, and cloud computing is a cost effective way to host and operate the entire PtMS.

5 Conclusions

Road transportation is one of the main sources of greenhouse gas emissions, which lead to global warming and climate change. Promoting the decarbonization of this sector through more efficient and greener mobility is a challenging task that can be achieved by intelligent transportation systems (ITS) enabled by vehicular communications. Intelligent Transportation Systems (ITS) have been developed for more than ten years in China. Furthermore, a new generation Intelligent Transportation Systems should be launched to meet the requirement of rapid development of transportation in China. For the last two decades, intelligent transportation systems (ITS) have emerged as an efficient way of improving the performance of transportation systems, enhancing travel security, and providing more choices to travellers. A significant change in ITS in recent years is that much more data are collected from a variety of sources and can be processed into various forms for different stakeholders. This paper presents an overview of the background, concepts, basic methods, major issues, and current applications of Parallel transportation Management Systems (PtMS). In essence, parallel control and management is a data-driven approach for modelling, analysis, and decision-making that considers both the engineering and social complexity in its processes. The developments and applications described here clearly indicate that PtMS is effective for use in networked complex traffic systems and is closely related to emerging technologies in cloud computing, social computing, and cyber physical social systems.

References

[1] Li L, Li X, Li Z, Zeng D D, Scherer W T 2010 *IEEE Trans Intell Transp Syst* **11**(2) 251-5
[2] Li L, Li X, Cheng C, Chen C, Ke G, Zeng D D, Scherer W T 2010 Research collaboration and ITS topic evolution 10 Years at T-ITS *IEEE Trans Intell Transp Syst* **11**(3) 517-23
[3] Wang F-Y, Broggi A, White C C 2009 Road to transactions on intelligent transportation systems A decade's success *IEEE Trans Intell Transp Syst* **10**(4) 553-6
[4] Wang F-Y 2010 Publication and impact A bibliographic analysis *IEEE Trans Intell Transp Syst* **11**(2) 250
[5] Sanchez-Medina J J, Galán-Moreno M J, Rubio-Royo E 2010 *IEEE Trans Intell Transp Syst* **11**(1) 132-41

[6] Wei] J, Wang A, Du N 2005 *IEEE Trans Veh. Technol* **54**(2) 744-8
[7] Kumar P, Ranganath S, Weimin H, Sengupta K 2005 *IEEE Trans Intell Transp Syst* **6**(1) 43-53
[8] Panwai S, Dia H 2005 *IEEE Trans Intell Transp Syst* **6**(3) 314-25
[9] Ma X, Andreasson I 2007 *IEEE Trans. Intell Transp Syst* **8**(1) 144-56
[10] Yao D, Zhang Y, Li L, Su Y, Cheng S, Xu W 2009 *IEEE Intell Transp Syst Mag* **1**(2) 25-30
[11] Zhao H, Tang S, Lv Y 2009 *IEEE Intell Transp Syst Mag* **1**(3) 22-8
[12] Chen B, Cheng H H 2010 *IEEE Trans Intell Transp Syst* **11**(2) 485-97
[13] Ma Y, Chowdhury M, Sadek A, Jeihani M 2009 *IEEE Trans Intell Transp Syst* **10**(4) 615-27
[14] Chen H, Wang F-Y, Zeng D 2004 *IEEE Trans Intell Transp Syst* **5**(4) 329-41
[15] Kaza S, Xu J, Marshall B, Chen H 2009 *IEEE Trans Intell Transp Syst* **10**(1) 83-91
[16] Burgoon J K, Twitchell D P, Jensen M L, Meservy T O, Adkins M, Kruse J, Deokar A V, Tsechpenakis G, Lu S, Metaxas D N, Younger R E 2009 *IEEE Trans Intell Transp Syst* **10**(1) 103-12
[17] Zografos K G, Androutsopoulos K N, Spitadakis V *IEEE Trans Intell. Transp Syst* **10**(2) 311-23
[18] Candamo J, Shreve M, Goldgof D B, Sapper D B, Kastur R 2010 *IEEE Trans Intell Transp Syst* **11**(1) 206-24
[19] Wang F-Y 2004 Computational theory and methods for complex systems *China Basic Sci* **6**(41)3–10
[20] Wang F-Y 2004 Artificial societies, computational experiments, and parallel systems: An investigation on computational theory of complex social-economic systems *Complex Syst Complexity Sci* **1**(4) 25-35

Author

Hongli Xiao, born in December 1969, Baicheng, Jilin Province, China.

Current position, grades: associate professor in Baicheng Vocational and Technical College, China.
University studies: computer science and technology.
Scientific interest: computer software.
Publications: 20 papers.

Research on the robustness of supply chain network with uncertainty

Xing Xu[1*], Yun Zhao[1], Jia Hu[2], Xinli Wu[3]

[1]*Zhejiang University of Science and Technology, Liuhe Road 318, Hangzhou, P.R. China*

[2]*University of Duisburg-Essen, Keetmanstr. 3-9, Duisburg, Deutschland*

[3]*Zhejiang Sci-Tech University, 2#Str.928, Hangzhou, P.R. China*

Received 1 September 2014, www.cmnt.lv

Abstract

A new method is proposed to enhance robustness of complex supply chain network structure in uncertain conditions of market, costs and others. The researcher constructs the topology of a supply chain network firstly, and then explains the concept of central nodes of network and robust supply chain network based on scale-free network model from the perspective of quality control and places research emphasis on the method of enhancing the robustness of complex supply chain network structure. Finally, the supply chain network of auto parts enterprises in Zhejiang Province is cited as an example to verify the validity of the method.

Keywords: supply chain network, network quality control, scale-free network, robustness

1 Introduction

With the increasingly fierce market competition and complex economic environment changes such as variation of the users' needs, the uncertainty of the supply chain is more and more prominent; however, the uncertainty may arise from within the supply chain such as demand, production, sales, management, and operation [1]. Robustness of the supply chain is a universal basic attribute of it. Robustness is the ability to maintain the functions of its system when faced with changes in the internal structure or external environment. At present, a group of scientists led by Dirk Helbing found in the study of complex supply chain network that bullwhip effect in supply chain management is relevant with the topological properties of network [2,3]. This paper focuses on construction of a robust supply chain network under uncertainty, modeling analysis of complex supply chain network based on network static geometric quantities and analytical methods in graph theories and social network analysis, and exploration of the geometric properties of the network topology, robustness of the structure, dynamic robustness in the process of network evolution, etc. from the perspective of the quality control based on statistical principles and analytical method.

However, there are no existing studies on complex supply chain network topology and its robustness in the process of evolution under uncertainty. Most studies only point out the robustness and vulnerability according to the nature of the complex network [4-6]. Research on the complex supply chain network at home and abroad is mainly about supply chain adaptability analysis based on the complexity theory and knowledge sharing between the nodes of the supply chain companies [7-9], without taking the properties of the complex supply chain network topology under uncertainty into consideration. Complex supply chain network modeling based on network graph theory clearly shows the relationship between the enterprise nodes within the supply chain network [10], but can't express explicitly the robustness and vulnerability of the supply chain network. As a matter of fact, the robustness of the supply chain network is not only related to and the stability of the node enterprises in the supply chain but also to the structure of the supply chain network.

2 Construct scale-free supply chain network

Uncertainty is in most of the supply chain system. Robustness, a universal basic attribute of the uncertainty of a supply chain network, is the ability to maintain the functions of its system when faced with changes in the internal structure or external environment. The system of the supply chain develops from scratch gradually. The supply chain network, which is spontaneous and market-driven, has strong uncertainty. Currently, some scholars at abroad have found in the supply chain bullwhip effect – the amplification effects are all related to the supply network topology. Therefore, the effective supply network topology can weaken the bullwhip effect while enhance the stability and resistance to attack of the complex supply chain network.

First, we build a complex supply chain network modeling; the modeling process is shown in Figure 1:

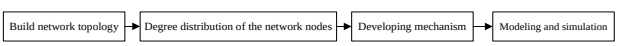


FIGURE 1 Complex supply chain network modeling process

[*]***Corresponding author's*** e-mail: xuxing3220@163.com

The formation, development and changes of the complex supply chain network can be seen as the result of interactions between the elements of the supply chain system and the external environment. Supply chain network is not just supply chain management of a single enterprise, but a complicated Multi-layer industry supply chain network. The construction method is as follows: the different business entities in the supply chain are taken as the network nodes, and the connections between the enterprises as the edges of the network. The edge direction is from the higher-level enterprise to the lower-level one; the edge weight is the number of the child nodes contained in the parent node, which is shown in Figure 2.

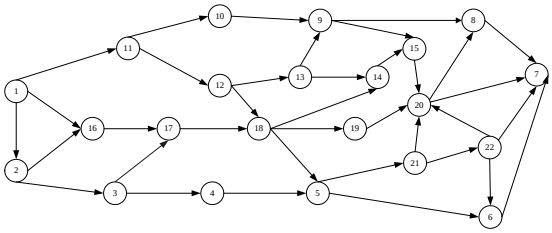


FIGURE 2 Supply chain network topology

In the supply chain network topology diagram shown in Figure 2, indegree refers to the number of edges pointing to the node, and outdegree refers to the number of edges starting from that point. We assume that x represents the node of complex enterprise cooperation network, and r represents the edge between the nodes in the network. i and j stand for any two nodes in the network. If there is an edge between i and j, the enterprise i and enterprise j are in cooperative relations, which is represented as $r(i, j)=1$. If N is the total number of the nodes in the network static statistics, the corresponding static statistics of the enterprise cooperation are defined as follows:

1) Degree of node: it is the number of edges the node connects with:

$$k_i = \sum_{i,j}^{N} r(i, j) . \tag{1}$$

2) Average path length of node: the distance between i and j in the network is d_{ij}, which is the number of the edges on the shortest path connecting the two nodes, namely:

$$L = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geq j} d_{ij} , \tag{2}$$

where N is the number of the nodes in the network.

3) Clustering coefficient: it is the parameter used to measure clustering situation of a node. We assume that i has k_i edges in the network, then the number of the edges between these k_i nodes may be up to k_i (k_i–1)/2. The ratio of E_i (the actual number of edges) between these k_i nodes and k_i (k_i–1)/2 (the total possible number of edges) is defined as the clustering coefficient of the node:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} . \tag{3}$$

Compile network processor based on the constructed supply chain network topology, and the aforementioned statistic eigenvalues of each node in the network can be achieved by applying the formulas mentioned above. Assuming that k_i (the degree of each node) is as shown in Table 1, the clustering coefficients of these nodes obtained by using Equation (3) are as follows.

TABLE 1 Node degree and clustering coefficient

No.	k_i	C_i	No.	k_i	C_i
1	3	0.06	12	1	0.00
2	22	0.01	13	1	0.00
3	1	0.03	14	23	0.00
4	1	0.00	15	45	0.07
5	39	0.09	16	8	0.13
6	56	0.12	17	4	0.02
7	86	0.15	18	5	0.01
8	1	0.13	19	12	0.01
9	12	0.00	20	23	0.02
10	112	0.04	21	20	0.04
11	1	0.04	22	1	0.03

By computing the clustering coefficient of each node in the supply chain network, it is found that the node with high node degree also has a high clustering coefficient; by performing statistical analysis on the distribution probability of each node degree, it is found that most nodes have a small number of connections in the network while a few nodes have a large number of connections, and there is no peak in the probability distribution of these node degrees, indicating that the node degree distribution follows the power-law distribution, in line with the characteristics of scale-free network.

3 Robustness analysis of the supply chain network under uncertainty

3.1 MODE AND FEATURES OF THE SUPPLY CHAIN UNDER UNCERTAINTY

The uncertainty of the supply chain contains two types. One is the uncertainty of the outer join. In scale-free supply chain network, partners often opt to quit due to reasons of their own. Such event, though occurs randomly, is considered to be attack on the network. For instance, the partners drop out of the supply chain network because they are in shortage of funds or stop producing some models of products. The other is uncertainty of emergency. Due to fierce market competition and inner motivation of achieving maximum interests, some important suppliers (i.e., core node) in the supply chain network are tempted by high profit from some other enterprises and some other enterprises or units are interested in inviting some important experienced intermediary retailers to enter into their network. Once these important suppliers and retailers are off the network, they will inevitably give a crushing blow to the core enterprises in raw material supply and sales of products. This is what vandalism means.

When the supply chain network is in uncertain conditions, the robustness of the scale-free supply chain network can be interpreted as: when a node in the network is des-

troyed, it continues to maintain its operational capacity, without causing significant impact to the entire network. A node which brings a catastrophic failure to the network when subjected to random external attack is judged to be unqualified. In accordance with the principles of quality control, only when the failure rates of the network nodes are all less than some expected number, this network is considered to be robust. Although Barabasi and some other scholars have pointed out that scale-free network shows unusual robustness when the node is under random attack [11,12], it is vulnerable to attack and leads to damage of the entire network. The robustness of scale-free supply chain network under conditions of uncertainty is to be analyzed from the perspective of network quality control in the following.

3.2 DEFINE THE CENTRAL NODE IN THE SUPPLY CHAIN NETWORK

First ($V(t)$, $E(t)$) represents a random complex supply chain network. $V(t)$ is a node set of the supply chain network at time t, that is, all enterprises in the supply chain; $E(t)$ is an edge set of the enterprises' nodes at time t; $N(t)$ is the number of enterprises in the network at time t; $E(N(t))$ is the average of the enterprises in the network at time t; $k_i(t)$ is the degree of the enterprises at time t; $P\{k_i(t)=k\}(k=1,2,...)$ is the degree distribution of enterprises at time t.

According to the robustness of the scale-free supply chain network, the central node of the scale supply chain network is defined firstly:

The central node of the scale-free supply chain network is defined as: if A and B are two proper subsets of $V(t)$, $A\neq\Phi$, $B\neq\Phi$, $A,B\subset V(t)$, $A\bigcup B=V(t)$, $A\bigcap B=\Phi$, $K>0$, when $k\geq K$, $\forall t>0$, $\forall i\in A$, $j\in B$, so that $P\{k_i(t)\geq k\}>P\{k_j(t)\geq k\}$, then A is named the central node of the scale-free the supply network. These nodes are the core enterprises in the supply chain.

3.3 DEFINITION OF ROBUST SUPPLY CHAIN NETWORK

This paper draws on SPC (Statistical Process Control) theory in quality control and mathematical statistical analysis theory to monitor and control the degree distribution of each node in the supply chain network, in order to realize the real-time regulation on node degree distribution and reduce variation of the robustness of the supply chain network. In the event that abnormal situation occurs to the central node of the network, i.e., the degree of the node exceeds the safety threshold σ, measures must be taken immediately to adjust it so as to ensure steady operation of the entire supply chain network [13].

The robust supply chain network refers to the network which is still able to maintain its functions when some nodes or edges in the network are under random attack. $p(t)=|A|/E(N(t))$ and $|A|$ is the potential of the set A. If $p(t)<\varepsilon$ and $\varepsilon\in(0,1)$, the failure rate of the central nodes in the network is less than the expected number ε. Such network ($V(t)$, $E(t)$) has certain robustness. The degree of each node in the supply chain network is often used in the study to facilitate the statistical analysis of robustness. In scale-free supply chain network, the degree of the central node is always very high, however, if it goes beyond a certain range and comes under attack, the entire network may malfunction. In accordance with quality control principles, the degrees of all nodes in the network are desired to be within the safety threshold. This kind of network has definite robustness [14,15].

The safety threshold is σ is divided into upper control limit σ_{UCL} and lower control limit σ_{LCL}, average is T, then:

Upper control limit: $\sigma_{\mathrm{UCL}}=T+A_2R$,

Lower control limit: $\sigma_{\mathrm{LCL}}=T-A_2R$,

where R is the average of the range of each node; A_2 is a constant. They vary with n (sample size). Please refer to Xbar-R coefficient table for the parameter values. In Table 1, n=22, A_2=*0.634*.

Therefore, for degree distribution of each node in the supply chain network, MSA (measurement system analysis) in SPC software is used for experimental statistics of the 24 subgroups. The following results can be obtained: control coefficient is 0; controls parameter: both upper control limit and the lower control limit are 22.27; stability judgment: the average number beyond the control limit is 8, which is judged to be unacceptable. Figure 3 is a diagram for stability analysis, from which the nodes beyond the control limits can be checked at any time.

4 Enhance the robustness of the supply chain network

Although the supply chain network is robust for random failure in the process of continuous preferential growth, it can't be ignored that the network is also vulnerable to attack. As long as the external attack impacts on the central high degree node, namely, the core enterprise in the supply chain network, it will lead to a destructive failure of the entire supply chain network, resulting in paralysis of the network. Therefore, how to enhance the stability of the network is essential. Two methods of enhancing the robustness in the process of supply chain network growth will be introduced in the following.

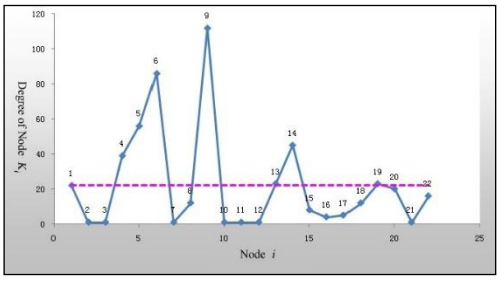


FIGURE 3 Diagram for stability analysis

4.1 CHANGE THE PREFERENTIAL MECHANISM

The first method: in the formation process of the entire supply chain network, the traditional way is to change the growth mechanism of the network so as to form a small-world network in the future development, in order to improve the stability of the entire network. In this paper, the referential mechanism is changed on the basis of two mechanisms growth mechanism and preferential attachment mechanism that the scale-free network follows so as to form a new network topology of exponential degree distribution. Network production model is as follows:

1) Growth: start from a network with m_0 nodes. Each time a new node is introduced, it is connected to m existing nodes, where $m \le m_0$.

2) Priority connection: the probability that a new node is connected to an existing node in the network i at time t is:

$$\prod_i(t) = 1/(m_0 + t - 1). \quad (4)$$

It is clear that in the network generated by the above-mentioned growth mechanism, the connection probability of its nodes has nothing to do with node degree. According to probability theory, the average degree distribution is $p(k) = \frac{e}{m} e^{-\frac{k}{m}}$, showing that its network nodes are exponentially distributed, so the role of core node in the network is weakened. Because of that, there is no major attack point in the entire network and the stability of the network is enhanced.

4.2 THE DYNAMIC ADJUSTMENT MODEL

The second method is: after an old node drops out, a new comparable point must be quickly found out to replace it. This requires real-time monitoring of the each node in the supply chain network to make different judgments in different situations.

1) In the operation process of supply chain network, a node may apply to quit the supply chain network due to some reason;
2) A node can't offer enough products because of its limited production capacity;
3) After a period of operation of the supply chain network, the clients' demand increases suddenly and the original supply chain network cannot satisfy the customers in time, quality and quantity;
4) After a period of operation of the supply chain network, the clients' demand decreases suddenly and the original supply chain network has overcapitalization, resulting in increase of operating costs in the supply chain network.

According to the aforementioned four situations, a dynamic adjustment model of scale-free supply chain network is constructed as follows, shown in Figure 4.

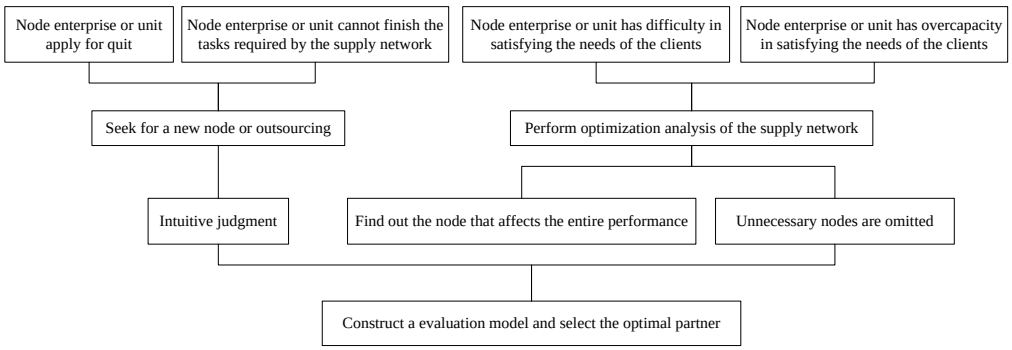


FIGURE 4 Scale-free supply chain network dynamic regulation model

The entire network needs to be adjusted continuously in the operation process. Only when the adjustment is finished and the new tasks of the adjusted supply chain network have been coordinated, can the new network realize effective conversion and start operation of new supply chain network. In this way, anti-risk capacity of the supply chain network is enhanced.

5 Application examples

Considering that components and parts of automotive industry have a wide range of distribution and are used in large quantity currently, auto parts enterprise clusters in Zhejiang province are taken as the object in this example, in order that automobile manufacturers can better manage the auto parts suppliers. By taking into account of various factors, with Matlab as the development tool, the researcher intuitively simulates the network distribution of the suppliers of automobile parts and motorcycle parts industry by studying the node degree distribution and random generation of point and edge, displays the central node of the supply chain network through the degree distribution of each node in the network, and checks if there are abnormalities in node distribution through network robustness test.

Here the information of 102 auto parts suppliers obtained from purchase department of China Youngman Automobile Group is taken as examples. We assume that the growth characteristics of the supply chain network are as follows: the initial number of nodes of the supply chain m_0=20; each time a new node is introduced, the number of

newly generated edges m=5; network size after growth N=102; network, m_0 nodes in the network are all isolated when the network is initialized.

Figure 5 is a diagram for supplier network topology generated by Matlab simulation software:

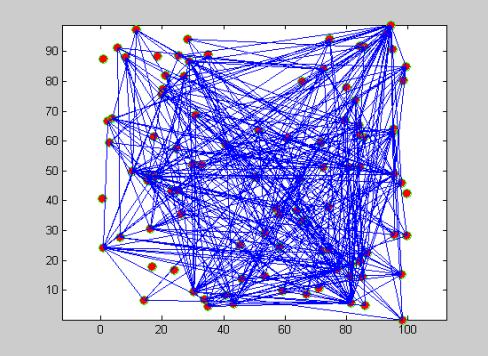
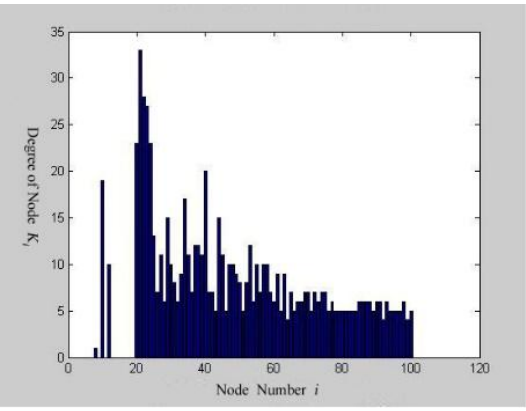


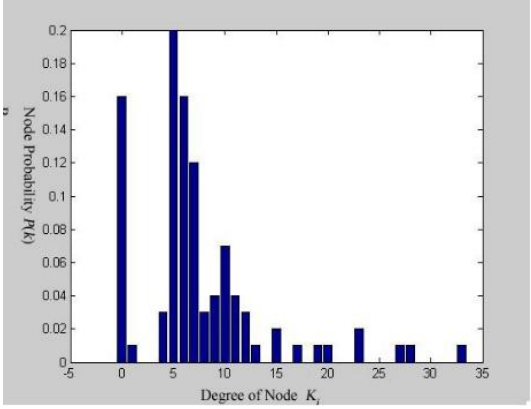
FIGURE 5 Diagram for network distribution

5.1 NETWORK DEGREE ANALYSIS

After growth in the above pattern, MATLAB program is used for statistics and analysis of the degree and distribution of each node. The final output of degree distribution and probability distribution of each node is shown in Figure 6.



a) Node degree distribution



b) Node probability distribution

FIGURE 6 Node degree distribution and node probability distribution

In Figure 6, there are 17 nodes with degree 5, 14 with degree 6, 10 with degree 7, 1 with degree 27, 1 with degree 28, 1 with degree 29, and 10 with degree 0. it shows that only few nodes have a high degree, for example, node 21, 22, and 23 meet the conditions in $P\{k_i(t) \geq k\} > P\{k_j(t) \geq k\}$ and become the central nodes of the supply chain network.

5.2 NETWORK ROBUSTNESS ANALYSIS

Previous data proves that the vast majority of nodes in the supply chain network have only a few connections, in line with the characteristics of scale-free network. By degree analysis of each node in Figure 3 and MSA analysis of degree distribution of the entire supply chain network, a few nodes have significantly exceeded the upper control limit. Therefore, if these enterprises are destroyed, the entire auto parts supply chain network will suffer a devastating injury. Attack on node 21, 22, or 23 will have an impact on the whole auto parts supply, while attack on node 1 or 0 will not.

It suggests that the supply chain network has a strong immunity to general random errors in the continuous growth process, but the central node in the supply chain network must be protected in order to maintain the stability of the network consistently, and as a result, the random network is shown to be in a steady state.

6 Conclusion

When the supply chain network is faced with many uncertain conditions, it is simple and intuitive to describe the supply chain network with the complex network theory. The theory of complex networks is a feasible method to complex supply chain network topology construction. The scale-free supply chain network is analyzed with emphasis through statistical analysis on each enterprise node in the supply chain network. The scale-free supply chain is observed from the perspective of quality control, and thus to propose the definition and quantitative analysis method of the robust supply chain network, providing a more scientific basis for scale-free analysis of the supply chain network. On that basis, two network optimization methods of enhancing the stability of the supply chain are proposed. Finally, practical examples show that the combination of network of quality control principles and complex network theory is a more practical and effective research method.

Acknowledgments

Project supported by the National Natural Science Foundation of China (Grant No. 61403346), Zhejiang Provincial Natural Science Foundation (Grant No. LY14F020023), Interdisciplinary research project of Zhejiang University of Science and Technology (Grant No. 2013JC05Y).

References

[9] Yeh C T, Lin Y K, Huang C F 2014 *IIE Transactions* **46**(10) 1066-1078.
[10] Dominguez R, Cannella S, Framinan J M 2014 *Computers & Industrial Engineering* **73** 85-95
[11] Liu G M, Che J G and Lei C C 2009 *Journal of Sichuan Ordnance* **30**(3) 44-5 *(in Chinese)*
[12] Santoso T, Ahmed S, Goetschlackx M, Shapiro A 2005 *European Journal of Operational Research* **167**(1) 96-115.
[13] Xu X, Li R W and Wu X L 2011 *Modern manufacturing engineering* **8** 62-9*(in Chinese)*
[14] Khalili-Dambghani K, Tavana M, Amirkhan M 2014 *International Journal of Advanced Manufacturing Technology* **73**(9-12) 1567-95
[15] Verma A, Seth N 2011 *International journal of Human and Social Sciences* 6(1) 5-10
[16] Zhang X, Zhao Q H and Xia G P 2012 *Journal of Convergence Information Technology* **7**(1) 45-53.
[17] Longinidis P and Georgiadis M C 2014 *Omega-International Journal of Management Science* **47** 73-89
[18] Guo J L 2006 *Acta Physica Sinica* 55(8), 3916-3920 *(in Chinese)*
[19] Li X L 2012 *Advances in Information Sciences and Service Sciences* **4**(8) 389-96
[20] Albert R and Barabási A L 2002 *Reviews of Modern Physics* **74** 47-97
[21] Lin J, Wang G, Hu Z J, Long Q Q 2012 *International Journal of digital content technology and its application* **6**(5) 225-34
[22] Xu X, Li R W, Wu X L and Liu H X 2012 *China Mechanical Engineering* 23(8) 941-6 *(in Chinese)*
[23] Huang E and Goetschalckx M 2014 *European Journal of Operational Research* **237**(2) 508-18

Authors

Xing Xu, born in March 1979, Hangzhou, P.R. China.

Current position, grades: associate professor.
University studies: PhD degree in mechanical design and theory at Zhejiang Sci-Tech University, P,R. China.
Scientific interests: supply chain management, system optimization and scheduling.
Publications: 9.

Yun Zhao, born in February 1981, Hangzhou, P.R. China.

Current position, grades: lecturer.
University studies: PhD degree in agricultural electrification and automation from Zhejiang University, P.R. China.
Scientific interest: signal processing, Hyperspectral data acquisition and processing.
Publications: 9.

Jia Hu, born in July 1977, Shenyang, P. R. China.

Current position, grades: scientific assistant.
University studies: University Duisburg-Essen.
Scientific interests: risk management, dynamic optimization of transportation.
Publications: 1.

Xinli Wu, born in May 1987, Hangzhou, P.R. China.

Current position, grades: assistant.
University studies: Master degree in mechanical engineer and automation major from Zhejiang Sci-Tech University, P,R. China.
Scientific interests: supply chain management.
Publications: 4.

Housing price forecast based on rough-set extreme learning machine

Benxue Wang*

Institute for Economic and Social Development, Quzhou University, Zhejiang, 324000, China

Received 1 March 2014, www.cmnt.lv

Abstract

The work, based on various factors affecting housing price in 31 provinces cities as research object, firstly adopted rough set theory to reduce those factors. Then, the main reduced influence factors were used as the input of extreme learning machine. On such basis, the housing price forecast model based on rough-set extreme learning machine was ultimately established. According to the simulation results, the algorithm in this work has good prediction effect, and its prediction precision is higher than that of BP neural network and RBF neural network. Therefore, this algorithm, with a certain practical and theoretical value, can be promoted to other areas for predication and classification.

Keywords: rough set theory, neural network, extreme learning machine, real estate, forecast error

1 Introduction

At present, Chinese real estate industry is booming rapidly, which is beneficial to the sustainable development of current economy and the increase of GDP. However, housing price has become the focus of government and the burden of people's life due to the high price of current commodity house. Too-high housing price becomes a huge threat to social stability and harmony. Therefore, studying the influence factors of housing prices contributes to the formulation and implementation of policies, with positive significance to the control of housing price.

Many famous experts and scholars have conducted researches on housing price. In 2002, Qiao Zhimin and his team did some researches on the fluctuation of housing price, finding out that the fluctuation of production cost affected the fluctuation of real estate price. In 2004, Ju Hong analyzed the influence of the holding cost and production cost on the price of commodity house. In the same year, Ping Xinqiao and Chen Minyan empirically analyzed the relationship of real estate investment, commodity house sale price, land price and government credit functions.

In 2004, Shen Yue and Liu Hongyu, adopting sample regression and city annul dummy variable, put forward the hypothesis that commodity house was not in conformity with the efficient market in China.

In 2007, Duan Zhongdong analyzed the relationship between real estate price and inflation output based on real estate data of China, discovering that inflation output contributed to the rising of housing price.

In 2012, Peter Wallstrom and his team applied support vector machine algorithm to real estate price evaluation, and the simulation effect was superior to the effect of other methods.

In 2013, W.K.Wong,Z.X.Guo and some other experts combined extreme learning machine with intelligent algorithm, and put forward housing price forecast model based on genetic-algorithm extreme learning machine.

2 Combined algorithm of rough set and extreme learning machine

2.1 ALGORITHM IDEA

The work combined rough set with neural network due to their strong advantage complementary. Then, extreme learning machine neural network was used to forecast housing price after the reduction of influence factors of housing price. Following that, a comparison analysis was conducted to extreme learning machine neural network and BP neural network and RBF neural network.

After the analysis of sample data, an initial information table was generated according to the known domain knowledge. Then, an appropriate discrete method was used to discretize continuous features, and rough set theory was adopted to achieve a quick feature reduction for the data. The reduced features were considered as input layer neurons. In the end, extreme learning machine neural network was applied to train and forecast the processed data. The process is shown as Figure 1.

* ***Corresponding author's*** e-mail: wbxdx0516@sina.com

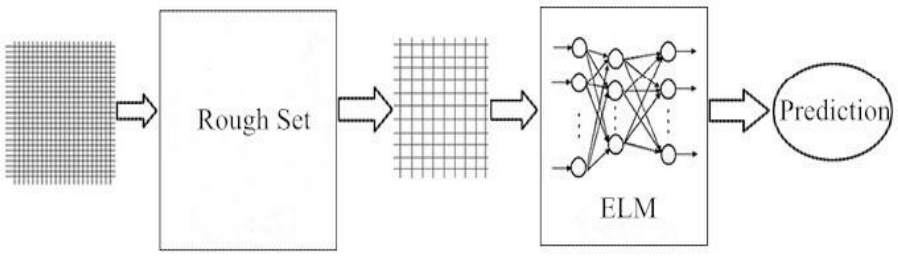


FIGURE 1 Process Diagram

2.2 ALGORITHM PROCESS

1) Discretization of Continuous Features: Continuous variables should be discretized before applying rough set method to data analysis. In essence, discretization is to divide N-dimensional space formed by condition feature into limited set of regions through selected breaking point, thus obtaining the same decision value of objects in each region.
2) Formation of Decision Tables: Quantized condition features and decision features were used to make a two-dimensional table with each line describing one object and each column showing one feature of corresponding object.
3) Reduction of Features: The reduction process of decision table feature is to eliminate unnecessary condition features of decision table system and figure out the decision principles of condition features in reduction to decision features.

Input: condition feature set $C = \{Y_{11}, Y_{12},\ldots, Y_{53}\}$, decision feature set $D = \{d\}$; output: reduction set of one feature REDU.

Step 1: Figure out the D positive domain $POSC$ (D) for condition feature C;

Step 2: As for feature $Y_{ij} \in C$, calculate and eliminate the D positive domain $POSC\backslash\{Y_{ij}\}$ (D) of its condition features subset $C\backslash\{Y_{ij}\}$;

Step 3: If $POSC\backslash\{Y_{ij}\}(D) = POSC$ (D), then feature Y_{ij} is unnecessary for decision feature D. Thus, it means $C = C\backslash\{Y_{ij}\}$ at this moment.

Go back to step 2; otherwise, the reduction of output feature REDU = C.

4) Reduction of Object: Eliminate inconsistent and redundant objects in data. Inconsistent objects are those with same condition feature but different decision features. Redundant objects are those with same condition features and same decision features.
5) Determination of Neural Network Model: The work adopted extreme learning machine neural network.
6) Network Learning and Testing: Choose corresponding training data and features from continuous features decision table according to neural network model input to train network and test it with corresponding test samples.

2.3 EXTREME LEARNING MACHINE NEURAL NETWORK

For N different samples $(\boldsymbol{x}_i, \boldsymbol{t}_i)$ where

$\boldsymbol{x}_i = [x_{i1}, x_{i2},..., x_{in}]^T \in \mathrm{R}^n$ and $\boldsymbol{t}_i = [t_{i1}, t_{i2},..., t_{im}]^T \in \mathrm{R}^m$, when the node number in one hidden layer is $\tilde{N}$, the unified SLFN model of activation function $g(x)$ will be:

$$\sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g_i(\boldsymbol{x}_j) = \sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g(\boldsymbol{a}_i \cdot \boldsymbol{x}_j + b_i) = \boldsymbol{t}_j, \quad j = 1,...., N, \quad (1)$$

where $\boldsymbol{a}_i = [a_{i1}, a_{i2},..., a_{in}]^T$ refers to the input weight of node connecting the i-th hidden layer; b_i the bias of i hidden layer nodes; $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2},..., \beta_{im}]^T$ the output weight of node connecting the i-th hidden layer; $\boldsymbol{a}_i \cdot \boldsymbol{x}_j$ the inner product of $\boldsymbol{a}_i$ and $\boldsymbol{x}_j$. The activation function can be "Sigmoid", "Sine" or "RBF".

The matrix form of N above functions can be: $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$, where

$$\mathbf{H}(\boldsymbol{a}_1,...,\boldsymbol{a}_{\tilde{N}}, b_1,...,b_{\tilde{N}}, \boldsymbol{x}_1,..., \boldsymbol{x}_N) = \begin{bmatrix} g(\boldsymbol{a}_1 \cdot \boldsymbol{x}_1 + b_1) & \cdots & g(\boldsymbol{a}_{\tilde{N}} \cdot \boldsymbol{x}_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(\boldsymbol{a}_1 \cdot \boldsymbol{x}_N + b_1) & \cdots & g(\boldsymbol{a}_{\tilde{N}} \cdot \boldsymbol{x}_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m}, \quad \mathbf{T} = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_N^T \end{bmatrix}_{N \times m}, \quad (2)$$

$E(\mathbf{W})$ is the square sum of error between expected value and actual value. The problem solution is to figure out the optimal weight $\mathbf{W} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta})$, making cost function $E(\mathbf{W})$ minimum. Its mathematical model can be expressed as:

$$\begin{aligned} &\underset{\mathbf{W}=(\boldsymbol{a},\boldsymbol{b},\boldsymbol{\beta})}{argmin}\, E(\mathbf{W}) = \underset{\mathbf{W}=(\boldsymbol{a},\boldsymbol{b},\boldsymbol{\beta})}{argmin} \| \boldsymbol{\varepsilon} \|^2 \\ &s.t. \sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g(\boldsymbol{a}_i \cdot \boldsymbol{x}_j + b_i) - \boldsymbol{t}_j = \boldsymbol{\varepsilon}_j, \quad j = 1,..., N \end{aligned}, \quad (3)$$

where $\boldsymbol{\varepsilon}_j = [\varepsilon_{j1}, \varepsilon_{j2},..., \varepsilon_{jm}]$ is the error of the j-th sample.

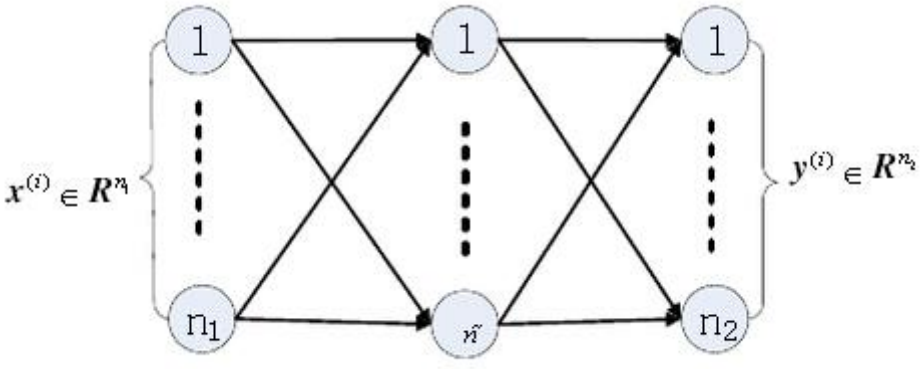


FIGURE 2 Extreme Learning Machine Schematic Diagram

3 Data sources

Data in this work, from *China Statistical Yearbook 2012*, cover all influence factors of housing price in 31 provinces, including building completion cost, land purchase expense, per capita disposable income, per capita GDP, population density, construction cost, consumer price index, land transaction price index and some other indexes.

There are various factors affecting housing price. Thus, this work firstly reduced those factors through rough set theory. Then, the reduced influence factors as the input and housing price as the output of extreme learning machine were applied to establish training model and forecast housing price. There are two sets of samples to validate the predication ability of extreme learning machine network: 31 training samples and 5 testing samples.

After rough-set reduction, the main factors affecting housing price refer to building completion cost, land purchase expense, per capita disposable income, per capita GDP and population density.

Figures 3-5 showed the predication results of extreme learning machine. Figure 3 represented the housing price forecast result of training samples, indicating a good predication and strong generalization ability. Figure 4 showed the forecast results of testing samples, mainly validating the effectiveness and correctness of extreme learning machine. It could be seen from the predication-error graph that the algorithm in this work has good prediction effect.

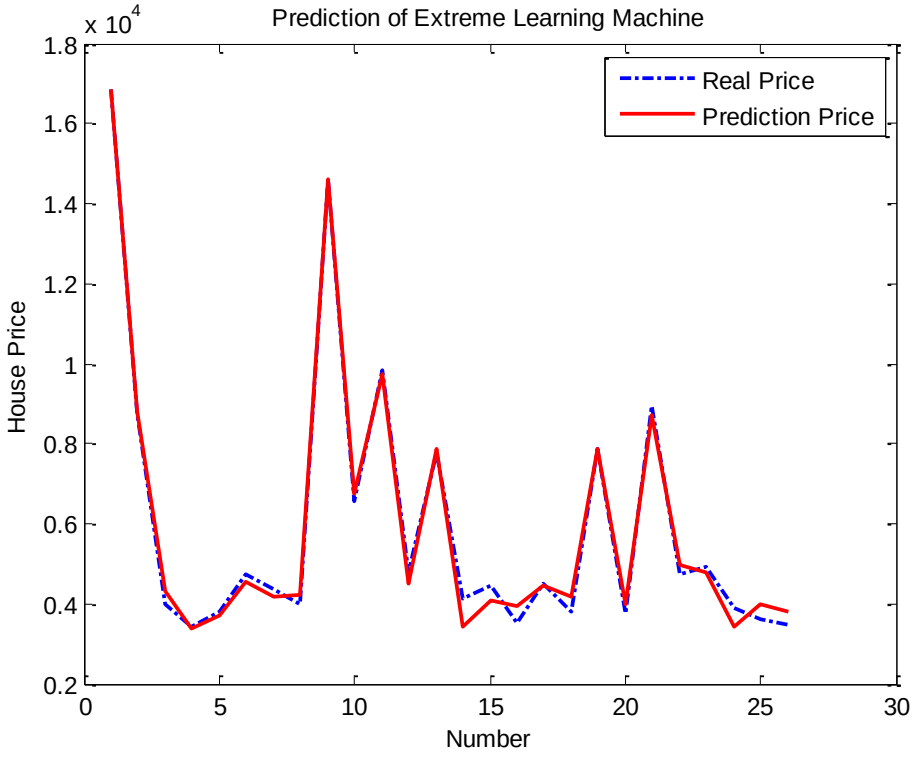


FIGURE 3 Housing Price Forecast Results of Training Samples Based on Extreme Learning Machine

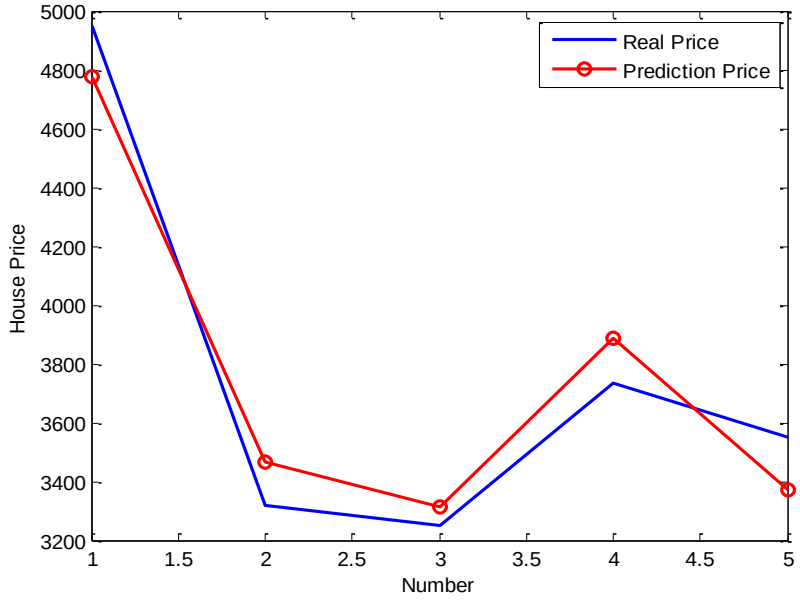
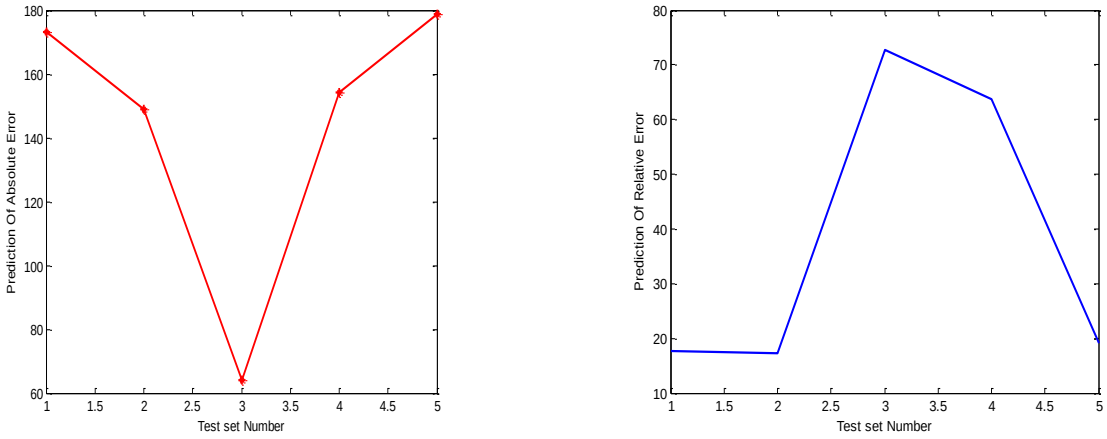


FIGURE 4 Forecast Results of Testing Samples Based on Extreme Learning Machine



a) Absolute Errors b) Relative Errors

FIGURE 5 Predication Errors

The predication effects of extreme learning machine and that of BP neural network and RBF neural network were compared to further prove the superiority of the algorithm. Figure 6 and 7 showed the comparison results. It can be seen that extreme learning machine has the best prediction results; RBF neural network has relatively inferior prediction results; BP neural network has the worst. The comparison of these different algorithms can effectively validate the effectiveness and correctness of the algorithm in this work. Therefore, this algorithm, with certain practical application value, can be applied to classify and predict problems.

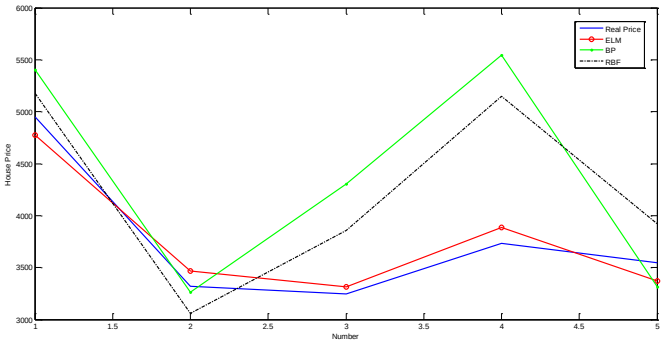


FIGURE 6 Comparison of the Predication Results of ELM, BP and RBF

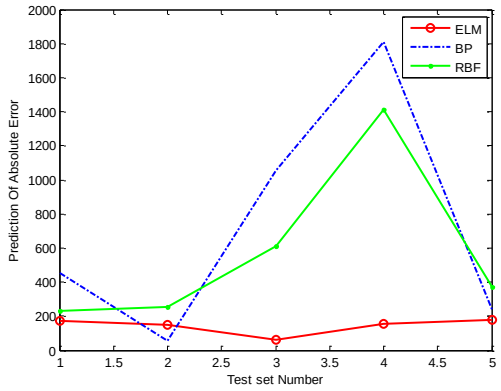


FIGURE 7 Comparison of the Predication Error of ELM, BP and RBF

4 Conclusions and analysis

With the combination of rough set theory and extreme learning machine, the work firstly reduced factors affecting housing price through rough set theory, thus obtaining the major influence factors. Then, the reduced influence factors (as the input of extreme learning machine) and housing price (as data) were used to establish the model to forecast housing price. Besides, The comparison of different algorithms – ELM, BP and RBF – proves the effectiveness and correctness of the algorithm in this work. Therefore, this algorithm, with certain practical application value, can be applied to the classification and predication of problems/

References

[24] Huang F 2007 Application of SPSS Correlation Analysis and Linear Regression Analysis *Foreign Language Teaching and Research* (10) 52-3 *(in Chinese)*
[25] Shen Y, Liu H 2004 Empirical Analysis of 1995-2002 Chinese Housing Price and Economy in 14 Cities *Economic Research Journal* (6) 42-9 *(in Chinese)*
[26] Shi W, Zhang K 2009 Political Analysis of Influence Factor of Housing Price in China *Northern Economy* (12) 25-7 *(in Chinese)*
[27] Yan J 2006 Housing Price and Land Price in China: Theoretical, Political and Policy Analysis *The Journal of Quantitative & Technical Economic*, (1) 17-26 *(in Chinese)*
[28] Xu H, Tang Z, Ba S 2010 Empirical Analysis of Periodical Fluctuation of Rising Housing Price in China *Journal of Shanghai University of Finance and Economics* **12**(4) *(in Chinese)*

Author

Benxue Wang, born in August 1969, Enshi, Hubei Province, China

Current position, grades: associate professor.
University studies: Master's degree in computer science and technology.
Scientific interests: regional economic research and computer software.
Publications: 22 papers.

Prediction of employment figures in the three main industries based on extreme learning machine

Xiaoyan Li*, Min Zhang

Xi'an Technological University, Shaanxi, 710032, China

Received 1 March 2014, www.cmnt.lv

Abstract

Extreme learning machine, with a fast speed of training, achieves globally optimal solutions and excellent generalization ability. This work is based on the production value and employment figure of the three main industries in China during 1996-2012 as research objects. An ELM employment figure prediction model was constructed with production value and the employment figure of the three main industries respectively as the input and output of extreme learning machine. Simulation test results proved a good effect and high accuracy of ELM employment figure prediction model. Besides, the comparison of ELM, BP and RBF algorithms further proved the effectiveness and precision of ELM, which has certain practical application value.

Keywords: three main industries, extreme learning machine, production value, neural network, training samples, test samples

1 Introduction

With the rapid development of economy and continuous upgrading of industrial structure in China, the total employment and structure of labor force have experienced dramatic changes. Under the combined effects of incoming labor force and transfer of laid-off labor and rural labor, employment in China becomes an increasingly severe problem.

Since there are some interconnection influences between industrial structure and employment, additional investment to some industries can directly create large amount of employment opportunities. Additional investments to some industries may not create more employment opportunities, but it can directly or indirectly influence consumption of products in other industries, thus providing employment opportunities. Therefore, the employment opportunities of the whole economic system will be increased accordingly [1].

Adjusting industrial structure is an important way to realize optimization of internal structure of economic system, solving employment problems and increasing employment opportunities in each industry. Therefore, researches on the potential of employment in the three major industries are of great significances to employment.

In recent decades, China has achieved high-speed growth in economy, but employment growth is relatively slow. At present, rural surplus labor and college graduates become main employment forces. They are faced with various employment pressures; some of them are even faced with unemployment. As a result, Chinese people are facing big pressure from employment at present. Besides, the total labor force exceeds the demands of employment, but some places are in a shortage of labors. Therefore, employment structural contradictions become increasingly fierce [2-3].

With huge employment pressure, it is of important theory value and practical significance to broaden visions and fully consider national situations while absorbing employment theory from western economics researches. Only on such basis can prediction of employment figure be achieved effectively.

2 Extreme learning machine neural network

It is clear that the learning speed of feed forward neural networks is in general far slower than required and it has been a major bottleneck in their applications for past decades. Two key reasons behind may be: (1) the slow gradient based learning algorithms are extensively used to train neural networks, and (2) all the parameters of the networks are tuned iteratively by using such learning algorithms. Unlike the traditional implementations, we applied a new learning algorithm called extreme learning machine (ELM) for SLFNs which randomly chooses the input weights and analytically determines the output weights of SLFNs. This algorithm tends to provide the best generalization performance at extremely fast learning speed.

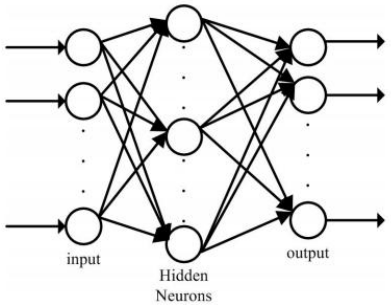


FIGURE 1 Extreme Learning Machine Schematic Diagram

* ***Corresponding author's*** e-mail: xiaoyanlixa@126.com

As pointed out by [4-9], for feed forward networks with many small weights but small squared error on the training examples, the Vapnik-Chervonenkis (VC) dimension (and hence number of parameters) is irrelevant to the generalization performance. Instead, the magnitude of the weights in the network is more important. The smaller the weights are, the better generalization performance the network tends to have. As analyzed above, our method not only reaches the smallest squared error on the training examples but also obtains the smallest weights. Thus, it is reasonable to think that our method tends to have better generalization performance. It should be worth pointing out that it may be difficult for gradient-based learning algorithms like back-propagation to reach the best generalization performance since they only try to obtain the smallest training errors without considering the magnitude of the weights.

3 Data sources and variable selection

In order to predict employment figure in the three main industries in China, this work selected the production value and employment data during 1996~2012 from each industry as samples (Table 1). Employment figure of each industry was considered as explanatory variable L_i, GDP of each industry as the explained variable Y_i. All data comes from *1996~2012 China Statistical Yearbook* [10,11].

TABLE 1 Production value and employment figure of each industry

Year	Y_1 / 100 Million RMB	L_1 / 10 Thousand People	Y_2 / 100 Million RMB	L_2/ 10 Thousand People	Y_3 / 100 Million RMB	L_3 / 10 Thousand People
1996	5866.6	38349	11699.5	14226	9357.3765	12979
1997	6963.7629	37434	16454.431	14868	11915.731	14071
1998	9572.6948	36489	22445.399	15254	16179.763	15456
1999	12135.811	35468	28679.458	15628	19978.46	16851
2000	14015.39	34769	33834.959	16180	23326.243	17901
2001	14441.886	34730	37543.002	16495	26988.147	18375
2002	14817.626	34838	39004.189	16440	30580.466	18679
2003	14770.028	35364	41033.582	16235	33873.445	18987
2004	14944.723	36043	45555.878	16217	38713.954	19823
2005	15781.269	36513	49512.291	16284	44361.611	20228
2006	16537.02	36870	53896.768	15780	49898.902	21090
2007	17381.718	36546	62436.312	16077	56004.726	21809
2008	21412.734	35269	73904.312	16920	64561.292	23011
2009	22420	33970	87364.579	18084	73432.866	23771
2010	24040	32561	103162	19226	84721.4	24614
2011	28627	31444	124799	20629	103879.59	24917
2012	34000	30654	146183.4	21109	120486.61	25717

Y_1 refers to the production value of the first industry; Y_2 the production value of the second industry; Y_3 the production value of the third industry. L_1 is the employment figure of the first industry; L_2 the employment figure of the second industry; L_3 the employment figure of the third industry.

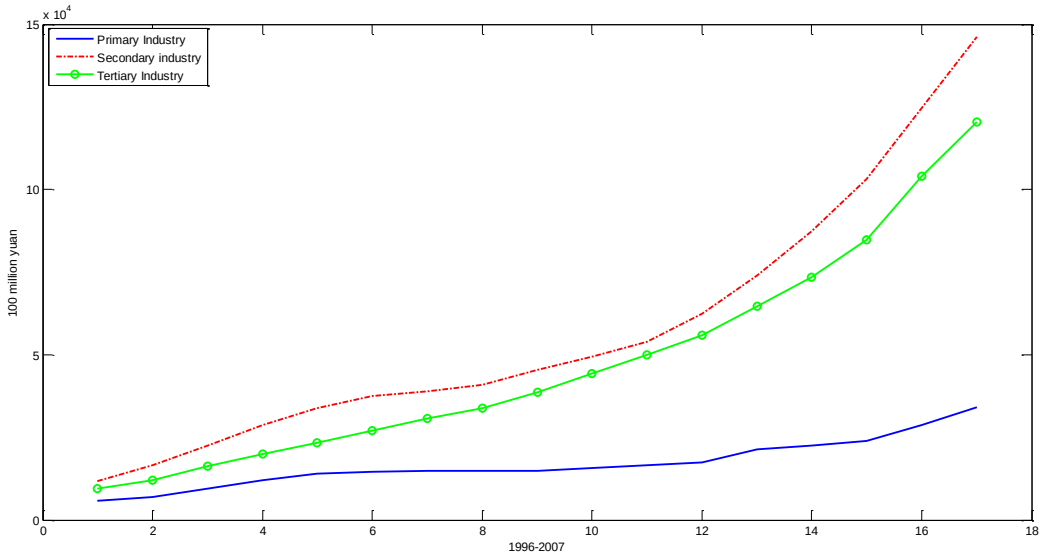


FIGURE 2 Change curve of production value in each industry over the year

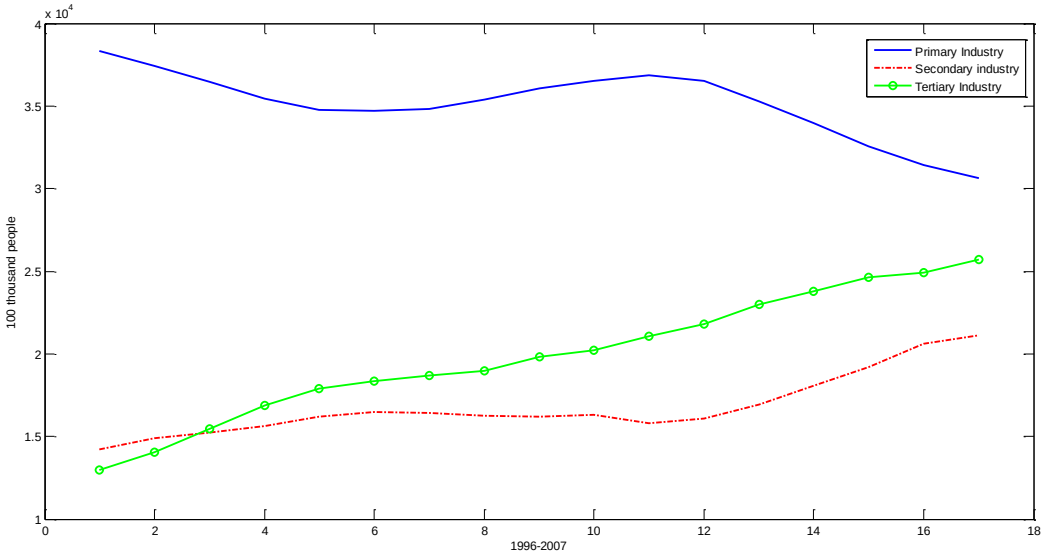
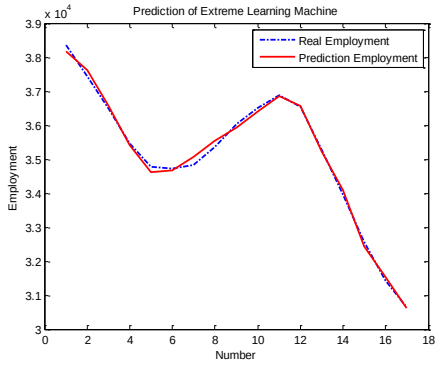


FIGURE 3 Change curve of employment figure in each industry over the year

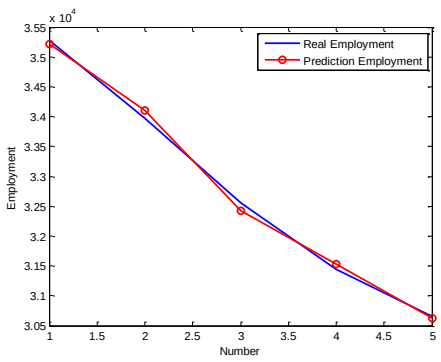
Figures 2 and 3 show that production values of the 2nd and 3rd industries increased year by year since 1996. And, the growth rate of the 2nd industry was greater than that of the 3rd industry. On the contrary, employment figure of the 1st industry declined year by year, implying that the 1st industry's allocation of labor ability was getting weaker. However, employment figures of the 2nd and 3rd industries also increased year by year. It means employment absorption capacity of the 2nd and 3rd industries is far greater than that of the 1st industry.

4 Employment figure prediction based on extreme learning machine

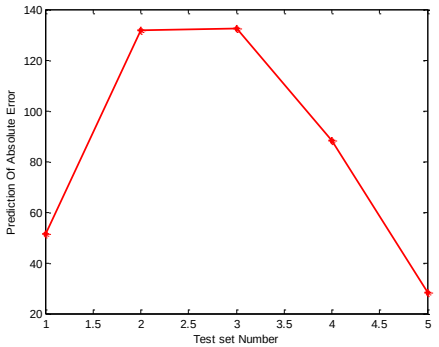
With production value and employment figure of the three main industries, respectively, as the input and output of extreme learning machine, a training model was established to predict employment figure. There were two sets of samples: 17 training samples and 5 test samples, all verifying the prediction ability of extreme learning machine network.



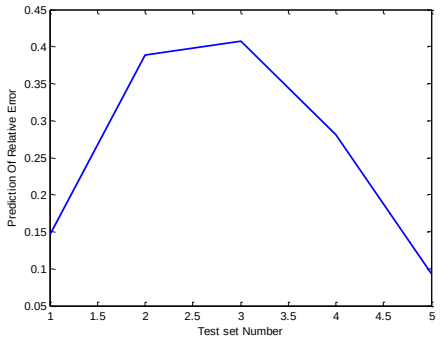
a) Prediction results of training samples



b) Prediction results of test samples

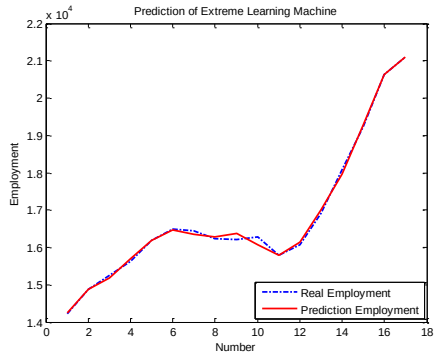


c) Absolute Error

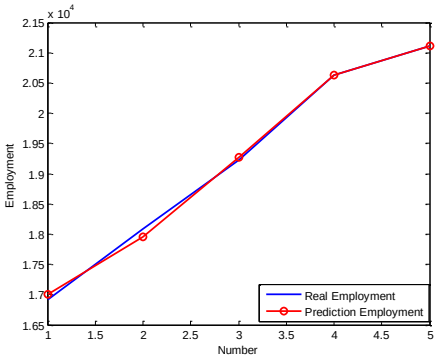


d) Relative Error

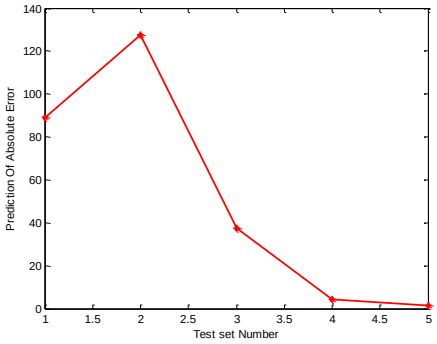
FIGURE 4 Prediction results and errors of employment figure of the 1st industry



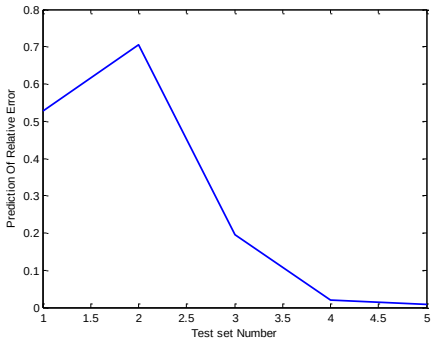
a) Prediction results of training samples



b) Prediction results of test samples

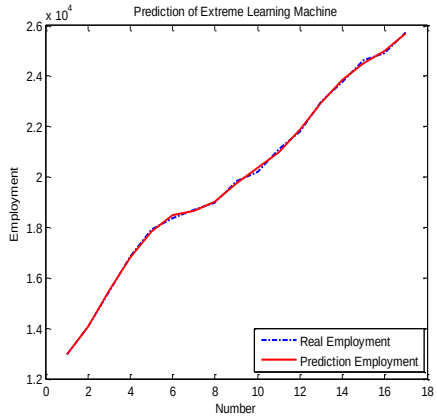


c) Absolute error

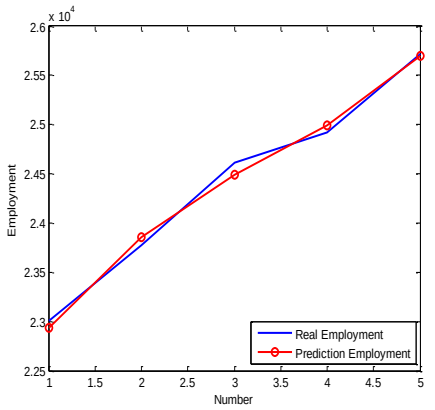


d) Relative error

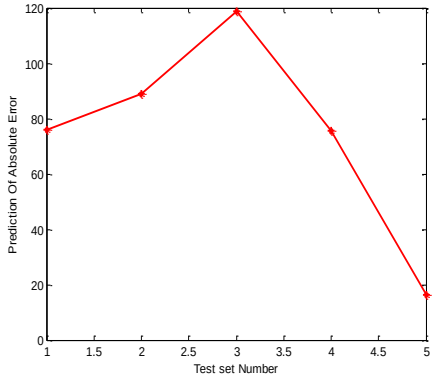
FIGURE 5 Prediction results and errors of employment figure of the 2nd Industry



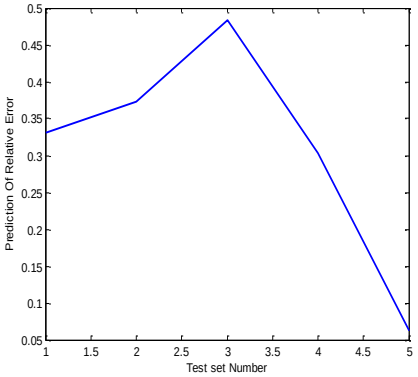
a) Prediction results of training samples



b) Prediction results of test samples



c) Absolute error



d) Relative error

FIGURE 6 Prediction results and errors of employment figure of the 3rd industry

Figure 3-5 respectively showed the prediction results of extreme learning machine. Figure 3a-5a, explained prediction results of training samples, presenting a good prediction and strong generalization ability. Figure 3 (b), Figure 4 (b) and Figure 5 (b) presented prediction results of test samples, mainly verifying the effectiveness and accuracy of extreme learning machine. It can be seen from Figure 3 (c) (d), Figure 4 (c) (d) and Figure 5 (c) (d) that this algorithm achieves good effects.

In order to compare and verify the superiority of this algorithm, the prediction results of ELM, BP neural network and RBF neural network were compared. It can be seen that ELM has the best prediction results; RBF neural has a worse prediction results compared with ELM; BP neural network has the worst. Comparison of those different algorithms verified the effectiveness and accuracy of ELM. Therefore, extreme learning machine, with certain practical application significance, can be used to classify and predict problems.

5 Conclusions

Based on production value and employment figure of the three main industries during 1996-2012 in China as research objects, this work established ELM employment figure prediction model. It was used to predict employment figures when production value and employment figure of the three main industries were respectively used as output and input of extreme learning machine. Comparison of ELM, BP and RBF further proved the effectiveness and high accuracy of ELM. Thus, this algorithm, with certain practical application value, can be used to classify and forecast problems.

References

[29] Wong W K, Guo Z X 2013 A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm *International Journal of Production Economics* 614-24

[30] Wallstrom P, Segeratedt A 2012 Evaluation of forecasting error measurements and techniques for intermittent demand *International Journal of production Economics* 625-36

[31] Collobert R, Bengio S, Bengio Y 2002 A parallel mixtures of SVMs for very large scale problems *Neural Computation* **14**(5) 1105-14

[32] Suykens J A K, De Brabanter J, Lukas L, Vandewaile J 2002 Weighted least squares support vector machines: Robustness and sparse approximation *Neurocomputing* **48**(1) 85-105

[33] Keerthi S S, Shevade S K 2011 SMO algorithm for least squares SVM formulations *Neural Computation* **15**(2) 487-507

[34] Hsu C W, Lin C J 2012 *IEEE Transaction on Neural Networks* **13**(2) 415-25

[35] Lei Y J, Zhang S W, Li X W, Zhou C M 2005 MATLAB Genetic Algorithms Toolbox and Its Applications *Xian China* Xidian University Press *(in Chinese)*

[36] Liang Z, Yang K, Sun Y 2006 Decision support for choice optimal power generation projects: Fuzzy comprehensive evaluation model based on the electricity market *Energy Policy* 34 3359−64

[37] Liu Z M, Wu M F, Xu Y 2006 A method of weight conformation based on genetic algorithm *Journal of Wuyi University (Natural Science Edition)* **20**(3) 45−8 *(in Chinese)*

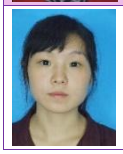
[38] Yu S, Yu M 2007 Fuzzy partial credit scaling: A valid approach for scoring the Beck Depression Inventory *Social Behavior and Personality* **35**(9) 1163−72

[39] Zhou M, Sun S D 1999 Genetic Algorithms: Theory and applications *Beijing China* National Defense Industry Press *(in Chinese)*

Authors

Xiaoyan Li, born in August 1973, Heyang, Shaanxi Province, China.

Current position, grades: master's degree associate professor at in Xi'an Technological University, Shaanxi, China.
Scientific interests: human resource management, mental health education.
Publications: 20 papers.



Min Zhang, born in September 1989, Zhumadian, Henan Province, China.

University studies: master's degree at Xi'an Technological University, Shaanxi, China.
Scientific interests: human resource management, modern Chinese history.
Publications: 5 journals.

Research on quality information integrated management for complex precision parts in multi-varieties and small batch manufacturing mode

Lihua Yu*, Qingsheng Xie

Key Laboratory of Advanced Manufacturing Technology, Guizhou University, China

Received 1 March 2014, www.cmnt.lv

Abstract

There were many problems such as quality information interaction low, real-time monitoring and tracing of production quality difficulty in workshops for multi-varieties and small-batch complicated precision parts. These problems restricted seriously the improvement of product quality. Therefore, a quality information integrated operation mode was built, combining with advanced management thoughts of lean management and integration. The main characteristics of quality information integrated management process were analysed. At last, the mode was applied in a manufacturing enterprise in CASTC 061 base for testing its effects. Results show that it is successful and satisfactory in practical application effect.

Keywords: workshop, integration, multi-varieties and small-batch, quality control

1 Introduction

With more and more various customer requirements and rapid product renewal, the production mode of manufacturer has been gradually changing from the traditional mass production into multi-varieties and small-batch production. More stringent requirements on quality control must be imposed for complex precision parts [1-4]. How to achieve quality information integrated management has become an urgent problem for the manufacturing enterprises.

A lot of research on quality information integration technology has achieved fruitful results at home and abroad. Gen-bao Zhang proposed a dynamic quality tracking model to achieve dynamic tracking and traceability of product quality in the manufacturing execution processes based on Computer Aided Process Planning (CAPP) and Manufacturing Execution System (MES)[5]. Liang Tong presented a function interoperation method of multiple heterogeneous systems in discrete manufacturing process according to problems [6]. Hui-ying Qi Proposed an integrated system scheme to realize uniform access of distributed and heterogeneous network resources [7].

In summary, domestic and foreign scholars have made many valuable results from the viewpoint of integration technology. However, research on quality information management is still relatively scarce from integration operation between departments and workshop. Therefore, the paper researched deeply the problem on quality information integrated management in the workshop.

2 Establishment of the quality information integrated operation mode

2.1 EXISTING PROBLEMS

Quality information has the characteristic of more sources, large volumes of data and frequent flow in workshops for multi-varieties and small-batch complex precision parts. Moreover, the traditional quality control mode often results in poor credibility, interactivity, utilization and traceability of quality information. Production quality cannot be controlled effectively. The main problems are as follows [8-16]:

1) Quality instruction information such as quality control tasks, CAD models and technology documentation was transmitted in hand-form in workshop, which made its adjustment and conversion inconvenient and often caused production confusion.
2) Dynamic quality information such as quality inspection, unqualified parts and equipment running information was collected in hand-form and recorded in paper-based way in workshop at present. The backward way made its consistency, confidence and systematization poor.
3) Various departments such as product design, materials, production and quality control department respectively managed their owner quality information. This kind of operating mode made quality information scattered, nonstandard, and isolated in the process relatively. Meanwhile, it became difficult to share, query and trace one for the workshop.
4) Quality information transmitting was by form and hand. This kind of ways not only made interactivity

* ***Corresponding author's*** e-mail: yulihua01@163.com

and utilization poor, but also showed abnormal quality-events not to be found in time and tackled efficiently. At the same time it was hard to monitor production quality effectively on line.

These problems above had not only seriously affected the comprehensive utilization and effects of quality information and reduced the quality control capacity, but also greatly restricted the improvement in quality management level.

2.1 THE GENERAL FRAMEWORK OF QUALITY INFORMATION INTEGRATED OPERATION MODE

According to quality information characteristics and management problems for this workshop, a quality information integrated operation mode was proposed. Its basic idea was that real-time interaction of quality information was extended to the workshop field, based on network multi-function interactive information terminal [17]. Quality information integrated operation has achieved among personnel, equipment, workshop scheduling, quality department, design department and enterprise senior by real-time transmission, adjustment and interactive feedback of quality information.

Based on the idea of establishing the operation mode, its general framework was constructed. As Figure 1 shown, it is mainly made up of five-level which is respectively objective, business execution, information interaction, status analysis and information collection level.

1) Objective level.
   In order to achieve operation mode above, according to quality information characteristics and management problems existed currently in workshop, an optimal operating objective set was established as follows:
   a) the collection efficiency, accuracy, interaction, comprehensive utilization and traceability for quality information was always expected,
   b) abnormal quality-events were processed as efficiently as possible,
   c) reject parts were as few as possible.
2) Business execution level.
   The achievement of operating objective needed the support for business activities under the operation mode. The departments related to business activities included design, production, quality, equipment and measuring instrument supplies department.
3) Information interaction level based on network multi-function interactive information terminals.
   Information interaction level described information interacting process of the operating mode. Network multi-function interactive information terminals were the key of information interaction. All quality information would eventually be pooled into interactive information terminals, which are from product and process design, production job scheduling arrangements, production information collection and quality management.

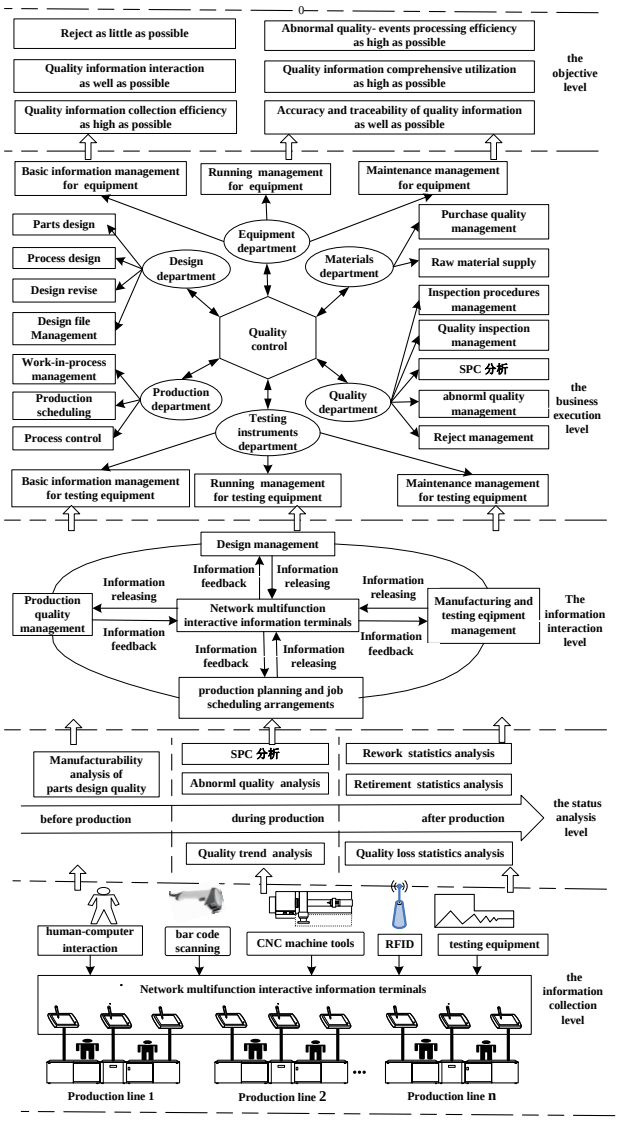


FIGURE 1 General framework of quality information integrated operation mode for workshop of multi-varieties and small-batch complex precision parts

4) Status analysis level.
   The status analysis level provided various analysis and decision-making tools of quality state for production quality control. The tools included as follow:
   a) manufacturability of parts design quality before production,
   b) SPC(Statistics process control)of key process, abnormal quality and quality trend analysis during production,
   c) rework, retirement and quality loss statistics analysis after production, and so on.
5) Information collection level.
   The information collection level obtained the quality-related information to support quality information integrated operation by various means during production in workshop. The information collecting was by means of testing equipment, bar code scanning, FRID, CNC machine tools, and human-computer interaction, and so on. All pro-

duction quality information collected were transferred and integrated into network multifunction interactive information terminals.

2.1 THE MAIN PROCESS FOR QUALITY INFORMATION INTEGRATED MANAGEMENT

The main purpose of quality information integrated management is to support the integrated operation of quality management activities. The management business includes the quality control plan made and issued, production quality information collected real time, quality inspection and control, statistical analysis and query for quality information, nonconforming product management, quality improvement, quality information feedback, etc. Quality information integrated management process is as shown in Figure 2.

The main characteristics of quality information integrated management process:

1) Based on the multi-function interactive information terminal, production workers and quality inspection personnel can receive quality control tasks, the machining process document and quality inspection procedures in real-time. They can collect and feedback quality data in real-time on site.
2) Based on the management terminal, the process management is realized easily. the process includes the disposition of nonconforming products, the process for scrap , the process for repair and so on. For another, the process management is integrated into the news system with a prompt, from "people work" into "thing for people". Then the efficiency of nonconforming products disposition is improved greatly.
3) Based on the management terminal, multi-function interactive information terminal and integrated interface, the organic connection of quality planning layer and workshop site and operation workers is established, Collaborative quality control capacity is improved through quality information interactive feedback and closed-loop control between departments.

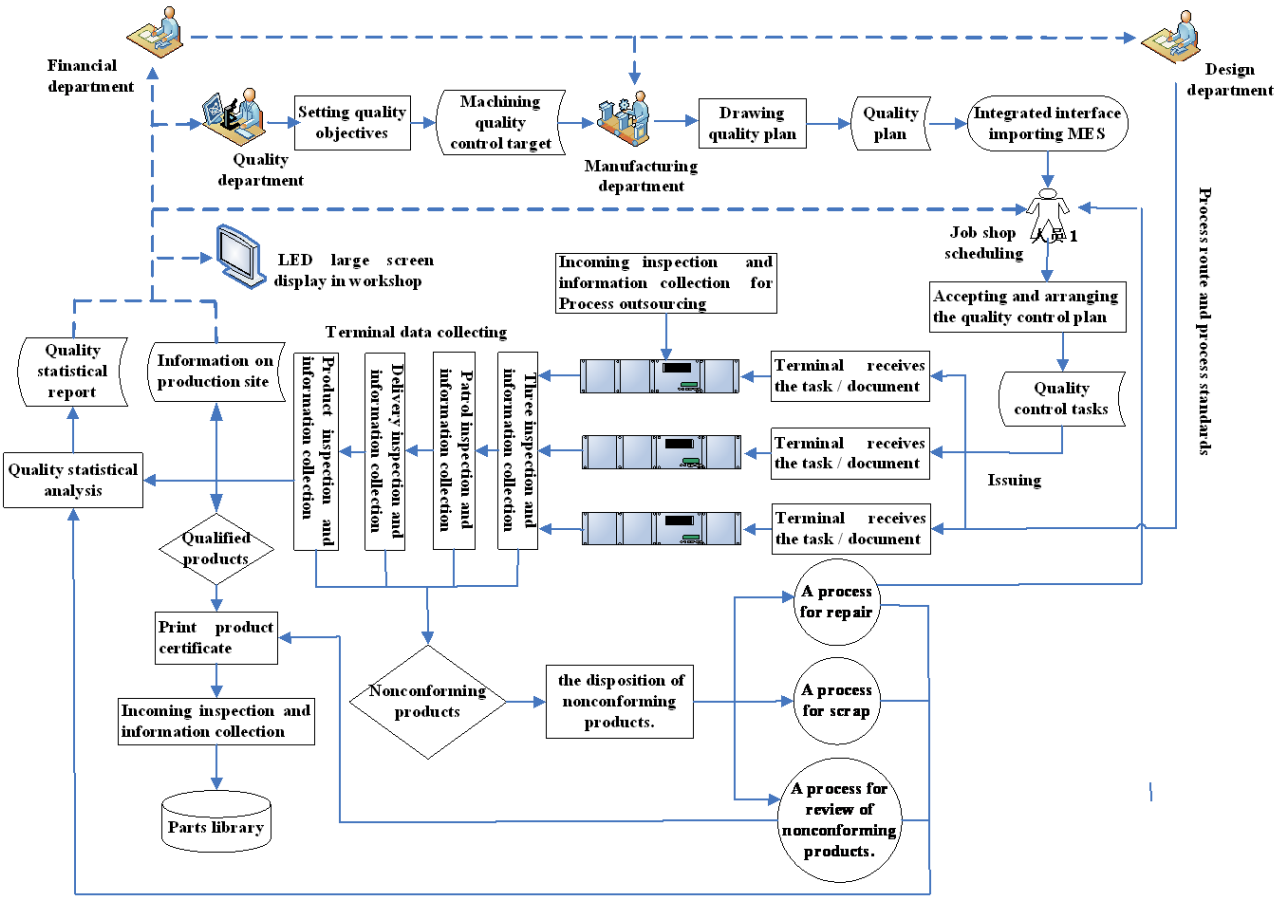


FIGURE 2 Main process of quality information integrated management for workshop of multi-varieties and small-batch complex precision parts

3 The application and practice of the mode

At present, the management mode above has been successfully applied in a manufacture enterprise in 061 manufacturing base of China Aerospace Science and Industry. The manufacturer is a large industrial enterprise, and has typical production mode for multi-varieties and small-batch complex precision parts. The characteristic of its workshop is diversified process rout and high-quality requirements, which makes information sources a lot, amount of quality information large and information flow frequent among departments. However the old mode had failed to meet the

confidence, interactive feedback and traceability requirements of quality information. So the manufacturer had implemented manufacturing process information systems lean production planning and manufacturing execution system (HTK-eMES). As a result, production management and manufacturing capabilities of the enterprise has improved effectively. Quality information integrated operation is an important part of the HTK-eMES in parts production workshop. Quality information integration and application is shown in Figure 3. The operating mode has achieved many processes integrated operation and improved the quality management. These achievements and improvements include as follows:

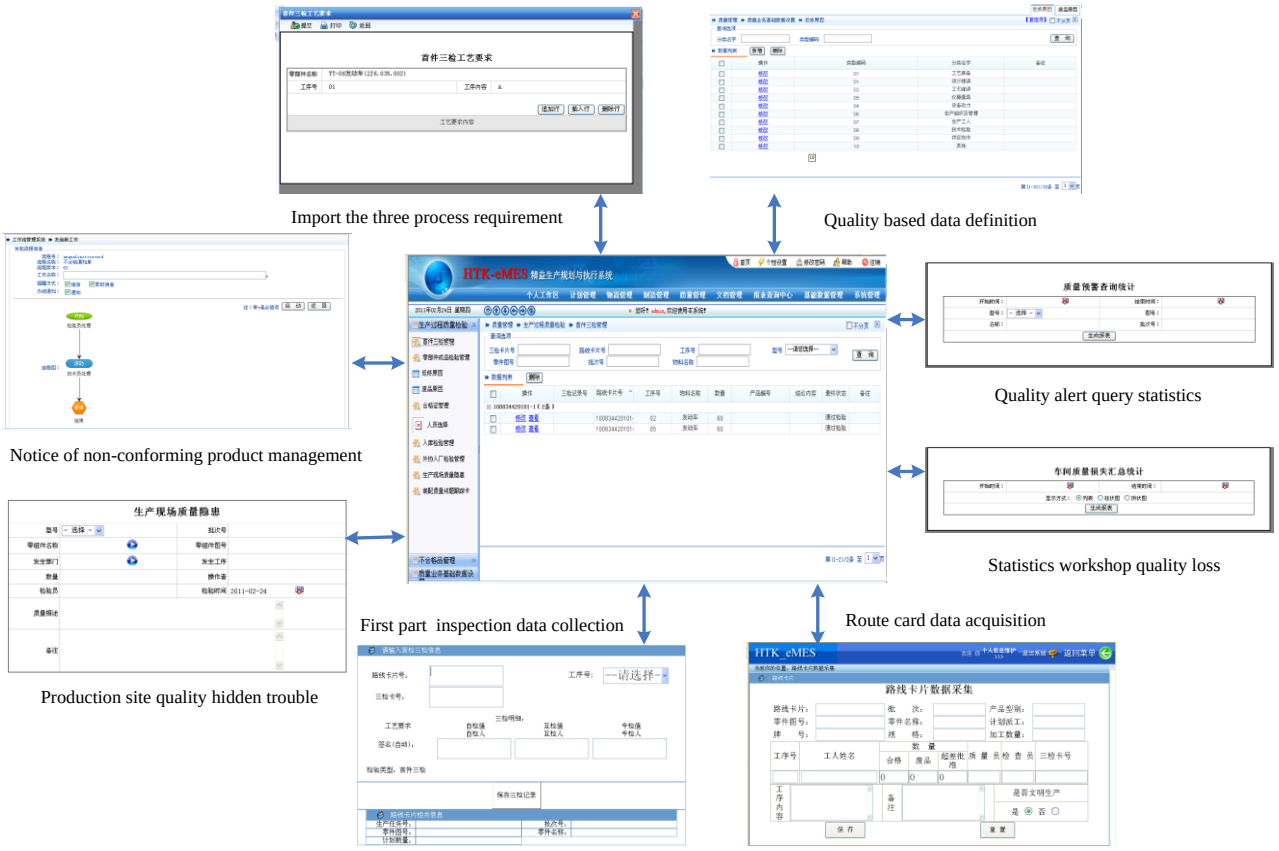


FIGURE 3 System integration and application diagram

1) The operating mode has achieved the efficient of quality information collection and improved its accuracy, real-time and systematic.
2) The operating mode has achieved the real-time, electronization of information transmission and improved information interaction and traceability.
3) The operating mode has achieved the real-time and hyalinization of quality control, and reduced reject and rework rate and quality loss.
4) The operating mode has achieved the interactive feedback of quality information and the close-loop operation of quality control, and improved the optimazed organization capacity of quality information management and cooperative quality control capacity.

Conclusions

There exits many problems on quality information management in workshops for multi-varieties and small-volume complex precision parts. These problems include low confidence, interactivity and utilization of quality information and difficulty in real-time monitoring effectively and tracing of production quality. In order to solve these problems, firstly a quality information integrated operating mode was proposed in this paper. Secondly, the general framework of the operating mode was build. And the main characteristics of quality information integrated management process were analysed. At last, the mode was applied in a manufacturing enterprise in CASTC 061 base for testing its effects. The operating mode gets many achievements including: realizing the accuracy, real-time and systematic of quality information collecting, achieving the real-time of information transmission, achieving the real-time and hyalinization of quality control, achieving the interactive feedback of quality information and the close-loop quality control. In a word, the mode had not only realized quality information integrated operation through whole quality information chain among departments, but also improved quality control efficient and reduced reject and rework rate and quality loss.

References

[40] Jiang X-y, Zhang Y, Zhao K 2008 Research on multi-type & small batch oriented process quality control system under network environment *Proceedings of the IEEE International Conference on Automation and Logistics, Qingdao, China* September 869-74 (*in Chinese*)

[41] Wu X-l, Li S-j 2009 Mass variety and small batch scheduling in the flexible job shop *Proceedings of the 2009 2nd International Conference on Biomedical Engineering and Informatics, BMEI 2009 Tianjin China* 1-7 (*in Chinese*)

[42] Wang S-l, Ren H-b, Cai B 2009 Automatic generation technology of the batch arrivals production rolling plan in multi-varieties and small-batch environment *Journal of Chongqing University* **32**(9) 1024-7 (*in Chinese*)

[43] Liu W-n, Zheng L-j 2010 RFID-based production operation management for multi-varieties and small-batch production *Proceedings of 2010 IEEE International Conference on RFID-Technology and Applications, RFID-TA* 1-6

[44] Zhang G-b, Ren X-l, Li M, et al. 2010 Dynamic quality traceable system based on MES and CAPP *Computer Integrated Manufacturing Systems* **16**(2) 349-55

[45] Tong L, Yan P, Liu F 2011 Function interoperation method of heterogeneous systems in discrete manufacturing process *Computer Integrated Manufacturing Systems* **17**(5) 971-9

[46] Qi H-y, Wang X 2010 Research on the realization of distributed and heterogeneous information resources integrated system *Journal of Harbin Institute of Technology* **42**(11) 1838-41 (*in Chinese*)

[47] Lu J 2008 Research and development and application on Workshop quality management system based on MES *Hangzhou Zhejiang University* (*in Chinese*)

[48] Zhang R-g 2008 Research on product quality management system oriented to machine shop *Shijiazhuang Hebei University of Technology* (*in Chinese*)

[49] Zhang G-f 2007 Application on quality management techniques in MES *Xiling: Northwestern University* (*in Chinese*)

[50] Zhang Y, Jiang P, Huang G Q 2012 Task-driven e-manufacturing resource configurable model *Journal of Intelligent Manufacturing* 23 1681–94

[51] Gao J, Yao Y, Zhu V C Y 2011 Service-oriented manufacturing: a new product pattern and manufacturing paradigm *Journal of Intelligent Manufacturing* 22 435–46

[52] Chiou R, Mookiah P, Kwon Y 2009 Manufacturing e-quality through integrated web-enabled computer vision and robotics *The International Journal of Advanced Manufacturing Technology* 43 720–30

[53] Yen C-T, Kao H-A, Wang S-M 2013 On-line quality inspection system for automotive component manufacturing process *Proceedings of the Institute of Industrial Engineers Asian Conference* 1031-8

[54] Yu Q-h, Zhu H-p, Yu H-c 2014 Design of the quality management system for manufacturing workshop based on the internet of things *Proceedings of 2013 4th International Asia Conference on Industrial Engineering and Management Innovation* 545-54

[55] Jin Z Research on solutions of cloud manufacturing in automotive industry *Proceedings of the FISITA 2012 World Automotive Congress Lecture Notes in Electrical Engineering Volume* 199 225-34

[56] Liu F, Yan P, He D-q 2003 Multi-function interactive information terminal in networked manufacture system *China, ZL02113585.1* 10-22 (*in Chinese*)

Authors

YU Lihua, born in November 1975, Tongren, China.

Current position: senior experimental division and PhD in Guizhou University.
University studies: mechanical design and manufacturing in Jiangsu University.
Scientific interest: manufacturing quality management, advanced manufacturing technology.
Publications: 1 patents, 10 papers, 1 book

XIE Qingsheng, born in October 1954, Guiyang, China.

Current position: doctor of Mechanical Engineering, professor and doctoral supervisor in Guizhou University.
University studies: railway vehicles in Southwest Jiao Tong University.
Scientific interest: advanced manufacturing mode and information system.
Publications: 2 patents, 50 papers, 3 books.

Incentive analysis of labour dispatch under asymmetric information theory

Yanli Yang*, Xianyu Wang

School of business, Sichuan University, Chengdu, 610000, Sichuan, China

Received 1 July 2014, www.cmnt.lv

Abstract

Under the circumstances of the fierce market competition, great importance is attached to labour dispatch as an innovative employment way since it saves more labour costs for enterprises. However, there exists obvious asymmetric information among the dispatched employees, employers and accepting entities. This paper makes an analysis of the incentive offered by accepting entities to the dispatched employees and employers under the circumstances of asymmetric information and symmetric information. It verifies through models that in the case of symmetric information accepting entities can offer incentive to the dispatched employees and employers so as to achieve Pareto optimality by means of linear contracts. In case of asymmetric information, the expected revenue of accepting entities will be influenced by the abilities, degree of risk aversion and effort costs of the dispatched employees and employers, which is of great reference value to the enterprise practice.

Keywords: Labour Dispatch, Symmetric Information, Asymmetric Information, Risk Aversion, Effort Costs

1 Introduction

As the present market competition increasingly intensifies, the profit margin gets lower and lower. In order to develop well, more and more enterprises opt to outsource their non-core business, which results in the business of labour dispatch. From the very beginning, labour dispatch has been valued greatly and widely applied in the practice of enterprises. The employment of Foreign Representative Office began to adopt the way of labour dispatch in China from 1979. In the 1990s, in order to tackle the employment of laid-off workers and migrant workers, labour dispatch experienced primary development. Since the promulgation of Labour Contract Law of the People's Republic of China in 2008, the business of labour dispatch saw a rapid development since enterprises tried to reduce the labour costs and avert non-fixed term contract. Up to now, the number of labour dispatch reaches about 10 million [1], widely spreading in the field of construction, telecommunications, banks, the power sector and public institutions.

Labour dispatch involves three parties of the dispatched employees, employers (the dispatching entities) and the accepting entities (the actual recruiting enterprises). In general, the dispatched employees and the employers have more information than accepting entities, such as the quality, credibility, professional competence, service level and the actual strength of the dispatched employees and the employers. In case of asymmetric information, the problem demanding prompt solution is how the accepting entities offer incentive to the dispatched employees and the employyers. From the perspective of symmetric information and asymmetric information, this paper probes into the incentive offered by accepting entities to the dispatched employees and employers. It verifies through models that in the case of symmetric information accepting entities can offer incentive to the dispatched employees and employers so as to achieve Pareto optimality by means of linear contracts. In case of asymmetric information, the expected revenue of accepting entities will be influenced by the abilities, degree of risk aversion and effort costs of the dispatched employees and employers, which is of great reference value to the enterprise practice.

2 Related literature

Labour dispatch is clearly stipulated by Labour Contract Law of the People's Republic of China. Labour dispatch is that employers sign the dispatching agreement with accepting entities in accordance with the demands of accepting entities and dispatch the employees with labour contract relations to accepting entities so as to establish a special labour relation that the dispatched employees provide labour under the management of accepting entities while the employers get dispatching payment from the accepting entities and pay the dispatched employees labour remuneration. The business of labour dispatch involves three parties and two contracts. Three parties refer to the dispatched employees, employers and accepting entities. Two contracts refer to the one between the dispatched employees and accepting entities and the other between employers and accepting entities. In general, when signing the dispatch agreement, the accepting entities cannot get a full understanding about the credibility, professional ability, service level and actual strength of the dispatched employees and employers. Therefore, there exists obvious asymmetric information among the three parties.

Current literature about the study of labour dispatch mainly focuses on qualitative analysis. Literature [2] to [4]

* ***Corresponding author's*** e-mail yydouble@sohu.com

mainly analyses the development and characteristics of labour dispatch. Literature [5] to [7] mainly elaborates the status quo and existing problems of labour dispatch in China and puts forward some suggestions. Literature [8] to [10] analyses the game relationship among the three parties of labour dispatch from a legal point of view. This paper makes a qualitative analysis of the incentive offered by accepting entities to the dispatched employees and employers.

3 Model description

The model involves three parties: the dispatched employees, the employers (the intermediary agent) and the accepting entities (the actual employing party). The dispatched employees and the employers sign the labour contract. The employers and the accepting entities sign the dispatching agreement. The income of the dispatched employee consists of regular wage plus performance bonus. The income of the employer consists of a fixed rate (which is related to the regular wage of the dispatched employee) plus performance bonus. The real income of the accepting entity is directly connected with the dispatched employee and the employer. For example, the personal quality and effort of the employee will influence the real income of the accepting entity. The employer's selection and training of the dispatched employee will directly influence the real income of the accepting entity. In accordance with the above explanation, the following hypotheses are made from the model:

The marginal revenue of the accepting entity's products is m, and the production is

$$Q = A_1 f(\mathrm{e}_1) + A_2 f(\mathrm{e}_2) + B + \varepsilon . \tag{1}$$

Here: A_1 is the dispatched employee's ability coefficient, $f(\mathrm{e}_1)$ is the output function of the dispatched employee's effort e_1, A_2 is the employer's ability coefficient, $f(\mathrm{e}_2)$ is the output function of the employer's effort e_2, B is the basic workload of the dispatched employee, and ε is a random variable. $f(\mathrm{e})$ is the output function of effort e. According to the actual situation, it should meet the requirement $f'(\mathrm{e}) \succ 0$, signifying that the more effort the dispatched employee and the employer make, the larger the output of the accepting entity. If the function is $f''(\mathrm{e}) \le 0$, it signifies that the effort's marginal output is digressive (The equal sign signifies the marginal output is invariant). The function of the dispatched employee's effort is $\mathrm{C}(\mathrm{e}_1) = \frac{b_1 \mathrm{e}_1^{\ 2}}{2}$, in which b_1 is the cost coefficient of the dispatched employee's effort. The function of the employer's effort is $\mathrm{C}(\mathrm{e}_2) = \frac{b_2 \mathrm{e}_2^{\ 2}}{2}$, in which b_2 is the cost coefficient of the employer's effort. ε is a random variable, and let it be normally distribute $(0, \sigma^2)$.

The income of the dispatched employee is:

$$S_{\text{the dispatched}} = w + \beta_1 Qm . \tag{2}$$

The income of the employer is :

$$S_{\text{the employer}} = \alpha w + \beta_2 Qm . \tag{3}$$

Here: w is the regular wage of the dispatched employee paid by the accepting entity; β_1 is the ratio of the dispatched employee's performance bonus sharing; α is the sharing ratio paid by the accepting entity to the employer, which is related to the dispatched employee's regular wage; and β_2 is the employer's performance bonus sharing ratio.

Thus the income of the accepting entity is:

$$\begin{aligned} S_{\text{the accepting}} &= Qm - w - \beta_1 Qm - \alpha w - \beta_2 Qm \\ &= (1 - \beta_1 - \beta_2) Qm - (1 + \alpha) w \end{aligned} . \tag{4}$$

To simplify the analysis, suppose $f(\mathrm{e}) = \mathrm{e}$, i.e., output is the linear function of effort. Here, suppose the accepting entity is risk neutral, and the dispatched employee and the employer are risk averse. Let the utility function of the dispatched employee and the employer be negative exponent function, and the utility function of the dispatched employee is $V_1(\mathrm{X}) = -e^{-r_1 x}$; the utility function of the employer is $V_2(\mathrm{X}) = -e^{-r_2 x}$. In the equation, r_1 and r_2 refer to the degree of risk aversion of the dispatched employee and the employer. If they are greater than zero, they stand for risk aversion; if equal to zero, risk neutrality; if less than zero, risk preference.

According to the above explanation, the utility functions of the three parties connecting with labour dispatch can be illustrated as follows, respectively:

$$\begin{aligned} &EUS_{\text{the accepting}} \\ &= (1 - \beta_1 - \beta_2) m(\mathrm{A}_1 \mathrm{e}_1 + \mathrm{A}_2 \mathrm{e}_2 + \mathrm{B}) - (1 + \alpha) w \end{aligned} , \tag{5}$$

$$EVS_{\text{the dispathced}} = EV[S_{\text{the dispathced}} - \mathrm{C}(\mathrm{e}_1)] , \tag{6}$$

$$EVS_{\text{the employer}} = EV[S_{\text{the employer}} - \mathrm{C}(\mathrm{e}_2)] . \tag{7}$$

According to the book Game Theory and Information Economics by Zhang Weiying, solution of equations (6) and (7) are equivalent to that of certainty equivalent value, then:

$$\begin{aligned} &CE_{\text{the dispathced}} \\ &= w + \beta_1 m(\mathrm{A}_1 \mathrm{e}_1 + \mathrm{A}_2 \mathrm{e}_2 + \mathrm{B}) - \frac{b_1 e_1^2}{2} - \frac{1}{2} r_1 \beta_1^2 m^2 \sigma^2 \end{aligned} , \tag{8}$$

$$\begin{aligned} &CE_{\text{the employer}} \\ &= \alpha w + \beta_2 m(\mathrm{A}_1 \mathrm{e}_1 + \mathrm{A}_2 \mathrm{e}_2 + \mathrm{B}) - \frac{b_2 e_2^2}{2} - \frac{1}{2} r_2 \beta_2^2 m^2 \sigma^2 \end{aligned} . \tag{9}$$

In the three parties of labour dispatch, the accepting entity is the Principal and have the advantage of priority selection in the game. While pursuing the maximizing of self-interest, the accepting entity must take into consideration the incentive compatibility constraint and participation constraint of the dispatched employee and the employer. Thus the following programming model is formulated:

$$(P)\max_{\{e_1,e_2,\beta_1,\beta_2,\alpha,w\}}(1-\beta_1-\beta_2)\,m(A_1e_1+A_2e_2+B)-(1+\alpha)\,w, \quad (10)$$

$$s.t.\ e_1\in\arg\max w$$

$$+\beta_1 m(A_1e_1+A_2e_2+B)-\frac{b_1e_1^2}{2}-\frac{1}{2}r_1\beta_1^2m^2\sigma^2, \quad (11)$$

$$w+\beta_1 m(A_1e_1+A_2e_2+B)-\frac{b_1e_1^2}{2}-\frac{1}{2}r_1\beta_1^2m^2\sigma^2\ge s_1, \quad (12)$$

$$e_2\in\arg\max \alpha w$$

$$+\beta_2 m(A_1e_1+A_2e_2+B)-\frac{b_2e_2^2}{2}-\frac{1}{2}r_2\beta_2^2m^2\sigma^2, \quad (13)$$

$$\alpha w+\beta_2 m(A_1e_1+A_2e_2+B)-\frac{b_2e_2^2}{2}-\frac{1}{2}r_2\beta_2^2m^2\sigma^2\ge s_2. \quad (14)$$

Here: s_1 and s_2 stand for the reservation utility of the dispatched employee and the employer respectively.

3.1 UNDER SYMMETRIC INFORMATION

Under the circumstances, the accepting entity, the dispatched employee and the employer possess symmetric information. The accepting entity can observe the action and effort degree of the dispatched employee and the employer and pursues the maximizing of self-interest, that is to say, expressions (12) and (14) can only take equality sign. So the above programming problem can be revised as:

$$(P')\max_{\{e_1,e_2,\beta_1,\beta_2,\alpha,w\}}(1-\beta_1-\beta_2)\,m(A_1e_1+A_2e_2+B)-(1+\alpha)\,w, \quad (15)$$

$$s.t.\ w+\beta_1 m(A_1e_1+A_2e_2+B)-\frac{b_1e_1^2}{2}-\frac{1}{2}r_1\beta_1^2m^2\sigma^2=s_1, \quad (16)$$

$$\alpha w+\beta_2 m(A_1e_1+A_2e_2+B)-\frac{b_2e_2^2}{2}-\frac{1}{2}r_2\beta_2^2m^2\sigma^2=s_2. \quad (17)$$

Substitute Eq.(16) and Eq.(17) into Eq.(15), we obtain:

$$e_1^*=\frac{mA_1}{b_1},\ e_2^*=\frac{mA_2}{b_2},\ \beta_1^*=0,\ \beta_2^*=0. \quad (18)$$

Substitute Eq. (18) into Eq. (16) and Eq. (17), we obtain:

$$w^*=s_1+\frac{m^2A_1^2}{2b_1},\ \alpha^*=\frac{2b_1b_2s_2+b_1m^2A_2^2}{2b_1b_2s_1+b_2m^2A_1^2}. \quad (19)$$

Now the marginal cost of the dispatched employee's effort is:

$$C'(e_1)=(\frac{b_1e_1^2}{2})'=b_1e_1=mA_1. \quad (20)$$

The marginal cost of the employer's effort is:

$$C'(e_2)=(\frac{b_2e_2^2}{2})'=b_2e_2=mA_2. \quad (21)$$

The marginal expected utility of the dispatched employee is:

$$\frac{\partial(EmQ)}{\partial e_1}=\frac{\partial[m(A_1e_1+A_2e_2+B)]}{\partial e_1}=mA_1. \quad (22)$$

The marginal expected utility of the employer is:

$$\frac{\partial(EmQ)}{\partial e_2}=\frac{\partial[m(A_1e_1+A_2e_2+B)]}{\partial e_2}=mA_2. \quad (23)$$

So it can be seen from above:

Eq.(20)= Eq.(22); and Eq.(21)= Eq.(23).

That is to say, the mariginal cost of the dispatched employee and the employer's respective effort is just equal to the effort's marginal expected utility, which conforms to Pareto optimality.

And the expected revenue of the accepting entity is:

$$EUS_{the\ accepting}{}^*=\frac{m^2A_1^2}{2b_1}+\frac{m^2A_2^2}{2b_2}+Bm-s_1-s_2. \quad (24)$$

The real income of the dispatched employee is:

$$S_{the\ dispathced}{}^*=s_1+\frac{b_1e_1^2}{2}=s_1+\frac{m^2A_1^2}{2b_1}. \quad (25)$$

The real income of the employer is:

$$S_{the\ employer}{}^*=s_2+\frac{b_2e_2^2}{2}=s_2+\frac{m^2A_2^2}{2b_2}. \quad (26)$$

Conclusion 1: Under symmetric information, the accepting entity may make the dispatched employee and the employer achieve Pareto optimality by means of linearity incentive contract.

3.2 UNDER ASYMMETRIC INFORMATION

In realitiy, it is difficult to achieve symmetric information among the three parties of the employer, the dispatched employee and the accepting entity. In general, the dispatched employee and the employer get hold of more information than that of the accepting entity, so the accepting entity cannot see clearly the effort degree of the dispatched employee and the employer. Then the dispatched employee and the employer will try to seek the maximi-

zing of self-interest – hence the following equivelent programme problem:

$$(\mathrm{P}^{''}) \max_{\{e_1,e_2,\beta_1,\beta_2,\alpha,w\}} (1-\beta_1-\beta_2)\,\mathrm{m}(\mathrm{A}_1\mathrm{e}_1+\mathrm{A}_2\mathrm{e}_2+\mathrm{B})-(1+\alpha)\,\mathrm{w}, \quad (27)$$

$$s.t. \quad e_1=\frac{\beta_1 m A_1}{b_1}, \quad (28)$$

$$w+\beta_1 m(\mathrm{A}_1\mathrm{e}_1+\mathrm{A}_2\mathrm{e}_2+\mathrm{B})-\frac{b_1 e_1^2}{2}-\frac{1}{2}r_1\beta_1^2 m^2\sigma^2\geq s_1, \quad (29)$$

$$e_2=\frac{\beta_2 m A_2}{b_2}, \quad (30)$$

$$\alpha w+\beta_2 m(\mathrm{A}_1\mathrm{e}_1+\mathrm{A}_2\mathrm{e}_2+\mathrm{B})-\frac{b_2 e_2^2}{2}-\frac{1}{2}r_2\beta_2^2 m^2\sigma^2\geq s_2, \quad (31)$$

To solve the programme problem P'':

$$\begin{cases} \beta_1^{SB}=\dfrac{A_1^2}{A_1^2+r_1 b_1\sigma^2} \\ \beta_2^{SB}=\dfrac{A_2^2}{A_2^2+r_2 b_2\sigma^2} \\ e_1^{SB}=\dfrac{\beta_1^{SB} m A_1}{b_1} \\ e_2^{SB}=\dfrac{\beta_2^{SB} m A_2}{b_2} \\ w^{SB}=s_1-\beta_1^{SB} m(\mathrm{A}_1\mathrm{e}_1^{SB}+\mathrm{A}_2\mathrm{e}_2^{SB}+\mathrm{B})+\dfrac{b_1 e_1^{SB2}}{2}+\dfrac{1}{2}r_1\beta_1^2 m^2\sigma^2 \\ \alpha=\dfrac{2s_2-2\beta_2^{SB} m(\mathrm{A}_1\mathrm{e}_1^{SB}+\mathrm{A}_2\mathrm{e}_2^{SB}+\mathrm{B})+b_2 e_2^{SB2}+r_2\beta_2^{SB2} m^2\sigma^2}{2s_1-2\beta_1^{SB} m(\mathrm{A}_1\mathrm{e}_1^{SB}+\mathrm{A}_2\mathrm{e}_2^{SB}+\mathrm{B})+b_1 e_1^{SB2}+r_1\beta_1^{SB2} m^2\sigma^2} \end{cases}. \quad (32)$$

Then, the expected income of the accepting entity is:

$$\begin{aligned} EUS_{the\ accepting}^{SB} &= m(A_1\mathrm{e}_1+A_2\mathrm{e}_2+\mathrm{B})-\mathrm{s}_1-\frac{b_1 e_1^2}{2} \\ &-\frac{1}{2}r_1\beta_1^2 m^2\sigma^2-\mathrm{s}_2-\frac{b_2 e_2^2}{2}-\frac{1}{2}r_2\beta_2^2 m^2\sigma^2 \\ &=\frac{A_1^2 m^2\beta_1}{b_1}+\frac{A_2^2 m^2\beta_2}{b_2}+Bm-s_1-\frac{A_1^2 m^2\beta_1^2}{2b_1} \\ &-\frac{1}{2}r_1\beta_1^2 m^2\sigma^2-s_2-\frac{A_2^2 m^2\beta_2^2}{2b_2}-\frac{1}{2}r_2\beta_2^2 m^2\sigma^2 \\ &=\frac{m^2 A_1^6}{2b_1(A_1^2+r_1\mathrm{b}_1\sigma^2)^2}+\frac{m^2 A_2^6}{2b_2(A_2^2+r_2\mathrm{b}_2\sigma^2)^2}+Bm-s_1-s_2 \end{aligned}.$$

The real income of the dispatched employee is:

$$\begin{aligned} S_{the\ dispathced}^{SB} &= w+\beta_1^{SB} m(A_1\mathrm{e}_1^{SB}+A_2\mathrm{e}_2^{SB}+\mathrm{B}) \\ &=\mathrm{s}_1+\frac{b_1 e_1^{SB2}}{2}+\frac{1}{2}r_1\beta_1^{SB2} m^2\sigma^2 \\ &=s_1+\frac{\beta_1^{SB2} m^2 A_1^2}{2b_1}+\frac{1}{2}r_1\beta_1^{SB2} m^2\sigma^2=s_1+\frac{m^2 A_1^4}{2b_1(A_1+r_1\mathrm{b}_1\sigma^2)} \end{aligned}.$$

The real income of the accepting entity is:

$$\begin{aligned} S_{the\ employer}^{SB} &= \alpha w+\beta_2^{SB} m(A_1\mathrm{e}_1^{SB}+A_2\mathrm{e}_2^{SB}+\mathrm{B}) \\ &=\mathrm{s}_2+\frac{b_2 e_2^{SB2}}{2}+\frac{1}{2}r_2\beta_2^{SB2} m^2\sigma^2 \\ &=s_2+\frac{\beta_2^{SB2} m^2 A_2^2}{2b_2}+\frac{1}{2}r_2\beta_2^{SB2} m^2\sigma^2=s_2+\frac{m^2 A_2^4}{2b_2(A_2+r_2\mathrm{b}_2\sigma^2)} \end{aligned}.$$

Conclusion 2: From the equations $\beta_1^{SB}=\dfrac{A_1^2}{A_1^2+r_1 b_1\sigma^2}$ and $\beta_2^{SB}=\dfrac{A_2^2}{A_2^2+r_2 b_2\sigma^2}$, it can be seen that when the dispatched employee and the employer are the type of risk aversion (i.e., $r_1\succ 0$ and $r_2\succ 0$), the value of β will be less with the value of r becoming bigger; that is to say, the lower the risk tolerance of the dispatched employee and the emplioyer, the less the commission ratio paid by the accepting entity to the dispatched employee and the emplioyer. When r tends to be infinite, $\beta=1$, and the effort degree of the dispatched employee and the emplioyer is same as that under symmetric information.

Conclusion 3: From the equation

$$EUS_{the\ accepting}^{SB}=\frac{m^2 A_1^6}{2b_1(A_1^2+r_1\mathrm{b}_1\sigma^2)^2}+\frac{m^2 A_2^6}{2b_2(A_2^2+r_2\mathrm{b}_2\sigma^2)^2}+Bm-s_1-s_2$$

it can be seen:

(1) $\dfrac{\partial EUS_{the\ accepting}}{\partial A_1}=\dfrac{m^2 A_1^5(\mathrm{A}_1^2+3r_1\mathrm{b}_1\sigma^2)}{b_1(\mathrm{A}_1^2+r_1\mathrm{b}_1\sigma^2)^3}\succ 0$,

$\dfrac{\partial EUS_{the\ accepting}}{\partial A_2}=\dfrac{m^2 A_2^5(\mathrm{A}_2^2+3r_2\mathrm{b}_2\sigma^2)}{b_2(\mathrm{A}_2^2+r_2\mathrm{b}_2\sigma^2)^3}\succ 0$,

which shows that the greater the ability of the dispatched employee and the employer, the higher the expected income of the accepting entity;

(2) $\dfrac{\partial EUS_{the\ accepting}}{\partial r_1}=-\dfrac{2m^2 A_1^7}{b_1(\mathrm{A}_1^2+r_1\mathrm{b}_1\sigma^2)^3}\prec 0$,

$\dfrac{\partial EUS_{the\ accepting}}{\partial r_2}=-\dfrac{2m^2 A_2^7}{b_2(\mathrm{A}_2^2+r_2\mathrm{b}_2\sigma^2)^3}\prec 0$ the greater the risk aversion degree of the dispatched employee and the employer, the lower the expected income of the accepting entity;

(3) $\dfrac{\partial EUS_{the\ accepting}}{\partial b_1}=-\dfrac{m^2 A_1^6(\mathrm{A}_1^2+3r_1\mathrm{b}_1\sigma^2)}{2b_1(\mathrm{A}_1^2+r_1\mathrm{b}_1\sigma^2)^3}\prec 0$,

$\dfrac{\partial EUS_{the\ accepting}}{\partial b_2}=-\dfrac{m^2 A_2^6(\mathrm{A}_2^2+3r_2\mathrm{b}_2\sigma^2)}{2b_2(\mathrm{A}_2^2+r_2\mathrm{b}_2\sigma^2)^3}\prec 0$

and the larger the effort cost coefficient of the dispatched employee and the employer, the the lower the expected income of the accepting entity.

4 Conclusion

This paper makes an analysis of the incentive offered by accepting entities to the dispatched employees and employers. It verifies through models that in the case of symmetric information accepting entities can offer incentive to the dispatched employees and employers so as to achieve Pareto optimality by means of linear contracts. In case of asymmetric information, the expected revenue of accepting entities will be influenced by the abilities, degree of risk aversion and effort costs of the dispatched employees and employers, which is of great reference value to the enterprise practice. The accepting entity should choose the dispatched employee and the employer with a greater ability, a weaker degree of risk aversion and a smaller effort coefficient so as to reach a better income level of the accepting entity.

References

[1] Zheng Dongliang 2010 *The development and regulations of labour dispatch – a survey based on international experience and domestic Practice* China labour and social security publishing press: Beijing
[2] Atkinson J 1984 Manpower strategies for flexible organizations *Personnel management* **8** 28-31
[3] Hogler R L 1996 Transforming employment relationship: implycation for human resource management *Human resource management review* **6** 75-88
[4] Cohany S R 1998 Workers in alternative employment arrangements: a second look *Monthly labour review* **121**(11) 3-21
[5] Su Quanyong 2009 An analysis of the problems and countermeasures of labour dispatch *Journal of Shanxi finance and economics university* **31**(2) 38
[6] Guo Gang 2004 Labour dispatch – a mode of management of human resources needing furthur improvement *Human resource development of China* **11** 90-2
[7] Wan Xiaofeng, Xu Guochun 2007 On the promotion of talent dispatch *Theory and contemporary* **1** 47-8
[8] Wang Qaunxing, Hou Lingling 2004 Legal thinking on the double operation of labour relations *China labour* **4** 23-5
[9] Sun Fangfei 2009 Game analysis of the behaviour of tripartite labour-capital relations under the employment mode of labour dispatch *Legal system and society* 7 243
[10] Wu Xiuqi 2009 On regulation and perfection of labour dispatch *Hebei law science* **27**(4) 130-6
[11] Zhang Kairu 2005 The principal-agent model analysis of labour contracts in enterprises – an adverse selection problem occurring before moral hazard *Journal of management sciences* **18**(1) 26-30
[12] Li Shanliang 2005 Principal-agent analysis of supply chain incentive contract with asymmetric information *Computer integrated manufacturing systems* **11**(12) 1758-65
[13] Laffont 2002 *The theory of incentives I – the principal-agent model* China renmin university press: Beijing
[14] Hart O 1983 Optimal labour contracts under asymmetric information: an introduction *Review* (50) 3-35

Authors

Yanli Yang, born in December 1982, China

Current position, grades: a doctor student at Sichuan university, China
University studies: master degree in management science and engineering from Hebei engineering university, china in 2007.
Scientific interests: human resource and incentive theory.

Xianyu Wang, born in May, 1947, China

Current position, grades: researcher at Sichuan university, China.
Scientific interests: incentive theory and supply chain management.

Establishment of comprehensive index model based on raster data and its application in the study of regional differentiation

Qiang Li*

Tourism and Environment College of Shaanxi Normal University, Xi'an, Shaanxi 710119, China

Environment and Resource Management Department of Shaanxi Xueqian Normal University, Xi'an, Shaanxi 710100, China

Abstract

The small grid calculation model was used as the theoretical basis; A variety of meteorological data in 1971-2007 of Chinese Loess Plateau, remote sensing image (TM), DEM and other maps in 2000 were used; factor analysis, raster calculator, variance analysis and other methods in ArcGIS were used; the mathematical model of the latitude and longitude related to the main meteorological elements of the loess plateau and comprehensive index model for regionalization were obtained; the north and south boundaries of the Loess Plateau were determined.

Keywords: factor analysis, analysis of variance, raster data, comprehensive index model, boundaries, loess plateau

1 Introduction

Regional differentiation refers to the mutual differentiation between the unequal-sized earth's crust and its internal similar sections and the resulting differences.[1, 2]. Differences reflect a certain rule, and grasping the rule can guide the production and life, and optimize the layout, plan the developing policies with important significance, thus the deep understanding of regional differentiation, directly or indirectly affects the regional development and sustainability. The study of the regional differentiation is one of the important signs of people's geographic environment cognition depth and natural environment research level, and the delineation of district boundaries has become an important and urgent work, especially in the region with complex, diverse terrain.[3]. Therefore, many scholars have carried out relevant research, and obtained some achievements. [4, 5]. But at present, the study of using the grid and GIS technology, combining with vegetation type, soil type and soil erosion less for natural regional division is quite rare.

2 Overview of the study area

Loess Plateau is located in north and central China, including the vast region of west of Taihang Mountains, north of the Qinling Mountains, east Wushaoling, south of the Great Wall. It is across Shanxi, Shaanxi, Gansu, Qinghai, Ningxia and Henan province, with an area of about 62.3299 million square kilometers. Apart from a few rocky mountains, the Loess Plateau is covered with a deep loess layer with a thickness between 50 to 80 meters, and the thickest layer is up to 150 to 180 meters. The loess are with fine particles, soft soil, rich in minerals, beneficial for farming, Agribusiness history of the basins and valleys is long, and it is the cradle of ancient Chinese culture. However, due to the lack of vegetation protection, and with the concentrated summer rains and storms, the ground has been divided in many broken pieces under long-term water erosion, forming highlands, hillocks, and loess hills in the gully interlocks (Figure 1).

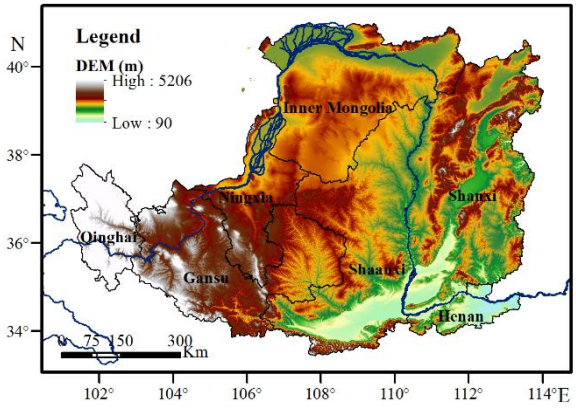


FIGURE 1 Overview of the Loess Plateau

3 Research Methods

3.1 DATA SOURCE AND PRE-PROCESSING METHOD

The data used included the daily precipitation, average speed, maximum speed, average temperature, average pressure, minimum temperature, maximum temperature, sunshine hours and relative humidity of 37 meteorological stations in the Loess Plateau and its surroundings in 1971-2000, the remote sensing (TM), the land use type map, the soil type map and the soil erosion map and DEM data of the Loess Plateau in 2000. First of all, the meteorological

* ***Corresponding author's*** e-mail: liqiangis@163.com

data of meteorological stations were processed in the aspect of the coarse grain into average monthly data for many years, the processing data of all the meteorological elements, and precipitation and sunshine time in April – September (growing seasons) were obtained. TM images were pre-processed in ERDAS8.7, including splicing, projection transformation, geometric correction and tailoring of the binary images. Then, the normalized model calculation of the vegetation index and soil information extraction index was established.

3.2 EVALUATION OF THE CONSTRUCTION OF FACTOR MODEL

3.2.1 Meteorological index

The data from the each station were used to conduct factor analysis for various meteorological factors to extract the main influence factors. The small grid calculation model was used as the basis, taking the effect of latitude and longitude and height above sea level on meteorological factors into account, so the three geographical factors of longitude, latitude, height above sea level were used as independent variables, climatic factors obtained by factor analysis were used as dependent variables to establish a multiple linear regression model. The model was as follows:

$$F = a_0 + a_1\phi + a_2 E + a_3 h, \tag{1}$$

where F is the dependent variable of meteorological model. φ, E, h are latitude and longitude and height above sea level respectively. a_0, a_1, a_2, a_3 are regression coefficient, which can be determined with the least square method according to some element value of each meteorological station and the data of latitude and longitude and height above sea level in this station. In which, φ, E, h are latitude layer, longitude layer and DEM respectively in layer calculation, the latitude and longitude layers were created by being imported into ArcGIS cross longitude and latitude data interpolation. The calculation results of various meteorological model layers represent the space distribution of meteorological factors [6].

Because the space change of terrain in the Loess Plateau had an increasing trend from southeast to northwest, the related analysis of meteorological model layers and DEM grid data layer were made, which was realized by the Band Collection Statistics Function of ArcGIS. According to the correlation coefficient obtained, each layer was standardized; negative correlation was standardized with lower effect measure, positive correlation was standardized with upper effect measure, to ensure that each layer changes in the same direction, the specific standard formula was as follows:

Standardization of upper effect measure:

$$Iscore_i = (x_i - x_{\min})/(x_{\max} - x_{\min}). \tag{2}$$

Standardization of lower effect measure:

$$Iscore_i = (x_{\max} - x_i)/(x_{\max} - x_{\min}), \tag{3}$$

where $Iscore_i$ represents the standardization value of an index in the i-th grid; x_i refers to the original data in an index in the i-th grid; $x_{\max}$ and $x_{\min}$ represents respecttively the maximum and minimum value of an index in all grid [7].

The meteorological index model was established as follows after the above models and standard:

$$MI = \sum_{i=1}^{m} a_i \times F_i, \tag{4}$$

where MI is the meteorological index; F_i is m meteorological elements. a_i is the corresponding weight of m meteorological elements. The weights are determined by the variation coefficient method.

3.2.2 Vegetation index

$$LCI = NDVI \times LT_i, \tag{5}$$

where LCI is the vegetation index; $NDVI$ is the normalized difference vegetation index of the cell; LT_i is the weight of each land use type. Land use types are obtained based on remote sensing image interpretation in 2010; the weight of each type can refer to the researches of Feng Zhiming and others [8].

3.2.3 Soil index

$$ESI = (TM5/TM7)/(TM4/TM3), \tag{6}$$

$$SI = ESI \times ST_i, \tag{7}$$

where ESI is soil information extraction index. TM_i is reflectance value of the corresponding band of remote sensing; SI is soil index. ST_i is the area percentage of each soil type.

Jin Ming et al thought (TM5/TM7)/ (TM4/TM3) not only highlighted the soil information, but also eliminated the influence of vegetation and terrain [9].

3.2.4 Erosion index

$$C = \sum A_i S_i, \tag{8}$$

where C is the comprehensive index of soil erosion in a cell. A_i is the area percentage of i-th soil erosion intensity within the cell. S_i is the corresponding index value of soil erosion intensity.

3.2.5 Comprehensive index

$$AI = \alpha \times NMI + \beta \times NLCI + \chi \times NEI + \delta \times NC, \tag{9}$$

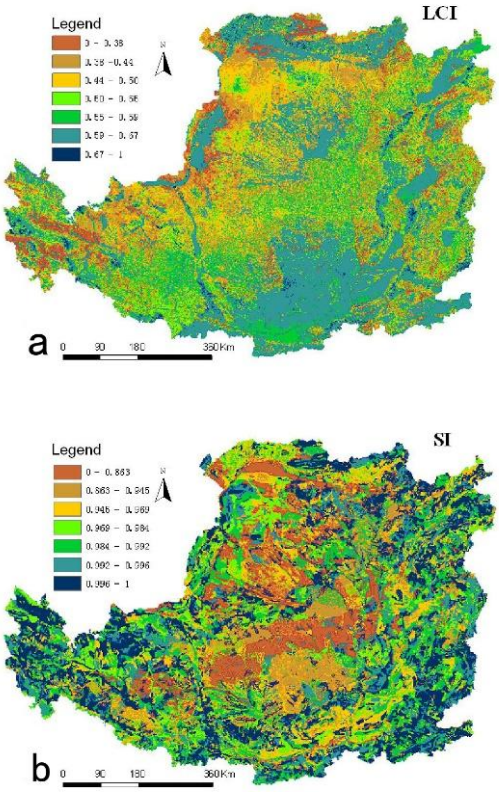
where AI is the comprehensive index for regional division; NMI is the standardized meteorological model

corresponding to various meteorological indexes; *NLCI*, *NEI* and *NC* are the standardized vegetation index, soil index and erosion index respectively. α, β, χ and δ is respectively the corresponding weight. Standard method was still calculated based on the correlation coefficient obtained from correlation analysis of DEM, and the weight was obtained by the AHP method.

4 Analysis of the Results

4.1 VEGETATION INDEX

Distribution of the vegetation in the space was a geographic unit of complete and continuous, and not repeated vegetation types and their combinations. As can be seen from Figure 2a, the vegetation distribution in the Loess Plateau had a stronger direction; vegetation was gradually in decline from southeast to northwest. Vegetation coverage of the Guanzhong Basin was quite good, in addition, Inner Mongolia, Shanxi, Yanan of Shaanxi and Henan also were distributed, and standardized vegetation index was mainly at 0.59 to 0.67; South and most of Shanxi, Shaanxi, south of Ningxia and southeast of Gansu had a worse vegetation coverage, standardized vegetation index was at 0.50 to 0.59; the vegetation coverage of Midwest of Erdos, Inner Mongolia, center of Ningxia and Dingxi, Gansu was relatively the worst, Qinghai province was also distributed occasionally, and the standardized vegetation index was at 0-0.50. This is mainly because terrain rises from the southeast to the northwest, and moisture decreases.



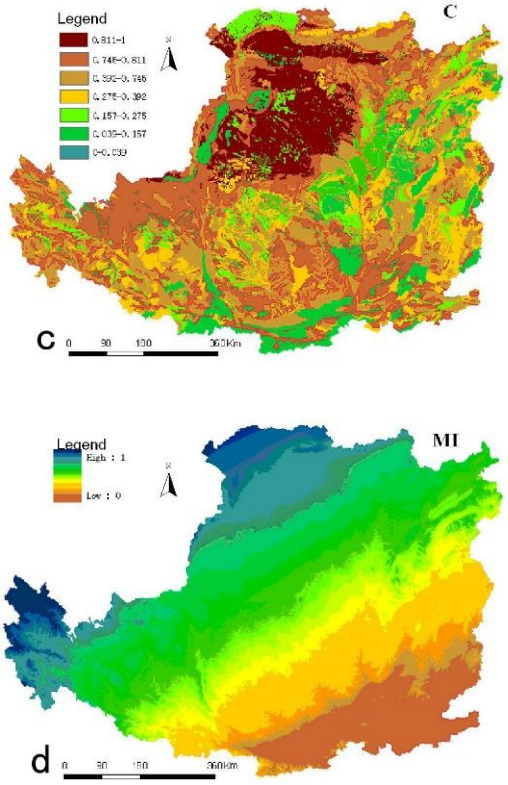


FIGURE 2 Standardized Index of Upper Effect Measure: a) Vegetation; b) Soil; c) Erosion; d) Meteorology

4.2 SOIL INDEX

According to Figure 2b, the spatial distribution of soil index was very scattered, but still has the difference in a certain direction, reducing the content of organic matter from southeast to northwest, and soil was pale in color, texture became light and crude. Among them, the standardized soil index of Shanxi, Guanzhong region of Shaanxi province and Henan, Inner Mongolia, north of Ningxia and Qinghai was at 0.945-1, and soil type was complicated; But the soil types in Shanxi, Guanzhong region of Shaanxi province and Henan had regosols, oil aquilegia soils, aquilegia clay soils, loessial soil, grey loess soils, good loess soils, light brown soils, cinnamon soils, leaching cinnamon soils; the soil types in Inner Mongolia, north of Ningxia and Qinghai had calcaric regosols, light brown soils, chestnut soils, light chestnut soils, light sierozem, chernozem, flow wind sands, cumulated irrigated soils, subalpine meadow soils, and alpine shrub meadow soils; the standardized soil index of Gansu, south central Ningxia, and Shaanxi was at 0-0.945, with a single soil type, mainly was sierozem, loessal soils, good loess soils, sandy loessial soils and grey loess soils.

4.3 EROSION INDEX

The purpose of the classification of soil erosion type is to reflect and reveal the erosion characteristics of different types and the law of regional differentiation, so that appro-

priate measures can be taken to prevent or mitigate erosion hazard. As can be reflected from Figure 2c, the intensity of Loess Plateau soil erosion also had a certain direction, increased from the southeast to the northwest, and the erosion type was also different. Among them, the erosion intensity of Guanzhong Basin of Shaanxi Province was the weakest, south central Shanxi was also sporadically distributed; the standardized erosion index was at 0.039-0.157; the erosion intensity at the junction of Shanxi and Northern Shaanxi was stronger, in addition, Inner Mongolia, Shanxi and Gansu also were distributed, and the standardized erosion index was at 0.157-0.275; the erosion intensity of Dingxi area, Gansu, most areas of Ningxia and Inner Mongolia obviously increased, and the standardized erosion index was at 0.745-0.811; the erosion intensity of the border between central south of Erdos, Inner Mongolia and Shaanxi was the strongest, and the distribution was concentrated.

4.4 METEOROLOGICAL INDEX

Due to obvious spatial difference of meteorological factors, it is necessary to take meteorological factors into consideration in the natural regionalization. There are many kinds of meteorological factors. However, the dominant factors are emphasized in the principle of natural regionalization, namely, the index of dominant factor reflecting the regional differentiation is selected as the main basis to determine the regional boundary. In this paper, the factors are analyzed, and annual average maximum wind speed and annual average wind speed are eliminated. Thus, other factors are used to establish the meteorological factor model (Figure 2d).

The multivariable linear regression model is established, with climatic factor which is obtained through factor analysis as the dependent variable as well as longitude, latitude and altitude as the independent variables, of which the upper limit measure effect is standardized. Then the other multivariable linear regression model is established, with the standardized climate factor as the dependent variable as well as longitude, latitude and altitude as the independent variable. The influence degree of the independent variable coefficient on dependent variable shall be achieved (Table 1), so as to compare the influence degree of the same independent variable corresponding to each dependent variable. It can be seen from Tab 1, the multiple correlation coefficient R of each model is more than 0.85, $F > F_{0.01}$, the model reaches the level of significance. Except for annual minimum temperature, annual maximum temperature and annual average temperature, the spatial and temporal variation of meteorological elements is very significant:

1) Under the same longitude and latitude, with the increase of altitude, annual average precipitation, precipitation in the growing season (April-September), annual average sunshine duration and relative humidity are increased, and their rates of change are 3.91mm/100m, 3.91mm/100m, 2.96h/100mand 0.13%/100m respectively, but other elements are decreased.
2) Uunder the same longitude and altitude, with the increase of latitude, annual average sunshine duration and sunshine duration in the growing season are increased, and their rates of change are 168.7411h/1°N and 89.3277h/1°N respectively, but other elements are decreased.
3) Uunder the same latitude and altitude, with the increase of longitude, the change of all elements is opposite to that of same longitude and altitude, namely, annual average sunshine duration and sunshine duration in the growing season are decreased, and their rates of change are 33.6667h/1°E and 15.8480h/1°E respectively, but other elements are increased.

In conclusion, combined with the influence degree of independent variables, the influence of longitude factor on annual average precipitation and precipitation of April-September is the most significant, followed by annual average relative humidity and atmospheric pressure. This is mainly because the terrain of high west and low east affects the degree of water vapor transport. With the increase of longitude, the higher the content of water vapor is, the easier it is cloudy and it rains. The influence of latitude factor on an annual average sunshine duration and sunshine duration of April-September is the most significant, followed by annual average precipitation and precipitation of April-September. This is primarily caused by the difference of north and south weather. Most are desert on the earth surface of the north, with no blockage and high wind, so there is a thin atmosphere, more sunny day and long sunshine time; the terrain of the south is gentle, so there is a thick atmosphere, the pollutants difficult to diffuse, low atmospheric visibility, more rainy day, short sunshine duration. In addition, the influence of altitude increasing gradually from the south to the north is also large. The influence of altitude factor on annual average temperature, average minimum and maximum temperature, as well as accumulated temperature of 10 or higher, followed by annual average precipitation and precipitation of April-September. The vertical lapse rate of annual average temperature, average minimum and maximum temperature is about 0.18°C /100m, quite different from the lapse rate of average temperature of free atmosphere (0.65°C/100m). This may be caused by the unique landform of the loess plateau as well as the different combinations of latitude, longitude and altitude. Therefore, it is concluded that the longitude, latitude and altitude affect greatly the natural factors of the loess plateau area.

TABLE 1 Projection model of meteorological factors

Regression coefficient/Degree of influence / Meteorological factors	a_0	a_1	a_2	a_3	Multiple correlation coefficient R	F (3.33)
Annual average precipitation	-467.2499	29.1348/0.67	-62.3635/0.77	0.0391/0.43	0.9023	48.2119
Precipitation of April-September	-573.7497	24.8187/0.72	-48.5759/0.76	0.0391/0.54	0.8750	35.9348
Annual average minimum temperature	2.3889	0.3840/0.37	-1.0346/0.53	-0.0018/0.81	0.8743	35.6973
Annual average maximum temperature	23.7370	0.1644/0.20	-0.6391/0.41	-0.0018/1.01	0.8561	30.1863
Annual average temperature	10.0353	0.2845/0.32	-0.7993/0.48	-0.0018/0.96	0.8729	35.2251
Accumulated temperature of 10°C or higher	8578.3357	730.2246/0.28	-1264.9502/0.26	-5.5423/1.02	0.8677	33.5060
Annual average atmospheric pressure	-46.4311	10.9787/0.49	-5.8929/0.14	-0.0316/0.67	0.9176	58.5876
Annual average sunshine duration	-124.9959	-33.6667/0.33	168.7411/0.89	0.0296/0.14	0.9318	72.5037
Sunshine duration of April-September	-191.4891	-15.8480/0.34	89.3277/1.02	-0.0103/0.10	0.9321	72.8853
Annual average relative humidity	49.7452	1.1864/0.49	-3.2894/0.73	0.0013/0.25	0.8817	38.4201

Based on meteorological data and the principle of variable coefficient method, the weight of each meteorological factor is obtained. Therefore, the meteorological index model is established. It is difficult to use the actual data to get the weight of ground cover index, soil index and erosion index, so each weight in the composite index model is calculated by using the AHP, among them, the comparative matrix is difficultly obtained. In this paper, based on the importance of each factor, first of all, the preliminary comparative results are obtained, and then the consistency ratio CR is calculated. Then various comparison results in the comparative matrix are adjusted according to the actual situation, until CR<0.1. At this moment, the consistency of comparative matrix is acceptable, as shown in Table 2.

TABLE 2 Comparative matrix of meteorological factors

	Ground cover index	Soil index	Erosion index	Meteorological index	Weight
Ground cover index	1	3	5	0.2	0.1999
Soil index	0.3333	1	2	0.125	0.0802
Erosion index	0.2	0.5	1	0.1111	0.0496
Meteorological index	5	8	9	1	0.6703

4.5 REGIONALIZATION RESULTS

The composite index layer is obtained through Equation 9, and the clustering method is used to divide the composite indexes into three levels from low to high, namely, I-III levels of different types of natural region (Figure 3). The overall distribution situation is a layered distribution from southeast to northwest, and then the regional statistics function of ArcGIS is used for different status values of different meteorological factors in three natural regions.

Level I natural region: The composite index is between 1.11 and 3.00. In this region, the lowest altitude is 91m, the highest altitude is 3681m, its average altitude is 1051.16m, and its pixel standard deviation is 469.435, indicating the change of topographic relief is small. Its land area is 19035376.2640hm^2, accounting for 30.6023% of the total area of Guanzhong region, mainly including Guanzhong region and the southern of Yan'an of Shaanxi, the south of Shanxi (the northern rim is basically the north end of Taiyuan Basin), the northwest of Henan, Tianshui City, Pingliang City and Dingxi area of Gansu Province. Among three natural regions, the precipitation, atmospheric temperature, atmospheric pressure and relative humidity of this region are the highest, but the sunshine duration is the shortest. The main vegetation is the cultivated vegetation, accounting for 54.2647% of the total vegetation area, followed by grasses and broad-leaved forest, and the vegetation coverage is the best. In terms of land use, the area of dry land is the largest, accounting for 40.7759% of the total area, followed by grassland, forest land and cultivated land. In terms of soil, the loessal soil is widely distributed, accounting for 37.5408% of the total soil area, followed by cinnamon soil and Lou soil. The erosion rate is the lowest, the typical erosion is surface erosion, accounting for 52.3126% of the total erosion area, followed by surface erosion-gully erosion-gravitational erosion, and the intensity of runoff erosion is smaller.

Level II natural region: The composite index is between 3.00 and 4.09. In this region, the lowest altitude is 367m, the highest altitude is 4914m, its average altitude is 1570.28m, and its pixel standard deviation is 540.374, indicating the change of topographic relief is the middle one among three natural regions. Its land area is 23810490.7610 hm^2, accounting for 38.2790% of the total area of Guanzhong region, mainly including the north of Yan'an and most parts of Yulin in Shaanxi Province, the north of Shanxi, the south of Ningxia and the southeast of Gansu Province. In addition, a small part of Level II natural region is also distributed in Qinghai Province. In three natural regions, the state of meteorological elements of this region is secondary. The main vegetation is the cultivated vegetation, accounting for 50.8289% of the total vegetation area, followed by grasses and grassland. The area of broad-leaved forest is smaller than that of Level I natural region, but its grassland is quite large, and its vegetation coverage is the better one. In terms of land use, the grassland area is the largest, accounting for 55.2701% of the total area. The area of dry land, forest land and cultivated land is much smaller than that of Level I natural region. In terms of soil, the loessal soil is widely distributed, accounting for 48.5126% of the total soil area, followed by cinnamon

soil, aeolian sandy soil, skeleton soil and dark loessial soil, and there is no Lou soil in this region. The erosion rate is secondary, the typical erosion is surface erosion-gully erosion-gravitational erosion, accounting for 52.3126% of the total erosion area, followed by surface erosion-gully erosion-gravitational erosion, and the intensity of gully erosion-gravitational erosion is smaller.

Level III natural region: The composite index is between 4.09 and 5.89. In this region, the lowest altitude is 783m, the highest altitude is 5067m, its average altitude is 1546.96m, and its pixel standard deviation is 655.918, indicating the change of topographic relief is the largest. Its land area is 19356553.9266hm^2, accounting for 31.1187% of the total area of Guanzhong region, mainly including Ordos Plateauof Inner Mongolia and the north of Ningxia. In addition, a small part of Level III natural region is also distributed in Shaanxi, Shanxi, Qinghai and Gansu. The state of all elements of this region is opposite to that of Level I natural region, namely, precipitation, atmospheric temperature, atmospheric pressure and relative humidity are the lowest, but the sunshine duration is the longest. The main vegetation is the grassland, accounting for 26.6676% of the total vegetation area, followed by cultivated land and meadow, and the vegetation coverage is the worst. In terms of land use, the area of grassland is the largest, accounting for 70.9113% of the total area. The area of dry land is much smaller than that of Level II natural region and the area of other land use is still increased. In terms of soil, the aeolian sandy soil is widely distributed, accounting for 25.9494% of the total soil area, followed by chestnut soil, Sierozem soil and skeleton soil. There is no cinnamon soil and Lou soil in this region, but a small part of desert soil and light brown soil is distributed. The erosion rate is the maximum, and the typical erosion is surface erosion, accounting for 29.4586% of the total erosion area, followed by runoff erosion (Figure 1).

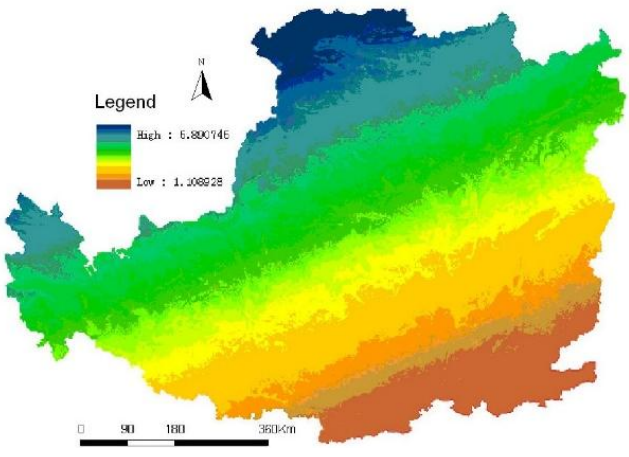


FIGURE 3 Diagram for Natural Regionalization of the Loess Plateau

5 Conclusions and discussion

Factor analysis and composite index are used for calculation to obtain the composite index model of the loess plateau, to determine the south boundary of the loess plateau. The conclusion is as follows:

1) The vegetation, soil and erosion rate have a consistent direction, so the selection of these three factors has a certain scientific nature.
2) The model set up based on meteorological elements, longitude, latitude and altitude is up to an extremely significant level, indicating that the accuracy of meteorological index model is higher. Through analysis, it is concluded that the longitude, latitude and altitude affect greatly the natural factors of the loess plateau area.
3) The overall distribution situation of the loess plateau composite index is a layered distribution from southeast to northwest. The size of meteorological elements is progressively increased or decreased in the sequence of regional level, gradually increasing with the direction of topographic relief. The vegetation coverage is reduced. The content of soil organic matter is decreesed, its color becomes lighter, and its texture becomes light and coarse. The erosion intensity is enhanced in turn. According to the results of the composite index regionalization, the south boundary of the Loess Plateau is basically determined, including Guanzhong region and the southern of Yan'an in Shaanxi Province, the south of Shanxi (the northern rim is basically the north end of Taiyuan Basin), the northwest of Henan, Tianshui City, Pingliang City and Dingxi area of Gansu Province. The north includes Ordos Plateau of Inner Mongolia and the north of Ningxia, Baiyin City and parts of Lanzhou in Gansu Province, Xining City and Haidong area in Qinghai and a few areas in the north of Shaanxi and Shanxi. The rest belongs to the central area of the Loess Plateau.

The research is involved in the modeling of meteorological factors. However, the meteorological data are limited, and only the data from 37 sites are available. Therefore, 12 meteorological factors are adopted to compensate for the inadequacy of data. In addition, the weights in the model of meteorological index and composite index are achieved respectively through two calculation methods, namely, the variable coefficient method and the AHP method. The variation coefficient method uses the actual data to obtain the weights, belonging to an objective weight method. But the partial actual data are difficultly obtained in the composite index model, so the AHP method is used to calculate the weight. Thus, in order to express the difference between regions accurately, besides the fine degree of the data should meet certain requirements, how to acquire the model parameters is also to be further studied.

6 Acknowledgement

National Natural Science Foundation of China(No.: 41301618); Shanxi Provincial Social Science Fund(No: 13D019); Shanxi Provincial Department of Education Science Research Program (No.: 2013JK0851).

References

[1] Zheng D, Yang Q, Zhao M 1997 Research on Natural Geographical System *Beijing China Environmental Science Press (in Chinese)*

[2] Liu C 2004 New Methodology Research on Chinese Comprehensive Physical Regionalization in the Support of Meso-scale Earth Observation System *Progress in Geography* **23** 1-9 *(in Chinese)*

[3] Hao C, Wu S, Li S 2008 Boundary Partition Methods based on SOFM *Progress in Geography* **27** 121-7 *(in Chinese)*

[4] Wang F, Ye H, Xing F 2005 Discussion of Hainan Island Subzone Demarcation *Journal of Beijing Forestry University* **27** 54-8 *(in Chinese)*

[5] Xu J, Yang Y, Zhu B 2009 Discussion of the Important Estuaries Sea demarcation in China *Journal of Marine Sciences* **27** 42-6 *(in Chinese)*

[6] Wang X 2006 Animal Husbandry and Agricultural Climate Resources and Regional Planning in Hulun Buir City, Inner Mongolia *Beijing China Meteorological Press (in Chinese)*

[7] Hao H, Ren Z 2009 Evaluation of Nature Suitability for Human Settlement in Shaanxi Province Based on Raster Data *Journal of Geographical Sciences* **64** 498-505 *(in Chinese)*

[8] Feng Z, Tang Y, Yang Y 2008 Establishment and Application of China Human Settlements Environment Index Model Based on GIS *Journal of Geographical Sciences* **63** 1327-36 *(in Chinese)*

[9] Sha J, Shi Z 2000 Image Processing and Classification of the Southeast Mountain Soil Monitoring by Remote Sensing *Journal of Soil and Water Conservation* **14** 38-47 *(in Chinese)*

Authors

Qiang Li, born in July 1980, Tianjin, China.

Current position, grades: doctor of Geographic Information Systems, associate professor and Deputy Director at Shaanxi Xueqian Normal University.
University studies: Shaanxi Normal University.
Scientific interests: resources environmental remote sensing and GIS applications.
Publications: 3 monographs, 17 papers.

Data mining applications research based on ISBN management

Yinglan Fang*, Bing Han, Binghui Chen

Department of Computer, North China University of Technology, Jinyuanzhuang Road 5, Beijing, China

Received 1 March 2014, www.cmnt.lv

Abstract

ISBN management systems existed irregularities publishing. How to summarizes the current book publishing rule according to publish information, it can better grasp the overall book market trends based on existing books. This paper introduced data mining ideas to the book publishing field through the study of real-name system to apply business processes. There were two data mining models. One was association rule analysis model based on subject field to book type distribution. It analyzed book type using classic Apriori association rule analysis algorithms to identify book publishing hot and the overall trend of publishing business. It can effectively regulate press book publishing behaviour and has great significant to the China's publishing industry healthy and orderly development.

Keywords: data mining, association rule analysis, ISBN

1 Introduction

The current society is in the period of an information technology, networks, data processing rapid development. How to apply suitable method extract valuable data from the large precious data, it has became an important issue. Data mining techniques emerged. Many universities, research institutions and companies worldwide are turning to data mining and knowledge discovery techniques and gain a lot of researches. So that data mining technology has been rapid development.

This issue research data is come from the ISBN real-name application system. ISBN real-name application system was major project which initiated and promoted by the Chinese General Administration of Press and Publication. It has improved the efficiency and quality of service book publishing industry. Over time, business data among the publishers and audit unit and the General Administration of the system have become rich. As a manager, the Press and Publication Administration eager to get useful part of the existing mass book publishing data to assist them to grasp book publishing direction and standard publishing industry. The research of this issue is to solve the book publishing industry currently facing problems using data mining technology. It would has great importance for the publishing industry.

2 ISBN management data mining analysis

ISBN real-name application system is an reform measures ISBN name to proposed by the Press and Publication Administration in 2006. Its rules is "See draft to get ISBN", "one book only one ISBS," "ISBN real name application." The purpose is to completely change the previous ISBN management and further strengthen publishing management and standardize the behavior of publication. With the accumulation of time, ISBN real-name application system has accumulated one million data, and it had to spend a lot of manpower and resources to maintain these data every year. Managers wanted to get useful information from these chaotic by data some means. Therefore, the data mining in ISBN real-name application system would have a very important significance.

2.1 DATA MINING SUBJECT FIELD

The issue starts from the existing ISBN data items, combining with the relevant provisions of how to optimize and book publishing administration reform publishing behaviour. It determines four sides subject area. They are ISBN distribution usage amount and book type distribution and audience and publishing books standard degrees. ISBN real-name application system's database table name were:

T_B_BOOKINFO (report books basic information table),
T_B_PUBLISHINFO (press basic information table),
T_D_ISBNINFO (audit and pass ISBN data table),
T_D_PLANADD (press add ISBN table),
T_D_PLANTOTAL (press publishers plan to distribute ISBN),
T_D_BOOKINFO_ STATUS (declaration book status table)

and so on nearly 10 table and more than 60 million data items. It creates corresponding data table around the data items needed by selected subject areas and ultimately establishment appropriate data warehouse.

2.2 ASSOCIATION RULE MINING MODEL ANALYSIS

Association rule analysis model is mainly comprehensive

* ***Corresponding author's*** e-mail: jlufangyl@163.com

analysis through the ISBN data items and ISBN real-name application system operation flow. It combines with the basic needs of decision makers, establishes the improved association rule analysis model. it includes selecting data items, filtering and integration, researches and implements Apriori algorithm, selects the appropriate threshold (support, confidence), analysis the resulting data and obtains corresponding conclusions or visualization rendering.

The association rule analysis model in this issue is created according to the press publishing book situation. Its purpose is to find the current book publishing rule within a certain period. For example, from the overall analysis 585 publishers, publishing arts (J category) and the history, geography (K category), 100% of publishers will be publish culture, science, education (G category) books. Which found that the country's book publishing trends and characteristics. The books in ISBN apply real-name system are classified according to the classification of Chinese Library Classification from A category to Z category.

- A Category: It represents the Marxism-Leninism, Mao Zedong Thought, Deng Xiaoping Theory and so on category.
- B Category: It represents the philosophy and religion category.
- C Category: It represents the social sciences category.
- D Category: It represents the political and legal category.
- E Category: It represents the military category.
- F Category: It represents the economic category.
- G Category: It represents the culture, science, sport and education category.
- H Category: It represents the language and writing category.
- I Category: It represents the literature category.
- J Category: It represents the art category.
- K Category: It represents the history and geography category.
- N Category: It represents the natural sciences category.
- O Category: It represents the mathematical science and chemistry category.
- P Category: It represents the astronomy and earth science category.
- Q Category: It represents the biological science category.
- R Category: It represents the medicine and health category.
- S Category: It represents the agricultural sciences category.
- T Category: It represents the industrial technology category.
- U Category: It represents the transportation category.
- V Category: It represents the aerospace category.
- X Category: It represents the environmental science and safety science category.
- Z Category: It represents the general books category.

Due to the every press's basic conditions is different, the different publishing book ability or some other causes leads to the published books type which is scattered. Some rare and poor sale books have small publishing amounts. Some popular and highest-selling books have high issue.

So only from single book publishing number and ISBN circulation amount, it cannot be seen in the overall trend and found publish association rule. Research data mining association rules model and data mining process is the same. They are divided into three processes: data extraction, data processing, association rule analysis.

The first step is data extraction. According to the selected subject areas, the book publishing type, ISBN real-name application system classify the book type, the basic situation of each country published book publishers can be obtained from the database system. It statistically analysis books information resource library and books type distribution of each publishing press. Combined with the actual needs, it select the appropriate data entry according to the association rules and create database table structure. Because it aims at the country's 585 presses and for nearly one million book information in six years, it can extract database in accordance with the certain period of time and different database tables corresponding data. based on these reasons it has created such a table structure: book the original information sheet named Original Book Type (include book serial number, ISBN number, book title, author, introduction, book type, publishers ID, extraction time), so it can prepare data items for the next step of data pre-processing.

The next step is data pre-processing. Book raw information obtained from last step data extraction, it needs for further processing and final processing to these data so as to obtain experimental data of books type distribution and use for analysis of Apriori association rules algorithm. Association rules was analyzed through books type distribution. It extracts book data according to the press group and statistics for each publishing house within a certain period of time books published by type. The statistics result was stored in the statistics books Information table named Book Type Count. Every presses are handled as a single transaction. Each transaction has a number of data items. Books type in the table is the statistics of the total book type within a certain period of time. Association rules algorithm requires incoming data is Boolean variable with the value 0 or 1. That the book publishing house types need to represent with value 0 or 1. These data distribution directly affect the results of association rules, so when it sets this threshold it needs repeated testing to find the most appropriate threshold and compare the various book type statistics value with the threshold. When the book type value is larger than the threshold value it is set to 1 and otherwise its value is set to 0. The handled books is stored to book conversion table named Book Type Static according to its rule. Association rules analysis data table structure is shown in Figure 1:

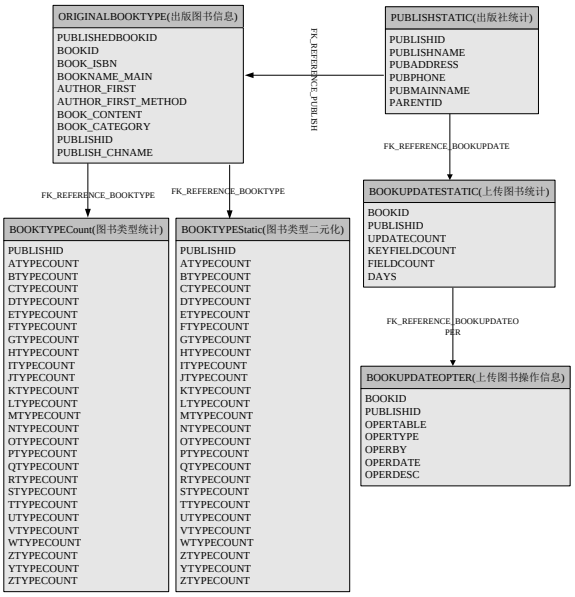


FIGURE 1 Association rules Analysis Database Design E-R

The next step is the results and analysis of association rules. There are two sides. The first side is the parameter setting in Apriori algorithm that is support and confidence threshold. The quality of selected value directly affects the accuracy of the association rules analysis, therefore it requires the combination of the actual situation and the repeated sampling test to help determine the final threshold. The second side is the result of the analysis. It derived some abstract data information after data mining. Researchers and professionals can clearly understand the meaning of the data represents. For general managers it need for appropriate process and convert the data so they can easily accept and understand with text presentation or graphic presentation, and ultimately achieve the purpose of data mining.

3 ISBN data mining design

It can be seen from the research background that the applicable targets to ISBN manage mining system is press administrative department and ISBN issuing departments, including the press and publication administration departments, barcode centres and 31 provincial bureau issued centres. Users can log in the system accordance with their own username and password and use their corresponding function. They can statistically classification analysis and supervise to published books and ISBN usage. It include four aspects. The first is the overall trend in publishing books to current national book press and timely tracking and analysis to book publishing rule. So it can adjustment and control to prevent overheating of some vulgar books published and some less famous books publishing phenolmenon occurs. The second is to supervise the whole process from the application publishing a book to publishing a book so as to discover problematic Press and eradicate illegal acts as "One multi-use", "trading ISBN" and so on. The third is manage publishers sample information cannot be uploaded books by press and the pending book cannot be audited by department. The fourth is the detailed statistical analysis to national book publishing situation. It divides from the time dimension, spatial dimensions and regional dimensions in order to understand the current overall situation of the book publishing industry.

3.1 SYSTEM NETWORK STRUCTURE DESIGN

The system uses B/S architecture model. The user can issue through the browser to many distributed servers on the network request to complete the corresponding business functions. The basic topology of the network structure is shown in Figure 2:

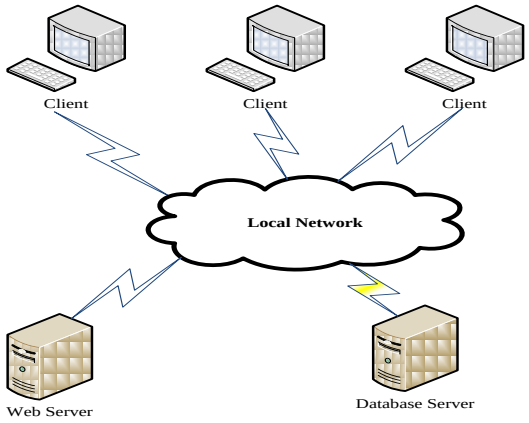


FIGURE 2 ISBN information data mining system network topology

Interaction between the user and the system are described as the below:

- User fills out the network address of the system in the browser and request a service to a Web server on the network via a Web browser. Web server user identity verification after transferring to the desired page using the HTTP protocol client and displayed in a Web browser.
- Web servers accept user requests and connect the database. Then put forward data processing request to the database server through SQL and transfer the results to the Web server.
- Web servers obtains visualization results from data service processing results and returns to the browser user.

3.2 SYSTEM MAIN CLASS DESIGN

In this paper, it has designed the interfaces and classes in association rules analysis algorithm (Apriori algorithm). Apriori algorithm is mainly read transaction data in the data warehouse by Get-Source method. Then through combination of internal processing transaction data, it fits

the data structure in line with the main function, and ultimately get frequent data mining item sets. The Apriori algorithm class design of association rules analysis are as follows (Table 1):

TABLE 1 Apriori Test class method function

Class name	method name	function
AprioriTest	Public String getRules ()	get related combination of frequent items
	Public Integer getconfidence ()	get a given minimum confidence
	Public Integer getSupport ()	get a given minimum support
	Public void mine ()	mining each frequent item sets
	Public void associationRule ()	mining frequent association rules
GetSource	getBookTypeCount ()	get the book information to data type
	getBookTypeStastic ()	get book type information
	getSet ()	get association rule analysis datasets

4 ISBN Data Mining Implementation

This section describes the main functions of association rules analysis modules. It includes raw data extraction, data conversion, data filtering, association rules analysis and graphical display.

FIGURE 3 Raw data statistics Schematic of association rules analysis

4.1 APRIORI ALGORITHM FOR DATA PRE-PROCESSING

1) Data statistics. It statistics and classifies source data extracted from the previous step and classifies according to classify method. Specific statistical results are shown in Figure 4.
2) Data filtering and data cleaning. It further processes data counted by the previous step, converts into binary data structure representation by a given threshold and provides analytical data for the association rule algorithm. It is as shown in Figure 5:

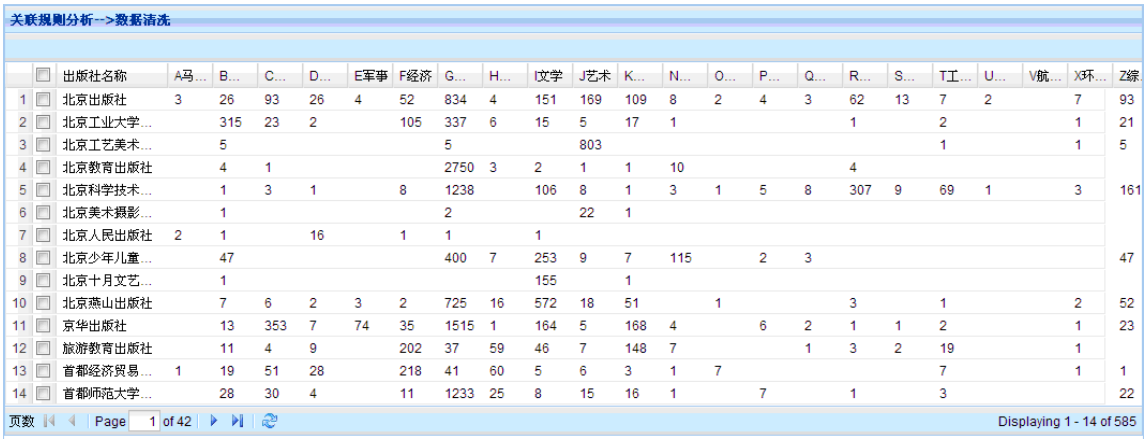


FIGURE 4 Association rules analysis data filtering schematic

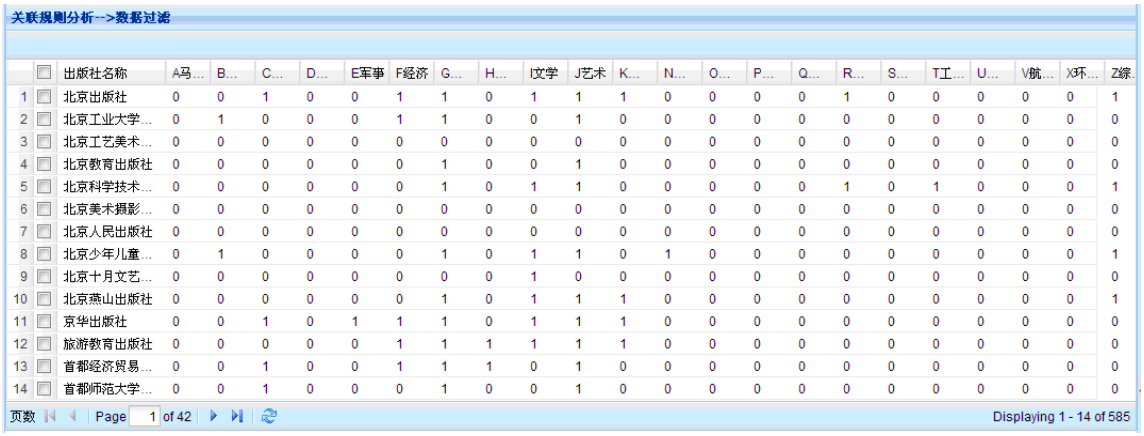


FIGURE 5 Data conversion schematic in association rules analysis

4.2 APRIORI ALGORITHM DATA ASSOCIATION ANALYSIS

The purpose of this analysis is the association rule 585 publishers across the country to conduct the type of book publishing association rules analysis, which found that the association between a certain period of time and the overall trend of analyzing summarizes book during this time.

It can be seen from the Figure 5 that it has 633,591 books and 585 press, namely the number of transaction in association rules analysis process is 585. In line with the specified degree of support minSupport equal to 0% and confidence minConfidence equal to 30%, there are frequent item sets conditions. Namely: {[I, J] =>G}, {K=> [G, J]}, {[G, K] => J}, {I => [G, J]}, {G =>J}, {[J, K] => G}, {[G, I] => J}, {J => G} which rules is [I, J] => G and confidence levelis 48% and degree of support is 100%.

It illustrates that 100% nationwide press publishes science books and economic books and culture, science, education, sports books. Among 48 percent press which published class I (social sciences) and J books (economy class) may also publish books of the class G (culture, science, education, sports).

From an overall perspective, we can see that the economic books, culture, science, education, sports books, literature books almost become mainstream during this time.

关联规则集	置信度	支持度
I,J => G	0.47692308	1.0
K => G,J	0.36923078	0.875
G,K => J	0.33333334	0.9692308
I => G,J	0.53675216	0.91719747
G => J	0.53675216	0.91719747
J,K => G	0.32307693	1.0
G,I => J	0.4923077	0.96875
J => G	0.774359	1.0

FIGURE 6 Association structure data shows in association rules analysis

4.3 ASSOCIATION RULES ANALYSIS RESULTS SHOW

The results association rules analysis is a string of characters and numbers. For which the system user is difficult to understand the meaning. It can be very clearly seen through the graphical display that the roughly distribution of books types and can help verify the accuracy results information of the association rules analysis (Figure 6-7).

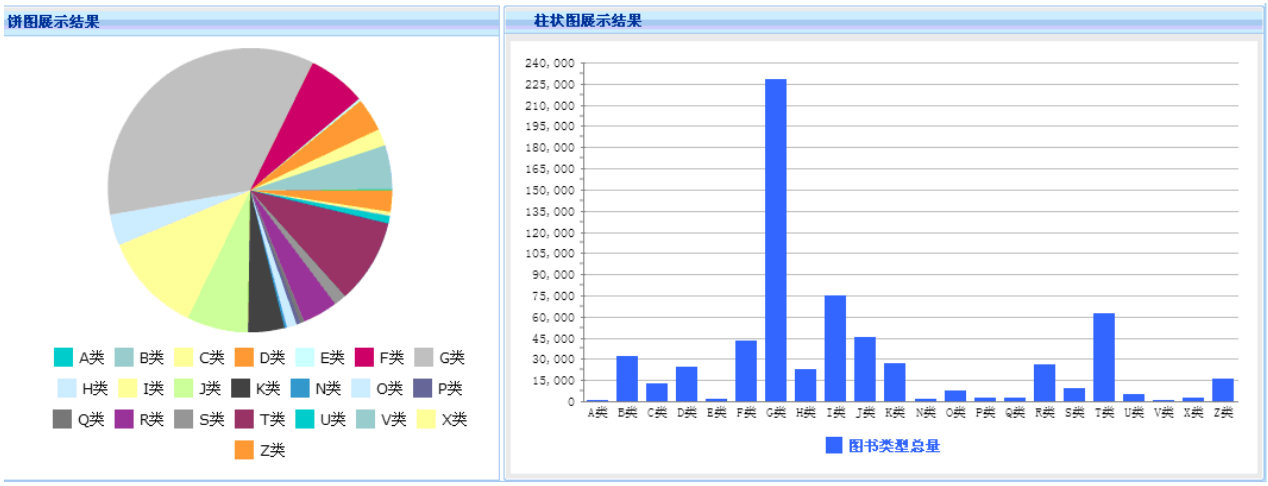


FIGURE 7 Graphical association rules analysis shows

5 Conclusions

In this paper, data mining technology has been imported into the book publishing industry, and conducted in-depth theoretical study and specific mining algorithm experiments. This paper mainly studied the association rules analysis Apriori algorithm. According to the needed analysis problems and actual application, it deeply sees the business of the ISBN real-name application system. Combined with existing data mining techniques, it has analyzed and created around the theme of type press books domain analysis model and association rules around behaviour publishing books degree of clustering model specification subject areas. From different angles ISBN statistical analysis of the data item information, the evaluation and mining results graphically is shown.

Acknowledgments

The work in this paper has been supported by Beijing Natural Science Foundation (No. 4131001). It is also supported by National Science and Technology Support program (No. 2012BAH04F00) and Beijing Natural Science Foundation (No. 4132026) and Beijing Education Science and Technology Development plan (No.KM 201210009006) and Beijing Image Information Processing and Intelligent Recognition Scientific Research Platform Construction (No.PXM2012-014212-000024).

References

[57] Liu Y 2006 Data mining in the management decisions of college students use pattern analysis *Chengdu University of Information Technology* **21**(3) 373-7 *(in Chinese)*
[58] Wu X 2009 Data mining Application of in Educational Administration technology *Harbin Engineering University master's thesis* 16-7 *(in Chinese)*
[59] Cai X 2007 Telecommunications data mining data preparation process standardized design *Computer Engineering* **33**(24) 44-5 *(in Chinese)*
[60] Jun X 2012 Data mining Application in the stock market technical analysis *Zhejiang University Thesis* 19-21 *(in Chinese)*
[61] Cheng J 2011 Research and analysis algorithms of multiple association rule mining *Changchun University of Technology (Natural Science)* 107-9 *(in Chinese)*
[62] Wei J, He Pi, Sun Y 2005 study text hierarchical K-Means clustering algorithm *Computer Applications* 23234 *(in Chinese)*
[63] Xia N, Su Y, Qin X 2010 An efficient k-medoids clustering algorithm *Application Research of Computers* 45178 *(in Chinese)*
[64] Li Q, Yuan J 2012 A text clustering algorithm based on optimal density DBSCA *Computer Engineering and Design* 1409-10 *(in Chinese)*

Authors

Yinglan Fang, born in August 1973, Shangcheng Henan, China.

Current position, grades: associate professor and master supervisor at North China University of Technology.
University studies: ME degree in Computer System Structure at Jinlin University of China in2002.
Scientific interests: computer application and computer security.
Publications: 12 papers.

Bing Han, born in August 1971, Jiaozuo Henan, China.

Current position, grades: associate researcher and master supervisor in North China University of Technology.
University studies: ME degree in Computer System Structure at Jinlin University of China in 2001.
Scientific interests: software engineering and computer control.
Publications: 2 patents, 13 papers.

Binghui Han, born in September 1987, Luoyang Henan, China.

University studies: ME degree in Computer Technology at North China University of Technology in 2013.
Scientific interests: software engineering and data mining.
Publications: 1 papers.

Recommendation system based on collaborative filtering in RapidMiner

Zhihang Tang, Zhonghua Wen

School of Computer and Communication, Hunan Institute of Engineering Xiangtan 411104, China

Abstract

Recommender systems facilitate decision-making processes through informed assistance and enhanced user experience. To aid in the decision-making process, recommender systems use the available data on the items themselves. Personalized recommender systems subsequently use this input data, and convert it to an output in the form of ordered lists or scores of items in which a user might be interested. These lists or scores are the final result the user will be presented with, and their goal is to assist the user in the decision-making process. The application of recommender systems outlined was just a small introduction to the possibilities of the extension. Recommender systems became essential in an information- and decision-overloaded world. They changed the way users make decisions, and helped their creators to increase revenue at the same time. Bringing recommender systems to a broader audience is essential in order to popularize them beyond the limits of scientific research and high technology entrepreneurship. The recommender systems will assist you in reaching quality, informed decisions.

Keywords: recommender systems, collaborative-based systems, nearest neighbour

1 Introduction

Making choices is an integral part of everyday life, especially today when users are over- whelmed with information, from the Internet and television, to shelves in local stores and bookshops. We cope with this information overload by relying on daily recommendations from family, friends, authoritative users, or users who are simply willing to offer such recommendations. This is especially important when we lack information to make a rational decision, for example, choosing a hotel for vacations in a new city, selecting a new movie to watch, or choosing which new restaurant to visit.

Recommender systems [1] facilitate decision-making processes through informed assistance and enhanced user experience. To aid in the decision-making process, recommender systems use the available data on the items themselves, such as item taxonomies, descriptions and other, and/or data on user experience with items of interest, for example, user choices, rankings, scores, tags, etc. Personalized recommender systems subsequently use this input data, and convert it to an output in the form of ordered lists or scores of items in which a user might be interested. These lists or scores are the final result the user will be presented with, and their goal is to assist the user in the decision-making process.

Recommender systems represent a broad and very active [2-4] field of research and were, from their origins in the 1990s [5], somewhat detached from the data mining field. This was mostly due to the specific form of data that recommender systems used. There are three basic types of recommender systems: collaborative filtering, content-based, and hybrid systems. Collaborative filtering recommender systems use social information, preferences, and experiences of other users in order to find users with similar taste. The assumption of these systems is that users of similar tastes might enjoy and consume similar items. Content filtering recommender systems use the available structured and unstructured information on users or items to recommend further items of interest. They assume that the user might be interested in items similar to the ones in which an interest has already been displayed. Both of these systems have their advantages, which are additionally reinforced, or disadvantages, which are diminished with the usage of hybrid systems-systems which combine both the collaborative and the content-based recommendation approach.

Recommender systems are ubiquitous, and an average Internet user has almost certainly had experiences with them, intentionally or not. For example, the well-known Internet commerce, Amazon.com employs a recommender system that recommends products its users might be interested in, based on the shopping habits of other users. Social networking sites like Facebook or LinkedIn use recommender systems for recommending new friends to users based on their social network structure. The music website Last.fm uses a recommender system to recommend new music to a user, based on the listening habits of users with similar music taste. The Internet Movie Database (IMDb) recommends similar movies, based on the content and style of the movies user previously browsed. Streaming provider Netflix tries to predict new movies a user might be interested in based on his watching habits and movie ratings, compared to other users. These, and numerous other examples like Stumble Upon, Google AdSense, YouTube, etc., which differ in services provided, like audio, video, general item, social network, other Internet content, books, etc., demonstrate the importance of these systems.

Recommender systems facilitate making choices, improve user experience, and increase revenue, therefore should be easily accessible for deployment to interested

parties. This led us to write a paper on recommender systems in a clearly understood and easily applied way through RapidMiner. We believe that RapidMiner's workflow approach entices systematic research and facilitates its implementation in combination with Rapid Analytics. The combination of research and production environment renders itself as an excellent environment for understanding recommender systems through practice. Throughout this paper, you will learn the basics of theory related to recommender systems, with a strong emphasis on practical implementation. This practical work will introduce you to all the necessary knowledge for rapid prototyping of recommender systems, thus enabling you to master them through application of your data. The implementation of recommender systems in RapidMiner has been additionally simplified through the Recommender Extension.

2 Basic conception

2.1 RECOMMENDATION OPERATORS

The Recommender Extension has a total 26 recommendation operators. These operators are grouped in the following categories: Item Recommendation, Item Rating Prediction, and Recommender Performance. The overview of operators supported in the Recommender Extension is given in the Table 1.

TABLE 1 Recommendation operators supported by the Recommender Extension.

Item Recommendation (IR)	Collaborative Filtering-based Item Recommendation	Item k-NN, Most Popular, Bayesian Personalized Ranking Matrix Factorization, User k-NN, Weighted Regularized Matrix Factorization, Random
	Attribute-based Item Recommendation	User Attribute k-NN, Item At-tribute k-NN
Item Rating Prediction (RP)	Collaborative Filtering-based Rating Prediction	Random, Global Average, User Item Baseline, User k-NN, Item k-NN, Slope One, Bi-Polar Slope One, Matrix Factorization, Biased Matrix Factorization, Factor Wise Matrix Factorization
	Attribute-based Rating Prediction	User Attribute k-NN, Item At-tribute k-NN

Item recommendation operators operate over large matrices that contain information about which user consumed which item. These input matrices often contain large numbers of empty entries, and can thus be used in a more space-efficient way. We describe the appropriate format used for efficient data loading and storage in the next subsection.

2.2 DATA

Typical recommender systems, operating on user usage data, are built on top of large matrices called utility matrices. These matrices usually contain elements from a limited set of numbers, for example from 0, 1, 2, 3, 4, 5, where 0 typically denotes that the user had no interaction with the item, and the rest describes the level of that interaction in a form of rating. Due to a large number of both users and items, these matrices are typically very large. In addition, since users mostly consume a very small portion of items out of the total number of items, these matrices tend to contain a lot of zeros – they tend to be very sparsely populated. This is why special data structures for handling sparse data need to be implemented. In RapidMiner, we can use AML reader operators to read such datasets. Input datasets used to learn a recommender system model must be formatted in two columns; for example, the first column can contain user IDs, while the second can contain item IDs. Attributes names, and their positioning can be arbitrary. Prior to applying recommendation operators to input datasets, proper roles have to be set for these attributes, as seen in Table 1. Any additional attributes will not be considered. An example of an AML, and a related DAT file for item recommendation operators is given in Figure 1.

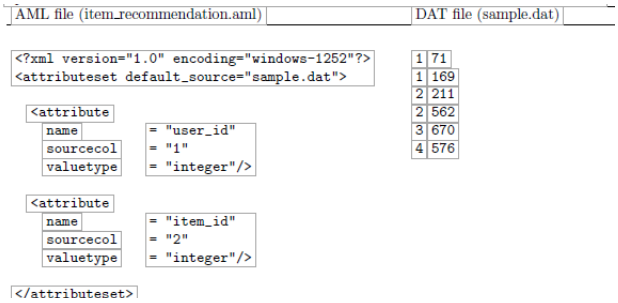


FIGURE 1 An example of an AML and a related DAT file for item recommendation operators

The recommender system datasets used throughout this paper consists of content and collaborative data. Content data was taken from the ALIDATA DISCOVERY (http://102.alibaba.com/competition/addDiscovery/index.htm)

The content dataset described contains the following content attributes for each item:

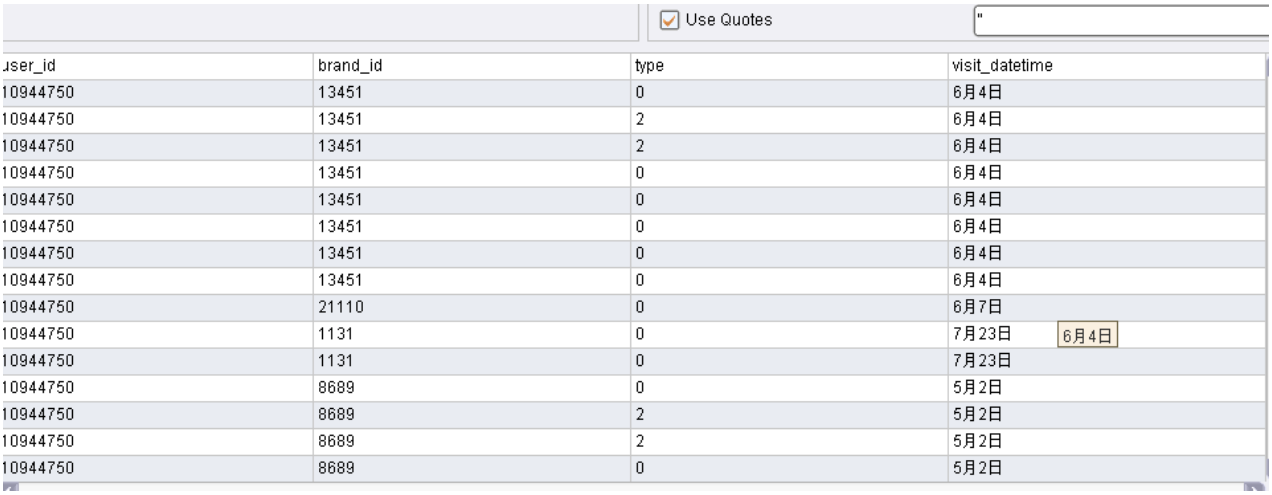
brand ID: a unique integer that represents a lecture;

brand name: a text string containing a name of a particular lecture;

brand description: a text string denoting a description of a particular lecture.

A particular item identifier is denoted by the small letter i and the set of all items is denoted by the capital letter I. Collaborative data contains synthetic click streams of users, where each click stream is a sequence of items viewed by a particular user in some time interval. In the following text, we refer to the synthetic users as users. A particular user identifier is denoted by the small letter u and the set of all users is denoted by the capital letter U. Click streams are transformed into the sparse matrix A, which is called the usage matrix. The non-zero elements of

the usage matrix (A (i, u)) tell us that the item i was consumed by the user u. Using this dataset, we construct collaborative and content recommender systems in the following sections. The collaborative recommender systems rely on the usage matrix A while the content recommender systems rely on items textual descriptions. We can get Figure 2.



Use Quotes "

user_id	brand_id	type	visit_datetime
10944750	13451	0	6月4日
10944750	13451	2	6月4日
10944750	13451	2	6月4日
10944750	13451	0	6月4日
10944750	13451	0	6月4日
10944750	13451	0	6月4日
10944750	13451	0	6月4日
10944750	13451	0	6月4日
10944750	21110	0	6月7日
10944750	1131	0	7月23日
10944750	1131	0	7月23日
10944750	8689	0	5月2日
10944750	8689	2	5月2日
10944750	8689	2	5月2日
10944750	8689	0	5月2日

6月4日

FIGURE 2 Initial data from Alibaba

2.3 COLLABORATIVE-BASED SYSTEMS

The main idea of collaborative recommendation approaches is to use information about the past behavior of existing users of the system for predicting which item the current user will most probably like and thus might consume. Collaborative approaches take a matrix of given user-item ratings or viewings as an input and produce a numerical prediction indicating to what degree the current user will like or dislike a certain item, or a list of n recommended items. The created list should not contain items the current user has already consumed.

Neighborhood-based recommender systems work by counting common items two users have viewed for every pair of users in the system, or the number of common users that viewed the same pair of items. Using this count, similarity between two users or items is calculated. Neighborhood systems use intuition that two users who have viewed a large number of common items have similar tastes. That information can be used to recommend items that one user consumed and the other one did not. We are interested in finding pairs of users having the most similar taste, or pairs of items having the most users that viewed both items. Those pairs of users/items are called "the closest neighbors". We describe two main approaches of the neighborhood-based recommender systems: user and item-based nearest neighbor recommendation.

2.3.1 User-based nearest neighbor recommendation

Given a user-item viewing matrix and the ID of the current user as input, identify other users having similar past preferences to those of the active user. Subsequently, for every product the active user has not yet consumed, compute prediction based on the product usage of the selected user subset. These methods assume that users, who have previously shared similar tastes, will share similar tastes in the future, and that user preferences remain stable and constant over time.

To calculate similarity between users, two typical similarity measures are used: the Pear- son correlation and the Cosine correlation [6]. In our item recommendation problem we used cosine correlation as a similarity measure. Typically, we do not consider all users in the database when calculating user similarity, rather the k most similar ones.

2.3.2 Item-based nearest neighbor recommendation

When dealing with large problems, consisting of millions of users, user-based collaborative filtering approaches lead to increased memory usage and execution time. Since the system is required to calculate a large volume of potential neighbors, it becomes impossible to compute predicttions in real time. In some cases, the number of users dominates the number of items in the system so it would be natural to try to use items for making recommendations. That is the reason for creating a second neighborhood-based recommender system based on items instead of users.

As opposed to the user-based approach, the item-based recommendation approach computes prediction using the similarity between items. We use a cosine similarity measure, as we did in the user-based approach. Likewise, as in the user-based approach, we use k-nearest neighbors, i.e. the k most similar items for prediction.

3 Personalizing Recommender Systems

Collaborative recommender operators use the user-item matrix to build a recommendation model. This user-item matrix is presented as an example set of user-item pairs describing user consumption history. The recommendation model built with this matrix is used to recommend items to

users from a query set. The query set is an example set containing identification numbers of users for which we want to make recommendations. For each user in the query set we recommend only the items not consumed by this user. Figure 3 depicts a basic collaborative recommender operator workflow.

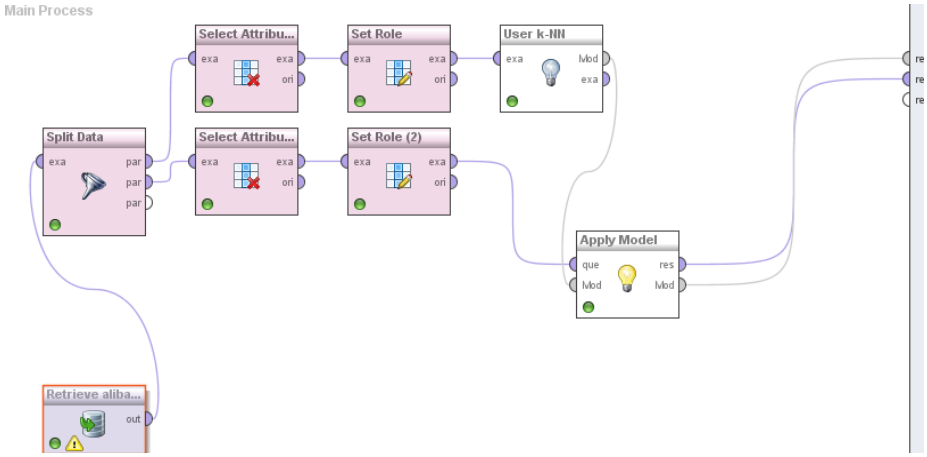
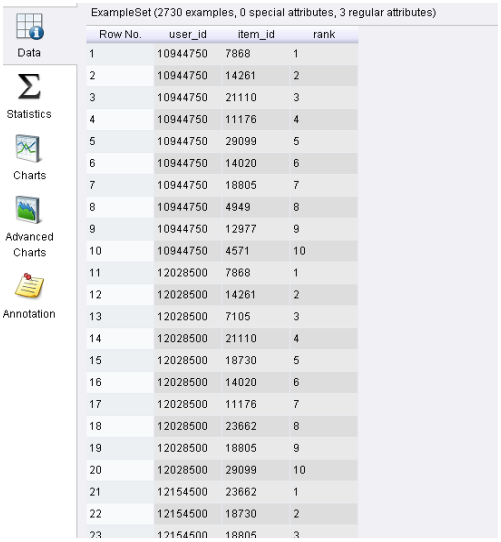


FIGURE 3 An example of an item recommendation workflow

The recommended results shown in Figure 4.


ExampleSet (2730 examples, 0 special attributes, 3 regular attributes)

Row No.	user_id	item_id	rank
1	10944750	7868	1
2	10944750	14261	2
3	10944750	21110	3
4	10944750	11176	4
5	10944750	29099	5
6	10944750	14020	6
7	10944750	18805	7
8	10944750	4949	8
9	10944750	12977	9
10	10944750	4571	10
11	12028500	7868	1
12	12028500	14261	2
13	12028500	7105	3
14	12028500	21110	4
15	12028500	18730	5
16	12028500	14020	6
17	12028500	11176	7
18	12028500	23662	8
19	12028500	18805	9
20	12028500	29099	10
21	12154500	23662	1
22	12154500	18730	2
23	12154500	18805	3

FIGURE 4 The Recommended results

In the item recommendation workflow, the first two operators read the train and the query example sets using the Read AML operators (1,4). Following, the appropriate roles are set to attributes using the Set Role operator (2). The user identification role was set to user id attribute and item identification role to item id attribute. Data attributes can have arbitrary names but roles for those attributes must be set. Next, we use the train data with the appropriately set roles to train an Item k-NN model (3). At this point we can use our trained model to recommend new items to users in the query set using the Apply Model operator (6). Prior to model application, the user identification role was set for the query set (5). The Apply Model operator (6) returns an example set containing the first n ranked recommendations for every user in a query set. In Figure 3 we have seen how to make recommendations for particular users. In the following figure, Figure 4, we show how to measure performance of a recommendation model.

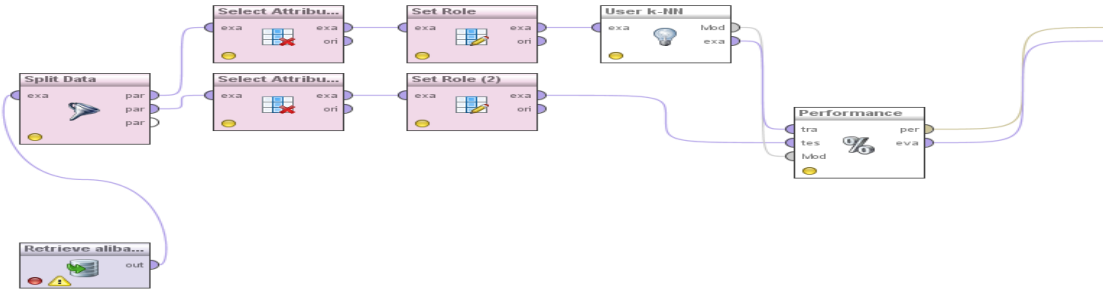


FIGURE 5 Measuring performance of a recommendation model.

The data management part of the workflow for measuring recommender model performance in Figure 5 is the same as in Figure 3. We use the Read AML operators (1,4) to load the data input, and the Set Role operators (2,5) to set the appropriate roles. In this workflow we use the test data (4) containing two attributes, the user id and the item id attribute and we set user identification and item identification roles to those at-tributes, respectively. The difference from the previous workflow is the need to calculate the performance of our built recommendation

model (3). We use the Performance operator (6) to measure standard recommendation error measures we previously defined: AUC, Prec@k, NDCG, and MAP. The Performance operator (6) returns a performance vector and an example set containing performance measures. This enables a user to choose which format suits his or her needs. We can get Figure 6.

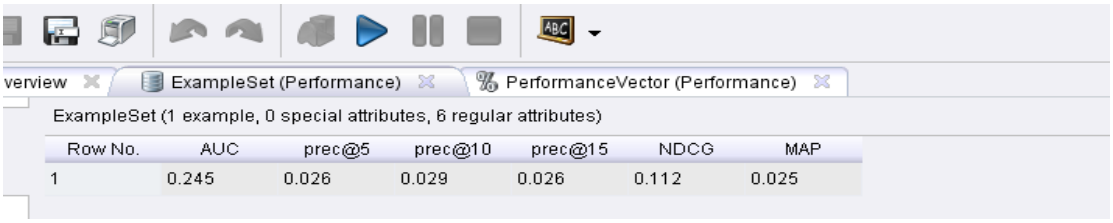


FIGURE 6 the performance of Recommender Systems

4 Conclusions

Recommender systems became essential in an information- and decision-overloaded world. They changed the way users make decisions, and helped their creators to increase revenue at the same time. Bringing recommender systems to a broader audience is essential in order to popularize them beyond the limits of scientific research and high technology entrepreneurship. The goal of the Recommender Extension for RapidMiner and this paper was to bring recommenders to a broad audience, in a theoretical, practical, and above all, application way.

In this paper we presented recommender systems and their different techniques: collaborative filtering, content-based recommender systems, and hybrid systems. We presented the advantages and disadvantages of each of those systems and demonstrated how they could be implemented easily in RapidMiner. The application of recommender systems outlined was just a small introduction to the possibilities of the extension. We hope you will use the knowledge obtained through this paper in your own applications, problems, and businesses, and that recommender systems will assist you in reaching quality, informed decisions.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61272295) and 2014 science and technology plan of Hunan province (No.2014GK3157).

Refrences

[1] Resnick P, Varian H R 1997 Recommender systems *Communications of the ACM* **40**(3) 56-8
[2] Adomavicius G, TuzhilinA 2005 *IEEE Transactions on Knowledge and Data Engineering* **17**(6) 734-49
[3] Lu L, Medo M, Zhang Y C, Yeung C H, Zhang Z K, Zhou T 2012 Recommender systems *Physics Reports* **519**(1) 1-49
[4] Rendle S, Tso-Sutter K, Huijsen W, Freudenthaler C, Gantner Z, Wartena C, Brussee R, Wibbels M 2011 Report on state of the art recommender algorithms *MyMedia public deliverable D4.1.2*. Technical report
[5] Goldberg D, Nichols D, Oki B M, Terry D 1992 Using collaborative ltering to weave an information tapestry *Communications of the ACM* **35**(12) 61-70
[6] Gunawardana A, Shani G 2009 A survey of accuracy evaluation metrics of recommendation tasks *Journal of Machine Learning Research* **10** 2935-62

Authors

Zhihang Tang, born in August 1974, Hunan, China.

Current position, grades: teacher at the department of computer and communication, Hunan Institute of Engineering (Xiangtan, China) since 2003.
University studies: PhD degree at Donghua University China in 2009.
Scientific interests: intelligent decision and knowledge management
Publications: 30 papers

Zhonghua Wen, born in May 1966-5, Hunan, China.

Current position, grades: professor at the Hunan Institute of Engineering.
University studies: PhD degree in computer engineering at Zhongshan University, China.
Scientific interests: intelligent decision and knowledge management.
Publications: 50 papers.

A study of region division based on spatial units fusion of clustering algorithm

Yu Wang*

School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, Sichuan, China

School of Traffic and Transportation Engineering, Dalian Jiaotong University, Dalian 116028, Liaoning, China

Received 1 March 2014, www.cmnt.lv

Abstract

Region division is the foundation of regional socio-economic development planning. Traditional regional division is only built upon the information of attributes in each spatial unit. However, the spatial relationships and their spatial interaction between units are ignored. In the study, based on the spatial linkage theory, the principle and steps of spatial units fusion of clustering algorithm is proposed. According the spatial range of Chinese railway as a case, determine the spatial units and the linkage model was built based on the proximity and attribute characters, the experiments of regional division for Chinese railway are achieved. The experiment shows the results are highly accordant with the real situations and have proved the feasibility of the algorithm.

Keywords: spatial linkage theory, region division, spatial units fusion of clustering algorithm, Chinese railway

1 Introduction

Regional planning is an important part of regional socio-economic development planning. Region division is the foundation of regional planning, and the division foundation in current research is classified into two types. One take the economic attribute as the main basis [1,2], as economic belt [3], economic region [4] and metropolitan area [5]. But the academe has not come to an agreement with it, the reason is the principles and methods are different and can also affect the regional planning establishment and implementation. Another take the spatial units attribute as the main basis [6]. But this method only consider the individual attribute data and the relationship are ignored in practice.

Based on the spatial linkage theory [7], starting with the relationship with the contiguous spatial units, the fusion of clustering algorithm [8,9] were studied and so as to realize the region division.

2 Spatial units fusion of clustering algorithm

2.1 ALGORITHM PRINCIPLE

Within the selected spatial range, the spatial set can be formed under the proper spatial units are selected, by computing the units correlation data, the correlation coefficient set is determined, fused with the two adjacent spatial units which has maximum correlation coefficient, connect the units in one direction according to the units attribute until there are not any external connection by the fusion units, so the units which have fused cluster a class. According to this algorithm, all the spatial units are clustered several classes in the spatial range.

2.2 ALGORITHM STEPS

Step 1: Determination of the spatial range *s*.

Step 2: Determination of the spatial units s_i:

$S = \{S_1, S_2, ..., S_i\}$.

Step 3: Determination of the model of the relation coefficient *f*.

Step 4: Take a spatial unit as the starting point S_1, computed the relation coefficient f_{1j} with the neighboring spatial units, fused with the two neighboring spatial units which has maximum relation coefficient max f_{1j}.

Step 5: Determination of the neighboring spatial units in one direction connection T_{1j}. The units attribute data is used to represent the (Q_1, Q_2), determine the connect direction T_{1j} by comparing Q_1 and Q_2, if $Q_1 / Q_j > 1$, $T_{1j} = 1$, otherwise $T_{1j} = 0$.

Thus, the two contiguous units have only one direction connection.

Step 6: In the same way, by computing as step4 and step5 until there are not any external connection by the fusion units, so the units which have fused cluster the first class C_1.

Step 7: According to this algorithm, all the spatial units are clustered several classes, $S = (C_1, C_2, ..., C_k)$.

* ***Corresponding author's*** e-mail: wangyudjtu@163.com

3 An example of region division of Chinese railway

3.1 DETERMINE THE SPATIAL RANGE

Regional railway planning is the effective link to the "Medium and Long-term Railway Network Planning" [10], so the spatial range of region division of railway transportation should be carried out within the scope of national railway space (except Hong Kong, Macau and Taiwan).

3.2 DETERMINE THE SPATIAL UNITS

Chinese railway transport mainly presented as inter-provincial transport relationship, and the volume of freight flow mostly completed between 1,000~2,000 *km* [11]. So the spatial units selection are based on the provinces, municipalities and autonomous regions. Considering the uniqueness of the adjacent space, fused the only adjacent units to each other at first and the number of spatial units are 28. As shown in Table 1.

TABLE 1 Spatial units of region division of Chinese railway

Region code	Region name	spatial unit code	Region code	Region name	spatial unit code
110000	Beijing	S_1	410000	Henan	S_{15}
120000	Tianjin	S_1	420000	Hubei	S_{16}
130000	Hebei	S_1	430000	Hunan	S_{17}
140000	Shanxi	S_2	440000	Guangdong	S_{18}
150000	Mongolia West	S_3	450000	Guangxi	S_{19}
150000	Mongolia East	S_4	460000	Hainan	S_{18}
210000	Liaoning	S_5	500000	Chongqing	S_{20}
220000	Jilin	S_6	510000	Sichuan	S_{20}
230000	Heilongjiang	S_7	520000	Guizhou	S_{21}
310000	Shanghai	S_8	530000	Yunnan	S_{22}
320000	Jiangsu	S_9	540000	Tibet	S_{23}
330000	Zhejiang	S_{10}	610000	Shaanxi	S_{24}
340000	Anhui	S_{11}	620000	Gansu	S_{25}
350000	Fujian	S_{12}	630000	Qinghai	S_{26}
360000	Jiangxi	S_{13}	640000	Ningxia	S_{27}
370000	Shandong	S_{14}	650000	Xingjian	S_{28}

3.2 SPATIAL UNITS CORRELATION MODEL

Spatial units correlation model was built from the method of spatial transport linkage theory [7], according to the data of $O-D$ matrix and considering the bidirectional transport data transfer. The model can be used to evaluate the correlation strength with the adjacent spatial units, as shown by Equation (1):

$$f_{ij} = a_i \frac{O_{ij}}{O_i} + a_i \frac{D_{ij}}{D_i} + a_j \frac{D_{ij}}{O_j} + a_j \frac{O_{ij}}{D_j}, (i, j = 1, 2, ..., n), \quad (1)$$

where f_{ij} is the correlation coefficient, O_{ij} is the freight flow from unit i to j, D_{ij} is the freight flow from unit j to i, O_i and O_j is the total output freight flow with unit i and j, D_i and D_j is the total input freight flow with unit i and j, a_i and a_j is the standardization coefficient, n is the number of spatial units.

Regional division of Chinese railway has based on the analysis on the freight flow in china. a_i and a_j were obtained from "China Statistical Yearbook for Regional Economy-2013" [12], O_{ij}, D_{ij}, O_i, O_j, D_i, D_j were obtained from "China Transportation Yearbook 2013"[13].

3.3 CALCULATION PROCESS

Take S_1 as the starting point, search for the adjacent spatial units as S_2, S_3, S_5, S_{14}, S_{15} and calculate the f_{1j}. As shown in Table 2.

TABLE 2 The f_{1j} between the adjacent spatial units

Variable / Adjacent units	a_1	O_{1j}/O_1	D_{1j}/D_1	a_j	D_{1j}/O_j	O_{1j}/D_j	f_{1j}
$S_1 - S_2$	0.96	0.03	0.53	0.88	0.49	0.05	1.00
$S_1 - S_3$	0.96	0.01	0.19	1.58	0.29	0.02	0.68
$S_1 - S_5$	0.96	0.01	0.01	1.41	0.04	0.02	0.10
$S_1 - S_{14}$	0.96	0.07	0.04	1.37	0.11	0.05	0.33
$S_1 - S_{15}$	0.96	0.02	0.01	0.81	0.01	0.03	0.02

From Table 2, we note that S_1 and S_2 have the maximum correlation coefficient, $\max f_{1j} = 1.00$. Determine S_1 and S_2 in one direction connection, the exchanges of freight volume can be regarded as (Q_1, Q_2), as $Q_i = O_{ij}$, $Q_j = D_{ij}$, $O_{12}/D_{12} = 0.02 < 1$, so the only direction connection between S_1 and S_2 was $S_1 \leftarrow S_2$.

Take S_2 as the starting point, search for the adjacent spatial units as S_1, S_3, S_{15}, S_{24} and calculate the f_{2j}. As shown in Table 3.

TABLE 3 The f_{2j} between the adjacent spatial units

Variable / Adjacent units	a_2	O_{2j}/O_2	D_{2j}/D_2	a_j	D_{2j}/O_j	O_{2j}/D_j	f_{2j}
S_2-S_1	0.88	0.49	0.05	0.96	0.03	0.53	1.00
S_2-S_3	0.88	0.01	0.01	1.58	0.01	0.02	0.06
S_2-S_{15}	0.88	0.04	0.01	0.82	0.01	0.20	0.22
S_2-S_{24}	0.88	0.01	0.02	0.90	0.02	0.04	0.07

From Table 3, we note that S_2 and S_1 have the maximum correlation coefficient, $\max f_{2j}=1.00$. Determine S_1 and S_2 in one direction connection, as $O_{21}/D_{21}=64.3>1$, so the only direction connection between S_2 and S_1 was $S_2 \to S_1$.
Take S_3 as the starting point, search for the adjacent spatial units as S_1, S_2, S_{24}, S_{25} and calculate the f_{3j}. As shown in Table 4. From Table 4, we note that S_3 and S_1 have the maximum correlation coefficient, $\max f_{3j}=0.68$.

TABLE 4 The f_{3j} between the adjacent spatial units

Variable / Adjacent units	a_3	O_{3j}/O_3	D_{3j}/D_3	a_j	D_{3j}/O_j	O_{3j}/D_j	f_{3j}
S_3-S_1	1.58	0.29	0.02	0.96	0.01	0.19	0.68
S_3-S_2	1.58	0.01	0.02	0.88	0.01	0.01	0.06
S_3-S_{24}	1.58	0.01	0.01	0.90	0.01	0.02	0.03
S_3-S_{25}	1.58	0.01	0.01	0.54	0.01	0.03	0.04

Determine S_3 and S_1 in one direction connection, as $O_{31}/D_{31}=57.6>1$, so the only direction connection between S_3 and S_1 was $S_3 \to S_1$.

Take S_{14} as the starting point, search for the adjacent spatial units as S_1, S_9, S_5 and calculate the f_{14j}. As shown in Table 5.

TABLE 5 The f_{14j} between the adjacent spatial units

Variable / Adjacent units	a_{14}	O_{14j}/O_{14}	D_{14j}/D_{14}	a_j	D_{14j}/O_j	O_{14j}/D_j	f_{14j}
$S_{14}-S_1$	1.37	0.11	0.05	0.96	0.07	0.04	0.33
$S_{14}-S_9$	1.37	0.03	0.01	1.76	0.04	0.06	0.24
$S_{14}-S_{15}$	1.37	0.10	0.02	0.82	0.04	0.13	0.30

From Table 5, we note that S_{14} and S_1 have the maximum correlation coefficient, $\max f_{14j}=0.33$. Determine S_{14} and S_1 in one direction connection, as $O_{141}/D_{114}=1.72>1$, so the one direction connection between S_3 and S_1 was $S_{14} \to S_1$. There are not any external connection by the unit S_1, so the units S_1, S_2, S_3, S_{14} which have fused cluster the first class $C_1=(S_1,S_2,S_3,S_{14})$, the fused path was $S_2 \to S_1; S_3 \to S_1; S_{14} \to S_1$.

As the calculation process is same, the other class calculation is not repeated, the fused cluster steps as shown in Table 6.

TABLE 6 The fused cluster steps of the class

Class	Step	Starting unit	Maximum relation unit	max f_{ij}	T_{ij}
C_1	1	S_1	S_2	1.01	$S_1 \to S_2$
	2	S_2	S_1	1.01	$S_2 \to S_1$
	3	S_3	S_1	0.68	$S_3 \to S_1$
	4	S_{14}	S_1	0.33	$S_{14} \to S_1$
C_2	1	S_4	S_5	0.66	$S_4 \to S_5$
	2	S_5	S_6	0.73	$S_5 \leftarrow S_6$
	3	S_6	S_5	0.73	$S_6 \to S_5$
	4	S_7	S_6	0.53	$S_7 \to S_6$
C_3	1	S_8	S_{10}	0.51	$S_8 \leftarrow S_{10}$
	2	S_9	S_8	0.48	$S_9 \to S_8$
	3	S_{10}	S_8	0.51	$S_{10} \to S_8$
	4	S_{11}	S_9	0.68	$S_{11} \to S_9$
C_4	1	S_{12}	S_{13}	0.66	$S_{12} \to S_{13}$
	2	S_{13}	S_{12}	0.66	$S_{13} \leftarrow S_{12}$
C_5	1	S_{15}	S_{16}	0.24	$S_{15} \leftarrow S_{16}$
	2	S_{16}	S_{15}	0.24	$S_{16} \to S_{15}$
	3	S_{24}	S_{16}	0.23	$S_{24} \to S_{16}$
C_6	1	S_{17}	S_{18}	0.81	$S_{17} \leftarrow S_{18}$
	2	S_{18}	S_{17}	0.81	$S_{18} \to S_{17}$
C_7	1	S_{19}	S_{21}	0.54	$S_{19} \to S_{21}$
	2	S_{20}	S_{21}	0.67	$S_{20} \leftarrow S_{21}$
	3	S_{21}	S_{20}	0.67	$S_{21} \to S_{20}$
	4	S_{22}	S_{20}	0.43	$S_{22} \to S_{20}$
C_8	1	S_{23}	S_{26}	0.11	$S_{23} \leftarrow S_{26}$
	2	S_{25}	S_{27}	0.29	$S_{25} \leftarrow S_{27}$
	3	S_{26}	S_{25}	0.13	$S_{26} \leftarrow S_{25}$
	4	S_{27}	S_{25}	0.29	$S_{27} \to S_{25}$
	5	S_{28}	S_{25}	0.1	$S_{28} \to S_{25}$

The transport relation with spatial units has its own characteristics and the characteristics distributed on the geospatial form the different transportation regions.Shown in Table 7.

TABLE 7 Name and composition of transportation regions

No.	Code	Name of region	Composition of region
1	C_1	Northern	Hebei, Beijing, Tianjin, Shanxi, Mongolia West, Shandong
2	C_2	Northeast	Heilongjiang, Jilin, Liaoning, Mongolia East
3	C_3	East	Shanghai, Jiangsu, Anhui, Zhejiang
4	C_4	Southeast	Fujian, Jiangxi
5	C_5	Central	Shaanxi, Henan, Hubei
6	C_6	Southern	Guangdong, Hainan, Hunan
7	C_7	Southwest	Guangxi, Guizhou, Sichuan, Chongqing, Yunnan
8	C_8	Northwest	Gansu, Ingria, Xingjian, Qinghai, Tibet

4 Conclusion

Traditional regional division model for spatial units is only built upon the information of attributes in every unit. However, the spatial relationships and their spatial interacttion between units are not considered sufficiently. In this study, based on scale-space theory, the spatial units cluster fusion algorithm is proposed so that the elements of spatial relationship between units could be integrated besides considering the information of attributes. By this approach, considered both of the spatial proximity and attribute similarity at certain spatial range, the spatial units could be clustered into one class if their connective direction is the same and achieve the regional division.

This paper uses the spatial transport linkage theory to construct the spatial units correlation model for railway freight transportation regional division. According to the data of $O-D$ matrix with the adjacent spatial units and considering economic parameters as limitation, the regional division is divided 8 parts of railway freight transporttation in china. As Northern, Northeast, East, Southeast, Central, Southern, Southwest and Northwest. As shown in Figure 1. The results are highly accordant with the real situations and can offer spatial background for regional railway freight transportation planning and development.

Further research is also proposed to improve spatial units correlation model. The data selection should not only reflect the present but also reflect the future trends of the regional development. Economic indicators should be further improved when considering the link between transportation and economic and also considering the impact of macroeconomic policies and geographic and other factors and make the model more comprehensive.

FIGURE 1 The spatial layout of railway freight transportation regional division

Acknowledgments

The project has been supported by National Natural Science Foundation ,China (Grant No. 71173177) and International Exchange Programs, Muroran Institute of Technology, Japan.

References

[1] Fujita M, Krugman P, Venables A J 2001 The spatial economy: cities, regions and international Trade Models of Disorder *MIT Press:* Edinburgh – Cambridge Chapter 6
[2] Capello R 2007 Regional Economics *Routledge Chapman & Hall Press:* London Chapter 4
[3] Zhang Z Z 2010 Evolution and evaluation of the Chinese economic regions division *Journal of Shanxi Finance and Economics University* **13** (2) 89-92 *(in Chinese)*
[4] Bureau of Economic Analysis 2010 *Regional economic measurement division: REIS Regional Economic Information System, 1969-2012"* US
[5] Yang L, Yang K Z 2007 Metropolitan-oriented regionalization of japan and lessons for china *Asia-pacific Economic Review* **33**(2) 113-7*(in Chinese)*
[6] Yuan C W, Wang J W, Lin W X 2006 Division method of transportation regions in china *Journal of Chang'an University* **26**(4) 79-82 *(in Chinese)*

[7] Zhang W C, Jin F J, Rong C H, Tang X F 1994 Analysis on spatial transport linkage *China railway Press:* Beijing Chapter 4
[8] Lindeberg T 1994 Scale-space theory in computer vision *Kluwer Press:* Netherlands Chapter 2
[9] Leung Y, Zhang J, Xu Z 2002 Clustering by Scale-space filtering *IEEE transactions on pattern analysis and machine Intelligence* **22**(12) 1396-410
[10] Ministry of Railways 2008 Medium and Long-Term Railway Network Plan (adjusted in 2008) *China railway Press:* Beijing Chapter 1
[11] Ren M 2007 Analysis on the railway freight exchange characteristics of provincial region *Chinese railways***23**(10)25-28*(in Chinese)*
[12] National Bureau of Statistics, The competent national transport ministries Bureau and validation 2013 China Transportation Yearbook *China statistical Press:* Beijing Chapter 4
[13] National Bureau of Statistics 2013 China Statistical Yearbook for Regional Economy – 2013 *China statistical Press:* Beijing Chapter 12

Author

Yu Wang, born in 1979, Dalian, Liaoning, China.

Current position, grades: lecturer at the School of Traffic and Transportation Engineering, Dalian Jiaotong University, China.
University studies: B.E. at Southwest Jiaotong University in China. ME at Southwest Jiaotong University in China.
Scientific interests: traffic and transportation planning and management.
Publications: more than 10 papers.
Experience: teaching experience of 5 years, presided 1 scientific research project.

A FAHP-based comprehensive evaluation on rural supermarket service quality: a case study of Jiangsu province

Tao Meng*

Department of Economics and Management, Huaiyin Normal University, No.71 Jiaotong Rd, Huai'an, China

Received 26 April 2014, www.cmnt.lv

Abstract

With the changes of market supply and demand, as well as increasing business competition, service quality has become a key factor constraining rural supermarket survival and development. This article targets at the rural supermarkets in Jiangsu Province. In the context of constructing an indicator system of rural supermarket service quality evaluation, it acquires data by questionnaires, uses AHP to establish the weights, and applies the multi-level fuzzy AHP to assess rural supermarket service quality. It concludes that there are some spaces for further improvement of rural supermarket service quality and proposes relevant solutions.

Keywords: FAHP, rural supermarkets, service quality

1 Introduction

In the urbanization process of China, the primary task should be the construction of rural infrastructure. In order to improve the life quality and convenience of rural residents, we should focus on the improvement of rural supermarket. However, in China the development of rural supermarket is still in its infancy. There are many problems in operations. Rural supermarket service quality needs to be improved. First of all, we must find out the factors affecting the quality of service and why consumers are not satisfied with rural supermarket service quality, which is the precondition of improving the quality of service and market competition. Make prospect for the improvement of rural supermarket service quality and promote the sustainable healthy development of rural supermarket [1]. Previous studies show that there is a linear positive correlation between customer satisfaction and customer loyalty (Chen Yuan, 2012). Besides, the quality of service affects customer loyalty, both directly and indirectly (Chen Shaojie, 2011). Lots of studies adopt a perceptual model or structure equation model to discuss the service quality of supermarket chains, but seldom specifically focus on rural supermarket service quality. Few articles concerning rural supermarket service quality use the grey relational evaluation method and grey cluster model (Yang Yanbo, et al., 2008), and service quality gap model (Liang Na, et al., 2009). In 2008, Guo Yuexian et al. proved that the FAHP model was appropriate for the evaluation of rural supermarket service quality. Therefore, this article targets at rural supermarkets in Jiangsu Province and adopts the FAHP model to assess their service quality, in the hope of finding out solutions for rural supermarkets improving service quality to keep same space with modern social development in the new urbanization process.

2 Generation of the data

2.1 FAHP COMPREHENSIVE EVALUATION MODEL

FAHP, namely fuzzy analytic hierarchy process, is a comprehensive evaluation method to assess the hierarchical condition of evaluation target from multiple factors, which is on the basis of fuzzy mathematics, applies the principle of fuzzy relationship synthesis, and quantifies some ill-defined and difficult-to-quantify factors [2]. As we analyze a large system, we should take lots of factors into consideration, while there are different levels of the hierarchy between these factors. Thus, we need to collect the evaluation factors and divide them into different categories according to some set of attributes. Perform a comprehensive evaluation on each category firstly, and then make the high-level comprehensive evaluation on the results, so that the final evaluation will be more comprehensive and scientific [3].

The FAHP evaluation procedures are shown as follows:

- Establish the evaluation factors sets. Assume the fist-level evaluation factors set $A=\{\alpha_1,\alpha_2,.....\alpha_n\}$ and the second-level evaluation indicator a_{ij}, where $i=1,2,...,n; j=1,2,...,m$ n the number of first-level indicators, and m the number of second-level indicators beneath each first-level indicator.

- Establish the evaluation results sets. Assume the set of all evaluation results of each level $B=\{\beta_1,\beta_2,\cdots,\beta_s\}$. Define them based on actual situations, such as "better", "good", "general", "bad", and "worse" the five levels.

* ***Corresponding author's*** e-mail: mtmxb@163.com

- Establish the weights of each evaluation indicator at each level by means of the analytic hierarchy process. Firstly make pair comparison of indicators based on judgment scale table [4] and expert scoring method and construct the judgment matrix of evaluation indicator factors X (Equation (1)). Calculate the sum of every row of judgment matrix and make it normalization. Get the weight vector δ_i (Equation (2)). In order to improve the reliability of weights, we need to apply the consistency ratio CR (Equation (3)) to conduct a consistency test on the judgment matrix.

$$X = \begin{bmatrix} \chi_{11} & \chi_{12} & \cdots & \chi_{1n} \\ \chi_{21} & \chi_{22} & \cdots & \chi_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ \chi_{n1} & \chi_{n2} & \cdots & \chi_{nn} \end{bmatrix}, \tag{1}$$

$$\delta_i = \frac{\sum_{j=1}^{n} \chi_{ij}}{\sum_{k=1}^{n}\sum_{j=1}^{n} \chi_{kj}} \quad i = 1,2,...,n, \tag{2}$$

$$CR = \frac{CI}{RI} = \frac{(\lambda_{\max} - n)/(n-1)}{RI}, \tag{3}$$

where, $\lambda_{\max}$ is the maximum characteristic value. RI is the average random consistency index. As $CR < 0.1$, the degree of inconsistency of X is within the allowable range. Otherwise, re-construct the matrix.

- Build the single-factor fuzzy evaluation matrix. Use the membership degree of fuzzy mathematics to make synthesis and get the single-factor fuzzy evaluation matrix (Equation (4)).

$$Y_t = \begin{bmatrix} \eta_{11} & \eta_{12} & \cdots & \eta_{1m} \\ \eta_{21} & \eta_{22} & \cdots & \eta_{2m} \\ \cdots & \cdots & \ddots & \cdots \\ \eta_{n1} & \eta_{n2} & \cdots & \eta_{nm} \end{bmatrix}, \tag{4}$$

where $\eta_{ij} = \frac{q_{ij}}{Q}$. The value is between 0 and 1. Here, η_{ij} refers to the membership degree of the comment β_j of indicator α_i, and $i = 1,2,...,n; j = 1,2,...,m$. q_{ij} is the number of the indicator i at the level j. Q is the total number of respondents in the survey[2].

- The multi-factor fuzzy comprehensive evaluation matrix. It is chiefly to construct the multi-level comprehensive evaluation matrix based on primary comprehensive evaluation matrix. The primary fuzzy comprehensive evaluation matrix is Equation (5), and the general fuzzy comprehensive evaluation matrix is Equation (6).

$$R_i = X_i \times Y_i, \tag{5}$$

$$R = X \times Y. \tag{6}$$

- Calculate the comprehensive score of three-level evaluation. Firstly get the score of each item C_j (Equation (7)). Then, get the score of each dimension F_i (Equation (8)). And finally get the total score of service quality Q (Equation (9)).

$$C_j = V^T \times Y_j, \tag{7}$$

$$F_i = V^T \times R_i, \tag{8}$$

$$Q = V^T \times F_i. \tag{9}$$

Classify the results and establish the membership degree of comprehensive scores on the principle of weighted average. Results are shown in Table 1.

TABLE 1 the level of comprehensive scores

Level	Evaluation Results	Comprehensive Scores
*E*1	better	[0, 1.5]
*E*2	good	(1.5, 2.5]
*E*3	general	(2.5, 3.5]
*E*4	bad	(3.5, 4.5]
*E*5	worse	(4.5, 5]

2.2 SELECT THE RIGHT EVALUATION INDICATORS

As early as 1955, Parasuraman Zeithaml & Berry defined service quality as the gap of customers' expectation for services and their feelings about actual services, namely the service quality = expected service – perceptual service. They further established the "five dimensions" for evaluating the level of service quality, i.e. the tangibles, responsiveness, reliability, assurance, and empathy of customers [5]. According to the five-dimension factors and considering the rural supermarket service conditions in Jiangsu Province and the expert investigation, we establish the indicator system for evaluating the rural supermarket service quality in Table 2.

TABLE 2 The indexes for rural supermarket service quality evaluation

Dimension	Number	The Second-level Evaluation Factors
Tangibles α_1	α_{11}	appropriate clothes for salesperson
	α_{12}	comfortable shopping environment
	α_{13}	abundant commodities
	α_{14}	size of shop
	α_{15}	diversified payment
Responsiveness α_2	α_{21}	solving problems with patience
	α_{22}	special accelerated service
	α_{23}	fast payment
	α_{24}	easy counsel from salesperson.
Reliability α_3	α_{31}	complete the committed service
	α_{32}	goods return or exchange rate
	α_{33}	convenient working time
	α_{34}	accurate and safe settlement.
Assurance α_4	α_{41}	commodity quality
	α_{42}	professional knowledge of employee
	α_{43}	employees' reliability
Empathy α_5	α_{51}	personalized service
	α_{52}	respect for customers
	α_{53}	understand customers' needs
	α_{54}	sincerely care about customers

2.3 QUESTIONNAIRE DESIGN AND COLLECTION

In this survey, we design each item in the questionnaire according to the second-level indicator in the evaluation indicator system, including 5 first-level indicators, 20 second-level indicators, i.e. the questionnaire contains 20 questions. Each question has five options, i.e. better, good, general, bad, and worse, and each option is given certain score, i.e. 5, 4, 3, 2, and 1 respectively. This questionnaire focuses on the customers in rural supermarkets in Jiangsu Province. We use a sample survey, handing out 150 copies, and finally collecting 145 copies. The recovery rate is 96.7%. Exclude 5 invalid questionnaires. Then, the recovery rate is 93.3%.

2.4 DATA PROCESSING

Use the fuzzy comprehensive evaluation method to analyze the data from the survey on rural supermarket service quality in Jiangsu Province. Detailed calculations and analyses are displayed as follow.

Establish the set of rural supermarket service quality evaluation factors, $A=\{\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5\}$. Here, the first-level evaluation factors are respectively: tangibles α_1, responsiveness α_2, reliability α_3, assurance α_4 and empathy α_5. The second-level evaluation factors α_{ij} refer to the specific evaluation indicators of the first-level evaluation factors. The tangibles α_1 contain five second-level indicators, namely five questions, i.e. α_{11} appropriate clothes for salesperson, α_{12} comfortable shopping environment, α_{13} abundant commodities, α_{14} size of shop, and α_{15} diversified payment. The responsiveness α_2 contains four second-level indicators, namely four questions, i.e. α_{21} solving problems with patience, α_{22} special accelerated service, α_{23} fast payment, and α_{24} easy counsel from salesperson. The reliability α_3 contains four second-level indicators, namely four questions, i.e. α_{31} complete the committed service, α_{32} goods return or exchange rate, α_{33} convenient working time, and α_{34} accurate and safe settlement. The assurance α_4 contains three second-level indicators, namely three questions, i.e. α_{41} commodity quality, α_{42} professional knowledge of employee, and α_{43} employees' reliability. The empathy α_5 contains four second-level indicators, namely four questions, i.e. α_{51} personalized service, α_{52} respect for customers, α_{53} understand customers' needs, and α_{54} sincerely care about customers.

Establish the evaluation set of rural supermarket service quality $V=\{better, good, general, bad, worse\}$, namely $V=\{5,4,3,2,1\}$.

Use the AHP to establish the weights. Respectively get the first-level evaluation matrix according to the Equation (1) and (3).

$$X=\begin{bmatrix} 1 & 3 & 5 & 2 & 3 \\ \frac{1}{3} & 1 & 2 & \frac{1}{2} & 2 \\ \frac{1}{5} & \frac{1}{5} & 1 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1 & 3 & 1 & 2 \end{bmatrix}.$$

Run the MCE comprehensive evaluation software and get $\lambda_{\max}=5.0755$, $CI=0.0189$, $RI=1.12$.

Then $CR=0.0169<0.1$, which meets the consistency test of evaluation matrix. According to the Equation (2), the weight of first-level evaluation indicator is $\delta=(0.413,0.155,0.074,0.241,0.117)$.

Similarly, get the weight vector of second-level indicator, i.e. respectively:

$$\delta_1=(0.096,0.214,0.479,0.157,0.054),$$

$$\delta_2=(0.488,0.089,0.248,0.175),$$

$$\delta_3=(0.274,0.158,0.065,0.503),$$

$$\delta_4=(0.630,0.218,0.152),$$

$$\delta_5=(0.136,0.280,0.114,0.470).$$

Construct the single-factor fuzzy evaluation matrix for rural supermarket service quality. According to Equation (4), respectively the fuzzy evaluation matrixes of the five dimensions of first-level indicators are gotten:

$$Y_1=\begin{bmatrix} 0.086 & 0.243 & 0.471 & 0.171 & 0.029 \\ 0.114 & 0.414 & 0.357 & 0.1 & 0.015 \\ 0.072 & 0.544 & 0.258 & 0.126 & 0 \\ 0.136 & 0.264 & 0.493 & 0.1 & 0.007 \\ 0.229 & 0.614 & 0.157 & 0 & 0 \end{bmatrix},$$

$$Y_2=\begin{bmatrix} 0.121 & 0.272 & 0.407 & 0.121 & 0.079 \\ 0.214 & 0.136 & 0.314 & 0.157 & 0.179 \\ 0.157 & 0.372 & 0.443 & 0.021 & 0.007 \\ 0.129 & 0.186 & 0.529 & 0.114 & 0.042 \end{bmatrix},$$

$$Y_3=\begin{bmatrix} 0.2 & 0.25 & 0.436 & 0.078 & 0.036 \\ 0.343 & 0.307 & 0.264 & 0.072 & 0.014 \\ 0.271 & 0.3 & 0.429 & 0 & 0 \\ 0.35 & 0.307 & 0.2 & 0.114 & 0.029 \end{bmatrix},$$

$$Y_4=\begin{bmatrix} 0.121 & 0.493 & 0.336 & 0.036 & 0.014 \\ 0.164 & 0.072 & 0.686 & 0.057 & 0.021 \\ 0.093 & 0.157 & 0.7 & 0.036 & 0.014 \end{bmatrix},$$

$$Y_5=\begin{bmatrix} 0.036 & 0.243 & 0.557 & 0.093 & 0.071 \\ 0.272 & 0.307 & 0.35 & 0.05 & 0.021 \\ 0.007 & 0.072 & 0.914 & 0.007 & 0 \\ 0.021 & 0.086 & 0.893 & 0 & 0 \end{bmatrix}.$$

Conduct the fuzzy comprehensive evaluation. According to Equations (5) and (6), get the fuzzy comprehensive evaluation matrix vector of each dimension and the fuzzy

comprehensive evaluation matrix vector of overall service quality, respectively:

$$R_1 = X_1 \times Y_1 = (0.101, 0.447, 0.331, 0.114, 0.007),$$

$$R_2 = X_2 \times Y_2 = (0.140, 0.270, 0.429, 0.098, 0.063),$$

$$R_4 = X_4 \times Y_4 = (0.126, 0.350, 0.468, 0.041, 0.015),$$

$$R_5 = X_5 \times Y_5 = (0.092, 0.168, 0.698, 0.027, 0.015),$$

$$R = X \times Y = (0.127, 0.352, 0.419, 0.082, 0.020).$$

Calculate the comprehensive score of three-level evaluation. Firstly get the score of each question C_j and introduce them into Equation (7). Obtain the matrix of each question score:

$$C_1 = (3.186, 3.512, 3.562, 3.422, 4.072),$$

$$C_2 = (3.235, 3.049, 3.651, 3.246),$$

$$C_3 = (3.500, 3.893, 3.842, 3.835),$$

$$C_4 = (3.671, 3.301, 3.279),$$

$$C_5 = (3.080, 3.759, 3.079, 3.128).$$

Then, calculate the score of each dimension F_i and get the matrix of each dimension score by Equation (8):

$F_1 = 3.521$,
$F_2 = 3.324$,
$F_3 = 3.753$,
$F_4 = 3.531$,
$F_5 = 3.292$.

Finally, calculate the total score of service quality by Equation (9): $Q = 3.484$.

3 Results analysis

According to the fuzzy comprehensive evaluation scores, we know that:

Firstly, the comprehensive score of current rural supermarket service quality in Jiangsu Province is 3.484, in the range of (2.5, 3.5), at the general service level of E3. Therefore, we need to make adjustment and reformation as a whole immediately, implement service management and innovation, and improve the overall service quality of rural supermarket.

Secondly, the comprehensive scores of five dimensions (tangibles, responsiveness, reliability, assurance, and empathy) are respectively 3.521, 3.324, 3.753, 3.531, and 3.292. Obviously, for rural supermarkets, the tangibles, reliability, and assurance win relatively better evaluations from customers. But in terms of responsiveness and empathy, customers' evaluation is just ordinary.

Finally, according to the detailed analysis of all items, rural supermarkets win better evaluation in terms of tangibles, which is represented by 3.512 grades for comfortable shopping environment, 3.562 for abundant commodity, and especially 4.072 for payment. Customers agree that the three aspects are better in general. As for the clothes of salesperson and the size of shop, the score is respectively 3.186 and 3.422, what are relatively lower. The scores of questions for reliability spread in the range of (3.5, 4.5]. The evaluation is relatively better. As for the assurance, the commodity quality is scored 3.671, which is better. Employees' professional knowledge and employyees' reliability is respectively scored 3.301 and 3.279, which is general. Although the evaluation on responseveness and empathy is both general, the payment speed under the responsiveness is scored 3.651 and the respect for customers under empathy is scored 3.759. The two items are better. In addition, the general evaluation on rural supermarket service quality also includes personalized service, understanding of customers' needs, sincerely caring about customers, solving problems with patience, providing fast service, and convenience of consulting salesperson.

4 Conclusions and suggestions

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads – the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar.

4.1 CONCLUSIONS

Whether from the comprehensive scores of dimensions or the score of each item, the evaluations are mainly at the level *E*3 and *E*4, which means "general" or "better", but no "good". Therefore, there is a large room for rural supermarkets improving their service management and innovation. As for the "better" aspects, the supermarkets have more comfortable shopping environment, compared with the traditional rural purchasing environment. The payment channels are diversified, such as card, cash, or membership, which makes shopping more convenient. Besides, rural supermarkets keep the good traditions, such as being loyal to commitment, quick for return or exchange goods, convenient hours, accurate settlement, and safety. All these characteristics make customers feel more reliable. Supermarkets give assurance for qualified products because of their inbound channels. And the fast settlement gives rural customers different shopping experiences. Thus, it gets higher score for the item of respecting for customers.

For some "general" evaluations on the clothes of salespersons and the size of shops, the reason is that rural supermarkets are still at the starting stage, plus the weak rural consumption capacities and management abilities, compared with cities. Therefore, the clothes of salespersons and the size of shops in rural supermarkets are lagged behind cities. The scores of the two items are relatively lower. In rural supermarkets, the employees are mainly from local. The civilization quality and professional knowledge are insufficient. However, by means of pre-service training, they can be equipped with basic techniques for the job. In serving customers, they are reliable and basically understand customers' needs. They can genuinely care about customers, help them solve problems with patience, and provide special fast services. The salespersons are familiar with their commodities and capable of responding to customers' various questions. Only under the circumstance of all aspects with better scores, can it trigger rural customers' needs for personalized services. In short, rural supermarkets perform not too bad or too excellent in rural area. Therefore, it is necessary to further strengthen the service management and implement service innovation in order to the sustainable development of rural supermarkets.

4.2 SUGGESTIONS

4.2.1 Perform management innovation and improve service image

By taking references from the successful management pattern and approaches of urban supermarkets, we can improve rural supermarkets' shopping environment and convenience on the basis of rural conditions, guaranteeing customers' satisfaction and loyalty. In rural supermarkets, the light should be bright, the signs should be lean and eye-catching, the salespersons should be wearing neat uniforms, and the shelves should be arranged appropriately. A comfortable environment could stimulate customers' desire for buying. Maintain diversified payment methods and update the hard facilities frequently, such as tickers and shopping carts. Aim at the "one-stop" purchase and provide all kinds of commodities, including food, clothing, decorations, and daily appliances. Ensure customers can enjoy all the convenience of rural supermarkets. Besides, rural supermarkets must keep to their commitments and accomplish the work with high requirements. Guarantee fast and timely commodity return and exchange. As the saying goes, "success depends on details". Make innovations from details. For example, adjust commodity structure periodically and meet customers' constantly changing needs and hobbies. Offer special channels for the weak. Present some small knowledge cards or magazines at the checkouts. Provide shuttle vehicles for some festivals in rural area. We should run the rural supermarkets with the ideology of serving customers and ensure the image of supermarkets in the minds of customers.

4.2.2 Pay attention to employees' knowledge reconstruction

In rural supermarkets, the employees are mostly from local area. Most of them did not get the chance of completing the college study for various reasons. However, they have the desire for knowledge deep down in the heart. Managers should pay attention to employees' development and cultivation, offering relevant trainings for special techniques, service management, and CRM. Perform target trainings. For example, help cashiers to accelerate the scanning speed and make them be familiar with all kinds of payments. Give special trainings for people displaying the shelves. Find out management talents based on special positions and give them a chance of promotion. By means of investting in employees and helping them achieve the reconstruction of knowledge, rural supermarkets can stimulate the employees' enthusiasm for their work, ensuring better services for customers.

4.2.3 Strengthen employees' awareness of service

Typically, in order to increase sales, supermarkets arrange more salespersons at the center or the front, where the most popular or promotional commodities locate. While at the front of shelves with common commodities, only one or two salespersons are necessary. They can give immediate responses to customers' needs. For this kind of arrangement, the precondition is that the salespersons must be familiar with supermarkets' overall structure and knowledge of commodities. Once they catch the confusing faces of customers, they must take initiatives to offer their help. Ask whether the customer need help and give answer for any question with patience. Help employees build the awareness of service. Everything must take customers' interests as the primary and fundamental point. Try to improve customers' satisfaction and loyalty. Solve customers' problems properly and timely. Each employee should focus on establish positive relationships with customers, serve them with passions, and communicate with them actively. Respect for customers and genuinely care about customers. Provide qualified services with positive attitudes. Train employees with capabilities of communicating and coordinating with customers. When customers give advices for certain employee's services, the employee should take the advice and promise to make an improvement. Supermarkets should provide convenient ways for customers' complaints. For example, give the free phone number for complaints and present detailed directions for customers how to make complaints. Improve the transparency of receiving and handling complaints and set up rewards for customer complaints. Construct the mechanism of encouraging employees to actively take and deal with customer complaints, so that customers will grow great enthusiasm for expressing complaints or getting consultation. By this way, supermarkets can collect more advices and serve customers better.

4.2.4 Provide personalized and humanized services

In the process of communicating with customers, the employees usually focus on promoting the commodity rather than emotional exchange, which might trigger the disgusting feelings among customers. The supermarkets should advocate the emotional services. As contacting with customers, employees should be active and enthusiastic, with smiles all the time, instead of blindly giving product bombing. Salespersons should establish a nice and friendly relationship with customers based or special knowledge and life experiences, cultivating a warm and kind communication atmosphere. Salespersons should introduce not only the name or the price of commodity, but also the ingredients, the main performance, the usage, and the maintenance. Different customers possess different expectations and requirements for commodities in supermarkets, which require supermarkets segment customer groups, analyze the purchase behaviors and motives of different groups, understand there practical and potential needs, provide the right commodities to satisfy or exceed their needs and expectations. Provide some convenient services. For example, set up a playing zone at the toys' area. Ask one or two employees to be in charge of customers' parking and look after the cars for free. Build a seating area and automatic lockers.

Acknowledgments

The research findings of this paper are supported by 2011 Young Talents Support Project from HNU (11HSQNS19) and 2012 MOE (Ministry of Education in China) Youth Project of Humanities and Social Sciences (No.12YJC63 0299).

References

[1] Liang N, An Y 2009 Study on the promotion of service quality of rural supermarket Chains *Agricultural Outlook* **12** 1-3 *(in Chinese)*
[2] Man J 2013 Evaluation of Green building based on fuzzy comprehensive analysis *Electronic Test* **16** 31-2 *(in Chinese)*
[3] Lv W 2013 A study of applying grey integrated assessment to cadre promotion *China Management Informationization* **16**(16) 60-1 *(in Chinese)*
[4] Guo Y, Yang Y 2008 Research on comprehensive assessment model of service quality in rural supermarkets based on FAHP *Journal of Shijiazhuang Railway Institute (Social Sciences)* **2**(4) 10-8 *(in Chinese)*
[5] Deng F 2011 Relationship of Service Quality and Customer Loyalty *Economy and Management Publishing House:* Beijing 56-7 *(in Chinese)*

Author

Meng Tao, born in July 1979, Xuzhou, China.

Current position, grades: lecturer at Huaiyin Normal University, Huai'an, Jiangsu, China.
University studies: master's degrees in Management Science and Engineering at Xi'an University of Technology of China in 2005.
Scientific interests: service quality, service management, service innovation.
Publications: 10 papers.

Face recognition algorithms for embedded entrance guard system

Jun Li *

Institute of Computer Science, Guangdong Polytechnic Normal University, 510665, Guangzhou, P.R. China

Received 1 March 2014, www.cmnt.lv

Abstract

Biometric identification technology is used to identify individuals based on their unique physiologica1 characteristics under sampling and measurement. The hardware platform of our entrance guard system is based on SAMSUNG S3C6410 embedded development board with ARM11 processor, and we adopt ordinary camera as facial image capture device, 7-inch touch-screen LCD display as an input, electronic door locks as executing components by the onboard port control. The software platform is based on Linux operating system to design and optimize face recognition algorithm, which is combined with local binary pattern (LBP) algorithm, principal component analysis (PCA) algorithm, and ridge regression algorithm. Also, the face recognition system is divided into two parts, namely, a face database training part and a face recognition part. Precisely, the face database training part acquires face information through the processes of face detection, feature extraction and stores the face information to serve as the detection basis of the face recognition part; then the face recognition part is connected with the entrance guard system and sends out an instruction for executing related action to the entrance guard system according to the face recognition result. Specially, the embedded face recognition entrance guard system is contactless and easy to collect the facial database, less in power consumption, low in cost, and has easy installation and stable performance.

Keywords: face recognition, local binary patterns, principal components analysis, embedded system

1 Introduction

The face recognition process is to identify or verify one or more persons in the scene using a stored database of faces for some given still or video images of a scene. Compared to fingerprint recognition, retina recognition, iris recognition, gene recognition, handwriting recognition, voiceprint recognition, the advantages of face recognition are hidden operation and contactless collection, and it accords with the human recognizing habits. By further analyzing the facial expression and pose, one can get some information that is not available on other recognition systems.

Biometric identification technology is used to identify individuals based on their unique physiologica1 characteristics under sampling and measurement. Since biometric characteristics are not easy to be stolen as the certificate holdings, and be forgotten or cracked as passwords, biometric identification has been extensively applied in public security for these unique benefits of biometric characteristics. And recently, face recognition technologies have been extensively studied and applied in China and abroad. In particular, the face recognition system based products lead to small volume and low power, meanwhile the face recognition technologies lead to contactless mode, thus one need not worry the situation whether they would be willing to touch the collection equipment because of public health. On the other hand, current application of face recognition technology has three main respects, i.e., criminal detection, identification inspection, and access control. For example, at the entrance of the building, district, factories, house and the parking lot, the frequent in-and-out causes high management difficulty in verifying procedure by security personnel. However, face recognition technology can be used in entrance guard system to realize the entrance controlling based on face recognition, supplemented by other recognition [1-3].

Traditional face recognition entrance guard system is based on the PC or server, which has some defects such as high cost and large size. With the rapid development of embedded technology, the face recognition for embedded application system is becoming increasingly intelligent, miniaturized and portable.

2 Overall design

Face recognition system consists of five modules: facial image acquisition module, image pre-processing module, face detection and localization module, facial feature extraction module, and face classification module. The overall design of entrance guard system is (Figure 1):

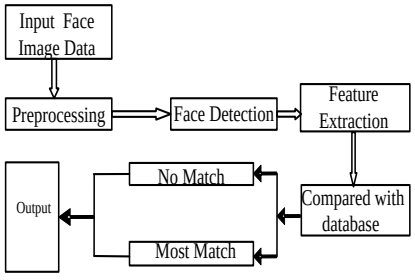


FIGURE 1 The overall design of Entrance Guard System

* ***Corresponding author's*** e-mail: lijunjanet@gmail.com

The face recognition for embedded entrance guard system is the aggregation of computer hardware and software, including the face recognition subsystem and entrance guard system, wherein the main function of the face recognition subsystem is to collect face information and execute and output the results. Additionally, main related hardware and software include camera, computer hardware system, Linux operating system, Open Source Computer Vision Library (Open CV), and some self-designed application software. Under such entrance guard system, one can realize the following modules:

1) Face Detection Module: in the system initialization phase, the computer will detect the image captured from the camera whether or not a human face exists.
2) Face Recognition Module: in the face detection phase, after collecting the face information, do the face localization, feature extraction and comparison with database, and output the results.
3) Face Register Module: the registration function is enabled to make a new record input after successfully recognizing a person's face. To improve the safety, only the users whose face information is collected in the database can be authorized to make a register.
4) Face Training Module: the training function is to train the face in the database. In fact, one extracts the face feature and then classifies and labels for saving. The purpose of the face training database is to compare with the extracted feature by face recognition process and then output the results.
5) Access Control Module: the entrance guard system is one part of the hardware platform, whose main function is to receive signals from face recognition software system, then it will drives the switch of electronic lock.

3 Hardware design

The hardware platform is based on the Samsung S3C6410 board with ARM11 processor,600MHz CPU frequency, 256MB RAM memory capacity, 2GB NAND Flash capacity, and we adopt ordinary USB camera as facial image capture device, 7-inch LCD screen as display device, touch screen as an input, electronic door locks as executing components by the onboard port control [4].

The drive of face recognition for entrance guard system is used to drive the electronic pin, whose function is to control the switch of the door. When the input is connected to the S3C6410 board, it will detect and successfully recognize the facial information, then the development board outputs high level signals, turns on the thyristor and suctions the electronic pin. Herein the electronic pin is a coil with one diode in each end in order to avoid the breakdown of thyristor by instant high pressure. The manual control switch is not only used to suction the electronic pin, but also reset it by making the signals of the anode and cathode of thyristor in the low level.

The layout of embedded entrance guard system is shown (see, Figure 2), in which the electronic pin is situated above the door.

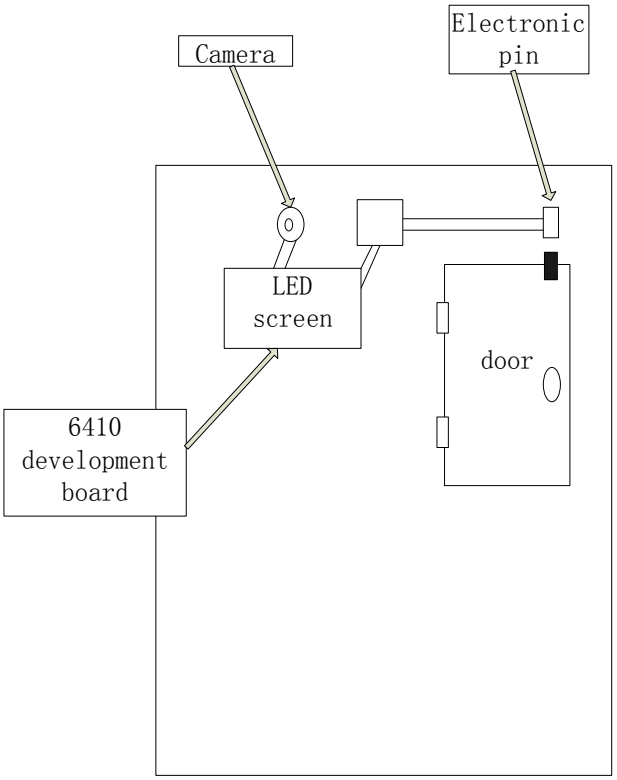


FIGURE 2 The layout of the embedded entrance guard

The door will open when the electronic pin suctions, while the door will close when the electronic pin resets.

4 Algorithm design

The face recognition system is composed of facial image acquisition, facial image pre-processing, face detection and localization, facial feature extraction, and face classification. The face recognition algorithm is combined with local binary pattern (LBP) algorithm, principal component analysis (PCA) algorithm, and ridge regression algorithm. The flow chart of face recognition system is shown (Fig. 3):

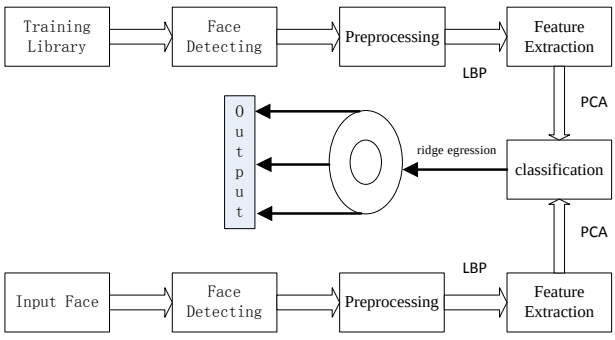


FIGURE 3 The flow chart of face recognition system

Firstly, the LBP algorithm can describe some tiny features (light spots, dark spots, stable area and the edge of various directions), and reveals the distribution of these features, which improves the performance of facial image recognition. Secondly, the PCA algorithm can map the high dimensional data to low dimensional to reduce the amount of calculation for one side and extract the key data

to minimal interference for the other. Finally, the ridge regression algorithm is applied to classify the mapping data by PCA algorithm in order to get the mapping vector of face training dataset and then save the training results.

4.1 LBP ALGORITHM

The basic idea of LBP algorithm is to label the pixels of an image by thresholding the neighborhood of each pixel and consider the result as a binary number [5-7]. For the face recognition system, we apply the LBP algorithm to each divided block of preprocessing image and analyze the LBP histogram, and then we obtain the feature histogram [4-6]. The so-called uniform patterns can be used to reduce the length of the feature vector and implement a simple rotation-invariant descriptor. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. For example, the patterns $U(00000000) = 1$, $U(11111111) = 2$ and $U(11001111) = 2$ are uniform whereas the patterns $U(11001001) = 4$ and $U(010100010) = 6$ are not.

The uniform patterns LBP algorithm is defined by

$$\mathrm{LBP}_{P,R}^{riu} = \begin{cases} \sum_{i=0}^{p-1} s(g_i - g_c), U\left(LBP_{P,R}\right) \le 2, \\ P+1, other, \end{cases} \quad (1)$$

where:

$$\mathrm{U}\left(LBP_{P,\ R}\right) = |\mathrm{s}(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{i=1}^{P-1} |s(g_i - g_c) - s(g_{i-1} - g_c)|. \quad (2)$$

From Equation (1), there are $P + 1$ uniform patterns when using P pixel neighborhoods. These uniform patterns are labelled as 0, 1, …, P, while the non-uniform pattern is labelled as $P + 1$. According to LBP operator feature, the face image is divided into 4×4 blocks. In this paper, the height and width are set at 96 pixels, and the LBP codes in an image are collected into a histogram.

4.2 PCA ALGORITHM

The eigenface approach began with a search for a low-dimensional representation of face images. In 1991 M. Turk and A. Pentland presented the eigenface method of face recognition [8-10].

Denote the facial data by:

$$X^i = \left(X_1^i, X_2^i, \cdots, X_d^i\right), i = 1,2,\cdots,k., \quad (3)$$

Then the covariance matrix is

$$\mathrm{S} = \frac{1}{k-1}\sum_{i=1}^{k}(X_i - \bar{X})(X_i - \bar{X})^T. \quad (4)$$

Next we adopt a computationally feasible method to determine the eigenvalues λ_i and eigenvectorsv_i of S, where i = 1, 2, …, d. Then we sort λ_i from largest to smallest and choose the first n ones, where $n < d$ to build a transfer matrix $\mathrm{A} = [v_1, v_2, \cdots, v_n]$. For the X data, $Z^{'} = A(X - \bar{X})$ its eigenvalues yields.

In our test, the face image is considered as a high dimensional vector, so that a typical image of size 100 by 100 becomes a vector of dimension 10,100, or, equivalently, a point in 10,100-dimensional space. We then adopt LBP algorithm for texture analysis, and output the histogram equalization with P=8, which implies that there may be $10(P + 2)$ uniform patterns. Also, the face image is divided into 4×4 blocks, and then this high dimensional vector reduces to a vector of dimension 160. Next, through the face training database, and save the LBP outcome as declared variable 'CvMat*', and face feature before reducing dimension as matrix 'm_pTrainingFaceX' with 160 rows and N columns, where N is the total number of images of face training database. Herein we collect 10 images for each person and save them to separate directory named by ID.

4.3 RIDGE REGRESSION ALGORITHM

The last processing flow of face training is to classify the face data by ridge regression algorithm [11,12]. We first introduce the main idea of this regularization method. Motivated by the fact that the m vertices of a regular simplex is the most balanced and symmetric separate points $v_i(i = 1,2,\cdots,m)$ in the m-dimensional space. We choose these vertices as the targets for m distinct face images, and apply ridge regression to train a linear model W, which maps the training face images $I_i(i = 1,2,\cdots,n)$ into the m-dimensional subspace such that the distance between each mapping result and the vertex is minimum. Herein each training sample$I$and each vertexvcorrespond to the independent variable x and dependent y variable of multivariate linear model, respectively.

Assume that there are n training samples for face classification $\{(x_i,\ y_i)|i = 1,2,\cdots,n\}$, where x_i denotes p-dimensional vector for face images of high dimension, and y_i denotes r-dimensional vector for face images of low dimension. As a result, the task of learning classifier in multivariable ridge regression algorithm is to find a matrix $\mathrm{W} \epsilon R^{P\times R}$ that can model the linear dependency between the image x_i and the label y_i to minimize the objective function:

$$J(W) = \sum_{i=1}^{n}\|W^T x_i - y_i\|^2 + \lambda\|W\|^2 = \mathrm{tr}(YY^T + WTXXTW - 2WTXYT + \lambda trWTW, \quad (5)$$

where $\mathrm{X} = [x_1, x_2, \cdots, x_n]$, $\mathrm{Y} = [y_1, y_2, \cdots, y_n]$ and tr stands for the trace of matrix.

Taking derivative of Equation (5) with respect to the W and equalling it to zero, we have:

$$\mathrm{W} = (XX^T + \lambda I)^{-1}XY^T. \quad (6)$$

Let x be a new image, the projecting value in the subspace is $W^T\boldsymbol{x}$, the class of minimum distance of x.

Suppose there are m face images to be recognized, we can construct m points in R^{m-1}such that these m points keep isometric under translation and rotation. In summary, the procedure of ridge regression algorithm for face recognition can be formally stated as follows:

1) Compute the mapping matrix W from (6).
2) Compute the map $y = W^T \boldsymbol{x}$ for a new face *x*.
3) Compute the distance $\|y - T_i\|$ between y and each label $T_i (i - 1,2,\cdots,m)$, and denote $\arg\min_i \|y - T_i\|$ as the class of *x*. The input is the data after reducing dimension, which is the matrix 'm_pTrainingDataX' with 70 rows and *N* columns, where *N* is the total number of images of face training database.

We firstly label the data after reducing dimension and construct the label function, and then compute the classifier by ridge regression algorithm. Here the input parameter of the function 'trainW()' is the directory of face training database. Also this function can save the key data such that it avoids the training procedure of face recognition.

The key data and saved variables and their description are listed in Table 1:

TABLE 1 The Variables and their file name and description

Variable	File Name	Description
m_pTrainingResW	TrainingRes.txt	TRAINING RESULTS
m_pRegSimVerT	T.txt	MATRIX MARK VECTOR ALL FACE CATEGORIES
m_pMeanOfFaceX	Mean.txt	AVERAGE VECTOR FOR PCA BEFORE REDUCTION DIMENSION
m_pEigVecs	EigenVectors.txt	FEATURE MATRIX PCA REDUCING DIMENSION
m_vecPersonName	PersonName.txt	THE ID OF FACE CATEGORIES
m_ifaceHeight **m_ifaceWidth**	ImageSize.txt	FACE IMAGE SIZE(HEIGHT AND WIDTH）

ID recognition can be carried out independently with face recognition, but the recognition must adopt the face training results, therefore one need to check whether the variable is null before face recognition. If yes, one must import the data according to variables and file name from Table 1.

Next combination with the LBP pre-processing, PCA reduction dimension and ridge regression mapping results in a map vector, and compare this vector with the training result in order to find the ID with minimum distance. Finally these operations are wrapped to drive the application program.

5 Application software design

Our software platform is based on Linux kernel (2.6.29), and we adopt Qt 4.73 as the user interface, Open CV 1.0 as face database. The framework of our application software is shown (Figure 4):

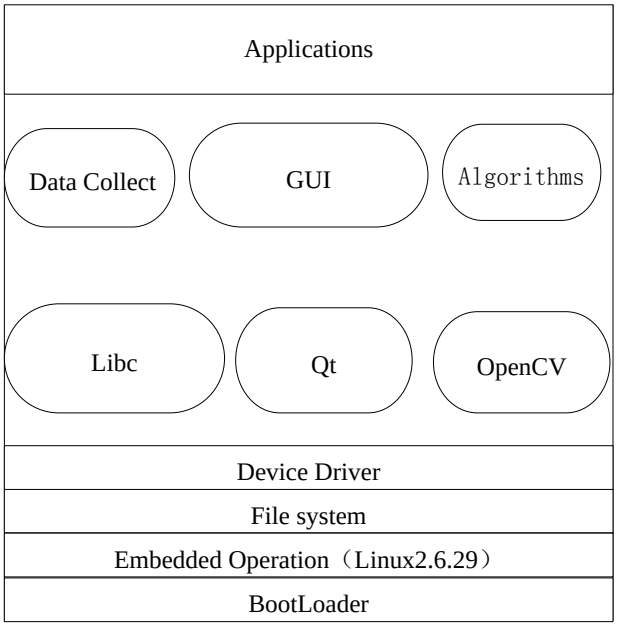


FIGURE 4 The framework of application

The applications are composed of Bootloader, Linux operation system (2.6.29 kernel), Qt libraries, Open CV library, embedded file system, driver program and GUI. Herein Open CV is a popular computer vision library started by Intel in 1999. The cross-platform library sets its focus on real-time image processing and includes patent-free implementations of the latest computer vision algorithms. Qt is a cross-platform application framework that is widely used for developing application software with a graphical user interface (GUI) (in which cases Qt is classified as a widget toolkit), and also used for developing non-GUI programs such as command-line tools and consoles for servers. Qt is fully object-oriented, easily extensible, and allows true component programming [13, 14].

In the Samsung S3C6140 development board, we compile the vivi bootloader, and configure Busybox with dev support. Licb library is the standard C libraries on Linux, and the compiler is installed in the user directory, and then it is copied into embedded root directory. Thus we complete the basic embedded root file system and make it as mapping commands by 'mkyaffsimage' program, and then to write vivi to flash root partition.

Additionally, the source codes need to be compiled before transplanting the Qt library in order to accommodate the Samsung S3C6410 platform. Since it takes the touch screen as input, one must import 'tslib' library to ensure that the touch screen driver can be used normally by GUI interface designed by Qt library.

Next we compile the Open CV library and V4L2 camera driver to get the video frames and map them to user memory, and then convert the frames encoded in YUV format to RGB format to generate the picture on a computer screen. According to the YUV/RGB conversion formula, we write the conversion function. So the RGB frames can display in Qt, and thus we can use some image processing function in Qt to process or save.

This system can achieve the following objectives: the recognition rate reaches more than ninety-eight percent; users can import a new member by the user registration function; the recognition speed can be five second or less; it provides a good interactive interface; it minimizes the energy consumption as much as possible; the whole system has high integration.

6 Face recognition system testing

The implementation of detection of face recognition system is divided into two parts, namely, a face training part and a face recognition part, and the latter use the parameters and outcome from the former, while both use the same face detection technology.

The testing program is to test the recognition rate and the accuracy of face recognition algorithm, where face database includes the internet open access resources and the face images appended by user. The design principle of our face database is to extract 16 people from in the face database from internet open access resources, each with 20 images, where we use 10 images as face training database and others for testing. On the other hand, we append two people (one with 20 images and another with 23 images) into the face database, and extract each with 10 images as face training database and others for testing.

Now we introduce the procedure of face recognition testing. Firstly, we activate the camera to capture face and recognize it. If that face to be recognized is not saved in the database then the algorithm program will display 'nobody'. If the face to be recognized is registered in the system, then the training function will build a folder with the name 'cj', and save 10 shots with different facial expression. Therefore, the face recognition system will activate the camera to capture image and compare it with the backend database, then it will recognize the information of 'cj', and save 10 shots with different facial expression, shown in Figure 5.

Therefore, the face recognition system will activate the camera to capture image and compare it with the backend database, then it will recognize the information of 'cj', so it will activate the embedded application system to open the door for this authorized person.

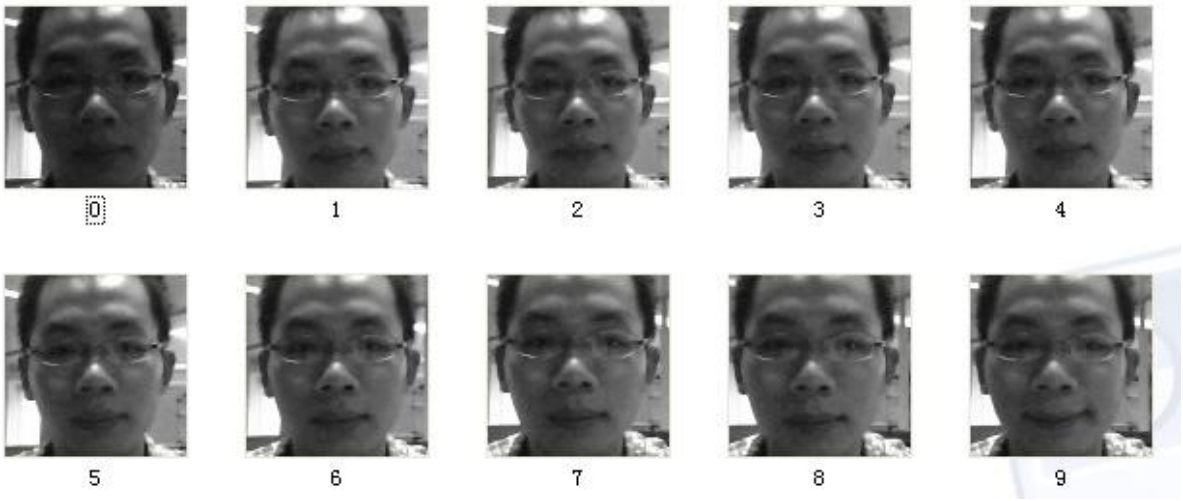


FIGURE 5 Face data

7 Conclusions

Face recognition technology is a hi-tech security way, which has superiority beyond comparing with other recognition approaches. Since it has a high accuracy rate, fast speed and difficultly counterfeited, the face recognition system is very suit for places where high safety is called for. Compared to some unique feature such as DNA, fingerprint, or iris, the face will change as one grows old, and collected images are disturbed by illumination, pose, expression, and makeup. But there exist some similar faces like the identical twins, so there may be uncertainty about face recognition. To improve the information security, one can adopt various recognition approaches to control the access.

For the current urgent need for security, we design an entrance guard system by face recognition and combine the ID registration technology to pre-register personnel in and out. Furthermore, we should update the samples in face database [15]. Finally, our system can be applied to finance, securities, social security, police, entry/exit port and other security authentication industry and sector.

Acknowledgments

This work was supported by Guangdong Provincial Scientific and Technical Project of China under Grant No. 2012 B010900089.

References

[1] Yan Y, Zhang Y 2009 State-of-the-art on video-based face recognition *Chinese Journal of Computers* **32**(5) 878-86 *(in Chinese)*
[2] Mitchell T 1997 Machine Learning *McGraw Hill*
[3] Wang J 2010 Research on Face Recognition Algorithm *PhD Thesis* Xi'an Electronic and Science University *(in Chinese)*
[4] Peng S 2011 Research of ARM Based Embedded Linux Development Platform *PhD Thesis* Harbin Engineering University *(in Chinese)*
[5] Wang X, Han T, Yan S 2009 An HOG-LBP human detector with partial occlusion handling *2009 IEEE 12th International Conference on Computer Vision* 32-9
[6] Zhang G, Huang X, Li S Z, Wang Y, Wu X 2005 Boosting local binary pattern (LBP)-based face recognition *Advances in Biometric Person Authentication Lecture Notes in Computer Science Volume* **3338** 179-86
[7] Huang J, Zhang H, Yan R 2012 Face recognition of LBP histogram PCA and Euclidean distance *Computer System Applications* **21**(6) 202-4*(in Chinese)*
[8] Turk M, Pentland A 1991 Eigenfaces for recognition *Journal of Cognitive Neuroscience* **3**(1) 71-86
[9] Brunelli R, Poggio T 1993 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(10) 1042-53
[10] Ahonen T. Hadid A, Pietikainen M 2006 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(12) 2037-41
[11] Wen Y, Shi P 2009 An approach to face recognition based on common vector and 2DPCA *Acta Automatic Sinica* **35**(2) 202-5*(in Chinese)*
[12] Teferi D, Bigun J 2007 Damascening video databases for evaluation of face tracking and recognition-The DXM2VTS database *Pattern Recogniion Letters* **28**(15) 2143-56
[13] Cai Z, Lu C, Li L 2008 Master Qt4 programming *Electronics Industry Press* Beijing
[14] Blanchette J, Summerfield M 2008 C++ GUI Programming with Qt4 *Electronics Industry Press* Beijing *(in Chinese)*
[15] Su Y, Shan S, Chen X, Gao W 2010 Integration of Global and Local Feature for Face Recognition *Journal of Software* **21**(8) 1849-62 *(in Chinese)*

Author

Jun Li, born in November 1964, Guangzhou, China.

Current position, grades: Master of Theoretical Physics, associate professor of Electrical and Computer Engineering in Guangdong Polytechnic Normal University
University studies: master's degree in Theoretical Physics at Sun Yat-sen University in Guangzhou in 1991.
Scientific interests: pattern recognition and machine intelligence and high performance computing in computer engineering.
Publications: 8 patents, 25 papers.

Control particle swarm optimization for unit commitment problem under emissions reduction

Xin Ma*, Fuxiaoxuan Liang, Wenbin Wang

School of Management and Economic, North China University of Water Resources and Electric Power, Zhengzhou, 450046, China

Received 29 May 2014, www.cmnt.lv

Abstract

The control particle swarm optimization (CPSO) algorithm is introduced to solve the unit commitment problem under the background of emissions reduction. Because the standard particle swarm optimization algorithm is easy to fall into local optimal solution. The closed loop control concept and feedback mechanism of classical control theory are posited, each particle is considered as controlled object to meet the changing needs in searching process, while dynamically adjust the inertia weight by proportion-Integra-derivative (PID) controllers according to the adaptation value of each step. These strategies greatly ensure the diversity of particles and improve the global search ability of the algorithm. The simulation results show that CPSO algorithm can reduce the dimension of the problem and ensure the feasibility of the particle in the optimization process, while it also has good convergence characteristics and global search ability.

Keywords: power system, emissions reduction, unit commitment, control particle swarm optimization

1 Introduction

From the view of mathematical, the unit combination problem is a multi-constraint non-deterministic polynomial (NP) hard combinatorial optimization problem (K. H. Abdul-Rahman et al.) [1]. It is difficult to obtain theoretically optimal solution. The typical algorithms for solving model include heuristics approach (S.K. Tong et al.) [2], dynamic programming (C.K. Pang) [3], branch and bound method (G.S. Lauer) [4], Lagrangian relaxation method (S.Dekrajangpetch et al., C.P. Cheng et al., W. Ongsakul et al.) [5-7] and the modern intelligent algorithm (S.S. Li et al., Z. Li et al., J.L. Lu, X.H. Zhang et al.) [8-11]. However, these algorithms have one or another defect (K. Han et al.) [12]. If the environmental constraints are considered into the model, the problem will become more complexities. But it is important to economic operation of power systems through improving the accuracy and speed of solution under various constraints. In this context, an objective function of emission reduction unit commitment is posed which considering the minimum cost and minimum emission targets together by introducing emission price factor, and then, the original multi-objective problem can be transferred into a single objective problem, while a variety of constraints can be easily considered, in order to obtain the reasonable compromise between energy consumption and emissions.

In the specific solving algorithm, this paper presents a new set of inertia restructuring strategy through introducing the concepts of feedback mechanisms and closed-loop control system of control theory into the PSO system, then a control particle swarm optimization (CPSO) algorithm is formed. In this algorithm, a closed-loop control system is constructed while each particle sets as a controlled object-tive. Fitness value of particle is used as the output variable and feed backed to closed-loop in the iterative process, and the updated inertia weight is calculated through a designed PID controller which has been widely used in industrial due to its single structural and robustness, then the velocity and position of the particle is adjusted. CPSO algorithm can satisfy the needs of each particle and greatly ensure the diversity of the population of particles, while improving the search capabilities of PSO algorithm.

2 Mathematic models

The goal of unit commitment optimization is to optimize the status and contributions of generators in the calculation scheduling cycle, while satisfying the some constraints, in order to achieve minimum energy cost or minimum pollutant emissions.

2.1 OBJECTIVE FUNCTION

2.1.1 Minimum Energy Cost

The objective function can be expressed as follows:

$$\min F\left(\eta, P^{gen}\right)=\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{f_i\left(P_{i,t}^{gen}\right)\eta_{i,t}+S_{i,t}^{up}+S_{i,t}^{down}\right\}, \quad (1)$$

$$f_i\left(P_{i,t}^{gen}\right)=\rho_i+\mu_i P_{i,t}^{gen}+\varphi_i\left(P_{i,t}^{gen}\right)^2, \quad (2)$$

$$S_{i,t}^{up}=S_i^{up}\eta_{i,t}\left(1-\eta_{i,t-1}\right), \quad (3)$$

$$S_{i,t}^{down}=S_i^{down}\eta_{i,t-1}1\left(-\eta_{i,t}\right), \quad (4)$$

* ***Corresponding author's*** e-mail: xm510@163.com

i is index for the number of units, *i*=1,2, …, *N*, *N* is the total number of units; *t* is index for time, *t*=1,2, …, *T*, *T* is the total number of periods in the scheduling period. $\eta_{i,t}$ is the start-up and shut-down statue of unit *i* at time *t*. The start-up of unit is 1 and shut-down of unit is to 0. $S_{i,t}^{up}$ is the cost of start-up -and $S_{i,t}^{down}$ is the cost of shut-down for unit *i* at time *t* respectively. S_i^{up} is the constant coefficient of cost of start-up and S_i^{down} is the constant coefficient of cost of shut-down for unit *i* at time *t* respectively. $f_i(P_{i,t}^{gen})$ is the generation costs of unit *i*, in \$, ρ_i, μ_i and φ_i are cost coefficients of unit *i*. $P_{i,t}^{gen}$ is the generation level of unit of unit *i* at time *t*, in MW.

2.1.2 Minimum Pollutant Emissions

The objective function can be expressed as follows:

$$\min E\left(\eta, P^{gen}\right)=\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{e_i\left(P_{i,t}^{gen}\right)\eta_{i,t}+S_{i,t}^{up}+S_{i,t}^{down}\right\}, \quad (5)$$

$$e_i P_{i,t}^{gen}=\alpha_i+\beta_i P_{i,t}^{gen}+\gamma_i\left(P_{i,t}^{gen}\right)^2, \quad (6)$$

$e_i(P_{i,t}^{gen})$ is the emission costs of unit *i*. α_i, β_i and γ_i are cost coefficients of unit *i*, in \$. The other symbols are same as those in last section.

2.1.3 Total Objective Function

Considering the minimum energy cost and minimum pollutant emissions together by introducing an emissions price factor $\theta_{i,t}$, the multi-objective problem turned into a single objective problem and the total objective function is defined mathematically as follows:

$$\min F^{total}\left(\eta, P^{gen}\right)=\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\left[f_i\left(P_{i,t}^{gen}\right)+\theta_{i,t}e_i\left(P_{i,t}^{gen}\right)\right]\eta_{i,t}+ S_{i,t}^{up}+S_i^{down}\right\}, \quad (7)$$

The emissions price factor $\theta_{i,t}$ is defined as the proportion between maximum energy cost and maximum pollution emission of unit *i*, $P^{gen}_{i,max}$ is the maximum generation capacity of unit *i*:

$$\theta_{i,t}=f_i\left(P_{i,\max}^{gen}\right)/e_i\left(P_{i,\max}^{gen}\right), \quad (8)$$

2.2 CONSTRAINTS CONDITION

Power balance constraints:

$$\sum_{i=1}^{N}P_{i,t}^{gen}\eta_{i,t}-L_t=0. \quad (9)$$

Spinning reserve constraints:

$$\sum_{i=1}^{N}P_{i,\max}^{gen}\eta_{i,t}\geq L_t+R_t, \quad (10)$$

Up and down generation constraints of unit:

$$P_{i,\min}^{gen}\leq P_{i,t}^{gen}\leq P_{i,\max}^{gen}, \quad (11)$$

Minimum operation time constraints:

$$t_i^{on}\geq M_i^{on}, \quad (12)$$

Minimum out of operation time constraints:

$$t_i^{off}\geq M_i^{off}, \quad (13)$$

Unit Ramp Rate Constraint:

$$\left|P_{i,t}^{gen}-P_{i,t-1}^{gen}\right|\leq Z_{i,\max}, \quad (14)$$

where, $P^{gen}_{i,max}$ and $P^{gen}_{i,min}$ are the maximum and minimum generation capacity of unit *i*, in MW. L_t is system load demand at time *t*, in MW. R_t is system spinning reserve demand at time *t*, in MW. M_i^{on} is the minimum number of periods unit *i*, it must remain on operation statue after it has been start-up at time *t*, in hours. M_i^{off} is the minimum number of periods unit *i*, it must remain on out of operation statue after it has been shut-down at time *t*, in hours. t_i^{on} is the continuing number of periods unit *i* remains on operation statue after it has been start-up at time *t*, in hours. t_i^{off} is the continuing number of periods unit *i* must remains on out of operation statue after it has been shut-down at time *t*, in hours. $Z_{i,max}$ is the maximum ramp rate of unit *i*, in MW/h.

3 Outline of PSO

PSO algorithm was proposed to solve optimization problems by Kennedy et al. [13], it was a kind of numerical optimization techniques which similarly with genetic algorithm. It is easy to know from the above literature that the forward movement of the particle swarm including the size and direction of the particles were decided based on the original speed, individual and global extreme weighted.

$$v_{j,b+1}=\omega v_{j,b}+\phi_1 c_{rand}\left(q_{j,b}^{best}-x_{j,b}\right)+\phi_2 c_{rand}\left(g_b^{best}-x_{j,b}\right), \quad (15)$$

where *j* is index of particles, *j*=1,2, …, *J*. *b* is index for iterations, *b*=1,2, …, *B*. $q^{best}_{j,b}$ is the individual best position of particle *j* in the iteration *b*. g^{best}_{b} is the best position of all the particles in the iteration *b*. $v_{j,b}$ and $x_{j,b}$ are the velocity and location of particle *j* in the iteration *b*. ω, ϕ_1 and ϕ_2 are weight coefficients when velocity renewed. c_{rand} is the random number between 0 and 1. The particle swarm will be:

$$x_{j,b+1}=x_{j,b}+v_{j,b+1}. \quad (16)$$

Each particle in swarm will execute parallel optimazation during the iterations according to its own inertial velocity and the best location in memory, while considering the best location which other particles experienced. Then the best optimization solution will be found through several iterations.

4 CPSO algorithm

The inertial weight of each particle is always set as a relatively large initial value at the beginning of searching process, in order to make the particle moves fast. But this approach obviously has a weak point, which will lead to some particles located at the better location move too large distance to neglect some valuable searching area. If the inertial weight adjusted to a small value at the final stage of searching process, in order to make the particle to do more detail search. But this method will prevent some particles located at the worst location to overtake the other particles. So the same inertial weight value of particle swarm is not appropriation for each particle. The feedback mechanism and closed loop concept are introduced into PSO algorithm in order to overcome the shortcoming described above.

Figure 1 shows a simple input and simple output (SISO) feed back control system, the controlled values $y(t)$ and input values $u(t)$ of object are usually defined as controlled variable and manipulated variable of object respectively. The controlled variable $y(t)$ is fed back through a sensor measurement H and transfer value $h(t)$ to the reference value or set-point $r(t)$.

Each particle actually can be treated as an object. Its dynamic characteristics include current location, fitness value of last iteration and velocity. So the closed loop control system can be built for each particle, the fitness value is defined as controlled variable which is fed back to controller in iterations. Then a new suitable inertial weight will be calculated by controller according to the current fitness of particle and used for renewing the velocity and location.

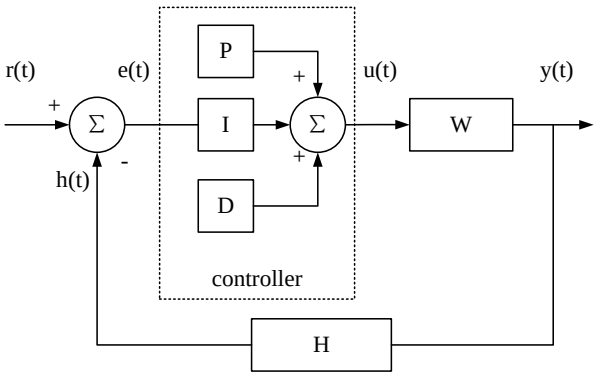


FIGURE 1 Structure of a SISO closed control system.

4.1 STRUCTURE OF PID

PID controller is a control loop feedback mechanism (controller) widely used in industrial control systems and has historically been considered to be the best controller with simple structure and best robustness. So the PID is selected as controller in this CPSO algorithm. There are three separate constant parameters which also be named as three-term control in the PID controller algorithm: the proportional values denoted as P, the integral values denoted as I and derivative values denoted as D, the structure of PID controller is show at Figure 1.

The output value Ξ^{out} which is proportional to the present error value $e(t)$ is produced by the proportional term. The proportional response value can be revised in term of multiplying the current error value through a constant defined as K^{prop}, this term is also named the proportional gain constant. The proportional term is set as follows:

$$\Xi^{out} = K^{prop} e(t). \tag{17}$$

The integral term I^{out} of a PID controller is the sum of the instantaneous error value during the time. This term will supply the accumulated offset value which should have been adjusted previously. The accumulated error term is then multiplied through the integral gain term defined as K^{iteg} and acted to the output value of controller. τ is the variable of integration which takes on values from time 0 to t. The integral term I^{out} is set as follows:

$$I^{out} = K^{iteg} \int_0^t e(\tau) d\tau . \tag{18}$$

The main purpose of the integral term is to accelerate the movement of the object towards the reference value and reduces the residual steady state error term which produced with a pure proportional controller.

The derivative term D^{out} of the object error is calculated through the determining the slope of the error value over the period of the time and multiplying this change magnitude according to the derivative gain constant K^{deri}. The role of the derivative term to the overall responding control action is effected on the basis of the derivative gain constant K^{deri}. The derivative term D^{out} is set as follows:

$$D^{out} = K^{deri} \frac{de(t)}{dt}. \tag{19}$$

The final standard form of the PID controller algorithm is given by:

$$u(t) = K^{prop}\left[e(t) + \frac{1}{T^I}\int_0^t e(\tau)dt + T^D \frac{de(t)}{dt} \right], \tag{20}$$

$$e(t) = r(t) - h(t), \tag{21}$$

where, each parameter has a clear physical meaning in this form. T^I is defined as the integral time constant and T^D is defined as the derivative time constant.

$$K^{iteg} = K^{prop} / T^I , \tag{22}$$

$$K^{deri} = K^{prop} T^D . \tag{23}$$

Sometimes it is useful to rewrite the PID controller in Laplace transform form. It is easy to determine the closed loop transfer function if this form is adopted.

$$G(s) = K^{prop}\left[1 + \left(1/T^I s\right) + T^D s\right]. \tag{24}$$

PID controller can be adjusted according to K^{prop}, T^I and T^D, these parameters are constructed as gain vector.

4.2 DEFINITION OF CPSO

There are J numbers of particles in whole swarm. W_j is the j particle, j=1,2, …, J. The manipulation variable u_j is defined as inertia weight factor. The output value of object is y_j defined as the fitness value of last iteration. H_j is the transfer function of feed back tunnel. It is also named as measurement sensor. Its output value is h_j. These parameters are given by the following equations:

$$u_j = \omega_j, \tag{25}$$

$$H_j = \varepsilon\left|1/y^{ave}\right|, \tag{26}$$

$$h_j = H_j y_j = \varepsilon\left|y_j / y^{ave}\right|, \tag{27}$$

where, ε is adjust factor and generally to be set as 1. y^{ave} is average value of current fitness for all particles. The parameter h_j is the evaluation term for particle W_j. The particle can be treated as a relative optimization location if h_j is larger than 1, on the contrary the particle is considered as lag location. The reference value is set as 1, and then the error term e_j is given as follows:

$$e(t) = r(t) - h(t) = 1 - \varepsilon\left|y_j / y^{ave}\right|. \tag{28}$$

The ζ is defined as gain vector of PID controller for j particle.

$$\zeta = \left[K_i^{prop}, \quad T_i^I \quad, T_i^D\right]^T = \left[-1, \quad \Re^{ite}/20, \quad 0\right]^T. \tag{29}$$

The gain proportion parameter K_i^{prop} will to be set lower than zero, because the movement magnitude of the particle which traps into local extreme value needs to be given larger inertia weight. The particle at a relatively optimization position need to be given a smaller inertia weight, in order to do more meticulous searching. Because the stability error will be introduced if the controller just include the proportion module, so the integral term should be added into the controller to eliminate the error term. The integral time constant T_i^I means it can play the same role as proportion term past T_i^I time. The specific problem is that the excess overshooting will be shown if integrator term is added into PID controller. Further more the integral windup will be caused. Some measures must be adopted to prevent this phenomenon. There is a simple anti integrator windup strategy for control PSO is given as follows:

$$T_i^I = \begin{cases} \Phi^{ite}/20, & \delta_j^{low} < \omega_j < \delta_j^{up} \\ 0, & \omega_j = \delta_j^{low} or \omega_j = \delta_j^{up} \end{cases}, \tag{30}$$

$\delta_j{}^{low}$ and $\delta_j{}^{up}$ are low and up limit of inertia weight ω_j. Φ^{ite} is the maximum iteration number. Because inertia weight and particle search area ranges are predefined between bounded set. Therefore, the output variables which are fitness value of the particle are bounded, the situation does not appear divergent. That means the closed-loop stability is guaranteed.

There will be more effective control if the derivative term added into controller, but at the same time the probability of unstable statute is also increased, so the derivative time constant is defined as zero in CPSO.

4.3 SOLVING STEP OF CPSO

The solving step of control PSO is as follow.

Step 1: Initialize the particles swarm and the initial iteration number φ_i is set as zero.

Step 2: Iteration calculation of particles.

Step 3: Renew the current individual particle best position $q_{best,j}$, the current best position g_{best} of all the particles swarm and the fitness value.

Step 4: If the iteration number φ_i does not reache the maximum iteration number Φ^{ite}, then the inertia weight ω is calculated through:

$$\omega = K^{prop}\left[1 + \left(1/T^I s\right) + T^D s\right]e. \tag{31}$$

Iteration calculation of particles by using new inertia weight ω, iteration number φ_i increases as follows:

$$\varphi_i = \varphi_i + 1. \tag{32}$$

Go to Step 2.

Step 5: If the iteration number φ_i reaches the maximum iteration number Φ^{ite}, then the evolution stops and the best position value is outputted.

5 Solving unit commitment problem under emissions reduction

5.1 CODING METHOD

The generation of all the units in different period will be connected as an individual of CPSO and represented by $N \times T$ matrix. This matrix is shown as follows:

$$P_k^{gen} = \left(Q_1 \;,\; Q_2 \;,\; \cdots \;,\; Q_t \;,\; \cdots \;,\; Q_T\right) = \left(R_1 \;,\; R_2 \;,\; \cdots \;,\; R_i \;,\; \cdots \;,\; R_N\right)^T = \begin{bmatrix} P_{1,1}^{gen} & P_{1,2}^{gen} & \cdots & P_{1,t}^{gen} & \cdots & P_{1,T}^{gen} \\ P_{2,1}^{gen} & P_{2,2}^{gen} & \cdots & P_{2,t}^{gen} & \cdots & P_{2,T}^{gen} \\ \vdots & \vdots & & \vdots & & \vdots \\ P_{i,1}^{gen} & P_{i,2}^{gen} & \cdots & P_{i,t}^{gen} & \cdots & P_{i,T}^{gen} \\ \vdots & \vdots & & \vdots & & \vdots \\ P_{N,1}^{gen} & P_{N,2}^{gen} & \cdots & P_{N,t}^{gen} & \cdots & P_{N,T}^{gen} \end{bmatrix} \tag{33}$$

where, P_k^{gen} is the k individual of particles swarm. $P_{i,k}$ is element of the i row and t column in the coding matrix, it represents the generation of unit i during the period t. Q_t is the t column vector in the coding matrix, it represents the unit statute of all units during the one period of all dispatch times, this vector can be used to calculate the cost of production and economic dispatch. R_i is the i row vector in the coding matrix, it represents the start-up and shut-down statute of all units during the all dispatch periods, this vector can be used to calculate the cost of start-up and shut-down for unit, but it must satisfy the minimum operation time constraints or out of operation time constraints.

The operation status of unit is decided by the element of matrix, furthermore the start-up or shut-down status is decided according to the generation of unit, this is given as following:

$$\eta_i(t)=\begin{cases}0, & P_{i,t}^{gen}=0\\ 1, & others\end{cases}. \tag{34}$$

Velocity matrix of CPSO algorithm is set as follows, where, V_k^{gen} is the velocity component of element P_k^{gen} in the particles swarm.

$$V_k^{gen}=\begin{bmatrix} V_{1,1}^{gen} & V_{1,2}^{gen} & \cdots & V_{1,t}^{gen} & \cdots & V_{1,T}^{gen}\\ V_{2,1}^{gen} & V_{2,2}^{gen} & \cdots & V_{2,t}^{gen} & \cdots & V_{2,T}^{gen}\\ \vdots & \vdots & & \vdots & & \vdots\\ V_{i,1}^{gen} & V_{i,2}^{gen} & \cdots & V_{i,t}^{gen} & \cdots & V_{i,T}^{gen}\\ \vdots & \vdots & & \vdots & & \vdots\\ V_{N,1}^{gen} & V_{N,2}^{gen} & \cdots & V_{N,t}^{gen} & \cdots & V_{N,T}^{gen}\end{bmatrix}. \tag{35}$$

5.2 FITNESS FUNCTION

The objective function of unit commitment optimization under emissions reduction is set as fitness function of CPSO algorithm. The lowest fitness value means the lowest operation cost, moreover it is the best solution.

5.3 INDIVIDUAL STRATEGY ADJUSTMENT

Because the individual particle may not satisfy the constraint conditions after initialization or renewable, so the individual particle needs to be adjusted in order to meet the constraints. The procedure is given as follows.

5.3.1 Reduce Dimension

Unit commitment is a high dimension optimization problem. The dimension is reduced in the searching process by using the implicit information through mining the constraints. The minimum number of operation unit during the period t is given by spinning reserve constraints Equation (10). And only the previous χ unit during period t under the operation status according to the priority sorting, that means the following constraint is satisfied:

$$P_k^{gen}\geq P_{i,\min}^{gen},\ 1\leq i\leq \chi. \tag{36}$$

The Equation (37) also can be obtained according to the power balance constraints Equation (9) and up and down generation constraints of unit Equation (11):

$$\sum_{i=1}^{N}P_{i,\min}\eta_{i,t}\leq L_t \tag{37}$$

There are γ units are out of operation during t, moreover, it means the maximum number of operation unit can also be obtained according to Equation (37). So only χ+1 to N-γ status *for* element $P_{i,t}^{gen}$ of matrix P_k^{gen} needs to be optimization:

$$P_{i,t}^{gen}=0,\ N-\gamma+1\leq i\leq N. \tag{38}$$

5.3.2 Constraints Handling

The individual of particle swarms adjusted to meet the constraints will be implemented as follows:

(i) Adjust the components of location matrix in order to make it meet the generation constraints of units, according to the (39).

(ii) Adjust the start-up and shutdown status of units to satisfy the spinning constraints. If the sum of maximum generation of units under the operation status is lower than the sum of total load and spinning at current time, then the units under the out of operation status while meet the minimum out of operation time constraints need to be start-up, in accordance with the order of priority until the total generation reaches the requirement of system. The unit start-up order of priority is decided according to the literature (S. Dekrajangpetch et al.) [5], it means the units which have the lower incremental rate of cost will be set as the higher priority level. On the contrary, the units will be treated as lower priority level.

$$P_{i,t}^{gen}=\begin{cases}P_{i,\max}^{gen} & P_{i,t}^{gen}\geq P_{i,\max}^{gen}, i\geq N-\gamma,\\ & t_i^{on}\geq M_i^{on}\ or\ \ t_i^{off}\geq M_i^{off}\\ P_{i,t}^{gen} & P_{i,\min}^{gen}\leq P_{i,t}^{gen}\leq P_{i,\max}^{gen}, i\geq N-\gamma,\\ & t_i^{on}\geq M_i^{on}\ or\ \ t_i^{off}\geq M_i^{off}\\ P_{i,\min}^{gen} & P_{i,t}^{gen}\leq P_{i,\min}^{gen}, i\geq N-\gamma,\\ & t_i^{on}\geq M_i^{on}\ or\ \ t_i^{off}\geq M_i^{off}\\ 0, & otherwise\end{cases}. \tag{39}$$

(iii) The total generation of all units probably not equals the total load demand during some periods. So the load needs to be adjusted. If the total technology generation of units under operation status is larger than the total load at current time, then these units which satisfy the minimum operation time constraints need to be shutdown according to the priority sorts.

(iv) If the total generation is lower than the total demand at current time, then the generation of units needs to be increased, in order to meet the power balance constraints. If the total generation is larger than the total demand at current time, then the generation of units under operation status should be reduced. But the ramp rate constraints and generation constraints of unit must be satisfied in the process of adjustment.

Some problems may be appeared after the above adjustment of individual. Furthermore, the total maximum of technology generation of units under operation status is still lower than the sum of total demand and spinning reserve at current period. Because the demand increases so fast, which induced by some units were shut-down during some periods, but this units cannot start-up according to

the change of demand for minimum start-up time constraints. So the shut-down decision of units need to be re-evaluated before this period, the shutdown decision should be replaced by reducing the technology generation of units, while satisfy the power balance constraints.

6 Simulation and analysis

The 9 units system is used to test the proposed method. The main parameters are given as following tables. Table 1 and Table 2 show the unit system data. The result of economic dispatch of unit commitment is shown as Table 3. The emission quantities of three different dispatch models are shown as Figure 2. The first scene represents the objective function is set as minimum economic dispatch and the average emission quantity is 346.7458lb, this is the highest scene. The second scene represents the objective function is set as minimum emission dispatch and the average emission quantity is 279.9292lb, this is the lowest scene.

TABLE 1 9-units system data (Unit cost data and emission data)

Unit	ρ_i($/h)	μ_i($/MWh)	φ_i($/MW^2-h)	Start cost($)	α_i(lb/h)	β_i(lb/MWh)	γ_i(lb/MW^2-h)
1	142.735	10.6940	0.00643	200	24.300	0.81	0.0036
2	230.00	19.1000	0.00712	115	27.023	0.10	0.0035
3	81.136	13.3272	0.00876	80	27.023	0.50	0.0330
4	81.298	13.3538	0.00895	80	22.070	0.30	0.0034
5	218.335	18.1000	0.00612	100	24.300	0.81	0.0380
6	87.136	19.3272	0.01036	80	29.040	0.03	0.0034
7	118.821	37.8896	0.01433	30	29.030	0.02	0.0039
8	128.821	39.8896	0.01633	30	27.050	0.02	0.0030
9	187.364	49.3272	0.02436	70	22.070	0.30	0.0034

TABLE 2 9-units system data (Unit operation data)

Unit	Up(h)	Down(h)	Initial(h)	R_{up}(MW/h)	R_{down}(MW/h)	P_{max}(MW)	P_{min}(MW)
1	5	3	5	78	78	155	54
2	4	2	-3	50	50	100	25
3	3	2	3	38	38	76	15
4	3	2	3	38	38	76	15
5	4	2	-3	50	50	100	25
6	3	2	3	25	25	50	10
7	1	1	-1	20	20	20	4
8	1	2	-1	20	20	20	4
9	3	2	3	25	25	50	10

TABLE 3 The results of unit commitment under different scenes

Hour	Scene1(Units 1-9)	Scene2(Units 1-9)	Scene3(Units 1-9)	Load(MW)	θ_{ec}($/lb)
1	101 110 000	110 101 001	111 101 000	355	17.78
2	101 110 000	110 101 001	111 101 000	327	8.07
3	101 110 000	011 111 001	111 101 000	309	8.07
4	101 110 000	011 111 001	111 101 000	290	8.07
5	101 110 000	010 111 001	101 100 000	281	8.07
6	101 110 000	010 111 001	101 100 000	281	8.07
7	101 110 000	011 111 001	101 100 000	290	8.07
8	111 100 000	110 101 001	101 110 000	318	8.07
9	111 110 000	110 101 001	111 110 000	364	17.78
10	111 110 000	110 101 001	111 110 000	400	17.78
11	111 110 000	110 101 111	111 110 000	409	27.63
12	111 110 000	110 101 111	111 111 000	414	27.63
13	111 110 000	110 101 111	111 101 000	409	27.63
14	111 110 000	110 101 001	111 101 000	400	17.78
15	111 110 000	110 101 101	111 101 000	396	17.78
16	111 110 000	110 101 111	111 101 000	396	17.78
17	111 110 000	110 101 111	111 101 000	414	27.63
18	111 110 000	111 101 111	111 111 000	455	27.63
19	111 110 000	111 101 111	111 111 000	450	27.63
20	111 110 000	111 101 111	111 111 000	441	27.63
21	111 110 000	111 111 111	111 111 000	428	27.63
22	111 110 000	110 111 111	111 101 000	418	27.63
23	111 110 000	110 111 111	111 101 000	396	17.78
24	101 110 000	010 111 111	111 101 000	368	17.78

The fuel cost of three different dispatch models are shown as Figure 3, the first scene represents the objective function is set as minimum economic dispatch and the average fuel cost is \$5861.38, this is lowest scene. The second scene represents the objective function is set as minimum emission dispatch and the average cost is \$9345.38, this is the highest scene. The third scene represents the objective function is set as comprehensive dispatch of unit commitment and the average cost is \$6227.13, this scene is interposed between the two cases described in the foregoing. All these problems are solved by control PSO algorithm. The total cost of three different dispatch models are shown as Figure 4, the first scene represents the objective function is set as minimum economic dispatch and the average cost is 12687.7625\$, this is interposed between the scene1 and scene 2.

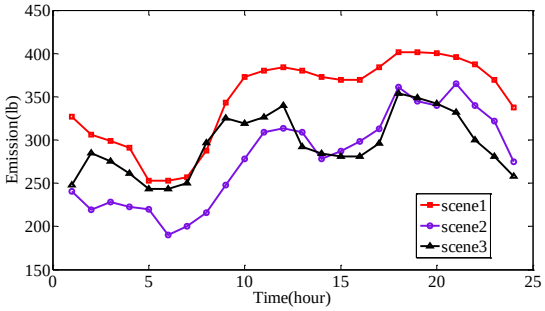


FIGURE 2 The emissions under different dispatch models by CPSO

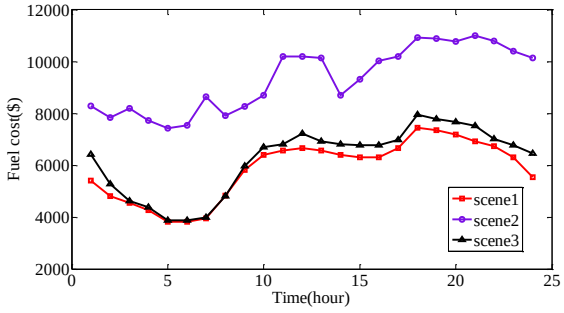


FIGURE 3 The fuel cost under different dispatch models by CPSO

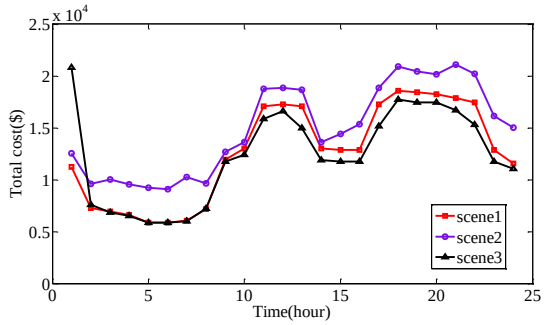


FIGURE 4 The total cost under different dispatch models by CPSO

The second scene represents the objective function is set as minimum emission dispatch and the average cost is \$14954.04, it is the highest scene. The third scene represents the objective function is set as comprehensive dispatch of unit commitment and the average cost is \$11918.88, it is lowest. Above results show that the total cost will increase so fast if too emphasize the emission reduction, so the relatively balance approach considering the emission reduction and economic dispatch together is needed.

Figure 5 shows the results distribution of total cost of comprehensive unit commitment under emission reduction by different algorithms including PSO, GA, and CPSO. Continuous simulating 30 times, the best solution of PSO is \$286882.55, the worst solution of PSO is \$294377.14. The deviation of the worst solution with respect to the best solution is 2.91%, and about 53.4% solution deviates from the best solution is lower than 1.5%.

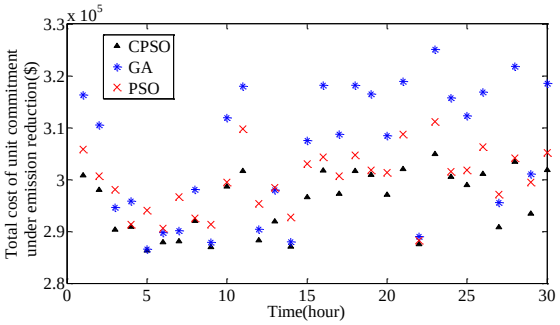


FIGURE 5 The results distribution of total cost of comprehensive unit commitment under emission reduction through different algorithms

The best solution of GA is \$286682.31, the worst solution of GA is £305275.76. The deviation of the worst solution with respect to the best solution is 6.34%, and about 26.3% solution deviates from the best solution is lower than 1.5%. The best solution of CPSO is \$286253.24, the worst solution of CPSO is \$291573.82, the deviation of the worst solution with respect to the best solution is 1.93%, and about 76.7% solution deviates from the best solution is lower than 1.5%. So the robustness and precise of CPSO are obviously best than the other algorithms.

7 Conclusions

There are following conclusions according to the simulation:

(i) The objective function considering the minimum cost and minimum emission targets together is more reasonable to discuss the unit commitment problem, because the environmental factor has same important role as economic factor.

(ii) The emission price factor is introduced into the model, this will helpful to transfer the multi-objective optimization problem to the single-objective optimization problem.

(iii) The total cost will increase so fast if too emphasize on the role of emission reduction, so the balance need to be selected between the target of economic and emission.

(iv) CPSO algorithm shows robustness and precise than the other algorithms.

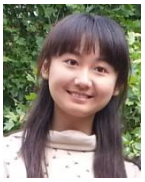
References

[1] Abdul-Rahman K H, Shahidepour S M 1996 *IEEE Transactions on Power System* **11**(1) 716–28
[2] Tong S K, Shahidepour S M, Ouyang Z 1991 *IEEE Transactions on Power System* **6**(3) 1210–6
[3] Pang C K, Sheble G B 1981 *IEEE Transactions on Power Apparatus and Systems*, **PAS-100**(3) 1212–8
[4] Lauer G S, Sandell N R 1982 *IEEE Transactions on Power Apparatus and Systems* **PAS-101**(3) 79–85
[5] Dekrajangpetch S, Sheble G B 1999 *IEEE Transactions on Power System* **14**(3) 82–8
[6] Cheng C P, Liu CW 2000 *IEEE Transactions on Power System* **15**(3) 707–14
[7] Ongsakul W, Petcharaks N 2004 *IEEE Transactions on Power System* **19**(1) 620–8
[8] Li S S, Li G, Cheng C T 2011 Thermal units energy conservation load dispatch method with combining dynamic unit commitment into equal incremental principle *Proceedings of the CSEE* **31**(7) 41-7 (*in Chinese*)
[9] Li Z, Tan W 2012 An improved dual particle swarm optimization algorithm for unit commitment problem *Proceedings of the CSEE* **32**(25), 189-195 (*in Chinese*)
[10] Lu J L 2013 Power system optimal dispatch considering wind farms based on improved multi-objective particle swarm algorithm *Power System Protection and Control* **41**(17) 189-195 (*in Chinese*)
[11] Zhang X H, Zhao J Q, Chen XY 2011 Multi-objective unit commitment modeling and optimization for energy-saving and emission reduction in wind power integrated system *Power System Protection and Control* **39**(17) 33-9 (*in Chinese*)
[12] Han K, Zhao J, Qian J X 2009 A Closed-loop particle swarm optimization algorithm for power system unit commitment *Automation of Electric Power Systems* **33**(1) 36-40 (*in Chinese*)
[13] Kennedy J, Eberhart R 1995 Particle swarm optimization *Proceedings of IEEE Conference on Neural Networks* Australia Perth 1942-8

Authors

Xin Ma, born in November 1972, Tianjin city, P.R. China.

Current position, grades: professor at the School of Management and Economic, North China University of Water Resources and Electric Power, China.
University studies: PhD degree in Electrical Engineering at Shanghai Jiaotong University, P. R. China
Scientific interests: simulation of complex system, environmental modeling, and energy modelling.
Publications: more than 50 papers.
Experience: over 20 years of working experience in electrical industry and in energy investment company.



Fuxiaoxuan Liang, born in January 1991, Zhengzhou city, P.R. China.

Current position, grades: research assistant at the School of Management and Economic, North China University of Water Resources and Electric Power, China.
University studies: master's degree at Queen Mary, University of London, UK.
Scientific interests: environmental modeling and simulation.
Publications: 2 papers.
Experience: 2 years of research experience.

Wenbin Wang, born in March 1983, Xinyang city, P.R. China.

Current position, grades: lecturer at the School of Management and Economic, North China University of Water Resources and Electric Power, China.
University studies: master's degree at North China University of Water Resources and Electric Power, China.
Scientific interests: information management and information system.
Publications: 5 papers.
Experience: 7 years of research experience.

Research on the cost estimation of transmission line project with case-based reasoning

Tao Yi, Shanshan Cui*, Yi Zhang

Department of Management science and Engineering, North China Electric Power University, Beijing, China

Received 28 May 2014, www.cmnt.lv

Abstract

Investment estimation phase of power transmission line project is an important stage of project cost control, the traditional unit investment forecasting methods, however, cannot meet the need of engineering rapid estimation. The research objective of this paper is to forecast the unit investment of the Transmission Line Project applying the case-based reasoning system. Based on the factor analysis tools of multivariate statistical software SPSS, the correlation coefficient matrix of characteristics factors are identified, and weight coefficient of characteristics factors is calculated. Based on the formula of European weighted distance, the similarity is calculated and case similarity is retrieved. Finally, the model is constructed to predict the unit investment of Transmission Line project, and the correctness and usefulness of the model is verified. It is shown that this case-based reasoning system of unit investment can achieve real-time updating of cost index database. The primary contribution of this research is the combination of the cost forecast and Case-based reasoning system of artificial intelligence. It is expected that this work will provide a reference and guidance for transmission line project investment decision in estimating phase.

Keywords: case-based reasoning, transmission line projects, forecasting

1 Introduction

Setting up the prediction model in line with the actuality, accurately and rapidly estimating the construction cost has dramatic theoretical significance and practical value. The accuracy of the transmission lines project cost prediction has a close relationship with the selection of prediction methods. The common predicting method can be divided into three categories:

The first is the approximate deduction method. According to the characteristics of all kinds of engineering projects, some uncertain factors are classified and made hypothesis. Then based on the mechanism of deductive reasoning, project cost prediction model is established. Such as artificial neural network, unit cost factor is the key to construct artificial neural network structure model. Only the various influence factors and influence mechanism are found, the artificial neural network model can be constructed, including the influence of the engineering geologic and topographic conditions, the influence of the transformer substation type, the influence of conductor cross section size of the transmission line, the influence of the cable laying mode, etc.

The second is statistical induction. Based on the finished engineering cost information, the engineering cost and index database is established. Through mathematical means, looking for the change rule of all kinds of construction project cost, project cost prediction model is established [1]. Such as engineering analogy method and fuzzy mathematics method, the difficulty of this method lies in the selection of fuzzy analogy estimate model category. Fuzzy system lacks self-study ability, the membership functions and fuzzy rule of which are chosen by experts subjectively, potentially skewing results.

The third kind is the time series method [2]. Such as moving average, exponential smoothing, autoregressive moving average method, the constant ratio regression method and Markov prediction method, regression analysis and grey prediction method. Using month as the unit to gather the cost information, and to generate a variety of cost index data sequence, fitting model is established on the basis of corresponding theory. Moving average method and exponential smoothing method usually lag and it is difficult to determine the weight coefficient. Autoregressive Moving Average (ARMA) model only adapts to the stationary time series. Constant ratio regression method requires regression analysis, however, missing individual data and uneven distribution will lead to the accuracy reduction or even failure of analysis result. Owing to the poor extrapolation, the model has the feature of high-precision and fitting unless the detailed scale and scope of the project are described in the selected completed project. It is difficult to decide the state transferring probability matrix which is the core. A function curve is hard to be established to describe the change trend of engineering cost, which brings limitations to the application of trend extrapolation.

Considering the shortage of the existing unit investment estimate method, this paper puts forward on estimating construction cost of transmission lines project which on the basis of case-based reasoning technology, and analyzes the necessity and feasibility of forecasting system based on the Case-Based Reasoning. The performance of this model is verified by an illustration.

* ***Corresponding author's*** e-mail: 961928275@qq.com

2 The establishment of the case-based reasoning system

2.1 THE PRINCIPLE OF CASE-BASED REASONING SYSTEM

The Case-Based Reasoning is a new strategy to demonstrate knowledge and its primary feature is to use previous similar experiences to solve current problems [3]. In CBR, to solve the problem or situation is known as the target case, and the memorized or completed case is called the source case. Case-Based Reasoning is a strategy which obtains the most similar source case in memory by the hints of the target case and finds the answers by the instruction of the source case. The artificial intelligence technology is often used in project cost system design [4], project cost risk control [5], real estate marketing [6], and project cost quickly estimation [7], etc.

2.2 ANALYSIS OF THE APPLICABILITY OF THE CASE-BASED REASONING SYSTEM

Compared to the existing measures, Case-based reasoning system shows more suitable for forecasting the unit investment of transmission lines. First of all, the knowledge of case-based reasoning system is organized in the form of case structure. Moreover, it is difficult to use the function model to describe the engineering cost of transmission lines and the multi-factors of complex misalignment relations characteristic. However, the form of case can provide more abundant information than that of a set of rules.

Second, fuzzy reasoning is a kind of incremental learning method, and neural network method has similar advantages. With the increase of cases, the coverage of the case (the range of solving the problem) and predictive accuracy are improved. Both fuzzy system and neural network are dynamic system. Neural network is suitable for processing unstructured information, while the fuzzy system deals with the structured knowledge more effecttively. Case-based reasoning shows better expandability, for the reason that it is a reference case experience, and requires little field knowledge, which avoids the integrity and consistency of the knowledge base with the increase of knowledge.

2.3 DESIGN OF THE CASE-BASED REASONING SYSTEM OF COMPLETED PROJECT

Based on the analysis of characteristics of transmission lines construction project, and combined with the working process of the case-based reasoning, a case-based reasonning unit investment forecasting system of transmission lines is built as follow.

The first step is acquiring knowledge from completed projects in the database, and extracting the representative attributes of engineering projects. The second step is establishing a completed project database, namely, to establish an extensible engineering database which saves project overview, the main technical and economic indices and unit investment and other attributes. The third step is confirming the target weigh coefficient, that is to say, factor analysis of SPSS is adopted and each attribute is assigned corresponding weights. The forth step is case retrieval. Based on the previous steps, the similarity between two cases is measured by the distance formula. Case retrieval includes two steps: firstly, searching the projects in the case base as similar as possible, secondly, matching up to the most similar item from the similar projects. The fifth step is case revision. Case revision aims at adapting to the problem to be solved, and on the basis of the predicting result it can set a certain minimum threshold to confirm the corresponding adjustment strategy. The sixth step is case study and maintenance. The existing case base can be complemented along with adding or removing case, adjusting the structure of the case, and resetting the indexes to meet the demand of continuous change and expending requirement of cost management.

3 The implementation of case-based reasoning system

3.1 CASE REPRESENTATION AND ORGANIZATION

Case representation is the premise and basis of case-based reasoning, and belongs to a kind of knowledge representation, which turns out to be “cases transition from our experience”. Problem to be solved are structured, and attributes are extracted. Case representation is a description of the current prediction problem, and the corresponding eigenvalue solutions needs to be obtained through reasoning system. On the one hand, it is the descriptive method of the case, on the other hand it is content contained in the case description. According to the basic data approved by a province in China in 2012 at feasibility research stage, the main factors influencing the unit investment of transmission lines engineering have two aspects. One is the project summary information, including voltage grade, overhead line type, topography, ice, and the path length of the overhead line, etc. The other is main economic indexes such as the proportion of corner, tower, ground wire, iron tower, steel, foundation pit concrete, earth rock and number of insulator string index, etc. Unit kilometer cost is needed to be predicted.

The common methods of case representation currently contain causality graph, object-oriented knowledge representation, logic representation, semantic network, frame representation, full-text representation. The framework representation uses a structure or organization to store the past experience, cases of which are structured in the form of “node-slot-value” of semantic network. So the general structure of framework is a semantic network, by the framework and a set of framework is used to describe various aspects of specific attributes of the slot. Frame-

work is a general structure of semantic network, consists of a framework and a set of slots used to describe all aspects of the specific properties. Each slot can be further equipped with "side" to be specified [8]. In the transmission line projects, slots are described formally as follow:

Item number:
Framework name: <Transmission lines project summary>
Slot 1. Project overview
Side 1. Completion date
Side 2. Construction site
Side 3. Construction type
Side 4. Voltage grade
Side 5. Overhead line types(normal, compact, shrunk, long-span)
Side 6. Altitude

Slot 2. Project properties
Side 1. Line length
Side 2. Tower type
Side 3. Tower base quantity
Side 4. Tower material quantity
Side 5. Wire type
Side 6. Conductors bundled number
Side 7. Cover ice
Side 8. Wind speed
Side 9. Terrain proportion
Side 10. Geological proportion
Side 11. Design of the joint and one standardization

Framework name: <Cost information>
Slot 1. (Property Value Inheritance) project profile inherited

Slot 2. Cost indicators
Side 1. Total cost
Side 2. Ontology construction cost
Side 3. Auxiliary construction cost
Side 4. Other expenses

Slot 3. Major material and labor consumption
Side 1. Artificial man-days
Side 2. Concrete
Side 3. Steel
Side 4. Wire
Side 5. Cement

Slot 4. Bill of civil part quantities
Side 1. Earthwork and rockwork
Side 2. Basic engineering

Slot 5. Bill of installation quantities
Side 1. Tower engineering
Side 2. Wiring project
Side 3. Attachment engineering

Slot 6. Result sets
Side 1. Cost of per kilometer
Side 2. Number of per kilometer tower base
Side 3. Wire consumption of per kilometre

Slot 7. Related instructions
Side 1. Analysis of representative case

The case number is the unique identification of engineering cases, represented by a string. Engineering properties are briefly introduced, including the profile information referred construction time, site, time schedule. Engineering metrics includes major staff and technology arrangement technical index, which corresponds case retrieving information. The result sets show better fidelity to the solutions, which includes engineering unit investment and major staff and technology consumption. Relevant specification is the analysis of the typical engineering cases, the relevant cost information of which will be explained or extended. The referred cost information reflected by the cases will be explained and extended.

In order to express case more methodically, cases are organized and arranged to form corresponding completed project database according to the characteristics of the case and retrieval need. Case organization directly affects the efficiency of case retrieval, whose key is the selection of key indicators of transmission line project. Key indicators are determined by the problem to be solved and the characteristics of certain cases.

3.2 CASE RETRIEVAL

3.2.1 Case index

Case indexing method has three types: nearest neighbor, inductive reasoning and knowledge-guided approach. According to the case representation and organization form, case retrieval is conducted by means of the combination of nearest neighbor approach and knowledge-guided approach in the engineering cost estimation system based on case-based reasoning.

Knowledge-guided strategy selects the optimum cases through searching the target case and completed cases in the database, whose attributes, including voltage grade, overhead line type, topography and other indicators, are determined by domain knowledge. Case sets in line with the features of pending cases domain can be output as the candidate sets to be selected in the nearest neighbor approach.

Nearest neighbor approach is a kind of strategy that target case to be resolved and completed cases in the case base are retrieved and traversed, respectively obtaining the similarity between cases of each index field. According to the established weighted vector, the similarity among cases and the problem to be solved are calculated. In addition, users can also get the cases whose similarity exceed the threshold, then modify and retrieve them.

The engineering unit investment factors of transmission lines will be divided into two categories: control attributes and comparison attributes. Among them, control attributes refers to the project profile and basic information of attributes, such as voltage grade and overhead line type, which is essential for judging the similarity of transmission lines projects. Primary retrieval is kind of searching the feasible

candidate sets by matching exactly the target case and the case base by the attribute index. Then, by applying the results of paired comparison, an index structure is established. When control properties of transmission line engineering match completely, comparing attributes can be retrieved through advanced searching. On the premise of same control properties, main characteristics of comparing attributes and cases of largest number of matched features are taken into first consideration.

In order to realize the initial matching of the control properties between the cases to be solved and the case base, c++.net is applied to set primary retrieval rules. Combined with representation and organization characteristics of engineering case, the primary search selects the voltage level for the first retrieval conditions, altitude for the second retrieval conditions, overhead line type for the third retrieval conditions, which finally form the candidate sets based on the nearest neighbor algorithm.

The main code of primary retrieval is as follows:

```
int JuniorSearch()
 {
 int count = 0;
 juniorData[count] = targetData[0];
count++;
 for (int index = 1;index<totalCount;index++)
 {
array<String^>^ rowData = tableData[index];
// Matching the first condition?
                if(Convert::ToInt32(rowData[1])!=Convert::ToInt32(
tarValue[1]))
                {
                continue;
}// Matching the second condition?
if((Convert::ToInt32(rowData[2])>Convert::ToInt32(tarValue[2])
)||(Convert::ToInt32(rowData[3])<Convert::ToInt32(tarValue[3]))
)
{
continue;
}// Matching the third conditions?
if(TransformType(rowData[4])!=LINE_TYPE(int(tarValue[4])))
{
continue;
}//Meeting all the three conditions?
juniorData[count] = rowData;
count++;
}
 return count;
 }
```

3.2.2 Similarity retrieval

Through primary index we can find candidate set of cases in the completed project as little as possible, which has reference significance to the target problem to be resolved. Then the similarity of comparing attributes are retrieved and evaluated, in order to find the relative matching case to forecasting unit investment. Due to the incompleteness of information, only the cases whose characterizing attribute are similar can be matched and retrieved, difficult to matching completely. So this section will use SPSS statistical analysis software[9], according to the correlation analysis between each variable, the main factors that influence the engineering cost are determined, furthermore the weight of each comparing attributes influencing the case similarity was defined.

Similarity is basis of the case-based reasoning, the measure method and standard of which determine success or failure of case-based reasoning and the accuracy of the prediction to a great extent. The similarity metrics between cases is represented by S. The closer the control attributes and comparing attributes of x and x' in two cases are, the higher the value of $S(x,x')$ will be. The further the control attributes and comparing attributes of x and x' in two cases are, the lower the value of $S(x,x')$ will be. Generally the similarity between cases is measured by the distance in feature space. The distance metrics which are frequently used include Euclidean distance, City Block, Minkowski distance, Chebyshev distance weighted variance distance and Markov distance.

The similarity of the prediction model in this paper is measured by Euclidean distance[10]. A case is represented with a data vector x. In this article $x_i = \{x_{i1}, x_{i2}, \cdots x_{in}\}$ represents case, where n is the dimension of vector, m is the case numbers in the space X. The calculation steps are as follow:

1) Standardization of data: $x_{mn} = \dfrac{x_{mn} - \min\limits_{m} x_{mn}}{\max\limits_{m} x_{mn} - \min\limits_{m} x_{mn}}$.

2) Determination of the weight: each influence of the comparison properties to transmission line project varies greatly. Some of attributes, including tower index, wire index, foundation pit concrete index, are crucial, while others, such as the number of the insulator string, have little effect. Similarity is sensitive to the change of the weight, so reasonably setting the weight ensures that the results of prediction accuracy. Many methods could be used to calculate the weight, such as expert discussion method, analytic hierarchy process, however, the strong subjectivity of which can lead to similarity deviation if the weight is allocated unreasonably. Factor analysis tools in SPSS software is used to assign the value of relevant weight, which greatly reduces the subjective influence on the prediction result. The article extracts 13 factors from 139 cases according to the engineering characteristics of transmission line projects of a province in 2012.

Based on the selected case database, cases of different voltage levels are selected, in order to examine the selected variable distribution. The unit investment varies greatly among different voltage levels, including 220kV, 110kV, 35kV. In order to identify the important deviation value of

the distribution of the variable based on the transmission line projects of different voltage levels, box plot in the software SPSS is used to remove the outlier spots. The result is shown below in Figure 1.

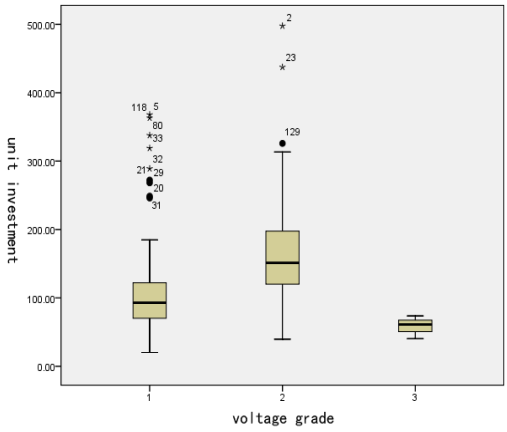


FIGURE 1 Box of voltage grade and unit investment

Based on the factor analysis tools of software SPSS, correlation matrix table is obtained after inputting the data of 13 factors for correlation. Unit investment is the factor to predict, the correlation between other attributes and the target attributes shall be strong, while the correlation between other factors shall be as weak as possible in order to minimize collinearity between them. After inspection, the correlation coefficient between tower index and the material of iron tower is 0.849 which means strong liner correlation and they are merged into tower index. Cast-in-situ foundation and other foundation pit concrete index, bored concrete pile foundation index, are merged as foundation pit concrete index. After calculation, the correlation coefficient between the foundation pit concrete index and foundation pit earthwork index is 0.793, which means strong correlation, and foundation pit concrete index is reserved. The correlation coefficients between unit investment conductor insulator composite index, conductor insulator index without composite are inferior to 0.4, which means weak correlation, and two indexes are deleted. After using factor analysis, the final correlation coefficient matrix is shown in the following Table 1.

TABLE 1 Correlation matrix

		Angle proportion	**Tower index**	**Wire index**	**Grounding line index**	**Foundation pit concrete index**	**Foundation steel index**	**Unit investment**
correlation	Angle proportion	1	-0.03	0.424	-0.136	-0.316	0.358	0.322
	Tower index	-0.03	1	0.234	0.196	0.027	0.276	0.416
	Wire index	0.424	0.234	1	-0.183	-0.282	0.725	0.434
	Grounding line index	-0.136	0.196	-0.183	1	-0.034	-0.168	-0.261
	Foundation pit concrete index	-0.316	0.027	-0.282	-0.034	1	-0.199	-0.202
	Foundation steel index	0.358	0.276	0.725	-0.168	-0.199	1	0.57
	Unit investment	0.322	0.416	0.434	-0.261	-0.202	0.57	1
Sig. (one-sided)	Angle proportion		0.434	0.006	0.221	0.034	0.019	0.032
	Tower index	0.434		0.091	0.133	0.441	0.057	0.007
	Wire index	0.006	0.091		0.15	0.053	0	0.005
	Grounding line index	0.221	0.133	0.15		0.425	0.172	0.068
	Foundation pit concrete index	0.034	0.441	0.053	0.425		0.129	0.125
	Foundation steel index	0.019	0.057	0	0.172	0.129		0
	Unit investment	0.032	0.007	0.005	0.068	0.125	0	

As shown in the Table 1, twelve attributes are extracted from the database, six of which are comparing indexes, including angle proportion, tower index, wire index, ground wire index, foundation pit concrete index, and foundation steel index. As the unit investment index is the target prediction index, the larger the correlation coefficient between comparing attributes and unit investment index is, the strong similarity between the target case and case set is. Based on the correlation coefficient between 6 comparing attributes and unit investment respectively in the correlation matrix table of SPSS factor analysis, the similarity weight of the cases are calculated [11], which is shown in Table 2.

TABLE 2 The weight of Case similarity impacted by characteristics factors

Engineering characteristics	The correlation coefficient to unit investment	The weight
Angle proportion	0.322	0.16
Tower index	0.416	0.19
Wire index	0.434	0.20
Ground wire index	0.216	0.10
Foundation pit concrete index	0.202	0.09
Foundation steel index	0.570	0.26
	2.16	1.00

3) Similarity calculation: to calculate the similarity, the users should input new cases, the attributes of which mainly include angle proportion, tower index, wire index, ground wire index, foundation pit concrete index, and foundation steel index. $d(x_i, y_j)$ represents the distance between the new case x_i and a certain case y_j from the database. The distance is defined by weighted Euclidean distance method [12].

$$d(x_i, y_j) = \left[\sum_{k=1}^{p} w_k (v_{ik} - v_{jk})^2 \right]^{\frac{1}{2}},$$

where p represents the total number of attributes. v_{ik} represents the value of k-th attributes of case i-th. w_k represents the weight of k-th attribute.

The target case and the candidate case base are traversed to retrieve with the maximum value $d_{\max}$, and three cases with the closest distance which is represented as d_1, d_2, d_3 respectively.

The similarity between new case and the case from the database is defined as $sim(x_i, y_j)$. The similarity calculation formula is as followed [13]:

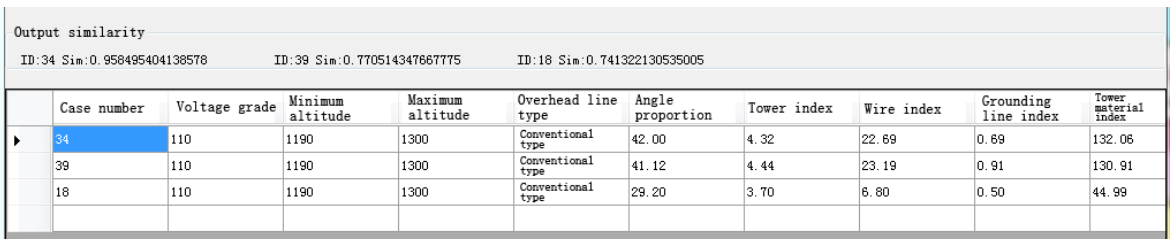
$$sim(x_i, y_j) = 1 - \frac{d_i}{d_{\max}} (i = 1, 2, 3).$$

The output of similarity value should between 0 and 1. The closer the sim value is to 1, the greater the similarity is.

4) Case similarity retrieval: firstly, Similarity is calculated as the followed program code:

```
double CalSim (int P, array <double> ^W,array
<double>^Vi,array<double>^Vk)
{
 double sumUp = 0;
for (int index=0;index<P;index++)
{
double w = W[index];
double vi = Vi[index];
double vk = Vk[index];
 //sum the squares of numerator
 sumUp += w * (vi - vk) * (vi - vk);
 }
 return (System::Math::Sqrt(sumUp));
 }
Advance search is conducted, and the program code is as follow,
int HighSearch()
 {
 if(juniorCount <= 1)
 {
 return 0;
 }
 int count = 0;
 highData[count] = juniorData[0];
 count++;
 int colTotal = juniorData[0]->Length;
 array<double>^ temp = gcnew array <double> (colTotal);
array<double>^ retList = gcnew array <double> (juniorCount -
1);
array<int>^ retIndex = gcnew array<int>(juniorCount - 1);
for (int index = 1;index<juniorCount;index++)
 {
 // The original data
 array<String^>^ rowData = juniorData[index];
 // Convert to processing dataTransformData (rowData,temp);
 // Calculate the similarity
 double sim = CalSim(colTotal,tarW,tarValue,temp);
 // Save the result?
 retList[index-1] = sim;
retIndex[index-1] = Convert::ToInt32(rowData[0]);
 }
 // sort
Array::Sort(retList,retIndex);
//Select the maximum data Y
 int length = retList->Length;
 Dmax = retList[length - 1 ];
 for (int index = 0;index < length && index < 3 ;index ++)
 {
 int id = retIndex[index];
highData[count] = tableData[id];
count++;
 }
 // show result?
showResult(retList,retIndex);
 return count;
 }
```

The case No.34, No.39, No.18 are selected with similar value of 98.85%, 77.05%, 74.13% respectively, which are the top three cases with the largest similarity. The result is shown in the following Figure 2:


Output similarity

ID:34 Sim:0.958495404138578 ID:39 Sim:0.770514347667775 ID:18 Sim:0.741322130535005

Case number	Voltage grade	Minimum altitude	Maximum altitude	Overhead line type	Angle proportion	Tower index	Wire index	Grounding line index	Tower material index
34	110	1190	1300	Conventional type	42.00	4.32	22.69	0.69	132.06
39	110	1190	1300	Conventional type	41.12	4.44	23.19	0.91	130.91
18	110	1190	1300	Conventional type	29.20	3.70	6.80	0.50	44.99

FIGURE 2 Prediction system based on case-based reasoning

3.3 CASE REVISION

Based on similarity calculation, the similarity between the target case and the candidate case in the case base is obtained. An appropriate minimum threshold is set according to the influence of similarity to the target case. Cases whose similarity are greater than or equal to the lower threshold are selected as the matching case of the target case, which provides decision support for new projects. If the similarity of the model doesn't reach the threshold, case will be amended according to the voltage level, overhead line type and other attributes in the case base.

If the similarity threshold is set to 0.9 in the above case-based reasoning system, the similarity between case 24 and the target case is 95.85%, above the threshold, which needs no adjustment. Case 39 and case 18 need appropriate amendments according to the expert experience and human-computer interaction, because the similarity of which are below the threshold. After case revision the similarity is above the threshold, which guarantees the precision of the prediction.

Comparison between main engineering characteristics of target case and similar case is shown in Table 3.

TABLE 3 The main engineering characteristics of target case and similar cases

Case No.	Angle ratio	Tower index	Wire index	Ground wire index	Iron tower material index	Basic steel index	Foundation pit concrete index	Unit investment
Target case	39.00	4.50	24.92	0.70	133.24	25.36	269.71	
34	42.00	4.32	22.69	0.69	132.06	28.00	99.78	279.63
39	41.12	4.44	23.19	0.91	130.91	24.94	270.99	153.16
18	29.20	3.70	6.80	0.50	44.99	17.68	161.28	178.84

As for case revision, the inconsistent between target case and senior retrieved case needs considering mainly, in order to find out the optimal solution. For example, the perception of angle steel tower, steel towers and steel rods needs to be considered in similar cases, for the iron tower indices differ greatly. The influence of different factors including wire type and wire section needs to be considered in order to correct the wire indicator. Casting type of basis needs to be considered to rectified foundation pit concrete index. After the above factors are considered comprehensively, the cases whose similarity are below the threshold are modified, and unit investment is derived from appropriate forecasting model.

Based on the comparison of the cost index and the actual situation of the project, the prices of case No.39 and No.18 increase 35%, namely, 206.766 and 236.07 ten thousand yuan/km, respectively. The prediction of unit investment is 2408200 using arithmetic average, which is 3.92% lower than actual estimate. The predicted results are accurate and convincing.

3.4 CASE LEARNING AND MAINTENANCE

Case learning takes a new round of case decision by adding the newly completed project to the corresponding decision-making environment through the case description and index. The case learning process can update the case base, so as to provide a rich case database for case reasoning. With the quantity of the case increasing, however, the efficient of the program may fall, emerging redundancy, repetition and obsolete. When the cost of retrieving similar case is greater than the benefit it offers, index mechanism of the case database should be adjusted and the case data should be maintained, including improving obsolete case and deleting the invalid case. The forecasting system's capacity to solving problems is gradually strengthened and improved through the case study and maintenance.

4 Conclusions

Based on artificial intelligence methods, case-based reasoning forecasting system of unit investment is analyzed, which provides a quick and effective approach for unit investment rapid estimation at estimate stage. Firstly, completed engineering data is organized, stored and managed using ACCESS. Secondly, the software SPSS is used to data denoising, and the cost index weight coefficient is built by factor analysis tools. Then, the programming algorithm of dynamic case reasoning system component is designed by C++/CLI language. The program dynamically associates the activities of the execution program with the completed engineering database, which realizes the primary search, advanced search and similar case output. Based on the completed transmission line projects in 2012, the essay accomplishes the rapid search of similar cases and optimizes the output of the program after adjustment.

At present, the case-based reasoning system prediction for transmission line projects is still at its initial stage. The uncertainty of the index weight setting still exists because of the strong dependence of historical database. Similarity distance formula is the key to the precision of prediction which directly affects the accuracy of the prediction results. The modification of the similar case is the difficulty of the research due to the objective of expert evaluation.

The combination of the unit transmission line project with the artificial intelligence greatly improves the speed of the unit investment fast calculation. The system provides reference for the foundation of the building cost index database, which also provides reference for the data accumulation of the owners' projects.

References

[1] Lu H M, Wang X Q 2013 Research on project cost prediction based on the fuzzy pattern recognition and BP neural network *Project Management Technology* **11**(5) 58-61

[2] Shahandashti S M, Ashuri B 2013 Forecasting Engineering News-Record Construction Cost Index Using Multivariate Time Series Models *Journal of Construction Engineering and Management* **139** 1237-43

[3] Li F G 2011 Technology of intelligent decision making based on case-based reasoning *Beijing Normal University Publishing Group* 11-2 *(in Chinese)*

[4] Liu Z T 2009 Research on evaluating construction cost system design based upon case-based fuzzy reasoning *Shanxi Architecture* **35**(27) 240-1*(in Chinese)*

[5] Wang C Z, Guo Y J, Yu Z J, Chen W 2007 Credit risk control system based on case-based reasoning *Journal of Northeastern Uuiversity Natural Science* **28**(3) 450-3 *(in Chinese)*

[6] Wang J 2008 Case-based Reasoning Research for Ontology-based Real Estate Marketing *Wuhan University of Technology (in Chinese)*

[7] Feng W M, Cao Y J, Ren H 2003 The Study on the case-based Reasoning Method of the cost-estimation in civil engineering *China Civil Engineering Journal* **36**(3) 51-6 *(in Chinese)*

[8] Qiu Z H 2010 Research on the CBR engineering project consulting system *Dalian University of Technology (in Chinese)*

[9] Ye Q 2011 Research and Application of the Artificial Intelligence Methods in the Field of Valuation *Huaqiao University (in Chinese)*

[10] Zheng K N, Li X Y, Yang K 2011 Gaussian Method in Case-Based Reasoning and Applications *Operations Research and Management Science* **20**(6) 99-104

[11] Kim S, Shim J H 2014 Combining case-based reasoning with genetic algorithm optimization for preliminary cost estimation in construction industry *Canadian Journal of Civil Engineering* **41**(1) 65-73

[12] Fu Y F, Liu X D, Li Y J 2012 Software cost estimation method based on the genetic algorithm and case-based reasoning *Computer Engineering and Applications* **48**(8) 87-8

[13] JiS H, Park M, Lee H S, Ahn J, Kim N, Son B 2011 Military Facility Cost Estimation System Using Case-Based Reasoning in Korea *Journal of Computing in Civil Engineering* **25**(3) 218-31

Authors

Tao Yi, born in October 1967, Beijing, China.

Current position, grades: associate professor at the School of Economics and Management at North China Electric Power University. Master instructor of Business Administration of North China Electric Power University.
University studies: Wuhan University.
Scientific interests: technical economics and management, theory and application of project management.
Publications: 18 papers.

Shanshan Cui, born in June 1990, Beijing, China.

Current position, grades: student of management science and Engineering at North China Electric Power University.
Scientific interests: engineering cost management in power grid project and constructional engineering.
Publications: 1 paper.

Yi Zhang, born in October 1990, Beijing, China.

Current position, grades: student of Management science and Engineering at North China Electric Power University.
Scientific interests: engineering cost management in power grid project and constructional engineering.
Publications: 1 paper.

Comparison on turnovers of agricultural products futures based on wavelet packet transform method

Qingbin Han*

School of Marxism, Wuhan University of Technology, 430070, Wuhan, China

Received 1 March 2014, www.cmnt.lv

Abstract

In order to distinguish the different patterns and evolving trends on turnovers of agricultural products futures between Zhengzhou Commodity Exchange and Dalian Commodity Exchange. A time-frequency analysis approach, i.e. wavelet packet energy spectrum, is investigated in this paper. Firstly, wavelet packet energy spectrum is briefly introduced. Secondly, two different non-stationary signals of turnover of agricultural products futures from 2009 to 2013 coming from Zhengzhou Commodity Exchange and Dalian Commodity Exchange are described in wavelet packet energy spectrum. First, reconstructed coefficients of main analysis wavelets packet of signals of turnover of agricultural products futures of Zhengzhou commodity exchange and Dalian Commodity Exchange are calculated. The obtained frequency band power ratios are used to show the different characteristics and evolving trends on turnovers of agricultural products futures between Zhengzhou Commodity Exchange and Dalian Commodity Exchange. With these results, the two signals are distinctly different from each other. It is proved that the technique of wavelet packet energy spectrum is effective for the purpose of distinction of turnover of agricultural products futures in commodity exchanges.

Keywords: turnover amounts, time-frequency analysis, non-stationary signals, wavelet packet energy spectrum, frequency band power ratios

1 Introduction

Turnover of agricultural products futures of Zhengzhou Commodity Exchange (Z-CE) and Dalian Commodity Exchange (D-CE) includes so much information that it can be used to diagnose, treat or predict economic trends. The data turnover of agricultural products futures has one or more of the following problems, (a) the total data span is too short, (b) the data is non-stationary and (c) the data represents nonlinear process [1-3].

In this paper, we choose turnover of agricultural products futures between Zhengzhou commodity exchange and Dalian commodity exchange to analysis. From their time domain waveforms and their power spectrum densities, they are not easily to be distinguished.

Wavelet packet energy spectrum analysis features multi-definition in time-frequency domain as follows: signal is decomposed into two parts in a wavelet transform, low frequency part and high frequency part. Then the low frequency part is decomposed as mentioned above again while the high frequency part remained. Repeatedly, multi-level frequency decomposition can be achieved. Wavelet packet energy spectrum is based on this, and it conducts an equal-band and orthogonal decomposition in the whole frequency domain. After that, energies of the reconstructed signal in the related frequency band are calculated, and frequency-band energy ratio (FBER) is obtained [4].

In this paper, wavelet packet energy spectrum is adopted to decompose the turnover of agricultural products futures between Zhengzhou commodity exchange and Dalian commodity exchange in the whole domain, and using proper frequency bands signals are reconstructed according to the turnover of agricultural products. Energies of the corresponding reconstructed signals are estimated by calculating the frequency-band energy ratio, which is used as a qualitative index for estimating different sleeping states. Plenty of experimental data [5] are used to validate the method, and the FBER is confirmed to be effective.

2 Basic algorithm procedure of wavelet packet energy spectrum

Wavelet transform is used to describe the signal, which is a 1-dimension signal in time domain originally, as a 2-dimension signal in time-frequency domain. Wavelet transform of function is defined as follows:

$$Wf(a,b) = (1/\sqrt{a})\Phi(\frac{t-b}{a})f(t)\mathrm{d}t, \tag{1}$$

where $\Phi(t)$ is called the mother wavelet, a and b are scale parameter and position shift parameter, respectively.

Wavelet packet features orthogonal, independent and multi-definite decomposition, the general form is

$$W_j = \bigoplus_{m=0}^{2^k-1} U_{j-k}^{2^k+m} \qquad (j,k,m \in Z), \tag{2}$$

where $j = 1,2,\cdots$; $k = 1,2,\cdots,j$; and $m = 0,1,2,\cdots, 2^k-1$.

According to multi-definition theory, wavelet packet transform treated Hilbert space as orthogonal summation of all the sub spaces, namely $L^2(R) = \oplus W_j$, $j \in Z$. Therefore, wavelet packet transform can decompose the signal into independent frequency bands orthogonally, and with

**Corresponding author's* e-mail: what_han@163.com

no redundancy or leakage. In a wavelet packet transform with a definition *j*, $x^{k,m}(i)$ represents the discrete signal in subspace $U_{j-k}^{2^k+m}$ corresponding to x_{2^k+m}.

The number of levels in the wavelet packet energy spectrum is determined according to the signal itself and the characteristic parameter. The frequency range of the bands obtained after decomposition is

$$f(j,i)=\left[\frac{f_s(j-1)}{2^{(i+1)}},\ \frac{f_s j}{2^{(i+1)}}\right], \tag{3}$$

where, $f(j,i)$ represents the range of band j in level $i(j=1,2,\cdots,2^i)$, f_s is sample frequency.

Schematic plan of wavelet packet decomposition is shown in Figure 1

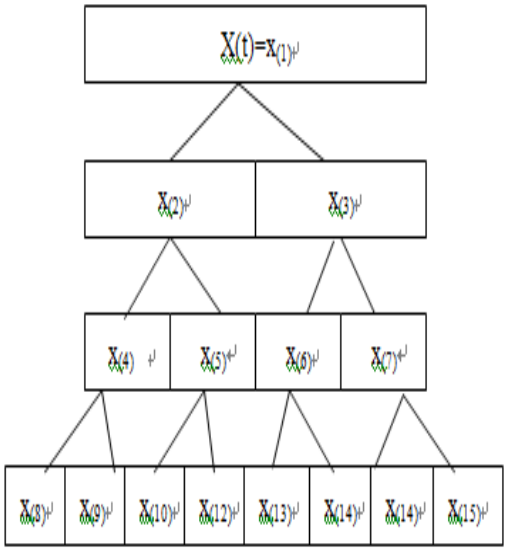


FIGURE 1 Schematic plan of wavelet packet decomposition

Assume that the data length of discrete signal $x^{k,m}(i)$ is *N*, the energy can be represented as

$$E_n\left(x^{k,m}(i)\right)=\frac{1}{N-1}\sum_{i=1}^{N}\left(x^{k,m}(i)\right)^2, \tag{4}$$

where, k is the times of decomposition, *m* is the sequence number of the frequency bands after decomposition.

According to conversation of energy theory, we have

$$E_n\left(x(t)\right)=\sum_{m=0}^{2^k-1}E_n\left(U_{j-k}^{2^k+m}\right)=\sum_{m=0}^{2^k-1}E_n\left(x_{2^k+m}\right)=\sum_{m=0}^{2^k-1}E_n\left(x^{k,m}(i)\right). \tag{5}$$

The ratio of energy of frequency band m to that of the signal, namely, frequency-band energy ratio in unitary wavelet packet transform is:

$$E_n(m)=\frac{E_n\left(x^{k,m}(i)\right)}{E_n\left(x(t)\right)}. \tag{6}$$

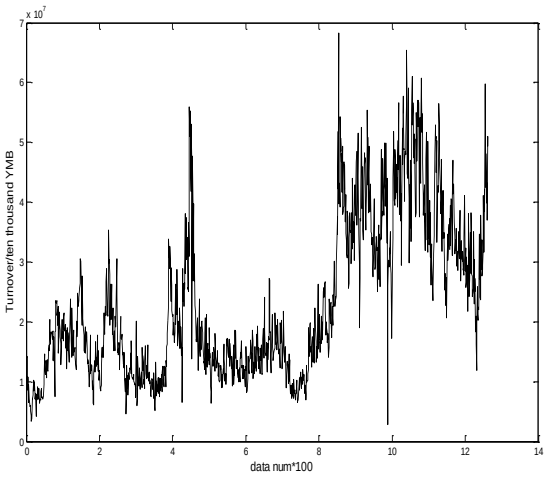
The summation of all the frequency-band energy ratios is 1, namely

$$\sum_{m=0}^{2^k-1}E_n(m)=1. \tag{7}$$

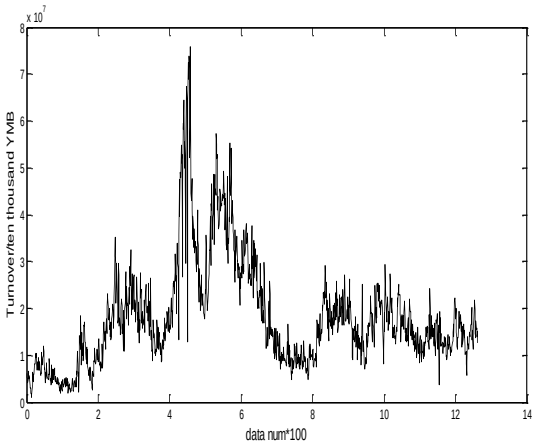
3 Results

We choose two time series of turnover of agricultural products futures between Zhengzhou commodity exchange and Dalian commodity exchange to be analyzed. These two time series are recorder as turnover of every day agricultural products futures of Zhengzhou commodity exchange and Dalian commodity exchange from the beginning of 2009 to the end of 2013.

There are two different non-stationary signals are shown in Figure 2. From Figure 2, the difference is obviously, but it is not easy to identify one from each other.



a) Turnover of agricultural products futures of Z-CE



b) Turnover of agricultural products futures of D-CE

FIGURE 2 Original signals of turnover of agricultural products futures of Zhengzhou commodity exchange and Dalian commodity exchange

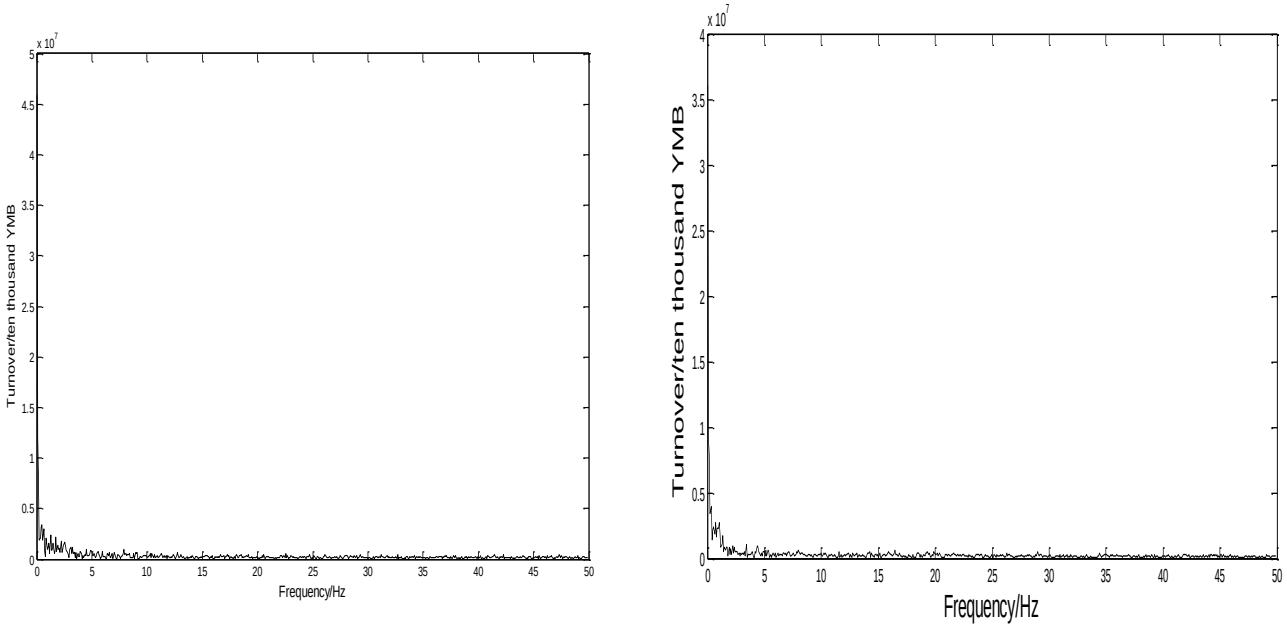
The method above is verified in signals of turnover of agricultural products futures of Zhengzhou commodity exchange and turnover of agricultural products futures of Dalian commodity exchange. The original signal is sampled in 500 Hz priory to make sure that the frequency range of wavelet packet analysis is consistent with the frequency bands of the turnover of agricultural products futures of Zhengzhou commodity exchange and turnover of agricultural products futures of Dalian commodity exchange. Frequency range of 4th level of turnover of agricultural products with wavelet packet decomposition is shown in Table 1. There are 16 wavelet packets in the high scale; each frequency width is 31.25 Hz.

TABLE 1 Frequency range of 4th level of turnover of agricultural products with wavelet packet decomposition

No.	4 layer number	Frequency sphere (Hz)	No.	4 layer number	Frequency sphere (Hz)
1	[4,0]	0~31.25	9	[4,12]	250~281.25
2	[4,1]	31.25~62.5	10	[4,13]	281.25~312.5
3	[4,3]	62.5~93.75	11	[4,15]	312.5~343.75
4	[4,2]	93.75~125	12	[4,14]	343.75~375
5	[4,6]	125~156.25	13	[4,10]	375~406.25
6	[4,7]	156.25~187.5	14	[4,11]	406.25~437.5
7	[4,5]	187.5~218.75	15	[4,9]	437.5~468.75
8	[4,4]	218.75~250	16	[4,8]	468.75~500

Frequency spectrum analysis of turnover of agricultural products futures of Zhengzhou commodity exchange and turnover of agricultural products futures of Dalian commodity exchange is shown in Figure 3. According to the main frequency, four wavelet packets of the signals are analyzed. The frequency bands are 31.25-62.5 Hz, 125-156.25 Hz, 187.5-218.75 Hz, 281.25-312.5 Hz.

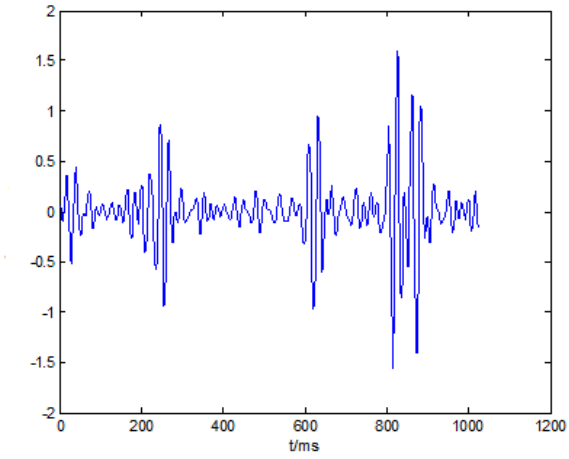
Figures 4 and 5 are the main analysis wavelet packets reconstruct figures according to the signal components.



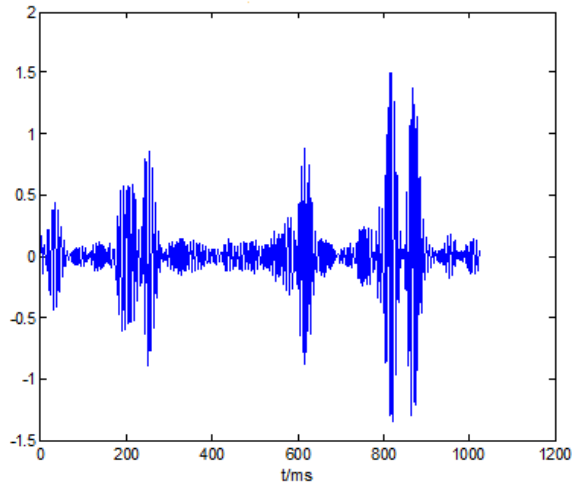
a) Turnover of agricultural products futures of Zhengzhou commodity exchange

b) Turnover of agricultural products futures of Dalian commodity exchange

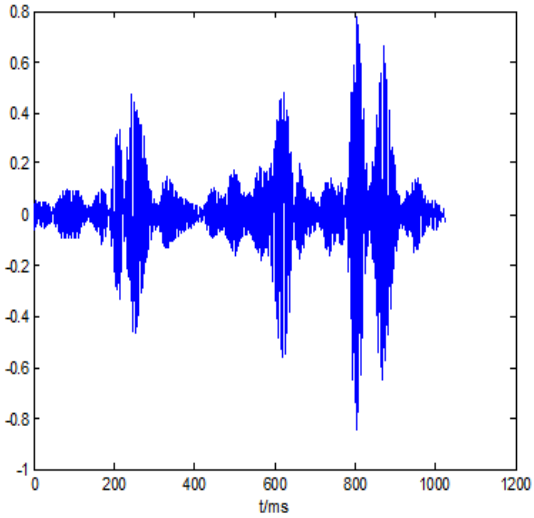
FIGURE 3 Frequency spectra of original signals of turnover of agricultural products futures of Zhengzhou commodity exchange and Dalian commodity exchange



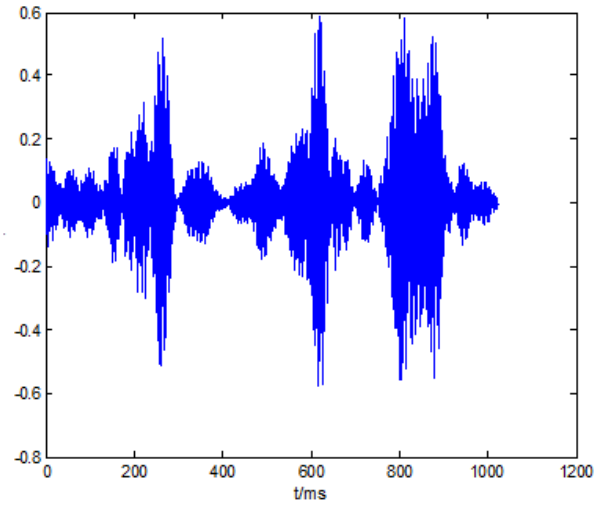
a) Reconstruct packet (4,1) 31.25-62.5 Hz



b) Reconstruct packet (4,6) 125-156.25 Hz

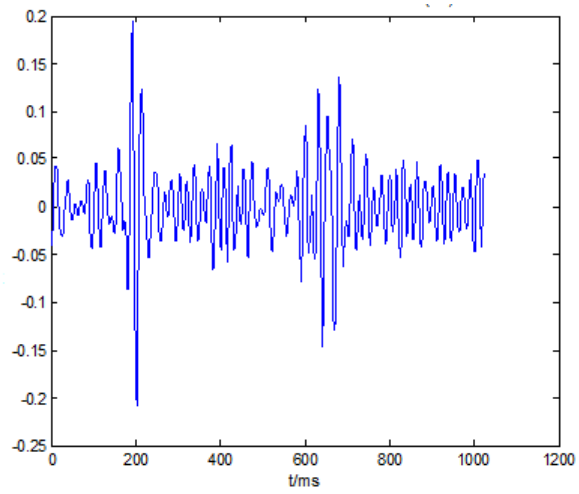


c) Reconstruct packet (4,5) 187.5-218.75 Hz

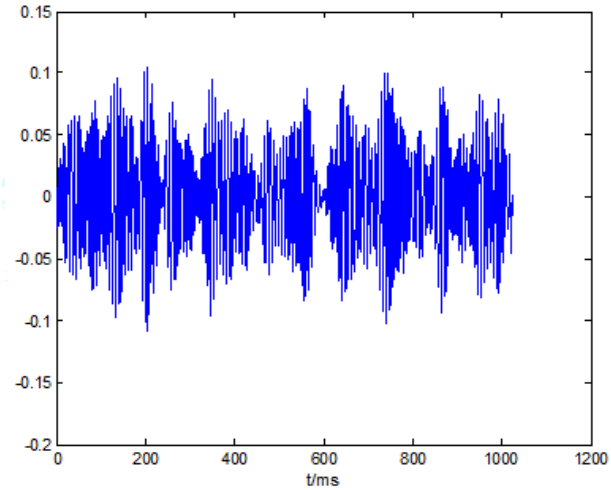


d) Reconstruct packet (4,13) 281.25-312. 5Hz

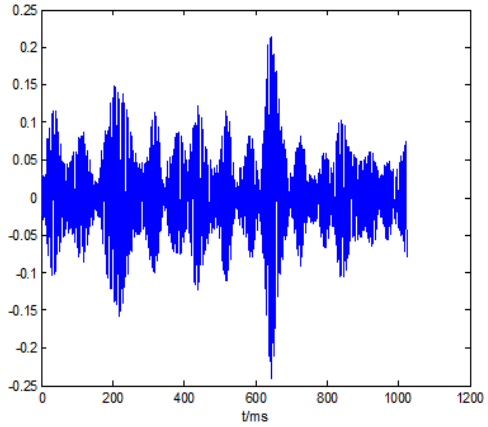
FIGURE 4 Reconstructed coefficients of main analysis wavelets packet of signals of turnover of agricultural products futures of Zhengzhou commodity exchange



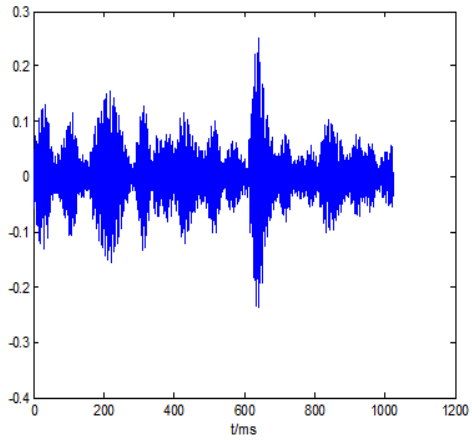
a) Reconstruct packet (4,1) 31.25-62.5 H



b) Reconstruct packet (4,6) 125-156.25 Hz



c) Reconstruct packet (4,5) 187.5-218.75 Hz



d) Reconstruct packet (4,13) 281.25-312.5 Hz

FIGURE 5 Reconstructed coefficients of main analysis wavelets packet of signals of turnover of agricultural products futures of Dalian commodity exchange

The FBERs of the 2 groups of signals obtained are shown in Figures 6 and 7.

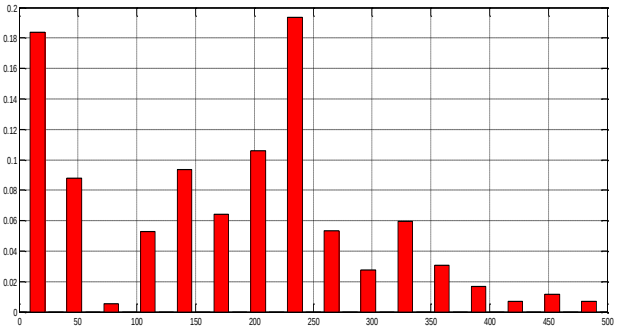


FIGURE 6 Power ratio bars for different frequency bands of turnover of agricultural products futures of Zhengzhou commodity exchange with wavelet packet decomposition

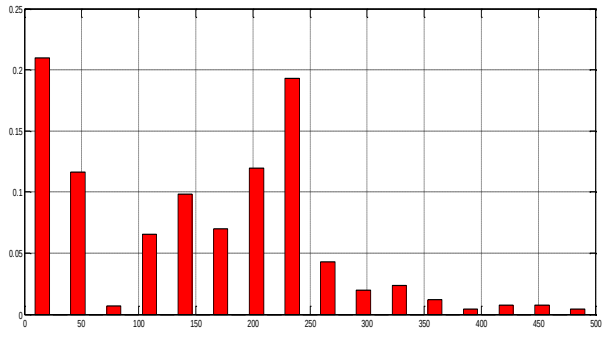


FIGURE 7 Power ratio bars for different frequency bands of turnover of agricultural products futures of Dalian commodity exchange with wavelet packet decomposition

The FBERs of the signals of turnover of agricultural products futures of Zhengzhou commodity exchange and turnover of agricultural products futures of Dalian commodity exchange are shown in Table 2 and Table 3. We can find the values and their tendency of the total amplitude contributions of turnover of agricultural products futures are remarkable different of Zhengzhou commodity exchange and Dalian exchange.

TABLE 2 Calculated results of scaled energy in frequency bands of turnover of agricultural products futures of Zhengzhou commodity exchange with wavelet package

No.	4 layer number	FBER	No.	4 layer number	FBER
1	[4,0]	0.1836	9	[4,12]	0.0532
2	[4,1]	0.0878	10	[4,13]	0.0275
3	[4,3]	0.0054	11	[4,15]	0.0597
4	[4,2]	0.0528	12	[4,14]	0.0305
5	[4,6]	0.0938	13	[4,10]	0.0166
6	[4,7]	0.0641	14	[4,11]	0.0067
7	[4,5]	0.1058	15	[4,9]	0.0118
8	[4,4]	0.1938	16	[4,8]	0.0068

TABLE 3 Calculated results of scaled energy in frequency bands of turnover of agricultural products futures of Dalian commodity exchange with wavelet package

No.	4 layer number	FBER	No.	4 layer umber	FBER
1	[4,0]	0.2096	9	[4,12]	0.0427
2	[4,1]	0.1160	10	[4,13]	0.0194
3	[4,3]	0.0071	11	[4,15]	0.0234
4	[4,2]	0.0657	12	[4,14]	0.0177
5	[4,6]	0.0986	13	[4,10]	0.0041
6	[4,7]	0.0699	14	[4,11]	0.0077
7	[4,5]	0.1193	15	[4,9]	0.0077
8	[4,4]	0.1927	16	[4,8]	0.0044

4 Conclusions

Turnover of agricultural products futures of Zhengzhou commodity exchange and Dalian commodity exchange as two different non-stationary signals are used as examples to be described and distinguished in the time-frequency analyses of wavelet package decomposition in the paper. The technique of wavelet package decomposition is proved to be effective on processing and distinguishing the differences of the non-stationary signals.

For the original non-stationary signals of turnover of agricultural products futures of Zhengzhou commodity

exchange and Dalian commodity exchange given here, the obtained wavelet package decomposition are easily distinguished from each other. The frequency spectra are also different between them. These results demonstrate that the wavelet package decomposition can offer a more effective way for identifying the different features of turnover of agricultural products futures of Zhengzhou commodity exchange and Dalian commodity exchange.

References

[1] Holger K, Thomas S 1999 Nonlinear Time Series Analysis *Cambridge University Press*
[2] He Z J1996 Time-frequency (scale) analysis and diagnosis for non-stationary dynamic signal of machinery *International Journal of Plant Engineering and Management* **1**(1) 40-7
[3] Meltzer G S, Ivanov Y Y 2003 Fault detection in gear drives with non-stationary rotational speed- Part I: the time-frequency approach *Mechanical Systems and Signal Processing* **17**(5) 1033-47
[4] Feng Y H, Schlindwein, Fernando S 2009 Normalized wavelet packets quantifiers for condition monitoring *Mechanical Systems and Signal Processing* **23**(3) 712-23
[5] Wong W K, Yuen C W M, Fan D D, Chan L K, Fung E H K 2009 Stitching defect detection and classification using wavelet transform and BP neural network *Expert Systems with Applications* **36**(2) 3845-56

Author

Qingbin Han, China.

Current positon, grades: PhD candidate in School of Marxism of Wuhan University of Technology, 2011.
University study: MSc in law in Northeastern University, 2007.
Research activities: marxist philosophy, quantitative economics.
Professional Activities and Memberships: Tutor in University of Science and Technology of Liaoning (2007-2011).

Research on Cooper ethical decision model based on Chinese opera

Yuan Qin*

Conservatory of Music, Suzhou Unierversity, Anhui, 234000, China

Received 1October 2014, www.cmnt.lv

Abstract

The effectiveness of Chinese opera decision directly affects the quality of traditional opera. In the era of rapid development of internet technology today, the transmission mode of opera has changed. Its spreading range is wider, and its propagation speed is faster. For not considering the ethical factors, traditional opera decision is easy to have decision deviation. With reference to the theory of Cooper ethical decision model, the basic framework of Chinese opera ethical decision is constructed, including its definition, criterion, flow block diagram, etc.

Keywords: Chinese opera decision, ethical decision, Cooper's ethical decision model

1 Introduction

The nature of opera determines the reciprocal interaction between opera activity and social development. From the transmission mode of word of mouth to printing transmission and to the electronic transmission until now, technique has brought tremendous impact and changes to the video record and lineage of opera. Everyone who engages in the performance, direction and even the practice and research of different operas will be more or less involved in it. It is a series of decision process [1-3] in nature about how to conduct identification and selection of all kinds of opera script, performing form, etc. To have opera activity and social development interact harmoniously, the quality of opera decision must be improved. However, the tendency of decision utilitarianism has appeared in many aspects, such as the change of transmission mode, appearance of decision deviation, opera compilation, opera performance, etc. In order to pursue economical benefits, the decision deviation appears. To some extent, it violates moral theory and affects the quality of opera. Therefore, traditional decision should transform into ethical decision. In decision-making process, it should comply with moral ethic and set up correct view of life value.

2 Overview of Cooper's ethical decision model

Terry L Cooper, American administrative ethicist, proposed the construction of responsible management model based on administrative ethical problem and divided the ethical thinking into four levels [4,5]: emotional expression level: to show one's own emotion tendency for ethical problems. This expression level is the initial form of one's value judgment which has no influence on the event itself; moral rule level: the first level of seriously raising question and seriously answering it. To think with the moral criteria of daily life; ethical analysis level: when the available moral rule is still unable to make the decision, then it is necessary to conduct basic rethink of our moral rule. Ethic code has clearly connected value and way of act together, which makes it more specific and explicit; after the ethic level: the last level of ethical thinking of administrative staff. Most of the administrators cannot reach this abstract philosophical thinking level. Only under the coerce of specific stubborn and grumpy opponent, or been tightly grasped by a kind of deeply disappointed feeling of disillusionment of dream, or facing comprehensive personal crisis, will we use this level.

From the perspective of practice, Cooper further put forward the process of ethical decision model: to cognize ethical problems, that is to determine ethical problems by describing facts; to demarcate alternative way, that is to assume as much ways as possible to solve ethical problems and to evaluate these ways; to imagine the consequences and to compare the "the profit and lose value" of each consequence; after comparison, to determine the combination of some or several ways and to make the decision.

At last Cooper ethical decision model was proposed: to combine the elements of responsible behavior and the elements of personal ethical autonomy, and to generate resultant force through interaction in practice.

3 Researches of traditional opera decision and opera ethic

Opera is a unique comprehensive national traditional art in China. It not only includes literature, music, dance, painting, etc, but also involves various performing arts, such as

* ***Corresponding author's*** e-mail: yuanqinqy@163.com

martial art, acrobatics, etc. It has more than 300 types of opera and numerous s [6]. Traditional opera decision generally includes two levels: the setting of the positioning, collection, ways of evaluation, etc is determined by the opera committee; the selection and adoption of opera script. At present, only a few of scholars conduct researches on opera ethical decision from many aspects, such as the ethical principle, ethical value, ethical basic content, ethical puzzle of opera [1].

Opera ethical principle refers to the basic principles that should be followed when opera decision makers adopting script. That is to combine social moral requests with practice, which mainly include subject free, principle of autonomy, principle of justice, equality and honesty, etc; opera decision ethical value refers to the actual contribution of decision maker to the society. The features and particularities of opera activities determine that serving the people is the basic principle of opera decision ethical value. The true, the good and the beautiful is the ideal state of opera decision ethical value. The basic content of opera ethics include devoting to professional dedication, sticking to principle, insisting on socialist, respecting opera maker and performer, etc. The ethical confusion of opera is caused by conflicts, such as opera goal, role, responsibility or benefits, etc. In the current opera, popularize and commercialize things are increasingly increased, while innovative and unique things are relatively insufficient.

The ethical research of opera in this paper mainly focuses on the analysis from decision maker. There is a close correlation between opera decision and opera ethics, and opera decision must be conducted that conform to the requirement of social ethics and morals. Therefore, the promotion of opera ethical decision is benefit to the perfection of opera decision theory.

4 The impact of the opera transmission method evolution to opera ethical decision

Chinese opera has a long history, and it is a shining card of Chinese nation. The transmission mode of opera has greatly changed, which is changed from word of mouth to printing and to the current electronic transmission mode. Thus, transmission, communication and sharing of opera become more convenient and efficient, which provides platform for improving the attention of whole nation to opera, stimulating the cultural creativity of opera worker and lifting spirituality of whole nation [6]. However, under interconnection technique, no limitation for delivering content has produced impact to opera ethics. According to personal likes and dislikes or to cater to audience tastes, opera creator and disseminator would spread immoral opera content. This situation would happen sometimes, which has caused adverse effect on society. The ordinary opera decision is unable to meet the requirement of new transmission situation. Therefore combined with the proposal application of Cooper's ethical opera decision, opera creator and transmitter conduct selection on opera in the core idea of opera. They also supervise and control the transmission of opera from internal and external two aspects. The external control refers to control the creator and transmitter from the outside through ethical legislation and regulations. The internal control refers to intensify the inner values and ethics of creator and transmitter and to perfect opera ethical decision through ethical training.

5 Enlightenment of the model theory of modern ethical decision and Cooper's ethical decision to opera decision

A. MODERN ETHICAL DECISION THEORY.

Modern ethical decision derives from the west and is applied to the management of business decision. The ethical decision model is established because of the business scandals that frequently outbreak in western society in the 1970s. It had attracted people's attention to the ethical problems of enterprises. Ethical conduct refers to the behavior that does harmful or good to others and society under the certain moral consciousness. Ethical conduct is produced after ethical decision process, of which refers to the appearance of corresponding ethical behavior is after the appearance of ethical or not decision process [7].

There exist certain commonalities between ethical decision and ordinary decision, which generally has the following characteristics: both were acted as cognitive activity, thus any decision is the behavior of process rather than isolation; for decision makers live in social relations and they have emotionality. Therefore, both decisions will be influenced by emotional and social factors; any decision should be measured the needs and preferences by decision maker.

There also exist certain distinctions between ethical decision and ordinary decision: ethical decision maker object will conduct ethical defense for decision plan and conduct ethical instruction for their behavior. Ethical decision includes the inner psychological activities of the ethical judgment of decision maker (behavior of right and wrong) and ethical evaluation (good or bad and good and evil of behavior), no matter whether the decision maker was asked to do that by the exist pressure in the outside world. In addition, ethical decision involves moral emotion. Ordinary decision maker will not facing moral choice and they do not have moral emotion, such as compunction, shame, etc. However, moral emotion is the by-product of ethical decision and the influencing factor throughout decision process.

Modern ethical decision divided it into four phases: ethical cognition: to realize the ethical issues; ethical judgment: to conduct moral evaluation for alternative behavior plan; ethical intention: what kind of behavior that the individual subjectively choose; ethical behavior, that is the implementation of intention. It shows the external moral or immoral behavior. The four phases describes the inner

process before the explicit of behavior. Though the four parts in practical decision are not appeared successively, on the whole process, the ethical decision process has shown these phases and there are influences among each other.

B. ENLIGHTENMENT ON OPERA ETHICAL DECISION.

Although ethical decision derives from business management and Cooper's ethical decision derives from administrative management decision, Chinese opera has the three conditions of ethical decisions: opera decision object and the opera of been decided involve ethical problems; the opera decision maker and the maker himself have the ethical subject of free will, they can realize the existence of ethical problems and make judgment and action; opera decision maker can judge the "ethical " and "not ethical" of decision result. Therefore, according to the above ethical decision, ethical decision theory can be applied to opera decision so as to construct opera ethical decision.

6 Framework of Chinese opera ethical decision

A. DEFINITION OF CHINESE OPERA ETHICAL DECISION.

Opera ethical decision refers to the process of evaluation, comparison and screening of opera script, opera performance way, opera transmission mode, etc by decision maker and decision team who have certain ethics, which is based on certain ethical criteria.

B. CRITERIA OF CHINESE OPERA ETHICAL DECISION.

The conduct of opera ethical decision needs certain ethical criteria, including the subject ethical criteria and the object ethical criteria of opera ethical decision.

The decision subject ethical criteria refer to have the moral and ethical requirement of ordinary decision maker and main principal with specific opera decision-making power. The subject ethical criteria of opera ethical decision are reduced to the following several aspects: autonomy. In respect for their own autonomies, decision maker should also.

C. PROCESS OF CHINESE OPERA ETHICAL DECISION

1) Judge the opera legality: examine whether the opera content is conforming to the requirements of national laws and regulations, veto those that do not conform that, and conduct authenticity judgment of those that conform that.
2) Judge the opera authenticity: examine whether there exist phenomena in opera creator, such as copying, plagiarism, etc. Conduct moral judgment if conform, otherwise to veto.
3) Judge the opera morality: conduct moral judgment of operas that meet the above requirements and examine whether it has adverse influence on society. If there has immoral content, then it should be vetoed. If it conforms to the requirement, then it will be as the alternative script.
4) Conduct comprehensive evaluation on the alternative script from aspects of society, technicality and standard ability, and select the best opera script for performance.

This ethical decision procedure is formulated mainly based on the object ethical criteria of opera decision. However, subject ethical decision often interacts with the object ethical decision. Therefore, much attention should be paid to opera ethic, and opera decision should convert from ordinary opera decision to opera ethical decision. Flow diagram of Chinese opera ethical decision is shown in Figure 1:

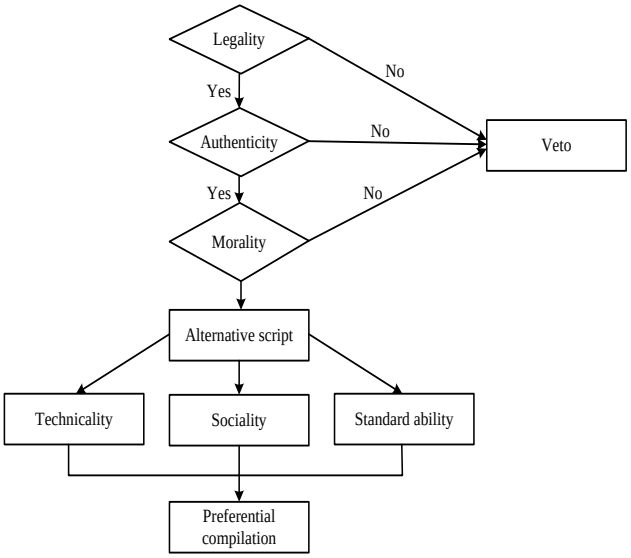


FIGURE 1 Flow diagram of Chinese opera ethical decision

7 Conclusions

Opera ethic is an important problem that opera decision maker need to face. Cooper's ethical decision has provided beneficial enlightenment for Chinese opera both in theory and practice. Much attention should be paid to the core "opera decision responsibility". Then the theoretical learning and the related skill training of opera ethical decision should be strengthened. Decision maker should be cultured to reflect the value problems and principle problems. Decision mechanism should be controlled and integrated from the inside and outside aspects, that is, the inter-coordination and development together of ethic and law. At last, personal moral quality is implanted into institutional framework and practical decision, which provides theoretical foundation and decision guidance for Chinese opera decision [4].

Acknowledgement

Research project of Suzhou Institute "diversified teaching approaches to promote quality-oriented education and cultivation of music talents" (project number szxyjyxm 201261).

References

[1] Xianghong L 2011 Research of Compilation of Ethic Decision *Journal of Guangxi University for Nationalities* 33(4) 183-6
[2] Boliang X 2010 Chinese Xiqu Opera under the International Context *Hundred Schools in Art* (1) 77-84
[3] Rui Y 2014 Network Communication of Traditional Chinese Opera *Shanxi: Shanxi Normal University*
[4] Zhao L 2010 Primary Investigation of Cooper's Administrative Ethical Theory *Morality and Civilization* (1) 139-43
[5] Qi L, Zhisheng X 2013 Research of Cooper's Administrative Responsible Ethical Theory and Its Contemporary Values *Seeker* (7) 91-4
[6] Xiaoxia Y 2013 Research on the Evolution of Traditional Opera Transmission Pattern *Jiangxi: Nanchang University*
[7] Hongmei W, Hong L 2006 Research Evaluations of Western Ethical Decision *Foreign Economics & Management* 28(12) 48-55

Author

Yuan Qin, born in 1981, Anhui Province, China.

Current position, grades: lecturer.
University studies: Bachelor degree in musical performance, Anhui Normal University in 2003; postgraduate degree in musicology, Anhui Normal University in 2010.
Scientific interest: art teaching and art studies.

A construction of an autonomous vocabulary self-study system aided by the web pages

Dongling Sun*

Changchun Institute of Technology, Changchun, China

Received 1 March 2014, www.cmnt.lv

Abstract

The thesis conducted a survey to confirm the hypothesis that vocabulary is the stepping stone in English Study to have an innovative application of modern educational technology to enrich the ways, methods and devices of enlarging the students' vocabulary. From the survey, it is easy to see that the hypothesis is true and the most students of the researcher felt vocabulary affected their motivation and interest in English study. So it is necessary to enlarge the vocabulary of the students in order to promote their application ability of foreign language. Vocabulary study is a process of self-study, and based on the theory of English study, the researcher has designed an autonomous vocabulary self-study system aided by the web to give flexible help for the students. The research is at its designing period and needs further assessments with a case study.

Keywords: construction, vocabulary, autonomous, system, web

1 A survey

English teaching and studying is comprehensive and it ranges from listening, speaking, reading, writing to translating. How to deal with the relations among the preceding five aspects in language skill improvement is what an English teacher should consider [1]. A survey was carried aiming to the question--"What is the most overwhelming element that affects your motivation and interest in English study, listening, speaking, reading, grammar or vocabulary?" Interviews of 123 students were made by the research to decide the teaching emphasis in the following term. The result shows 64 persons (48.12%) thought vocabulary was their big headache; 21 persons (17.07%) listening; 18 persons (14.63%) speaking; 13 persons (10.57%) reading; 7 persons (5.69%) grammar. From the investigation it can be easily seen that most of the students felt vocabulary was the impeding element in their English study. According to the interview result, the emphasis would be put on vocabulary study in the following teaching. However, memorization in most cases needs autonomous completion, and the teacher's task is to provide students environments and guidance. Technology is the productive force of education. Modern technology means the application of modern educational theory and the results of modern science and technology, especially the information and Web technology on the basis of the computer. Modern educational technology is the re-modification, reformation and derivation [2]. In this research the vocabulary study automation is designed based on the web pages, the integration of the teacher's personal web pages and online dictionary web pages.

2 The cutting point of the vocabulary source

According to *the General Requirements of College English* the College students should at least have a vocabulary of 4,795 words and 700 phrases, among the 4,795 words, 2,000 words are active words which are required to be applied skill fully in speaking and writing. *The Requirements* also gives the standard which the students should achieve in listening, speaking, reading, writing and translating. It means the vocabulary *the Requirements* regulate is reasonable for certain English level. For example, the regulations for writing in *the requirements* like the following:

1. To be able to complete the general writing task and give a description of personal experience, feelings and happenings etc.
2. To be able to give common practical writings.
3. To be able to write a complete coherent composition of not less than 120 exact words within half an hour based on the outline and general topic, which has a clear theme.
4. To be able to master the basic writing skills [3].
5. From teaching experience of many years, many students cannot even remember the words in the text, let alone the flexible application. So the research designs vocabulary tasks from the actual conditions of the students. In the textbook, *Experiencing English*, there are 16 passages containing 8,320 words. In the 8,320 words, about 600 words are the new words [4]. From *youdao* software the 600 words can be enriched with large amount of information for the student's reference. The teacher will design practice which will be sent to

* ***Corresponding author's*** e-mail: 1229058230@qq.com

the teacher's personal Web Pages. Generally there are two kinds of practice, multiple choices and gap filling. In the practice the stem is from the text, and the choices are from the word lists of the text book. The students are advised to do multiple choices before the text study and do gap filling after the text study. When they do practice on the platform of the teacher's personal Webpage, they are suggested that *youdao* software be used.

3 Theoretical background

Human recognition is closely connected with context. In broad sense context indicates the social atmosphere of humans and the physical environment of survival, and in narrow sense it indicates the structure of the passage and the relationship between words. The contextual elements are connected with the target study materials in a meaningful way. When the target information is coded, the contextual information is also coded in the same way, so the contextual information becomes the explicit clue. According to the theory of context dependent memory, the environment of memory occurrence has a big effect on the effectiveness of memory [5]. A word can be remembered alone in one second and forgot in one minute. If the context of the word occurrence is provided, the time spent to memorize it is a little longer, but it can be changed into long-term memory, so the effectiveness of word memorization would improve. Because the students know the context of the word occurrence which serves as the explicit clue in memory, they can always associate the context with the target word. Thus what the students remember is not the single word, but the context which is coded in the same way in reminiscence. In this research, the design of the vocabulary tasks is based on the theory of context dependent memory and the contexts of the new word occurrence are offered via the teacher's personal Web Pages and *youdao* software.

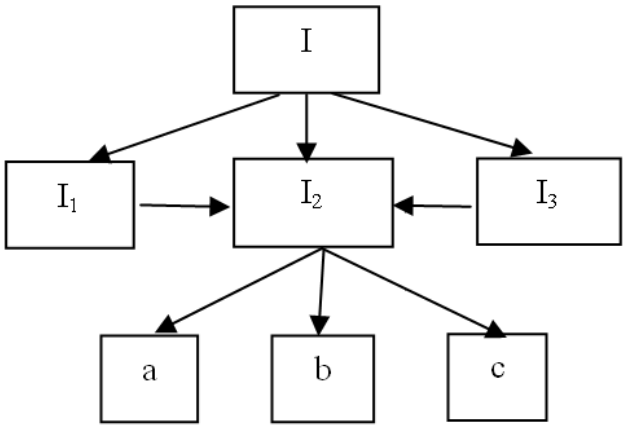
4 The construction of an autonomous vocabulary self-study system aided by the Web Pages

Vocabulary study automation in this research means students study vocabulary aided by automatic Web Pages without the supervision of the teacher. The rapid development of the Web provides the powerful tool for language learning. In vocabulary study Web serves as both information carrier and information source which is a platform that information can be put on and gained from. The research catches the characteristic of Web to design a system to assist the students to remember vocabulary in and out of the text book. In this system two kinds of Web Pages are involved, online dictionary Web Pages and the teacher's personal Web Pages. The employment of the online dictionary can be used directly, but as for the teacher's personal Web Pages, the technical problem is still needed solving.

4.1 THE PERSONAL WEBPAGES ARE CONNECTED

Not all the people can have a construction of personal website, but they can have their own Personal Web pages which depend on the whole Web system. With the popularization of the Web, the personal Web Pages can reach everyone. Personal Web Pages are popular, convenient, and flexible which provide good conditions for their application in teaching. It is easy to have an account of your own Personal Web Pages; however, there are technical problems, too. The personal Web Pages cannot be altered freely to suit your own needs. For example, the layout of it cannot be freely adjusted. This is why some teachers just use them to send e-mails or talk with the students via QQ. Therefore the glaring problem is that the personal Web Pages are scattered, and apparently they are not in a system.

Teaching and learning are interactive which are in a system [6]. How to make the scattered Web Pages into a system is a point in case to solve the technical problems. The personal Web Pages are in fact in a system, too, which escape the people's eyes because their apparent scattered traits. The following figure shows the connection among the Web Pages.



Connections among the WebPages

FIGURE 1
I=QQ account passageway; I_1=contact list page;
I_2 =Q zone homepage; I_3 =email-box page;
a =the owner's words; b=the message board; c=QQdiary

From the above figure, we can see the personal Web Pages are interrelated that construct a recessive system that needs to be discovered the technical connections. In the recessive system, we cannot change the general layout, but we can make a creative use and add items.

4.2 YOUDAO ONLINE SOFTWARE AUTOMATICALLY PROVIDES NEEDED INFORMATION

Youdao translation is China's largest free online translation which carries 23 million examples and 1.8 million encyclopedia articles with one-stop knowledge inquiry platform [7]. *Youdao* translation provides new video examples, including international prestigious public class and America and European classic film video examples make users to experience pure English. The recurring audio examples are from the English original radio recording are authoritative.

It is the first time that *Youdao software* supports IE9 browser bingmuquci function. Cross word definition function in the software is strengthened comprehensively, and, you can easily get translated results and more information with the mouse choosing paragraphs, long sentences and phrases. It is more helpful that there are bilingual examples with explanations word by word and controls sentence by sentence. *youdao* translation provides an ocean of information.

4.2 THE DESIGN OF THE VOCABULARY STUDY AUTOMATION SYSTEM BASED ON THE INTEGRATION OF THE TEACHER'S PERSONAL WEBPAGES AND YOUDAO ONLINE TRANSLATION SOFTWARE [8]

The scattered Web Pages can be built into a recessive system which is built the further connection in some way with the online dictionary Web Pages into a vocabulary study automation system.

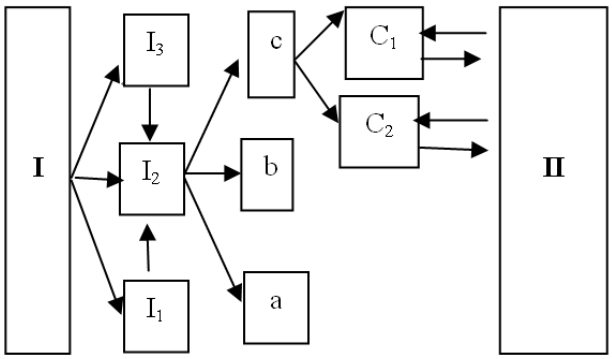


FIGURE 2
I=QQ account passageway; I_1=contact list page;
I_2 =Q zone homepage; I_3 =email-box page;
a=the owner's words; b=the message board; c=QQdiary;
C_1=multiple choice; C_2=gap filling;
II=*youdao* online dictionary interface

In this system the personal Web Pages and *youdao* online dictionary Web Pages can all be clicked. According to the concrete demand the needed Webpage is shown and restremain inactive. The teacher's personal Web Pages are the main part of the system and *youdao* online dictionary Web Pages are ready for use as assistance. Web Pages Qzone can serve as a center, and the column of "the owner's words" can serve as a notification board which provides vocabulary practice information; the message board for communication or left message. QQdiary for the vocabulary practice carrier. On the contact list page, communication and submission can be done. Email-box page submission and off-line communication.

In this system C1 (multiple choices) and C2 (gap filling) are concrete practice contents sent by the teacher. Multiple choices should be done by the students before the text study. The disturbance terms are from the new word list in other texts, so the students can consult *youdao* online dictionary to find out the equivalent meaning of the choices and choose the best one responding to the sentence context. Gap fills for enhancement of the new words after the text study. The students are supposed to master the words in the text, but it is possible that the students still need to consult the online dictionary for enhancement.

5 Assessment of the design

The research aims to exploit the Web technology to innovatively enrich the ways, methods and devices of enlarging the students' vocabulary. The scientific, practicability, efficiency and popularization of the design of the vocabulary study automation system based on the integration of the teacher's personal Web Pages and *youdao* online translation software should be checked in the practical study. In order to prove the value of the study, questionnaires, observations, visits, diaries, comparisons, and the analyses and summaries will be done through the device of materiallization and quantization. The scientificity of the design of vocabulary tasks will be analyzed by using social science statistics software.

6 Conclusion

The present research is going to be carried out in the following term. The subjects are the freshmen the researcher will teach. During the process of implementation it is possible that the necessary improvements of the design will be made, though the proof materials haven't been made. However the expected effects are in following aspects: The English vocabulary of subjects is enlarged much; Vocabulary expansion indirectly promotes students' language application ability; the original application cases of the online vocabulary study automation system resources are formed; the results of the study can be shared and discussed with the fraternity. With the implementation of the study, the research will provide further reports.

References

[1] Liu R 1999 On English Language Teaching Beijing Foreign Language Teaching and Research Press 227 (in Chinese)
[2] Wang Y, Guo H, Li Z 2006 The application of modern educational technology for the improvement of teaching quality North-east Agr Univ (Social Science Edition) l(1) (in Chinese)
[3] Department of Higher Education, The Basic Requirements of College English Course Beijing: Higher Education Press 11 (in Chinese)
[4] Wu Z 2012 Experiencing English Integrated Coursebook III Beijing Higher Education Press (in Chinese)
[5] Adams M J 1985A Schema-theoretic View of Reading
[6] Douglas Brown H 2001 Teaching by Principles: An Interactive Approach to Language Pedagogy Beijing Foreign Language Teaching and Research Press 201
[7] http://www.youdao.html
[8] Xia J 2003 The Theory and Practice of Modern Foreign Language Course Design Shanghai foreign Language Education Publishing House 1-13 (in Chinese)

Author

Dongling Sun, China.

Current position, grades: associate professor in Changchun Institute of Technology, China.
University study: Changchun Normal University, Major in English, 1988, MA degree in Applied Linguistics, Changchun University of Science and Technology, Changchun, Jilin, 2006.
Research activities: English teaching, language application and Modern educational technology.
Professional Activities and Memberships: translation network course.

A design for simulation model and algorithm of rail transport of molten iron in steel enterprise

Bin Ge*, Kai Wang, Yue Han

School of Computer Science and Engineering, Anhui University of Science & Technology, Huainan, Anhui, 232001, China

Received 28 August 2014, www.cmnt.lv

Abstract

An emulation method can be chosen according to the characteristics in rail transport of molten iron in steel companies and a simulation model can be established based on the basic conditions for model, distributed tank mode and path selection, etc. By studying the automatic collision avoidance algorithm, a method of shortest path optimization for rail transport scheduling of molten iron was proposed based on ant colony algorithm. Simulation results show the validity and the feasibility of the algorithm. Programs and strategies of implementing visual simulation platform are proposed laying the theoretical foundation of further research and application for rail transport scheduling model of molten iron and intelligent optimization algorithm.

Keywords: dynamic routing, selection model, ant colony optimization algorithm, collision avoidance algorithm, mode of allocated tank, rail transport of molten iron.

1 Introduction

Rail transport of molten iron is an important step to guarantee steel production. Rail transport of molten iron system transports TPC (Torpedo Car, TPC) which has been ironed under the blast furnace from the blast furnace (BF) to the station before arriving specified point to pre-processed before-steak, desulfurization and after-steak, then it will be transported to the steel factory to pour molten iron. After that, the empty TPC will be hauled back to blast furnace, which completes one course of transporting molten iron [1,2]. In order to better carry out research on the railway transport system of molten iron in steel enterprises, it requires to mimicking the operation of existing systems and simulating progress of future system, and providing a theoretical basis for the simulation and optimazation of transport by improving and perfecting the design.

The research on models and algorithms about rail transport of molten iron has been carried out currently. Some results have also been achieved. Such as: in literature [3], relying on plant railway network. A road collision detection mechanism and reservation model are established and a neighbor distribution. Process scheduling and packet-optimized scheduling for rail transport of molten iron are proposed the with rail transport of molten iron in steel companies as the background and the locomotive and torpedo car TPC as means of transport. The adoption of dynamic programming to determine the optimal number of vehicles is related to the needs of molten iron and transportation distance and other factors. In literature [4] through systems analysis of the molten iron transport, this problem is solved by computer simulation method. The project and the time during rail transport of molten iron are measured. The distribution and the main parameters of random variables are determined by statistical test. Furthermore, there is precision analysis of the computer system simulation and simulation results. In literature [5] using modeling and system simulation technology, the corresponding object model is established through analysis of system objects. The transport path selection of rail transport of molten iron in steel enterprises and automatic collision avoidance algorithm implemented by virtual detection and competition rules are achieved based on the strategy of hierarchical event scheduling simulation.

In order to overcome some shortcomings of static path selection, this paper introduces a dynamic path selection algorithm to solve the problem of adjusting the path of collision avoidance in train operation. When conducting path selection, we change road weights for dynamic under the new constraints reselect path according to the needs of adjusting collision avoidance.

In literature [6,7] using linear programming methods to build analytic model of transportation systems. In literature [8] introducing a heuristic algorithm to study the transportation scheduling problem. Because of a lot of thinking and reasoning process in transport schedule of molten iron, including task allocation, process coordination, time arrangement, vehicle arrangement, route selection and intelligent decision-making behavior of analog and motorists based on dynamic traffic, these mathematical models cannot meet the requirements of simulation research. In recent year's artificial intelligence and expert systems technology is widely used in scheduling research providing a new adjunct of real-time and intelligence, especially the artificial ant colony algorithm for solving scheduling problems of rail transport of molten iron can provide a better solution.

Ant colony algorithm is a new intelligent optimization algorithms proposed by Italian scholar M. Dorigo, V. Maniezzo and A. Colorni [9-11] inspired by the shortest

* ***Corresponding author's*** e-mail: bge@aust.edu.cn

path from the nest to a food source when ants foraging. Currently people working on ant colony optimization algorithm have penetrated from the original single traveling salesman problem into a number of applications. Now the algorithm having achieved good results has been optimized in function, combinatorial optimization, graph coloring, vehicle scheduling and so on [12-14].

In this paper, simulation models rail transport of molten and algorithms iron are studied. Pathway selection model collision avoidance algorithm and improved ant colony optimization algorithm are proposed which has been proved to be effective and feasible by the experiment.

2 Simulation method

Depending on whether the continuous state variables change over time Simulation is consisted of continuous system simulation, discrete event simulation and continuous-discrete hybrid system simulation [15]. Continuous system simulation model can be parsed into differential equation; Discrete event simulation can be expressed with logical conditions or flowcharts [16,17]; Continuous-discrete simulation hybrid system is expressed with the first two mixed. However, in the simulation design of molten iron transport system, state of the system only changes at discrete moments. So, discrete event simulation is chosen.

Nevertheless, as there are lots of factors in actual production influencing transportation and production of molten metal, some factors may be left out. Main factors affecting the production are as follows: blast furnace fluctuations, half a deal in principle, car transfer system features, and road traffic. For other factors, such as traffic scheduling, simulation pathways need to consider when selecting models. Thus, the simulative design of molten iron transport system is so complex that we need to make a systematic plan.

2.1 THE ESTABLISHMENT OF SIMULATION MODEL

2.1.1 Model input

The basic conditions for model input are volume X (m^3) of ironworks blast furnace, the average designed tapping speed Y (t/min) of blast furnace, the designed maximum fluctuation coefficient C, maximum and instantaneous tapping speed Y max (t/min), average time Z (min) of two overlapping iron tap hole and tapping of each iron times W (t). It ought to establish the corresponding transport organization program and allocate reasonable transportation equipment based on design parameters of blast furnace and actual production situation.

2.1.2 Mode of allocated tank

General configuration mode is the "3-2-1" mode. Namely, "tap hole at one time" allocated with 3 TPC", tap hole at next time" allocated with 2 TPC", tap hole at third time" allocated with 1 TPC.

Definite: suppose in an engineering company, there are 6 sets of locomotive cars, assuming the length of locomotive car is 15 cm and locomotive speed is 3-5 km/h, the capacity of mix stickers car is 400t, the length of mix stickers car is 30 cm, load factor is 0.9 [18,19].

Division of locomotive. There are 5 sets of locomotive invested initially with 1 sets of locomotive available. (Locomotive distribution: 2 locomotives for each station of blast furnaces, 1 locomotive for intrusion station), locomotive operation using transport practices with a large loop (BF→ intrusion blast station→ BF). If there is not enough capacity of locomotives, then the sixth car will be put into production.

Mode of allocated tank. Assuming the sequence of using iron mouth is 1#→2#→3#→4#, 4# as a spare overhaul iron mouth, as shown in Figure 1.

Step1. 1#iron mouth railway line 1 ① torpedo car during iron, machine A in line 1 waiting for hanging heavy, machine B hang two cans in a field waiting with empty.

Step2. 1# iron mouth railway line 1 ① torpedocar finished by iron, then swing to iron TPC ② on line 2, TPC ② by iron; machine A hang on heavy TPC ①, changing lanes to a safe location waiting outside the line 2;machine B send ⑦, ⑥ empty TPC to line 1, empty TPC ⑥ registration uncoupling, machine B tow empty TPC ⑦ waiting outside.

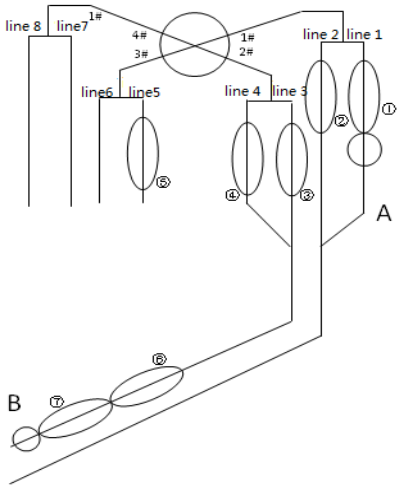


FIGURE 1 The locomotive schedule mode

Step 3. 1#iron mouth railway line ② iron by iron mixed with iron chariots is completed, the nozzle swing after swing stream is mixed with iron chariots of iron a line ⑥ by iron, A machine push ① heavy torpedo car into the iron ② TPC, tow ①, ② TPC to steelmaking intrusion station; machines B and ⑦ empty TPC transfer ran from line 1 to line 2 on the bit, and then turn to line 1 waiting outside after the completion of the ⑥ car to take iron; machine C push two empty torpedo car station to the intrusion from a steel blast furnace operation to its original position of machine B.

At this time, machine A, B, C have completed one bit position conversion, A→C, B→A, C→B.

3 Routing selection model

Regarding the valuable cut-off points at the apex of the rail network as nodes constituting a new network, denoted as G=(v, E), where v={v_1, v_2,...,v_m} is a set of m nodes; E={e_{ij}|1≤i, j≤m} is a set of directed edges(between the junction section).

When searching for the shortest path in the railway network diagram with the help of algorithm Dijkstra, due to the excessive number of network nodes and locomotives and the huge amount of computation, which cause more computation time than simulation step size, the static pathway selection model cannot meet the requirements of simulation system. Therefore, a dynamic pathway selection model is needed to adopt.

Let q denote locomotive number, P denotes a set of locomotives priority; d_{ij} denote the length of directed edge e_{ij}, γ^{kt}_{pij} denote flow from node k to node t of locomotives with priority of p through the directed edge e_{ij}. T_i denotes travelled distance when locomotive i avoid. H_{ij} denote time during which locomotives wait on the directed edge e_{ij}. Gij denotes the driving speed of locomotive on the edge e_{ij}. Dynamic Routing selection model can be defined as follows:

The objective function: opt Z={Z1, Z2, Z3, Z4}.

The shortest total driving distance of locomotive:

$$\min Z_1 = \sum_{k\in V}\sum_{t\in V}\sum_{p\in P}\sum_{n=1} d_{ij} r^{kt}_{pij} . \tag{1}$$

The shortest total mileage of vacant locomotive:

$$\min Z_2 = \sum_{n=1}^{q} T_n . \tag{2}$$

The least total waiting time of locomotive:

$$\min Z_3 = \sum_{n=1}^{q} H_{ij} . \tag{3}$$

The least total driving time of motorcycle:

$$\min Z_4 = \sum_{k\in V}\sum_{t\in V}\sum_{p\in P}\sum_{n=1}^{q} (d_{ij}/G_{ij}) r^{kt}_{pij} + \sum_{k\in V}\sum_{t\in V}\sum_{p\in P}\sum_{n=1}^{q} \mathrm{H}_{ij} r^{kt}_{pij} . \tag{4}$$

4 Research of collision avoidance algorithm

During the study of molten iron transport system, the trains can follow certain traffic laws and avoid colliding with each other through the establishment of collision detection mechanism. According to collision avoidance algorithms and priority of task, high priority trains go first. Meanwhile, every train need to get basic information of others providing more information to support decision-making for the transportation safety.

The length of train T in the simulation model is set n. During driving, the reservation section number of railway is m ($0 \le m \le n$), the current section of railway is S_i ($0 \le i \le n$), occupancy and reservation path is represented by a collection of array $A(S_i, S_{i+1}, S_{i+2}, \ldots, S_{i+m})$. It can be seen that there are many reservation section when $m=n$ and railways are smooth. However, the utilization of the railway and efficiency of running trains are very low. When $m=0$, there is no reservation section, while trains are easily collided. Given the actuality of transport system in steel company and the need of experiment and simulation, we select $m=2$.

Set two trains T_1, T_2, occupation and reservation paths are represented by a collection of array $A_1(S_i, S_{i+1}, S_{i+2})$, $A_2(S_i, S_{i+1}, S_{i+2})$. The task priority numbers are respectively P_1, P_2. T_1 is treated as the research object. S_{i+2} denotes the reservation section in the current road sections S_i of train T_1. Collision avoidance algorithm is as follows:

1) Comparing the size of P_1 and P_2. if $P_1=P_2$ then obeying the rules of FCFS, first appointment S_{i+2} with high default task priority. If $P_1 < P_2$, there will be the same travel in path S_i and T_2 , then T1 need to transfer the right of using S_i to T_2.
2) Appointment success: T_1 traveling on the path S_i.
3) Appointment failed:
   - **a.** If $S_{i+2}=S_j$, T_2 is running. P_1 is lower than the b default P_2. Set T_1 is the waiting time. The value of *time* denotes ratio of the traveling distance of train T_2 from current position to the end of path S_{i+2} and speed of the train T_2. At the moment by adjusting the speed mechanism of the train, such as speeding up T_2 to reduce the waiting time of T_1.
   - **b.** If $S_{i+2}=S_{j+1}$ or $S_{i+2}=S_{j+2}$, sections S_{i+2} has been reserved by T_2. Appointment T_1 failed.

For the research of multiple trains collision avoidance we can use priority algorithm based on the contents of the train running task space to let high-priority trains go first.

5 The optimization of shortest path algorithm

5.1 DEFINITION OF ALGORITHM

Definition:

Assuming that there are m ants, d_{ij} $(i, j = 1, 2, ..., n)$ denotes the distance between the city i, and the city j, $b_i(t)$ denotes the number of ants located in the city i at the moment t, $\tau_{ij}(t)$ denotes remaining pheromone concentration in attachment between city i and city t. It is equal of initial pheromone concentration on all paths, set $\tau_{ij}(0) = C$ (C is constant), the ants k, $(k = 1, 2, ..., m)$) choose its path in the process of movement according to the pheromone concentration on the path. $p^k_{ij}(t)$ denotes the probability of ants k transferring from city i to city j at the moment

$$p^k_{ij}(t) = \begin{cases} \dfrac{\left[\tau_{ij}(t)\right]^{\alpha}\left[\eta_{ij}\right]^{\beta}}{\sum_{k\notin tabuLk}\left[\tau_{ik}(t)\right]^{\alpha}\left[\eta_{ij}\right]^{\beta}}, & j \notin tabu_{k,} \\ 0, & otherwise \end{cases} , \tag{5}$$

where η_{ij} denotes heuristic information transferring from city i to city j. Its value is generally $1/d_{ij}$. α denotes the importance of the residual pheromone on path ij; β denotes the importance of the heuristic information; List $tabu_k$ record cities which ants pass through, it dynamically change as the moving process of the ants.

When all the ants traverse city *n*, we need to calculate path L_k the ant k, then to solve the shortest path $L_{\min} = \min\{L_k \mid k = 1,2,...,m\}$, and update pheromone concentration on every path.

$$\tau_{ij}(t+1) = \max(\rho\tau_0(t) + \sum_{i=1}^{n}\Delta\tau_{ij}^k + \sigma\Delta\tau_{ij}^*, \tau_{\min}(t)), \quad (6)$$

$$\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{Length}, & (i,j)\in T \\ 0, & (i,j)\notin T \end{cases}, \quad (7)$$

$$\Delta\tau_{ij}^* = \begin{cases} \frac{Q}{L_{opt}}, & (i,j)\in T_{opt} \\ 0, & (i,j)\notin T_{opt} \end{cases}, \quad (8)$$

where ρ denotes the pheromone sustained. Parameters, $\Delta\tau_{ij}^k$ denotes the increased concentration of pheromone of ant k on the (i,j) edge from time t to $t+1$, Q is a constant that denotes the total amount of hormone released of the ants by a search. T denotes the exploring path of the ants, *Length* denotes the exploring path length of the ants, T_{opt} denotes the exploring optimal path of the ants, L_{opt} denotes the exploring optimal path length of the ants.

5.2 THE HYBRID ANT COLONY OPTIMIZATION (HACO) ALGORITHM FOR SOLVING OPTIMAL PROCESS ROUTE OF VEHICLE

A colony consists of N ants and all the ants distributed in continuous space S. The point x_i denotes the location of ant i in space S and the corresponding to a random sequence $S_i = \{C_{i(1)}, C_{i(2)}, ..., C_{i(n)}\}$, its adaptive values is $f(S_i) = \sum_{k=1}^{n-1} d(C_{i(k)}, C_{i(k+1)}) + d(C_{i(n)}, C_{i(1)})$.

The new sequence can be acquired by sequence generation mechanism. Algorithm is as follows:

BEGIN

1) Configuring the population size as N, the parameters as ψ_d, a, b, r_i, the number of initializing iterations as t, the position of each ant as $x_i(0)$, its corresponding sequence as S_i and fitness as $f_{i,0}(S_i)$;

2) To calculate the optimal positions $p_i(0)$ of the ants and their optimal adaptive value $f_{i,0}^{best}(S_i)$ at the next step $t=1$, $f_{i,0}^{best}(S_i)$ is the optimal adaptive value between adaptive value $f_{i,0}^{best}(S_i)$ of ant i and adaptive value $f_{k,0}(S_k)$ of its neighbor;

3) While (the number of iterations <= maximum number of iterations)

[a] Calculating the position $x_i(t)$ of each ant, reversing the sequence S_i to a sequence S_i', calculate its adaptation value to get $x_i(t) \to f_{i,t}(S_i')$;

[b] Calculating the optimal positions $p_i(t)$ of the ants and its optimal adaptive value $f_{i,(t-1)}^{best}(S_i')$, $f_{i,(t)}^{best}(S_i')$ is the adaptive value of ant i. $f_{i,0}^{best}(S_i)$ is the optimal adaptive value of ant i between the optimal adaptive value $f_{i,(t-1)}^{best}(S_i')$ in last step and adaptive value $f_{k,(t)}(S_k')$ of its neighbour (ant k and ant i are neighbors each other), (if $t=1$, the optimal adaptive value is $f_{i,0}^{best}(S_i)$ in last step);

[c] calculating $\Delta f_{i.t} = f_{i,t}(S_i') - f_{i,(t-1)}(S_i')$;

[d] If $\Delta f_{i.t} \le 0$. Then

Setting $S_i = S_i'$;

End if

[e] While (cross operation condition is met)

[i] Settings $p_i = S_i, f_i(p_i) = f_i(S_i)$;

[ii] Do crossover operation to get $p_i', f_i(p')$ and calculating $\Delta f_i^p = f_i(p_i') - f_i(p_i)$;

[iii] If $\Delta f_i^p \le 0$. Then

Setting $S_i = p_i'$;

End the if

End the While

End the While

4) To choose the optimal solution

END

5.3 THE SIMULATION OF ALGORITHM

In this section, Matlab7.0 was used as the simulation platform to verify the feasibility and performance of the improved algorithm of HACO. In this paper, there are only two instances of the simulation results, which are Eil50 and KroA100. Figure 2 and Figure 3 shows respectively the optimal solution of vehicle route.

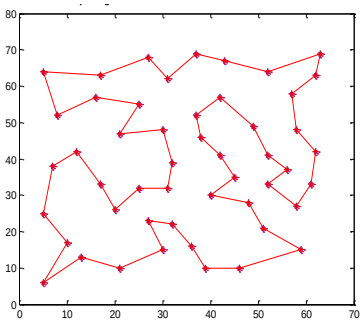


FIGURE2 Optimal solution 428.73 of Eil50

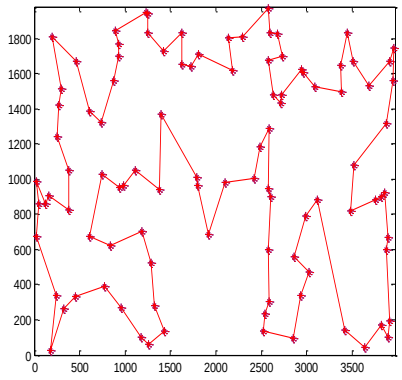


FIGURE 3 Optimal solution 21289.56 of KroA100

The convergence of HACO-TSP algorithm can be seen clearly and it is feasible that solved TSP problem together among ants exchange information each other from Figure 4 and Figure 5. Figure 4 and 5 show the convergence and the good performance of the improved HACO algorithm.

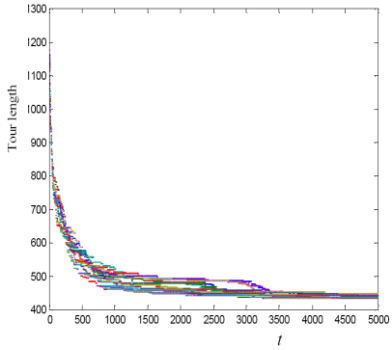


FIGURE 4 Eil50 the dynamic evolution over time of all ants over time

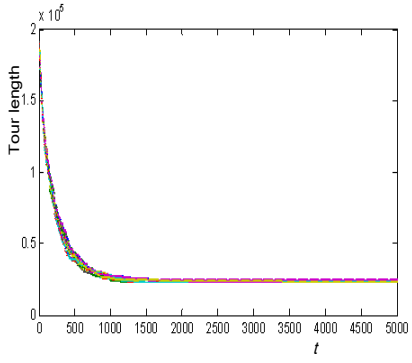


FIGURE 5 KroA100 the dynamic evolution of all ants over time

6 Analysis and implementation of designing

There are many factors to take into account in the process of designing. For instance, the locomotive operating rate, cycle and turnover rate of TPC, utilization of the railway line, the ability of level crossing, etc are analyzed.

It is viable to adopt programming languages such as C, Visual C++ to create model for rail transport of molten iron according to above pathway selection model based on design simulation system. Three-dimensional virtual simulation software Flexsim is used to achieve dynamic pathway selection and conflict-free driving.

Via regarding a simulation system of distributed intelligent hot metal scheduling as experiment platform and recording the raw data adopted to verify scheduling process of molten iron transportation. Intelligent control is applied before and after the key technology of hot metal scheduling, locomotive and TPC running are all got optimization to some degree.

7 Conclusions

Rail transport of molten iron is an important part in steel production. The scheduling simulation model and algorithm research also has important theoretical significance and practical value. This paper presents a simulation model of molten iron scheduling and strategy. Firstly, we propose input conditions of model and then the feasibility of scheme is proposed as complement through simulation studies. Finally, concluding the simulation results. This paper provides an effective and feasible method of simulation programs through research to optimize the in-depth study of intelligent scheduling of molten iron and practical application of the algorithm.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (61070220, 61170060), Anhui Provincial Natural Science Foundation of China (1408085ME110), Anhui Provincial Major Project of Colleges and universities Natural Science (KJ2013ZD09) and Anhui Provincial Key Project of Colleagues and universities Natural Science (KJ2012A096).

References

[1] Wang Y 2008 Virtual Simulation Modeling for Production Process of Continuous-discrete Hybrid Manufacturing System *Journal of system simulation* **20**(9)2445-53 (*in Chinese*)
[2] Yang X, Cui B 2013 Optimization and simulation of transportation scheduling for molten iron in steel enterprise *Journal of Computer Application* **33**(10) 2977-80 (*in Chinese*)
[3] Zhao H, Li M 2009 Intelligent scheduling transport of molten iron based on dynamic programming *Journal of Qingdao University of Science and Technology (Natural Science Edition)* **30**(1) 70-3 (*in Chinese*)
[4] Qi M, Wang L, Yuan C 2010 eM-Plant-Based Simulation Model for Hot Metal Transportation in Steel Plant *Industrial Engineering Journal* **13**(2) 76-81 (*in Chinese*)
[5] Cui J, Hu K, Xu X 2007 A research of modeling and simulation in transport system of melton iron *Information and Control* **36**(4) 103-8 (*in Chinese*)
[6] Chen Z, Li Z, Lu M 2012 Research on dynamic balance and real-time scheduling of hmlogistice in large steel enterprise *Computer Applications and Software* **29**(8) 115-7 (*in Chinese*)
[7] Huang B, He D, Tian N 2010 Operational control of torpedo ladles *Journal of University of Science and Technology Beijing* **32**(7) 933-7

[8] Zhang W, Li X, Deng H 2011 Modeling and analysis of roll motion for high temperature ladle carrier vehicles based on relative motion theory *Journal of Central South University (Natural Science Edition)* **42**(11)3347-51
[9] Dorigo M, Maniezzo V, Colorni A 1991 Positive feedback as a search strategy. *Technical Report 91-016 Dipartimento di Elettronica, Politecnico di Milano, IT*
[10] Colorni A, Dorigo M, Maniezzo V 1992 Distributed optimization by ant colonies *In Proceedings of the First European Conference on Artificial Life*
[11] Dorigo M 1992 Optimiztion Learning and Natural Algorithm (in Italian) *PhD thesis Dipartimento di Elettronica Politecnico di Milano* IT
[12] Xing L, Rohlfshagen P, Chen Y 2011 *IEEE Transactions on Systems, Man, and Cybernetics Part B Cybernetics* **41**(4) 1110-23
[13] Ho S L, Yang S, Bai Y 2013 *IEEE Transactions on Magnetics* **49**(5) 2077-80
[14] Zhang Z, Zhang N, Feng Z 2014 Multi-satellite control resource scheduling based on ant colony optimization *Expert Systems with Applications* **41**(6) 2816-26
[15] Huang H, Chai T 2012 Design and implementation for simulation of rail transport of molten iron *Journal of System Simulation* **24**(6) 1192-9
[16] Atighehchian A, Bijari M, Tarkesh H 2009 A Novel Hybrid Algorithm for Scheduling Steel-Making Continuous Casting Production *Computers & Operations Research* **36**(8) 2450-61
[17] Xu G, Zheng L 2009 Application science of digital simulation *Computer Simulation* **26**(12) 1-4
[18] Tang L, Wang G, Liu J 2007 A branch-and-price algorithm to solve the molten iron allocation problem in iron and steel industry *Computers & Operations Research* **34**(10) 3001-15
[19] Tang L, Wang X 2009 Simultaneously scheduling multiple turns for steel color-coating production *European Journal of Operational Research* **198**(3) 715-25

Authors

Bin Ge, China.

Current position, grades: associate professor in Anhui University of Science & Technology, Huainan, Anhui, China.
University study: PhD candidate in computer application technology, Hefei University of Technology, Hefei, Anhui, China, from 2010.
Scientific interest: Intelligent control technology and Internet of things technology.

Kai Wang, China.

Current position, grades: postgraduate student in Anhui University of Science & Technology, Huainan, Anhui, China.
University study: MSc candidate in computer application technology, Anhui University of Science and Technology, Huainan, Anhui, China, from 2012.
Scientific interest: internet of things technology.

Yue Han, China.

Current position, grades: postgraduate student Anhui University of Science & Technology, Huainan, Anhui, China,
University study: MSc candidate in computer application technology, Anhui University of Science and Technology, Huainan, Anhui, China, from 2013.
Scientific interest: artificial intelligence and Intelligent control technology.

The design and implement of alarm processing system for large-scale railway maintenance equipment diesel engine combustion control based on multi-agent

Hairui Wang, Ya Li*

Department of Computer Science Kunming University of Science and Technology, Kunming, China

Received 1 March 2014, www.cmnt.lv

Abstract

The Alarm Processing System reference model for large-scale railway maintenance equipment diesel engine combustion control, which is based on multi-agent, is an important system for high speed and automation. Alarm Processing System are vital aspects in M large-scale railway maintenance equipment diesel engine combustion control, in this sense, alarm processing system should support decision-making tools as known as decision support system(DSS), new maintenance approaches and techniques, the enterprise thinking and flexibility. In this paper a Multi-agents based alarm processing System reference model for large-scale railway maintenance equipment diesel engine combustion control is presented which combines the existing models and multi-agents. This model is based on a generic framework using multi-agent systems for the diesel engine combustion control the diesel engine combustion control on-line monitoring system. In this sense, the alarm processing system is viewed like a feedback control process and the actions are related to the decision-making in the scheduling of the preventive maintenance task and the running of preventive and corrective specific maintenance tasks. The result of an evaluation of the Multi-agents based alarm processing system reference model for the diesel engine combustion control are presented. This new model is compared to some important existing models and applied to a real investigation.

Keywords: diesel engine combustion control, alarm processing systems, multi-agent system, automation

1 Diesel engine combustion control technology

In the innovation progress of diesel engine, new technology such as turbocharger technology, the three element catalytic processing, the lean combustion, homogeneous charge compression ignition (HCCI) comes out. In the conventional altitude conditions, the engine has reached a higher level of technology. But in the plateau and the special circumstances, the engine is still weak. In addition, with the requirements of increasingly stringent emission regulations, to explore a new combustion control method has become a necessity. Figure 1 shows the Schematic Structure of the diesel engine combustion.

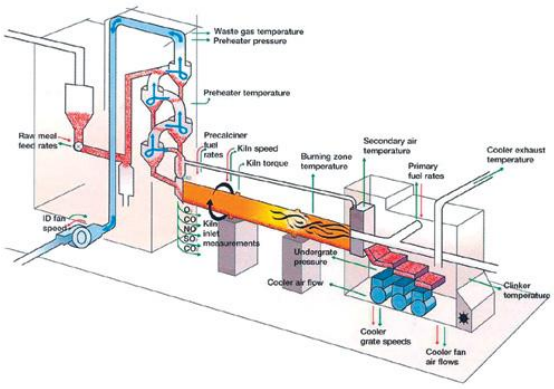


FIGURE 1 Schematic Structure of the diesel engine combustion.

With development of the catalytic converter processing technology, gasoline engine emissions are reduced, however, HC at starting stage is a major obstacle to achieve ultra-low emission [1]. To solve the problem of NO2 diesel engine emissions has been facing challenges; there are contradictions and difficulties in the trade-off between NO2 and particle. Although the existence of complex factors of engine combustion process, but the combustion is using oxygen as oxidant, controlling the share of air intake oxygen combustion control is the most fundamental way.

During combustion a control system must therefore maintain the following conditions:

1) Control of the O2 concentration in the flue gas to keep it at a constant figure.
2) Maintenance of a uniform thermal output.
3) Maintenance of optimum flow conditions in the furnace and the first boiler pass with as little variation as possible so that the desired conditions for attaining the lowest possible degree of emissions and thus preventing corrosion are always maintained [2].

These conditions can only be fulfilled by optimized combustion at a stable operation point. Conventional control systems are incapable of reacting to the inevitable local and intermit and inhomogeneity of the refuse fed to the grate, which are attributable to varying calorific values and

* ***Corresponding author's*** e-mail: sherly2001@sina.com

ignition properties. It is as a result impossible to avoid pronounced variations in the combustion process and such variations are always associated with unfavorable emission figures.

The engine combustion is mainly controlled by the contents of oxygen, nitrogen [3,4]. One is the intake combustion, the other is flame retardant, the two shares become the basic elements of the combustion process control. To implement the combustion and emission control is by the way of reforming the intake charge component. Oxygen Enriched Air (OEA) can significantly reduce the non-full combustion emissions (except NO2), can enhance the power performance and shorten the ignition delay period, which is conducive to low quality fuel [5,12]. Nitrogen Enriched Air (NEA) can inhibit the combustion process, can be beneficial to relieve the No emission, which is considered to be an effective method.

A possible method of automating this process is the registration of combustion by infrared thermography. This offers the immediate advantage that the plant operator is in a position to observe combustion directly from the control room [6,9]. Because of the geometry of the furnace in the plant under consideration the area which must be observed lies mainly in grate zone 3, but also to some extent in zones 2 and 4, which cannot be so completely observed.

This is however more than adequate in order to determine the position of the fire. Additionally it is possible, by statistical evaluation of the infrared picture, to determine the width of the combustion zone and derive from this information on asymmetric positions of the fire or secondary combustion zones.

Even if all the information given above is available it is still necessary to development an adequate mathematical model which will allow the information the information to be applied to a conventional control system. However combustion processes as a rule are of a highly non-linear nature and represent multi-variant problems. For conventional control system strategies the only feasible method is therefore to influence controllers by heuristic programming [8,10].

New solutions to problems of that kind are offered by the use of advanced control techniques. In particular fuzzy logic control has been applied to similar combustion processes [7,13]. Complex interactions between the various items of information evaluated require a methodology of structured information analysis.

Diesel engine combustion control is a capital intensive process due to increasingly severe environmental regulations. Complying with the environmental rules demands the adoption of new and expensive technologies for both the combustion process and treatment of the residues (gaseous, liquid). The investment and operation costs have to be regained through adoption of a number of measures, primarily more efficient combustion, which should minimize the residues and maximize the energy recovery.

With the advent of Diesel engine combustion control's facilities toward high speed and automation, the on-line monitoring becomes important [11]. Since a facility of the Diesel engine combustion control may break down unexpectedly, the unscheduled shutdown will cause the losses and deteriorate of the environment. The traditional time-based maintenance strategy does not take into account the practical machine condition and only performs the maintenance practice at a fixed time interval from local personnel. Therefore, accidental breakdowns will unavoidably happen, when the chosen time interval is too long. Otherwise, it will lead to over-maintenance, when the chosen time interval is too short.

In order to increase automation in Diesel engine combustion control, an on-line monitoring system for facilities based on computer's technology gains popularity. In general, facilities monitoring or supervision is carried out by measuring some parameters from some facilities of Diesel engine combustion control. The justification for on-line facility monitoring is prompted by the needs to achieve the availability of facilities and to protect facilities from some problems of running.

A five level's architecture is considered, in general, for MSW incineration automation schemes: process, local control, supervision, planning and administrative management, each one with specific tasks. In general, the couple model–control has constrained suppositions about the process' dynamic and its causes are not thoroughgoing modelled, thus, in case of not well working, the process performance should be deteriorated. These modelling suppositions put up with the development of Alarm processing system proposing fault detection, diagnosis and prediction tasks, and preventive maintenance plans; these tasks living together in the supervision and control levels. A multi-agent based automation model is proposed, taking into account the general objects of the production processes: production planning, production factors management, processes control, Alarm processing and abnormal situations management. The agents-based reference model for distributed control systems (DCS) has provided a generic conceptual framework permitting to see any system like a feedback control systems [14].

Alarm processing and maintenance are vital aspects in industrial process. They not only include the knowledge about the failure modes and their causes, but an awareness of the extent to which equipment failure affects safety, product quality, costs and plant availability. Changes in the maintenance approaches, recent knowledge, skills of maintenance people, and so on, must be incorporated into the maintenance systems in order to attain high performances. In this sense, maintenance systems should be developed in an open framework permitting to take into account the changes in the maintenance paradigms. In the maintenance framework, the challenges are related to the selection of the appropriate techniques for fault detection, isolation and diagnosis; the building of the appropriate models for detection, isolation, diagnosis and prediction; the decision-making in the design of feasible preventive maintenance plans; the decision-making in the emergency situations.

In this work, a multi-agents based Alarm Processing System (APS) reference model for MSW incineration process is proposed. This model outlines the management of the preventive maintenance plans permitting to perform detection, isolation and diagnosis tasks. This model takes

into account the following tasks: local or global abnormal behavior detection of the controlled process; fault isolation and diagnosis; fault causes identification; preventive and corrective maintenance task running; maintenance task planning; fault prediction models development.

The APS objectives are achieved by the coordinated interaction of the agents. Reasoning aspects as well as adaptive and learning aspects are incorporated into the multi-agent behavior, helping in the interpretation of available data from different sources and characteristics. This way, the agent-based APS provides the assistance in the detection-diagnosis-decision process, as well as in the planning and running of maintenance tasks.

Alarm processing refers to the interpretation of a large number of alarms in a computerized control system. It is one of the important functions of the computerized supervisory control and data acquisition (SCADA) systems. An alarm is a structured signal from the SCADA system which is used in distribution power substation control centers [15].

Alarm processing thus relies on status information or measurements gathered at a large number of points distributed throughout those substations. The advances in computer and communications technology in the last three decades have made it possible for a large SCADA system to scan 20,000 to 50,000 points every few seconds. Such a SCADA system will have the ability of displaying 500/1000 or even more alarm messages per minute.

The enormity of the system and rate at which messages are displayed on the console increases the complexity of the problem multifold. This complexity leaves an operator in the control center that has to analyze these alarm messages a highly constrained time frame suffering from high stress and cognitive overload. In early years, prioritizing the alarms and arranging alarm messages have been used to reduce the number of alarms listed to the operator's console [16]. However, the operator still needs to analyze the real-time messages to diagnose the system problem. In general, alarms can be caused by different events and an event can result in multiple alarms.

Uncontrollable events such as sudden load fluctuations, equipment failures and atmospheric perturbations can propel the system from a stable and secure state to an insecure and unstable state. The operator in these circumstances has to take immediate action in order to pull back the system into an acceptable state. Failure to respond quickly, can lead to catastrophic circumstances like the state of the system may continue to deteriorate with some loads getting disconnected or in extreme cases, and the entire system may collapse, leading to a blackout for hours. Under abnormal conditions, trigged alarms responsive to abnormal conditions will disturb the operator to take corrective actions in handling the fault. Therefore, alarm processing issues are worth studying and have received significant interests of many researchers.

In this paper, an agent-based alarm processing architecture is presented to solve the alarm problems inherent in computerized and centralized SCADA systems. For operations that require timely action, the proposed intelligent alarm processor is an ideal way to assist operators to take corrective actions.

The definition of an alarm is described as follows:

a. A measured analogue value exceeds the preset level.

b. A measured digital value changes the status.

2 The multi-agent technology

2.1 ORIGINS AND CONCEPTION OF MULTI-AGENT

The direct reason for MAS (Multi-Agent System) coming into being is that the ability of MAS in dealing with problems is better than that of single Agent, and MAS can communicate with current systems and software, it can enhance the efficiency and robustness of systems, and it is distributed. MAS are Agents that can work together or act autonomously in their environment to complete a set of goals. An Agent as well as a Multi-Agent system can be developed efficiently [17]. The study on Multi-Agents technology mainly lies in two fields: studying the architecture, communication and learning mechanism of MAS in theory, and studying scientific computing, computer network, robot and traffic controlling in application.

2.2 THE FRAME OF MULTI-AGENT SYSTEM

2.2.1 Centralized multi-agent system

The center is responsible for the control and cooperation of the system. This frame would keep the information coherent in the system; it can be managed, controlled and maneuvered easily. The disadvantage is that as every Agent becomes complicated and dynamic; the bottleneck in controlling becomes outstanding. The deadly shortcoming is that, once the Agent that controls the local part or all parts collapses, all parts will collapse.

2.2.2 Distributed multi-agent system.

This frame contains many service organizations, it offers services when Agents cooperate with others, and every Agent is equal. The virtue is to increase the flexibility and stability of the system, and thus the bottleneck in controlling is abated [18]. The shortcoming is that since every Agent is restricted by local and incomplete information, it is hard to implement the coincident behavior.

A multi-agent system is a system where multiple agents are acting in the same environment. Running in the same environment implies that they get the same inputs and that there actions affect the same environment. This automatically leads to more complex issues as soon as there are multiple agents in the same environment the action of one action affects all the other agents in the system as well. This clearly shows the need for some sort of communication between these agents. There are two basic modes of multi-agent systems: cooperation and competition. In a cooperative mode, agents cooperate to reach a common goal. Agents in such a system might for example each have a very specialized function so that they can only together achieve a goal that requires several of specialized functions. In competitive MAS, on the other hand, agents compete with each other for limited resources like computation

time [19]. MASIAP is designed as a cooperative one. Agents in MASIAP are specialized agents that are each pursuing their own goals. To reach these design objectives those agents most cooperate with other agents in the system because they can't achieve their goal without cooperation.

Traditional SCADA system attempts to monitor and control everything from a control center. Remote Terminal Units (RTUs) are widely used for centralized monitoring and control. A fixed communication channel has to be established for data exchange between control center and remote terminal units, and the failure of a remote terminal unit or of the centralized monitoring server could cause the whole system to fail. To conquer the limitations of centralized monitoring, a multi-agent approach is employed. The rationale and justification for the choice of multi-agents for distributed transformers on-site monitoring is the capability of an agent to respond to changes in its environment, act in anticipation of such changes, and interact with other agents to solve problems in a cooperative manner. Multi-agent systems are ideally suited to the characterization or design of distributed computing systems. With reference to the presentation of information and resource control in the distributed monitoring system, agents are appropriate because:

1) They are intelligent agents.
2) The autonomy of an agent enables it to encapsulate the identity of a particular user, and tailor its response to reflect the user's responsibilities, such as monitoring the state of relays.
3) Agents allow the integration of existing systems. Therefore agents can negotiate with each other to obtain required information.

2.3 THE FRAME OF MULTI-AGENT DECISION-MAKER

Agent Decision-Maker is used to hold the diesel engine combustion control of DSS, which is based on A-Team (asynchronous team) agent interaction structure. As shown in Figure 2, all of the diesel engine combustion control tasks which are need to process are store in a "Population". The sub-agents of the Decision-Maker are discussed as below:

1) Analyzer, which is used to analyze the diesel engine combustion control tasks and target the attributes of all the tasks.
2) Inquirer, which acquires the information from the system source database to support Analyzer for analyzing and targeting the attributes.
3) Updater, which updates the information changed by Analyzer to the system source database.
4) Specilizer, which constructs the rules of tasks from "Population".
5) Decomposer, which decomposes the tasks from Specilizer.

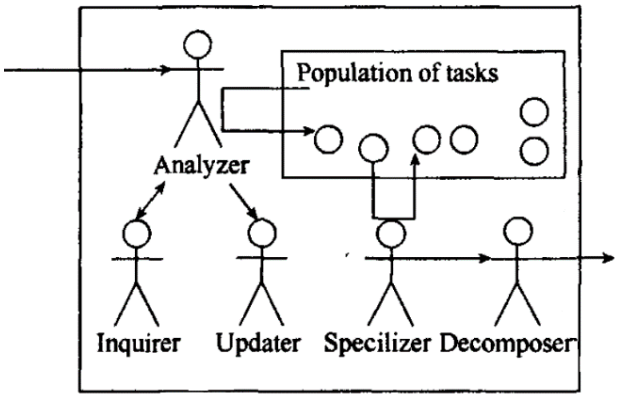


FIGURE 2 The frame of decision-maker

3 The proposed architecture of multi-agent based alarm processing system

This section will outline the general architecture of the multi-agent based alarm processing system (MASAPS). The architecture is presented from two perspectives: one is the software perspective and the other is the system perspective. The system perspective shows the relationship between different parts of the system and the physical location of the different software parts. The software perspective describes the way in which the system is embedded into different agents and the style in which they interact with each other [20]. The MASAPS mentioned above has been evaluated in the simulation experiment. The MASAPS is under development using the Java language and can run under any computational platform that supports Java byte code and GUI.

The system is under development using the Java language and can run under any computational platform that supports Java byte code and GUI (Graphic User Interface). The system architecture is depicted in Figure 3. It consists of five major components, Java-based human machine interface (HMI), data analysis agent, monitoring agent, control agent and database agent. We choose JADE (Java Agent Development framework) as development tools because it can provide an agent environment to communicate with each other and migrate from one substation to another. We will discuss the function of these components in the following sections.

In each substation, there exists database used for persistent storage of historical real-time data which connected to database agent. The real-time information is gathered from substation remote terminal units by monitor agent over serial RS232 or TCP/IP. The data analysis agent is not only responsible for analyzing the receiving measured values, but also can generate diagnosis information via its reasoning process. One of the advantages to detect data directly from each local remote terminal unit is the ability to suppress redundant alarms. After perceiving information from substation remote terminal units, the monitoring agent provides a copy of the receiving information to database agent for persistent data storage. Each user can access the system information through the Java-based HMI. In

order to effectively analysis system alarms, all agents are capable of communicating with other agents in different substations.

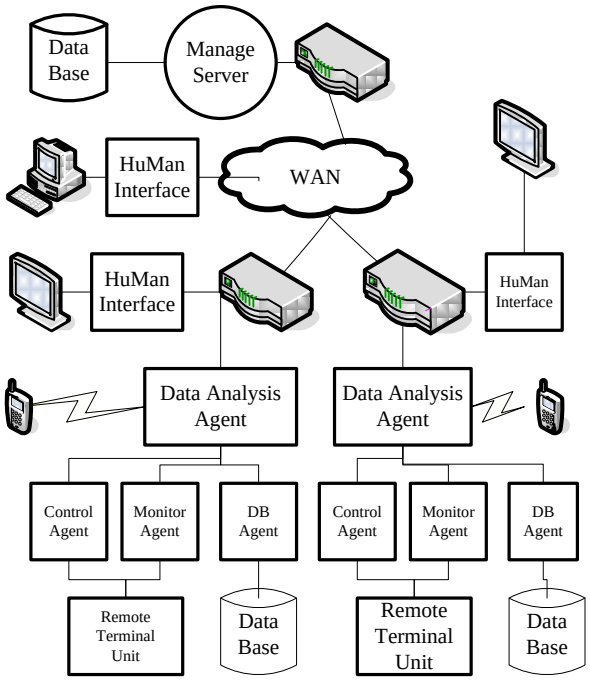


FIGURE 3 The proposed architecture multi-agent based alarm processing system

4 The function of the multi-agent

4.1 DATA ANALYSIS AGENT (DAA)

The Data Analysis Agent (DAA) plays an important role in the multi-agent systems. Hich task is to gather information from substation remote terminal unit and perform data analysis, and here the control task is executed by the control agent. Hile exist three software agents: messenger agent, fault analyzer agent, and learner agent. The messenger is an information centre that allows sending and receiving messages to other agents via FIPA compliant agent communication language. The incoming real-time data which come from monitoring agent is analyzed and fault sections using cause-effect networks to determine the relevant event is identified by the fault analyzer agent. The learner agent plays the role for enhancing accuracy of event identification by using the post experiences from interaction with user.

4.2 MONITORING AGENT (MA)

The main function of monitoring agent is to continuously monitor measured analogue values that exceed the preset limitation and/or digital values that change state from remote terminal unit. The collected data will be sent to data analysis and store in the database via database agent.

4.3 CONTROL AGENT (CA)

The control agent is responding to manages and performs a low-level control action in a secure way. By clicking a component for operation requests in the diagram window, a pop-up dialog box is displayed for confirming an executable operation. Through check before execution and software interlock functions to avoid critical operation process.

4.4 DATABASE AGENT (DA)

Several reasons lead us to make the decision that data about the activity in the substations, its environment and the systems activity should be logged.

These reasons were:

1) Requirement to make statistical tests about the behavior of the system.
2) To have a learning base, where the system could learn from experience.
3) An ability to plot charts about the behavior of the system.

4.5 USER INTERFACE AGENT (UIA)

The User Interface Agent is fully implemented by using Java, which provides a graphic operation interface with cross-platform feature that can run under different platforms with Java Virtual Machine (JVM).

4.6 MANAGEMENT SERVER

The management server, located in control center, contains a facilitator that provides with agent service lookup, user authentication, ontology database. The directory facilitator, connected to a recovery database, is designed for system recovery and migrate information from each substation. User authentication is the way of restricting access to database by asking for a username and password. Ontology is a description of a conceptualization and has properties for knowledge sharing. For pragmatic reasons, we choose to provide an ontology database as a set of definitions of formal vocabulary.

The proposed approach adopts a rational agent model as human practical reasoning and to represent human's mentalist attitudes. Belief, desire, and intention, representing the informative, motivational, and deliberative states of the agent, respectively, are the three attitudes of an agent.

When an event occurs, the goal of the proposed agent system will try to reach the event from overwhelming alarm messages by agent cooperation in related substations. The definition of events and their corresponding alarms, according to substation configuration, is used as the beliefs. The plan, which is deliberative attitude, is the way that agents achieve their goal. Each agent has to store the facts as beliefs that make up the agent's knowledge.

5 Conclusions

This Multi-agent based alarm processing system for MSW incineration could better realize heterogeneous enterprise data, application systems and business processes. Due to a distributed system, MAS modular design not only reduces the complexity of the system, but also is easy to expand, as well as has good robustness and reliability. It also can be integrated to the existing SCADA system. We have designned and implemented a multi-agent based alarm processing system framework, MASAPS, Which has cross-platform architecture, distributed computing capability, and intelligent inference mechanism. The system may be a suitable training simulator for system operators in the future.

Acknowledgments

This paper is supported by National Natural Science Foundation of China (Grand No. 61263023, No. 61064010 and No. 61063006), Yunnan Provincial Department of Education Research Fund (Grand No. 14051522) and Kunming University of Science and Technology Discipline Direction Team Fund (Grand No. 14078265).

References

[1] Feng J Q, Bise D P, Wu Q H, Fitch J 2002 A multi-agent based intelligent monitoring system for power transformers in distributed substations IEEE *International Conference on Power System Technology (PowerCon)* **3** 1962-5
[2] Buse D P, Sun P, Wu Q H, Fitch J 2003 IEEE *Power and Energy Magazine* **1**(2) 50-5
[3] Wang H, Wang T 2009 Multi agent based on-line Monitoring System for Diesel engine combustion control *2009 International Workshop on Advanced topic in Embedded Computing Technology and Systems* (ATECTS09)
[4] Wang H, Wang H 2008 Research and implementation of multi-agent based test paper generation algorithm *International Conference on Computer Science and Software Engineering, CSSE* 2008a
[5] Wang H, Wang H 2008 Research and design of multi-agent based intrusion detection system on wireless network 2008 *International Symposium on Computational Intelligence and Design ISCID* 2008b
[6] Wang H, Zhang Y 2008 Multi-agent based chemical plant process monitoring and management system 2008 *International Conference on Wireless Communications, Networking and Mobile Computing WiCOM* 2008
[7] Zhou J, Sun Y 2003 Study on distributed measuring and controlling system based on multi-agent technology Control & Automation **19**(9)
[8] Luo Y, Liu K, Davis D N 2002 IEEE Network **16**(1) 20-7
[9] Bryson J 2001 Intelligence by design: principles of modularity and coordination for engineering complex adaptive agents *PhD Thesis Massachusetts Institute of Technology*
[10] Yu R, Iung B, Panetto H 2003 A multi-agents based e-maintenance system with case-based reasoning decision support *Engineering Application of Artificial Intelligence* 16
[11] Iung B 2003 From remote maintenance to MAS-based e-maintenance of an industrial process *International Journal of Intelligent Manufacturing* **14**(1) 59-82
[12] Jennings N, Bussmann S 2003 IEEE *Control Systems Magazine* **23**(3) 61-73
[13] Marik V, McFarlane D 2005 IEEE *Intelligent Systems* **20**(1)
[14] Wagner T 2003 Applying agents for engineering of industrial automation systems In: M Schillo (Eds) *Lecture Notes of Artificial Intelligence MATES* **2831** 62–73
[15] Cerrada M, Aguilar J, Cardillo J, Faneite R 2002 Agent-based reference model for fault management Technical Report Grant Project I-621-98-02-A *Universidad de Los Andes Mérida, Venezuela*
[16] Wang H R, Li Y, Wang J Y, Liao Y Y 2013 Application of Siemens 840Di system in the reconstruction of KISC roll collar grinder *Applied Mechanics and Materials* **2773**(437) 740-3
[17] Wang H R, Li Y, Wang J Y, Liao Y Y 2013 The process improvement of KISC roll collar grinder *Applied Mechanics and Materials* **2773**(437) 744-7
[18] Zhang S H, Wang HR, He XL 2013 Pose-invariant face synthesis and recognition via sparse coding and symmetrical information *Applied Mechanics and Materials* (437) 894-900
[19] He X L, Bi G H, Wang H R 2013 A dynamic bipartite scale-free agent network model for analyzing heterosexual HIV transmission *25th Chinese Control and Decision Con (CCDC) Guizhou, China* 4842-8
[20] Wang H R, Li Y, Wang J Y 2013 A novel power saving algorithm for wireless sensor networks *International Journal of Online Engineering* **9**(5) 78-81

Authors

Hairui Wang, China.

Current position, grades: professor in Kunming University of Science and Technology, China.
University study: PhD degree in Metallurgical Engineering Control, Kunming University of Science and Technology 2007.
Research activities: automatic control, signal detection and processing, instrumentation development.
Professional Activities and Memberships: National Natural Science Foundation of China (on Tamping Wagon automatic control).

Ya Li, China.

Current position, grades: associate professor in Kunming University of Science and Technology, China.
University study: PhD degree in Metallurgical Engineering Control, Kunming University of Science and Technology 2012.
Research activities: automatic control, signal detection and processing, instrumentation development.
Professional Activities and Memberships: National Natural Science Foundation of China (on Tamping Wagon automatic control).

Fault diagnosis expert system for large-scale railway maintenance equipment based on BP neural network

Ming Pu[1], Hairui Wang[2*]

School of information engineering and automation Kunming University of Science and Technology, No. 727, Jingming South Road, Kunming, China

Received 1 March 2014, www.cmnt.lv

Abstract

According to the characteristics of the neural network and expert system, a fault diagnosis method for large-scale railway maintenance equipment based on Back Propagation (BP) neural network and expert system is proposed. Fault diagnosis for large-scale railway maintenance equipment on BP neural network and expert system model are constructed. A weak of collection and expression of knowledge on traditional expert system is overcome. Availability of the method based on BP neural network system and expert system is verified by experimental results of large-scale railway maintenance equipment fault.

Keywords: BP neural network; neural network-knowledge base; expert system; fault diagnosis

1 Introduction

Tamping machine is a complex large-scale railway maintenance equipment, it often appears various faults inside and outside the tamping machine because tamping machine always work in field with influence of weather, temperature, humidity, dust, etc. [8]. Once a part of tamping machine fails, the whole system will cause malfunction, therefore the system performance will decline, serious faults will lead to a loss of most of the original features of system. Expert system and neural network methods both can be used in fault diagnosis, but single expert system and neural network fault diagnosis both have certain limitations. Theoretical analysis and experimental studies have shown that expert system based on symbolic logic specializes in reasoning work and artificial neural network based on practice specializes in information perception, both complement each other in function. A combination of expert system and neural network can make themselves mutually reinforcing. In order to make full use of the advantages of expert system and neural network, this paper will organically integrated diagnosis method based on rule-based reasoning of expert system and artificial neural network. Thus we can solve problems that existed in system such as "Knowledge Acquisition Bottleneck", "Match Conflict", "Combinatorial Explosion", "Infinite Recursion" and apply in fault diagnosis of tamping machine.

2 Principles and characteristics of expert system based on BP network

Since the system of tamping machine is very complex, working environment is very bad and the existence of other common phenomenon such as "high fault rate" and "hard to find fault reason" [10]. The system measures working parameters through monitoring status of tamping machine in real-time and uses these parameters to make a contrast with normal condition value [7]. When the measured working parameter does not match normal condition value, it can be shown that there exist at least one or more faults in tamping machine. We can consider using functions which conclude learning and memory in a neural network to replace the role of knowledge engineers. The main content is to learn from faults and complete automatic acquisition of fault diagnosis expert system and automatic updates. The reasoning process of building expert system in a neural network is actually a process of computing, it means that the neuron of neural network constantly parallels until convergence on solving space problems through the relationship among knowledge which are stored in a neural network. Because the parallel mathematical reasoning process in neural network replaces the matching search and backtracking problems in the processing of traditional artificial intelligence (AI) symbol, neural network has a higher efficiency of reasoning.

BP neural network will change diagnosis information into the weights and thresholds, and make them stored in network separately [12]. A combination of BP network and traditional expert system can overcome many defects in expert diagnosis system. BP network has a strong ability of adaption because diagnosis information which are distributed stored in the connected weights of the whole network, and each neuron stores various kinds of information, even though parts of neurons lost, the impact on BP network storage diagnosis information is not much, thus BP network can handle large disturbed input information, furthermore, BP network can solve problems about the bottleneck of knowledge acquisition in traditional expert system – self-learning [11].

* ***Corresponding author's*** e-mail: hrwang88@163.com

3 Technical basis of BP neural network

In system based on neural networks, method of knowledge representation is completely different from method of traditional AI system. Traditional AI system uses the method of explicit knowledge representation, on the contrary, neural network uses implicit knowledge representation [9]. In Artificial Neural Network (ANN), knowledge is not expressed as a set of rules, unlike traditional methods, it is expressed a certain knowledge as the weights distribution in the same network. Figure 1 and 3-layer threshold BP network show the following 4 "XOR" logical formula rules:

IF x_1 =0 AND x_2=0 THEN y=0;
IF x_1 =0 AND x_2=1 THEN y=1;
IF x_1 =1 AND x_2=0 THEN y=1;
IF x_1 =1 AND x_2=1 THEN y=0

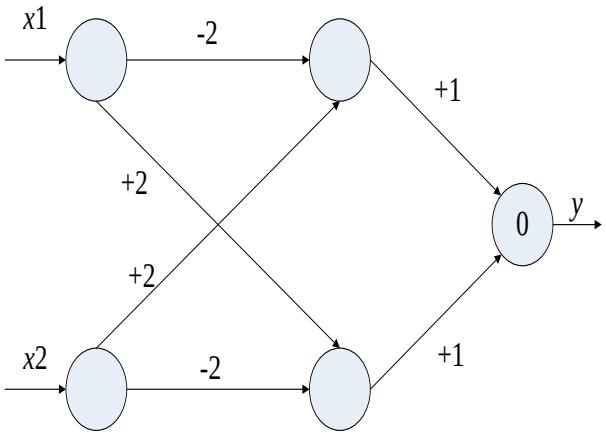


FIGURE 1 ANN representation of "XOR" logic

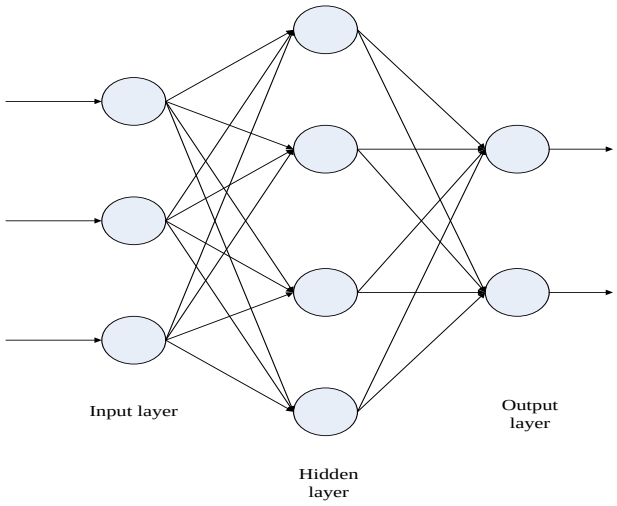


FIGURE 2 Network structure

Based on this network knowledge representation structure, its BP network structure is shown in Figure 2. Experts provide examples of problems and corresponding solution, BP network learns samples through learning network modules and modifies the weights until required learning through adaptive algorithm in network is accurate [12]. The learning process is as follows: Firstly, input fault symptoms and spread them forward to hidden nodes, secondly, spread the information of hidden nodes to the output nodes through output transfer function, thirdly, give results of output. Sigmoid function is chosen for the output transfer function. The learning process of this algorithm is composed of forward-propagation and back-propagation.

4 Corresponding formula

4.1 ORWARD-PROPAGATION STAGE

Set *k-th* sample of input vector x_k =(x_{k1}, x_{k2},…, x_{kl}), get the output of y_j of *j*-th nodes in the hidden layer:

$$y_j = f\left(\sum_{i=1}^{L} W_{ij} x_{ki} + \theta_j\right). \quad (1)$$

The output of O_r of *r-th* nodes in the output layer is:

$$Q_r = f\left(\sum_{j=1}^{M} W_{rj} y_j + \theta_r\right). \quad (2)$$

4.2 BACK-PROPAGATION STAGE

Set L pairs of learning samples (X_k, O_k) (K=1,2,…,l), the mean square error of actual output Y_k and required output O_k is:

$$E_k = \frac{1}{2}\sum_{p}^{m}\left(Y_{K,p} - O_{K,p}\right)^2. \quad (3)$$

The total error of sample set is:

$$\mathrm{E}=\sum_{k=1}^{l} E_k \ . \quad (4)$$

Modifying weights along negative gradient direction of error function so that network convergence, the weights in the output layer is modified as ΔW_{rj} :

$$\Delta W_{rj} = \frac{\partial \mathrm{E}(t)}{\partial W_{rj(y)}} = -\left[O_r'(t) - O_r(t)\right]\cdot\left[1 - O_r(t)\right] y_i(t). \quad (5)$$

Weighs in the output layer which is modified is:

$$W_{rj}(t+1) = W_{rj}(t) - \eta \Delta W_{rj} \quad (6)$$

Similarly, weights in the hidden layer which is modified is ΔW_{ji} :

$$\Delta W_{ji} = \frac{\partial \mathrm{E}(t)}{\partial W_{ji(t)}} = -\left[y_j(t)\left[1 - y_j(t)\right]\right]\cdot x_{ki}(t)\sum\nolimits_r \delta_r W_{rj} \ . \quad (7)$$

Weights in the hidden layer is modified as:

$$W_{ji}(t+1) = W_{ji}(t) - \eta \Delta W_{ji}. \quad (8)$$

Let us introduce the following terms:

f - sigmoid function;

L - numbers of neurons in the input layer;

W_{ij} - connection weights between input layer and hidden layer;

x_{ki} - the input of k-th sample;

k - the number of sample;

θ_j - thresholds of j-th neurons in the input layer;

M - numbers of neurons in the hidden layer;

W_{rj} - connection weights between hidden layer and output layer;

θ_r - thresholds of j-th neurons in the hidden layer;

m - numbers of units in the output layer;

E(t) - error function;

$O_r(t)$ - desired output;

η - steps, experience;

δ_r - sensitivity of r-th nodes correspond to output error function E(t);

5 Design of fault diagnosis expert system

Based on the analysis of fault diagnosis expert system and BP neural network technology above, expert system and neural network must be combined to play their strong point when problems can't be solved using expert system or neural network alone. So I design a fault diagnosis expert system. The basic structure of the process is shown in Figure 3.

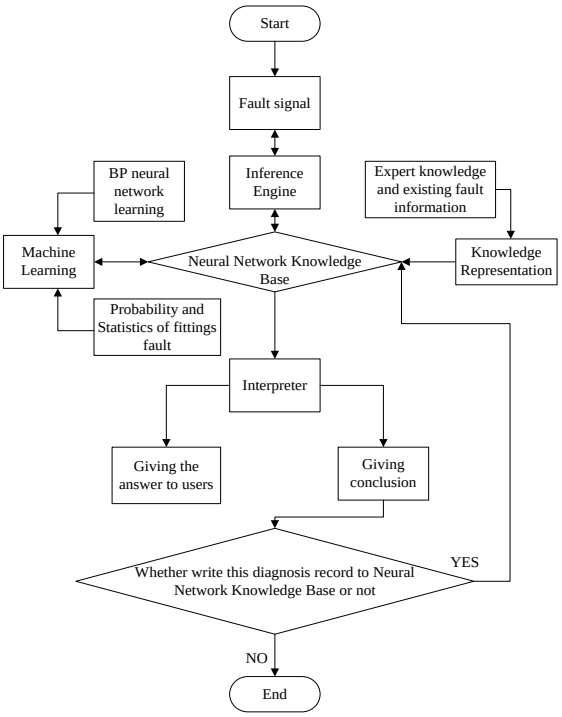


FIGURE 3 Fault diagnosis process of tamping machine

6 Simulation of tamping machine fault diagnosis expert system

The size of neural network is determined by need, the number of fault symptoms is corresponded to the number of input nodes; the number of fault is corresponded to the number of output nodes. The select of the number of hidden nodes in the hidden layer doesn't have theoretical guidance, the selected number is merely to meet capacity and requirements of certain learning speed.

This paper selects common engine fault diagnosis of 09-32 taming machine as the research object, illustrating the practice and diagnosis process of expert system and neural network fault diagnosis system [8].

Tamping machine mainly uses diesel engine, diesel engine often appears many faults such as electrical fault, mechanical fault and oil path fault [8]. Electrical fault shows mainly the line-to-ground fault, line empty pick up fault, damaged components fault and unreliable grounding fault. Line-to-ground fault is most common, as long as grounding can cause a short circuit and automatically cut off the power supply. Virtual connect circuit fault is mainly due to loose caused by vibrations in long-term work, effective way to avoid this is to strengthen the line inspection and tight of terminals. Mechanical faults are relatively less, but there are always several typical faults such as governor throttle trolley balance spring results in only idling, bolts of governor transmission shaft sheared causes failure of work, nozzle jammed results in a mass of diesel leaking into the machine and diluting the oil. The typical fault of oil path fault is diesel tank blocked, the main reason is that diesel tank not cleaned for a long time or some foreign matters get into the tank. When suction pipe appears loosen or the installation of filter element is unqualified, it will result in air suction, low back pressure of suction pipe, air into high-pressure oil pump.

The engine of this diagnosis system has 5 symptoms and 7 faults (Table 1). The 5 symptoms and 7 faults are designed for 5 input nodes and 7 output nodes, hidden nodes are calculated as 5. To calculate the answer, using the formula $h = \sqrt{n+m} + a$ (h is the number of neurons in the hidden layer; n is the number of input neurons; m is the number of output neurons; a is the network layer, which is a constant number between 1-10, here a is 2).

TABLE 1 Neural network output table

Input	Symptom	Output	Fault
x_1	DG08 charging current is high, table shows the maximum and the next to the diesel engine has screaming voice.	y_1	The ground wire of B13box is bad contact.
x_2	5e9 self-healing insurance will jump and can't start when the engine starts.	y_2	The damage of 317 lines in dryer working power supply results in grounding.
x_3	The diesel engine suddenly stalled when working.	y_3	5e9 and 5e7 self-healing are jumping.
x_4	Diesel tank is blocked.	y_4	Engine is damaged.
x_5	Throttle can't lift.	y_5	The damaged 81Re3 results in adjustment can't be passed to post - amplifier.
		y_6	The damaged Si1insurance of 81u1 results in power supply.
		y_7	Diesel tank is not cleaned for a long time.

In the Table 1, x_i represents fault symptom, x_i=0 (i=1,2,…,5) represents no symptom, x_i=1 (i=1,2,…,5) indicates the symptom, y_j =0 (j=1,2,…,7) represents a fault that doesn't exist, y_j =1 (j=1,2,…,7) represents a fault that existed. Based on the above, we selected 7 samples of examples to learn, when the network structure and learning sample are selected, you can train the neural network. First of Table 2 is trouble-free sample.

TABLE 2 Learning sample table

Sample	x_1	x_2	x_3	x_4	x_5	y_1	y_2	y_3	y_4	y_5	y_6	y_7
1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	1	1	0	0	0	0	0
3	0	1	0	0	0	0	0	1	1	1	0	0
4	0	1	0	0	0	0	0	1	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	1
6	0	0	1	0	0	0	0	0	0	1	0	0
7	0	0	0	1	0	0	0	1	0	0	1	0

Table 3 shows the desired output and actual output through training.

TABLE 3 Desired output and Actual output

Desired output							Actual output						
y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_1	y_2	y_3	y_4	y_5	y_6	y_7
0	0	0	0	0	0	0	0.0266	0.0268	0.0381	0.0038	0.0337	0.0093	0.0032
1	1	0	0	0	0	0	0.9663	0.9663	0.0042	0.0043	0.0027	0.0170	0.0152
0	0	1	1	1	0	0	0.0024	0.0037	0.9635	0.9342	0.9795	0.0068	0.0091
0	0	1	0	0	0	0	0.0023	0.0024	0.9923	0.0043	0.0199	0.0041	0.0108
0	0	0	0	0	0	1	0.0016	0.0160	0.0110	0.0073	0.0053	0.0233	0.9586
0	0	0	0	1	0	0	0.0027	0.0034	0.0087	0.0110	0.9694	0.0026	0.0152
0	0	1	0	0	1	0	0.0102	0.0093	0.9762	0.0046	0.0860	0.9508	0.0128

In fault diagnosis, according to the input fault symptom, we call the trained weights between the layers to calculate forward and display the results of diagnosis based on output of the output layer.

TABLE 4 Fault diagnosis example

Input					Output						
0	1	0	0	0	0.0019	0.9900	0.0152	0.0427	0.0025	0.0014	0.0

Table 4 shows that fault can be located in the damage of 317 lines in dryer working power supply results in grounding when 5e9 self-healing insurance jumps with diesel engine started. Table 3 also shows that the result of fault is consistent with the actual and this method is feasible in fault diagnosis application of tamping machine.

7 Conclusions

BP neural network has a feature of highly nonlinear function mapping, so it is very convenient for tamping machine fault diagnosis. It can not only improve the reliability of diagnosis, but also develop tamping machine fault diagnosis expert system by using BP neural network technology. This system has a self-learning function, with the growth of time investment, the system will automatically record the result of each test into knowledge base and update network weights constantly based on knowledge base, as a result, the system and the diagnosis accuracy will constantly growing and improving.

Acknowledgements

This paper is supported by National Natural Science Foundation of China (Grand No.61263023, No. 61064010 and No. 61063006), Yunnan Provincial Department of Education Research Fund (Grand No.14051522) and Kunming University of Science and Technology Discipline Direction Team Fund (Grand No.14078265).

Reference

[1] Wang H R, Li Y, Jiang Y 2011 Web service and multi-agent based fault diagnosis system for MSW incinerators *Advanced Materials Research* **179–180** 580-5
[2] Wang H R, Li Y 2010 Multi-agent based alarm processing system for MSW incineration *Int Conf Technology Management and Innovation TMI Wuhan China*
[3] Wang H R, Li Y, Wang J Y, Liao Y Y 2013 Application of Siemens 840Di system in the reconstruction of KISC roll collar grinder, *Applied Mechanics and Materials* **2773**(437) 740-3
[4] Wang H R, Li Y, Wang J Y, Liao Y Y 2013 The process improvement of KISC roll collar grinder *Applied Mechanics and Materials* **2773**(437) 744-747
[5] Zhang S H, Wang HR, He X L 2013 Pose-invariant face synthesis and recognition via sparse coding and symmetrical information *Applied Mechanics and Materials* (437) 894-900
[6] Wang H R, Li Y, Wang J Y 2013 A novel power saving algorithm for wireless sensor networks *International Journal of Online Engineering* **9**(5) 78-81
[7] Wu G, Zhang L, Hu X, Dai Y 2012 Design and realization of condition monitoring and fault diagnosis system for railway tamping machine Changsha *Journal of Hunan University*
[8] Han Z, Tan D 2004 Hydraulic Tamping Wagon *Beijing China Railway Publishing House* 1-4 (*in Chinese*)
[9] Wu J, Kuo J 2009 An automotive generator fault diagnosis system using discrete wavelet transform and artificial neural network *Expert System with Application* (6) 9776-83
[10] Rosiek S, Batlles F J, Battlesf J A 2008 Microcontroller based data-acquisition system for meteorological station monitoring *Energy Conversion and Management* (12) 3746-54
[11] Nan Cen, Khan Faisal, Iqbal Tariq M 2008 Real-time fault diagnosis using knowledge based expert system *Process Safety and Environmental Protection* (1) 55-71
[12] Tang H 2004 Network Fault Diagnosis System based on BP neural network *Nanjing Nanjing University of Science and Technology* (*in Chinese*)
[13] Guarnaccia C, Neri F 2013 An introduction to the special issue on recent methods on physical polluting agents and environment modelling and simulation *WSEAS Transactions on Systems* **12**(2) 53-4
[14] Pekar L, Neri F 2013 An introduction to the special issue on advanced control methods: Theory and application *WSEAS Transactions on Systems* **12**(6) 301-3
[15] Muntean M, Neri F 2012 Foreward to the special issue on collaborative systems *WSEAS Transactions on Systems* 11(11) 617
[16] Mahmoud Q H 2007 Getting started with the java rule engine API(JSR 94): toward rule-based applications http://java.sun.com/developer/technicalArticles/J2SE/ accessed on Aug. 27 2007

Authors

Ming Pu, China.

Current position, grades: postgraduate student in Kunming University of Science and Technology, China.
University study: Bachelor degree in network engineering, Xi'an University of Post & Telecommunications, 2013.
Research activities: main research direction for automatic control.

Hairui Wang, China.

Current position: grades professor in Kunming University of Science and Technology, China.
University study: PhD degree in Metallurgical Engineering Control, Kunming University of Science and Technology 2007.
Research activities: Main research direction for automatic control, signal detection and processing, instrumentation development.
Professional Activities and Memberships: National Natural Science Foundation of China (on Tamping Wagon automatic control).

Innovation model for college ideological and political education in network environment

Jiahua Zhou*

Institute of Management Engineering, Su Zhou University, Anhui, 234000, China

Received 1 October 2014, www.cmnt.lv

Abstract

Ideological and political education is impacted by the rapid development of network technology, computer technology and high-speed communication technology, in aspects of concept, content, ways and means, and can not meet requirements of current network age development. Therefore, the ideological and political education is placed in network technology to form new pattern. This paper, first, made brief overview of the concept and characteristics of ideological and political education in network environment, then summarize college ideological and political education at home and abroad. Afterwards, it analyzed the challenges and opportunities after combined network technology and college ideological and political education, then proposed relative solutions in the perspectives of concept, content, means and position. At last, it put forward the construction of shared community pattern for ideological and political education in network environment, consisting of the organization structure and operation mechanism of shared community.

Keywords: network platform, college ideological and political education, teaching model research, innovation

1 Introduction

College is a place with high informatization degree in our country. The wide application of information technology and network technology brings not only rich resource and teaching platform for college ideological and political education, but also the unprecedented impact and challenge. Network, as the emerging fourth media after broadcast, newspaper and television, changes the idea and value orientation of leaders and students from college [1]. Under the network environment, the urgent task and efforts direction for current college ideological and political educator is to carry out ideological and political education work, innovate education model, and better perform the function of ideological and political education [1].

2 Overview of ideological and political education in network environment

The combination of information technology and communication technology promote the formation of network environment. Academic world defines network environment as the socialized visual environment. Ideological and political education is placed into network environment. With the principle of declaring idea and communication and the basis of stereo spatiality and interactive communication, ideological and political education relies on various newly-developing social tools and platforms supplemented with the behavior and ethics constraints of ideological concept, political concept and moral rule on network audience, thus promote the comprehensive development [2,3].

The combination of ideological education and network environment makes the emerging education model has the following characteristic:

1) Openness and timeliness. Resources in network environment can receive the information from various fields, then integrate, absorb and update it. Thus, ideological and political education excesses the boundary of time, space, nation and country in the open environment.
2) Subjectivity and anonymity. Ideological and political education transforms from single one-way knowledge transportation into bilateral interaction communication mode. Educates gradually become the other subject in education relationship. A new interpersonal model in ideological and political education based on network environment open under the effect of dual subjects. In the other hand, visual technology in network technology make the identity information and behavior purpose of netizen no longer transparent. It can bring unexpected effect for ideological and political educator as well as the troubles generated by anonymity of network.
3) Equality and interactivity. Ideological and political education in network environment ensures the equality of educator and educate. Meanwhile, the discussing mode between these two becomes more equal and domestic. Passiveness of traditional education will be replaced by two-way and multi-way communication means, which can provides conditions for the facticity, depth and effectiveness of communication.

* ***Corresponding author's*** e-mail:zjhzhoujiahua@163.com

4) Richness of content. Network can possess huge advantages, a large amount of information resource and various forms of expression through digital storage of information technology, thus further wide the ideological and political education in network environment.

Ideological and political education is extended and expanded under the network environment. Relatively, traditional ideological and political education is single. The media are most newspaper, television, etc. that is filtered by official media; in addition, it adopts inculcation education mode and political-centered education contents. It often neglects the individual difference of students. Ideological and political education in network environment is relatively complex, and has obvious diversification, openness and more frequent interactivity [4].

3 The research status at home and abroad

Ideological and political education is one of characteristics in our country. Corresponding subject does not exist abroad. However, network ideological education which has certain relevance to that is always in an important position for teenagers abroad. The research status of ideological and political education in network environment is shown in Table 1.

TABLE 1 Research status of ideological and political education at home and abroad

Country	Research status
At home	In 2000, Several Suggestions for Strengthening Network Work of College Ideological and Political Education pointed out to combine network technology with ideological and political education and emphasize education work situation IO In 2003, Information Network and Innovation of College Ideological and Political Education written by Hu Jue, Doctor of Law from Tsinghua University proposed that network has become thesis and basic method for new platform of ideological and political education innovation, and put forward feasible practical suggestions IO In 2004, Ministry of Education and The Central Committee of the Communist Young League proposed higher standard and requirements, in order to further enhance college campus network management to deeper combined ideological and political education with network technology IO In 2005, in order to further combined college ideological and political education with network technology, Beijing held national working conference for strengthening and improving ideological and political education of university students, and in the same year, Research on Network Ideological and Political Education edited by Dr. Wei Jifeng marked the initial construction of research on network ideological and political education in out country IO In 2013, Ministry of Education proposed to implement comprehensive deployment on network ideological and political education and campus network management in college.
Abroad	There is no relative subject abroad since ideological and political education is the product of Chinese characteristics. Foreign Countries mainly focus on the network ideological research that has certain relationship with ideological and political education in our country. Network ethics is a part that foreign countries emphasize, consisting of the specific problems generated in network application: definition of network behavior and norm, cross problem generated because of the mutual effect of visualized network and real society, etc.. It covers economy, culture, society, etc.

4 Challenges and opportunities faced by college ideological and political education in network environment

4.1 OPPORTUNITIES BROUGHT BY NETWORK TO COLLEGE IDEOLOGICAL AND POLITICAL EDUCATION WORK

Ideological and political education with new media communication mode, overcomes the drawbacks of traditional media such as small amount of information, obvious time and space limitation and slow update of information. In addition, single one-way inculcation mode of "teacher teaching and student listening" has transformed into more vivid and charming teaching means consisting of words, voice, animation, image, etc. Meanwhile, time and space limitation of traditional education is improved; completely release the constraints of education coverage and education reception and form a global information system with broad region, large amount of information and convenient teaching communication. That is, college education space has not only been restricted within campus but also has realized the extension from internal campus to social open education space.

4.2 CHALLENGES PROPOSED BY NETWORK TO COLLEGE IDEOLOGICAL AND POLITICAL EDUCATION

Every coin has two sides, so does network. Cross combination of network technology and education not only brings opportunities for college ideological and political education, but also put forward challenges in other aspects. Differing from senior school and junior school, teachers from college do not often restrict students, thus it proposes requirements for the self-discipline of students. The quality of education content will be affected by the time university students spent on network, the screening of network information, the restraint ability for bad information in network [6]. First, many university students probably neglect excise, since they are addicted in network. In addition, facing computer screen for long time may lead to diminution of vision. Irregular timetable can cause poor physical reaction such as chronobiology disorders,

insomnia and dreaminess, etc. On the other hand, students with weak self control and strong curiosity may be attracted by the junk information as well as wrong moral and notions in network, which will affect the physical and mental health of university students and lead to bad result; moreover, one of the characteristics of network is anonymity and concealment, thus university students can plays roles different from in daily life in network; this drawback is easy to corrupt the moral character of university students and make them to release speech violate morality and ethics; furthermore, the development of network brings huge changes for the interpersonal communication way, which impact the previous though of university students.

5 Exploration on innovation solution for college ideological and political education in network environment

5.1 INNOVATION OF COLLEGE IDEOLOGICAL AND POLITICAL EDUCATION IDEA

Global and open network platform requires college ideological and political education to establish global and democratic education idea [7,8]. It should be improved in aspects of college and students. On the one hand, college ideological and political education should effectively apply network platform to make education promotion, widen education area and get rid of the constraints of regional consciousness; on the other hand, it requires university students to perform their subjective initiative, autonomously search, analyze, settle and screen network information, accept positive information, resist bad information and finally internalize the positive thought as the individual behavior and quality.

5.2 INNOVATION OF COLLEGE IDEOLOGICAL AND POLITICAL EDUCATION CONTENT

The development of computer network technology can provide technical support for information circulation, update and communication, thus it covers various fields in people's life. Information resources of network become more comprehensive and rich. How college ideological and political education use rich network resources to innovate content should be considered by the reform for college ideological and political education content. Educator obtains the newest message, research achievement as well as the newest ideological and political education content from a large amount of network information resources, and presents them in form of words, music, image, etc. In addition, one of the disadvantages brought by students is that students are too easy to be addicted in network and away from right outlook of right and wrong and mortality; furthermore, university students should establish network legal system consciousness. Therefore, ideological and political education should strengthen students' network psychological education, guide right cultural value and culture good network mentality. Furthermore, university students should establish network legal system consciousness and restrict self behavior while strengthening self-protection.

5.3 INNOVATION OF COLLEGE IDEOLOGICAL AND POLITICAL EDUCATION METHOD

In college ideological and political education, traditional duck-stuffing one-way teaching method is that teachers teach students and students accept teaching content passively, which will block the stimulation of students' enthusiasm and initiative to some extent [7]. Elicitation teaching of college ideological and political education is realized by network education platform. On the one hand, educator digitizes education resources and presents it in forms of words, music, image, animation, etc. Students no longer accept passively. Teaching students becomes more interesting in aspects of students' sense and acceptance mode, thus improve teaching quality and enhance the acceptance degree of students on information. On the other hand, educator can change his position by network platforms such as blog, BBS, etc., and interact with students with more equal means. Educator can understand the thought and opinion making use of anonymity of network. In addition, channel of ideological education can be further widened by network technology, for example, set psychological guidance and online employment guidance services in network platform.

5.4 INNOVATION OF COLLEGE IDEOLOGICAL AND POLITICAL EDUCATION POSITION

Books, newspaper, broadcast and television are the main transmission ways. The development of network technology forms a new way of transmission, and replaces the traditional one to some extent. However, the popularity of network platform conforms to the requirements of age development and provides a new education position. The development of new position for ideological and political education need the extension of network space as well as the expansion and support from ideological and political education function of network mainstream media [7]. On the one hand, establishing ideological and political education website has become one of education approaches. College should strengthen establishment of red thematic website, and introduce rich red culture resource, historic culture, cultural relic of revolutionary spirit and revolutionary remains as teaching materials, in order to attract students, exert a subtle influence on students' character. By doing this, students can impressively feel the lofty ideal, attic faith and strong patriotic spirit, thus to awake their own patriotism and national spirit and better perform the function of ideological and political educa-

tion. On the other hand, establishing college mainstream network media becomes another approach. Currently, the main stream media includes government websites. Government spread and generalize their proposition and voice through mainstream media, thus to guide network consensus. Besides, business websites such as Baidu also has the function of spreading mainstream culture. College can generalize ideological and political education through the education function of mainstream media.

6 Construction of college ideological and political education shared community pattern in network environment

The word "society" first appeared in Community and Society written by the German sociologist Tnnies in 1887. Interpersonal relationship is the starting point of community. The same characteristics of societal community are to possess same value orientation, belief, custom and strong population homogeneity and indicate intimate state [9]. The formation and extension of visual community is realized with the support of network technology. No regional restriction is the biggest characteristics of visual community. Meanwhile, network provides a more convenient platform in the process of contributing and acquiring knowledge. Knowledge sharing is involved and generalized in network environment and form concept of share. Community sharing refers to community established based on common interests and hobbies. The important elements are shared resources, shared values, reciprocal behavior and shared rules. Shared community pattern for college ideological and political education should possess knowledge sharing, life sharing, resource sharing and process sharing.

6.1 ORGANIZATION STRUCTURE OF COLLEGE IDEOLOGICAL AND POLITICAL EDUCATION SHARED COMMUNITY IN NETWORK ENVIRONMENT

Core leadership level, management executive's level and learning community level together form the college ideological and political education sharing community in network environment. Its organization structure is shown in Figure 1:

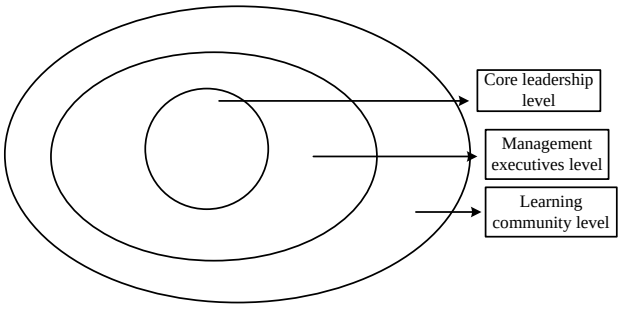


FIGURE 1 Three organization levels of college ideological and political education shared community in network environment

We can see from the Figure 1 that the core leadership level is located in the center, and meanwhile, also is person or organization for defining frontier problem in the whole community. Therefore, staffs in this level have rich experience, highly respected quality. The main function is to formulate the communication content and communication means of ideological and political education in network environment. Management executive's level is in the middle. It is the most important level in shared community, playing the maintenance function of community daily operation. It is composed of information coordinator and network support services staff. Learning community is the outermost. It is the relaying body and basis for shared community, consisting of members with same interests and professional background. They can autonomously discuss topics such as social hotspot, life fun, etc.

6.2 OPERATING MECHANISM OF COLLEGE IDEOLOGICAL AND POLITICAL EDUCATION SHARED COMMUNITY PATTERN IN NETWORK ENVIRONMENT

1) Leadership mechanism: college ideological and political education mechanism under the management, organization and implementation of party committee has certain defect when implementing ideological and political education teaching, management and logistics services, thus the overall efficiency can not be realized. Therefore, we should first innovate leadership mechanism. First, mechanism innovation should perfect party committee and government, school and students. Party committee of colleges and universities formulates overall planning after analyzing ideological and political education situation; schools further combine ideological and political education with other works and implements in various departments. Grassroots party and league organizations focus on organization work. Secondary, mechanism innovation should balance the relationship between organization structure of ideological and political education shared community in the perspective of network characteristics. At last, mechanism innovation should establish and improve level responsibility system, stabilize party cadres team, establish ideological and political education theoretical teacher team and improve structure of student cadre team.
2) Early warning mechanism: schools should understand thought dynamic of student group within the campus as well outside the campus, in order to ensure the successful operation of college ideological and political education shared community pattern and strengthen forward looking and pertinence of college ideological and political education. By doing that, the early warning mechanism for shared community is established.
3) Regulatory mechanism: regulatory mechanism of college ideological and political education shared community should insist the policy that combines heteronomy and self-discipline as well as the policy that pay equal attention to technical monitoring and personnel monitoring. Regulatory mechanism is a purposeful

education practice [9]. Regulation includes heteronomy and self-discipline. Heteronomy is to establish and perfect regulation and rules of shared community and form norm, thus to purify network space and monitor network content. Self-discipline is to culture self consciousness of legal system, sense of duty and network moral consciousness to achieve self-regulation. In addition, technology monitoring is combined with personnel monitoring to develop monitoring software for ideological and political education shared community, supporting with professional monitoring controller. Moreover, monitoring standard for network content should be formulated, and the plans and schemes for shared community should be properly revised, regulated, supplemented and improved combined with practice.

4) Security mechanism: establishment of ideological and political education shared community security should develop in aspects of content, technology, material and environment. The content of ideological and political education looks at the practice of college, and meanwhile, ensures the pertinence and attraction. For example, establish college ideological and political education in the perspective of ideal and faith, patriotism education, basic moral rule, quality-oriented education, etc., set up special ideological and political education channel and strengthen infectivity of college ideological and political education. College ideological and political education should make technical security on health of network content. In addition, certain material basis is needed for ensuring the normal operating of network education. Moreover, positive campus environment, scientific legal environment and effective moral environment are also the security of ideological and political education shared community.

7 Conclusions

Network society is characterized by openness, virtuality, universality, democracy and sharing compared with traditional society. In today's China, network forms complete social structure and function for its peculiarity, enters into fields such as education, industry, agriculture, information society, etc., and changes people's previous life style and thinking pattern in interactive process. Ideological and political education in network environment is a relatively new research field. We should remove the conventional part in traditional concepts, refer and innovate emerging technology, keep pace with the times, positively promote the innovation of college ideological and political education in aspects of concept, content, method and position, thus expand the space, channel and position of ideological and political education [7].

Acknowledgement

Provincial research projects of Suzhou Institute- college students' deficiency of moral responsibility education in the colleges and universities teaching and its strategies (Project number2012 jyxm542).

References

[1] Zhang J 2011 Research on Ideological and Political Education Innovation in Internet Environment *Sichuan Chengdu University of Technology* (*in Chinese*)
[2] Zhang Z, Wen S Z 2013 Though, Morality and Self-cultivation: Analysis on University and College Moral Education *Journal of Hebei University of Science and Technology (Social Sciences)* (6) 130 (*in Chinese*)
[3] He C 2014 Exploration of College Ideological and Political Education Theory in Internet Environment *School Party Construction and Ideological Education* (12) 80-2 (*in Chinese*)
[4] Wang H 2012 Path Choice of College Ideological and Political Education Innovation in Internet Environment *Academic Journal of Zhongzhou* (5) 78-9 (*in Chinese*)
[5] Zhao H 2013 Research on Ideological and Political Education Innovation in Internet Environment *Hebei: Hebei University of Science and Technology* (*in Chinese*)
[6] Mao Y 2012 Problems and Solutions Faced by College Ideological and Political Education in Internet Environment *Examinations* (15) 8 (*in Chinese*)
[7] Yan X, Ning J, Jia Y, Gou Z 2013 Discussion on College Ideological and Political Education Innovation in Internet Environment *China Newspaper Industry* (4) 30-1 (*in Chinese*)
[8] Xu Y 2012 A New Idea for the Teaching Pattern of Political Course in Universities in Network Situation *China Adult Education* (12) 140-2 (*in Chinese*)
[9] Ji H 2013 The study of Ideological and Political Education in Colleges and Universities in the New Media Era *Jiangsu Najing Normal University* (*in Chinese*)

Author

Jiahua Zhou, 10. 1976, Anhui Province of China.

Current position, grades: associate professor.
University studies: BSc of history, Anhui University in 2000, MSc of historical theory and historiography, the School of History and Culture, Shandong University in 2008.
Scientific interest: university culture and regional economic and social development.

Management on the recombination of manufacturing whole process logistics operation based on MES

Juan Zhang[*1], Guanghui Chen[1], Boqin Liu[2]

[1]*Department of Business Management, Xi'an Technology University, Xi'an, 710032, China*

[2]*The unit of 69029*

Received 1 October 2014, www.cmnt.lv

Abstract

With the constant intensification of market competition, manufacturing enterprises are racking their minds to lower the cost so as to obtain more profit. As the enterprise's third profit source, logistics has attached more and more enterprises' attention. According to the relevant data statistics, in the current production system, trucking expense accounts for 30% to 75% of total expense. However, after optimal management of logistics, it can save 15% to 30% of expense for enterprises. This paper discussed the current situation and research results of the workshop logistics of machine manufacturing, and analyzed the loopholes existed in the whole process of traditional logistics management of manufacturing. It also established a kind of manufacturing whole process logistics lean management operation model based on MES, which provided a kind of feasible management mode for machinery industry to lower the logistics cost of the whole manufacturing process.

Keywords: MES, manufacturing whole process, logistics lean management, operation model

1 Introduction

As we all know, the main aim of enterprise manufacturing is to make a profit, and one of the methods of obtaining profit is to lower production cost. However, with the continuous innovation of science and technology, and after the sufficient excavation of the first profit source and the second profit source, people gradually realized that the cost saving of materialized labor and direct labor seems to have come to the end [1]. Therefore, in recent years, not only the enterprise external logistics (supply chain management), but also the enterprise internal logistics (workshop logistics) have become research hotspots. In the article Research on the Operation Mode and Mechanism of Logistics Network System Based on Dissipative Network [2], Shan Lihui of Beijing Jiaotong University started from the entirety of logistics system and conducted recognition on logistics network from systematic and macro perspective. Through researches on the dissipative structure and entropy model of logistics network, she analyzed that resources integration, dynamic alliance and collaborative optimization are beneficial to the coordinated operation of logistics network. In addition, the operation mode of logistics network was established based on it, which optimally utilized the limited logistics resource and provided brand new exploration and theoretical security. In the article A Study on the Collaborative Operation of Logistics Service Supply Chain Based on Process Management [3], based on relevant theory commentaries, Li Yibin, Dong Qianli and Sun Haojie conducted research on the logistics service supply chain coordination with the idea of business process management and modeling approach of workflow for reference. They pointed out that the logistics service supply chain coordination should base on the perspective of process optimization and take the coordination of technology and management into consideration, thus to promote the performance of logistics service supply chain coordination. In the article Exploration and Analysis on the Logistics Process of Guangdong Donghong Metal Equipment Manufacturing Enterprise [4], Huang Yubin and Qian Yuanfang conducted field research on Guangdong Donghong Metal Equipment Manufacturing Company. Through the understanding of the company logistics operating process, they combined the theory of enterprise logistics management and found some problems existed in the company. They also combined the theory of logistics process reengineering, which provided a good referential experience for the remodeling of enterprise logistics process and the operation of other enterprise logistics management models in future.

This paper established a kind of manufacturing whole process logistics lean management operation model based on MES, which conducted redefinition on the department responsibility of manufacturing enterprise and recombination on the logistics operation process of the whole manufacturing process. In addition, this paper described the support of MES and other relevant information system to the logistics lean management operation mode of the whole manufacturing process. It also further perfected and managed enterprise logistics in technology level so as to promote the enterprise economic benefit, which has important economic and practical significance.

[*] ***Correponding author's*** e-mail: zjjuanz@163.com

2 Traditional logistics management problems

Manufacturing whole process logistics management refers that after the manufacturing enterprise accepts productive task execution instruction, the management on the dispatching, carrying, loading and unloading, and storing of materials and half-finished products during the process of starting from the recipient of raw materials, through the process and assembly of intermediate products and finally the production of qualified finished products.

With the increasingly fierce market competition, manufacturing enterprises in China are increasingly highlighting the optimal management of logistics. However, in recent years, most of manufacturing enterprises focus on the optimal management of enterprise external logistics (that is supply chain management), but neglect the enterprise internal logistics optimal management (that is manufacturing whole process logistics management). Therefore, manufacturing whole process logistics management has become the weak link among most of manufacturing enterprises in China. At present, under the circumstances of non-transparent workshop logistics information, the typical logistics management process of manufacturing enterprise is shown as follows:

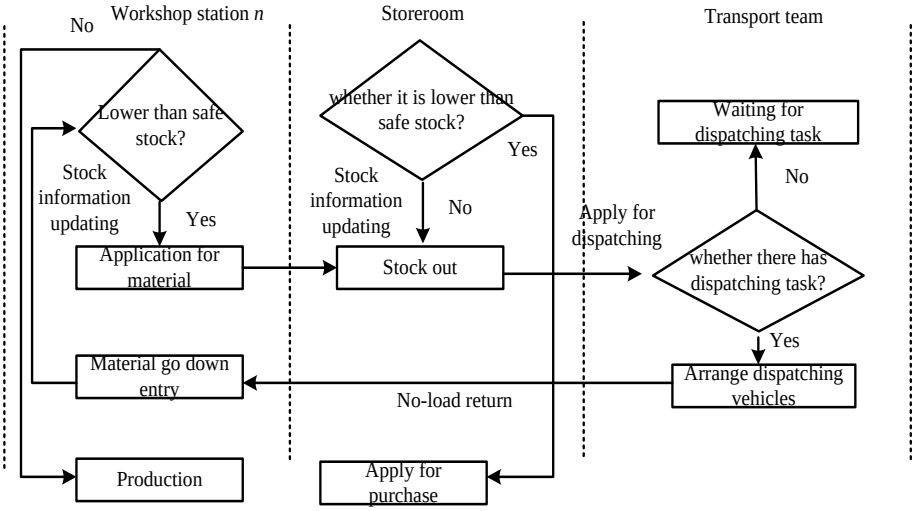


FIGURE 1 Manufacturing whole process traditional logistics management process

From the above Figure 1, the disadvantages of traditional logistical management can be summarized as follows: waste of space, large occupation of material capital; high no-load ratio of logistics transportation; repeated handling; unreasonable logistics path; untimely waste debris recycling.

3 Logistics lean management mode

3.1 MANUFACTURING EXECUTION SYSTEM (MES)

MES is the executive layer between the operation control system (SFC) of planning player (ERP) and workshop layer, which is mainly responsible for the production management and scheduling execution of production site. MES can transmit the management information in management layer like production plan, production instruction, etc and production instruction information like working diagram, process flow diagram, etc to lower levels of each station of production site. At the same time, MES also can real time collect and process the production data of production site, and submit them to ERP [5]. Therefore, manufacturing enterprises can realize the visual and digital management on the manufacturing whole process by implementing MES, which can provide timely and comprehensive information support for the manufacturing whole process logistics management.

3.2 LOGISTICS LEAN MANAGEMENT

As for manufacturing enterprise, in products total cost, the logistics cost of manufacturing whole process often occupies a large proportion. Therefore, the realization of manufacturing whole process logistics lean management is one of the core links of enterprise to reduce cost and improve market competitiveness. Manufacturing execution system can provide all static and dynamic data of manufacturing whole process for enterprises, a set of real time comprehensive logistics information and abundant data foundation for the logistics tracing, statistics, analysis and optimazation of manufacturing whole process [6,7]. At present, MES has obtained more good graces of Chinese manufacturing enterprises as well as enough reusing. Therefore, this paper proposed a kind of logistics lean management mode based on MES. The main content of manufacturing whole process logistics lean management based on MES is "material storage" and "material flow". Thus the purpose of manufacturing whole process logistics lean management based on MES is to realize the stock lean management and dispatching lean management.

4 Operation model

4.1 TOTAL FLOW DIAGRAM

Manufacturing whole process logistics lean management mode based on MES not only needs to realize the logistics lean management to manufacturing whole process, but also needs to perform recombination on traditional manufacturing whole process logistics management process. The recombined process will make full use of the powerful logistics real time information in MES, so as to maximize embody the value of real time logistics information in MES. The specific operation mode is shown as follows (Figure 2):

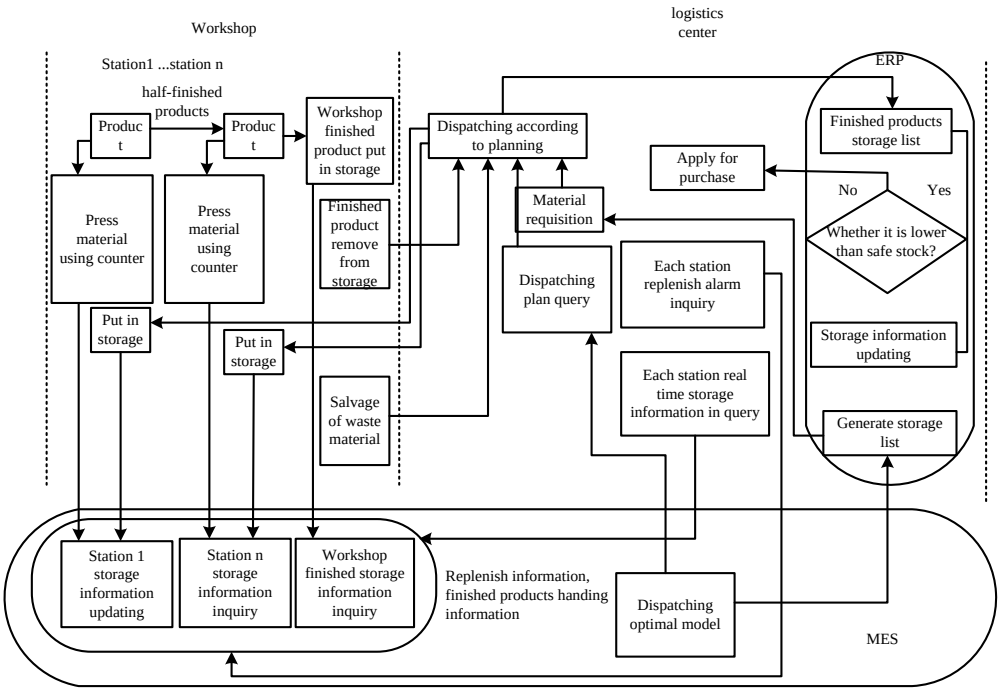


FIGURE 2 Logistics lean management operation mode based on MES

4.2 INFORMATION SUPPORT SYSTEM

Manufacturing whole process logistics lean management based on MES can not leave the enterprise informatization supporting environment. Besides MES system, it also needs ERP system of MES integration and SFC system of control layer. The information supporting system designed based on the above three systems are shown as follows in Figure 3:

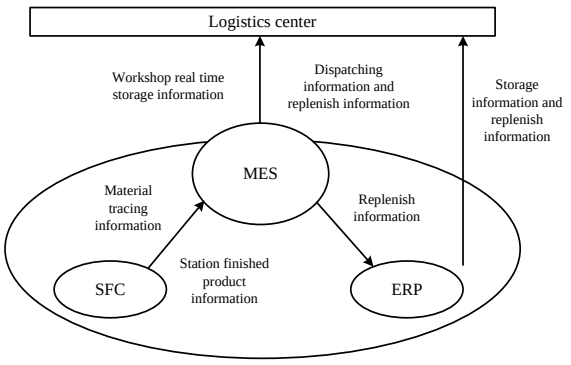


FIGURE 3 The informatization supporting environment of logistics lean management

Conclusion

Modern enterprise is a complex and large scale system. The generation of logistics is to meet the demand for purchase, production and sales departments, which is the inter communication bridge among various subsystems. It plays an important role in enhancing the quality of products and reducing production cost. Therefore, an increasing number of enterprises turn their attention to the enterprise logistics optimization. As a kind of advanced organization mode and management technology, modern logistics has become the third enterprise profit source of creation besides reducing material consumption, raising labor productivity. In terms of essence, the optimization of logistics management model is an important strategic decision for enterprises in coping with increasingly fierce market competition [8,9]. Base on the analysis of the shortages of traditional logistics model, this paper proposed the manufacturing whole process logistics lean management operation mode based on MES, thus to have the enterprise MES plays terrific value in logistics management. Its assistance in reducing production cost has better maintained enterprise profit, and created superior environment for further driving the economic development of society.

References

[1] Hu J, Liu Y 2010 Research on the Application of RFID-based MES in Dispersing Manufacture Enterprise *Machine Design and Manufacturing Engineering* 39(1) 34-6
[2] Shan L 2012 Research on the Operation Mode and Mechanism of Logistics Network System Based on Dissipative Network *Beijing Jiaotong University* 6 (*in Chinese*)
[3] Li Y, Dong Q, Sun H 2012 A Study on the Collaborative Operation of Logistics Service Supply Chain Based on Process Management *Logistics Technology* 31(5) 174-7 (*in Chinese*)
[4] Huang Y, Qian Y 2013 Exploration and Analysis on the Logistics Process of Guangdong Donghong Metal Equipment Manufacturing Enterprise *Science & Technology Economy Market* (6) 54-5 (*in Chinese*)
[5] Cui L, Hertz S 2011 Networks and Capabilities as Characteristics of Logistics Firms *Industrial Marketing Management* 40(6) 1004-11
[6] Bratukhin A 2011 *IEEE Transactions on Industrial Informatics* 7(4) 740-2
[7] Ji L, Yuan H, Ye S, et al 2012 Research and Application of MES in Logistics Laboratory *Research and Exploration in Laboratory* 31(5) 180-2
[8] Zhuang S, Ge Y 2012 Design, Implementation and Application of Management and Control Integrative System Based on RFID *China Mechanical Engineering* 23(13) 1551-4
[9] Wu W, Yang J 2012 Research on MES Data Collection of Manufacturing Execution System Based on OPC Technology *Machine Building & Automation* (4) 137-9

Authors

Juan Zhang, born in January 1977, Shaanxi Province of China.

Current position, grades: lecturer.
University studies: MSc of management science and engineering, Xidian University in 2006.
Scientific interest: administrative decision, logistics management.

Guanghui Chen, Sichuan Province of China.

Current position, grades: associate professor.
University studies: BSc of computer application, Xi'an Technology University in 2009.
Scientific interest: logistics management.

Baoqin Liu, born in August 1977, Shaanxi Province of China.

Current position, grades: engineer.
University studies: MSc of CIS, Dalian University of Technology in 2011.
Scientific interest: electronic information system management.

Design and construction of college English teaching model based on multi-dimensional integrated technology

Jingdong Hao*

Suzhou University, Anhui, 234000, China

Received 1 October 2014, www.cmnt.lv

Abstract

How to enhance language teaching efficiency, especially college English is always the focus of people. People pay more and more attention to the progress of the CAT with the development of linguistics and improvement of computer and network technology, at the same time, the combination of language teaching and computer science is also a major breakthrough in the history of the development of language teaching. This paper used the C3I-CAT integrated technology into college English curriculum design and research and development of independent study system, and developed more college English courseware and application methods that is matching to the need of teaching practice, thus to improve the learning efficiency of college English course.

Keywords: college English course, information technology, integration, design of model

1 Introduction

With more and more common application of information technology nowadays, education information technology represented by computer science and communications technology is playing an increasingly important role and influence on Chinese foreign language teaching [1]. Information technology and foreign language teaching are belonging to two different subject fields, and they have entirely different academic tradition and paradigm. On the other hand, interdisciplinary conformity is becoming a tendency of scientific development in 21st century. Under this macro-background, more and more scholars devote themselves into the study. Jiasheng Hu from Shanghai International Studies University used intelligent research [2] methods based on metatheory basis to prove some rationality and validity of microscopic view and infrastructure in Research on Construction of Foreign Language Education Technology Based on Paradigm Shift. He tentatively finished the structure of foreign language education technology based on normal form transformation. It became a subject rudiment in reality from an existence in theory. Thus the problem of disciplinary development and teaching actual effect of Chinese foreign language teaching were solved effectively. In the article of Integration of Information Technology and College English Course, Shengnan Zhang studied the significance of integration of modern information technology and college English teaching and discussed the strategy for the integration [3], which provides some feasibility idea for the further integration of college English teaching and modern information technology. Analysis of College Education in Network Times [4] wrote by Feng Cao discussed about the contradiction and countermeasure between network and college education and the opportunity of college given by network, centering on the theme of how the college education seize the opportunity and overcome the challenges in network times, which provides a reference for correctly handling the relationship between network and college education.

2 Significance of integration of college English course and information technology

With the development of Information technology and network technology, information technology provides more space for university English teaching. The combination of information technology and English lesson is not just a fixed model, but a concept. Teacher should catch the point of the lesson, and express it in efficient way. Same text book, different background and different teacher and student can give teaching and study different life [5,6]. The English teaching will become efficient, if combine the information technology and university English. In the aspect of information technology and college English course integration teaching, our country accumulate full of experience. However, in the process of teaching, many problems come out. The researchers should focus on the problem which appear in the combination and discuss the strategy in the combination of information technology and university English, to provide reference for the further development integration of information technology and college English course.

* ***Corresponding author's*** e-mail: hjdszxy@163.com

3 The application of C3i-Cat integrated technology in college English course

3.1 THE THEORETICAL FOUNDATION OF C3I-CAT INTEGRATED TECHNOLOGY

In the perspective of the cognitive psychology, learning about new knowledge should accord with the mechanism of human brain to accept and process information, and this mechanism tells us the way of information acquisition and storage in our human brain. The grasp of a new skill need three stages: acquisition, combination and spontaneous stage. The best way to get the declarative knowledge is to build relevance with past knowledge, but the best way to get procedural knowledge is to practice with systematic and suggestive. The cognitive theories show that, it is a proactive involvement process of people in a series of activities from incipient stage to enter a certain state and at last to get the new information.

We can use the C3I-CAT integrated technology in teaching and learning of college English based on this theory and regard the students as the subject and teachers as the leading. We should make full use of the practical type courses and the characters of general education class. This course assessment used the form of " PPT exhibition of research result in groups+ course work+ communication between 100 people about the note+ work+ classroom discussion+ final", that is, "3+2+1" appraisal mode, in order to encourage the students to proceed the course practice. The aim is to examine the master degree and application level of communication skill of students.

3.2 PATTERN DESIGN OF COLLEGE ENGLISH CURRICULUM BASED ON C3I-CAT INTEGRATED TECHNOLOGY

There have three ways to improve the college English teaching mode by using C3I-CAT integrated technology based on the study of cognitive theories, they are:

Strategy teaching model: this model thinks the objective is the core of learning. New information should establish relevance with the past knowledge, and learning needs the organization of knowledge. C3I-CAT integrated technology is to organize the effective resource of Computer-College-class and Internet, etc. The students are helped to gain new knowledge by the way of computer-assisted instruction, and they can make a correlation between new information and past knowledge.

Language study model of cognitive theories: the aim of this model is to gain procedural knowledge and use it to prompt the study of related discipline, and its applicative object is learners with middle rank of English. It contains three aspects, that is, theme and content, skills development and policy guidance, and they all can be realized by C3I-CAT integrated technology.

Cognitive strategy and society/ emotion strategies: modern multi-media and network communication tools such as MSN, QQ, BBS, E-mail and Blog can be used to improve the learning efficiency of students supported by C3I-CAT integrated technology.

3.3 DEVELOPMENT OF COLLEGE ENGLISH COURSE DESIGN AND AUTONOMOUS LEARNING SYSTEM BASED ON C3I-CAT INTEGRATED TECHNOLOGY

According to the spirit of College English Curriculum Requirements and Information of Launching College English Teaching Reform Pilot Work, we should train students' comprehensive proficiency in applying English, especially their listening and speaking abilities. It can make them perform effective verbal and written information exchange in English in the future work and social communication. At the same time, it can strengthen their autonomous learning ability, improve their comprehensive cultural literacy and make them adapt to the requirement of our social development and international communication. College English teaching reform of Henan University of Science and Technology implemented step by step according to phases like "beforehand pilot, gradual popularization, overall development", etc. Three experimental classes were established in class of 2010. According to experimental effect, they were gradually improved and promoted. At present, they have already developed overall among students of 2012 grade. The present new curriculum design is exactly based on C3I-CAT integrated technology and adopting new teaching mode. It has gradually improved the traditional teaching mode that centered on teaching, established the mode that centered on students' autonomous learning, which integrated the brand new teaching mode that the multimedia network teaching of C3I-CAT integrated technology combined with classroom teaching.

The realization of effective share among valid resources in various aspects between computer-college-class and Internet has effectively solved problems like failure of students' listening and speaking ability and "difficult to talk". Schools are trying their best to create learning time on the net for students. Students conduct self-study and discussion on the net, and teacher conducts mentoring, tutoring and correcting of students' homework on the net. In addition, teachers also guide students to carry out the preparation and representation of multimedia courseware work. It can realize online test, require students to finish the exercise and test prescribed online on time, and conduct duly record as curriculum evaluation basis.

4 Problems that C3i-Cat Integrated technology solve in college English course

English course teaching in colleges and universities in our country exists many problems because of the influence of Chinese traditional English teaching concept means. These problems seriously affect the development and progress of college English course teaching. Overall, the problems exists in college English course teaching in our country reflect on the following aspects. First, the concept and pattern of college English course falls behind. Many college English teachers have misunderstanding on English course teaching concept. They think that, English teaching is only limited within the class, thus neglect the importance

of English extracurricular teaching [7,8]. What's more, teachers may think that English teaching is only limited in teaching material, thus neglect the application of other English material. Second, students are lack of interest in English learning. Teaching forms adopted by college English teachers so far are always cramming education and duck-stuffing. This kind of teaching form neglects the individual difference of students, killing the enthusiasm and intuitive of English learning of students. Third, college English course teaching and evaluation system is not perfect. It is necessary to improve relative teaching evaluation system to improve college English course teaching quality, which can better promote teaching skills and level of teachers themselves. However, the establishment of English course teaching evaluation system is not perfect, thus the English course teaching quality and level of teachers is difficult to improve. It is also an important problem existing in college English course in our country.

However, this paper organically combined C3I-CAT integrated technology and college English teaching course and effectively solved the above problems. Therefore, students can learn college English more effectively under the help of C3I-CAT integrated technology, exchange in oral or written information by English in future job and social interaction, strengthen autonomous learning ability, enhance comprehensive cultural quality and adopt to the need of social development and international exchange in our country.

5 Conclusions

Bruner, American famous psychologist, said that leaning was an active process, and the best way to attract students is to let students actively and autonomously involve in study, and realize they are able to deal with the outer world. There is no doubt that language learning is a complex cognitive behavioral process. In addition, the control of all the knowledge and skills must depend on the psychological experience of the individual learner. The cognition assimilation theory can help the English teachers to achieve the individual guidance, and create an effective, close and fast contact with the English learners, thus to improve the efficiency of college English teaching.

Cognitive learning theory held that there is a thinking process behind human behavior, and based on this hypothesis, they thought that the behavior change was observable; meanwhile we can infer the learner's inner heart activities by the changing behavior. Good technology and new media means play an important role in human cognitive activities [9]. We used the latest research results in language teaching theory and cognitive science, and explored new approach of college English course design based on C3I-CAT integrated technology, autonomous learning system development, and manufacture and application research of college English courseware.

References

[1] Li D 2010 Exploration on the Integration of Information Technology and Tourism English Teaching *China Science and Technology Information* (19) 238-9 (*in Chinese*)
[2] Hu J S 2011 Research on the Construction of Foreign Language Educational Technology Discipline *Shanghai: Shanghai International Studies Universities* 11 (*in Chinese*)
[3] Zhang S N 2013 Discussion on the Integration of Information Technology and University English Curriculum *Science and Technology Consulting Herald* (1) 176 (*in Chinese*)
[4] Cao F 2012 Elementary Analysis on University Education in the Internet Age *China Education Innovation Herald* (34) 173-4 (*in Chinese*)
[5] Nan G N 2010 The New Development of the Theory on the Research of Educational Technology *E-education Research* (01) 8-10 (*in Chinese*)
[6] Liu C Y 2010 The Application of Design Research in On-line English Writing Platform *Distance Education Journal* (5) 74-8 (*in Chinese*)
[7] Hu J S, Feng Q L, Li Y 2010 An Analysis to the Role of Information Technology Integrated into Foreign Language Curriculum *The Modern Education Technology* (12) 72-7 (*in Chinese*)
[8] Fu R 2010 Reflection on Foreign Language in Digital Environment *Journal of Sichuan International Studies Universities* (2) 127-30 (*in Chinese*)
[9] Shen Y Z 2011 The Application of Modern Multimedia Technology to Improve the Level of University English Teaching *Commercial Culture (Second Half)* (7) 198 (*in Chinese*)

Author

Jingdong Hao, born in 1973, Anhui Province of China.

Current position, grades: lecturer.
University studies: Master's degree of English Linguistics, Anhui University in 2007.
Scientific interest: foreign language teaching and research.

Comprehensive evaluation of education mode of university English model and swot analysis

Ying Bai*

Longdong University, Gansu, 745000, China

Received 1October 2014, www.cmnt.lv

Abstract

College English is a compulsory course for non-English majors, which is an important public class. The pros and cons of university English education are directly related to the improvement of comprehensive quality of college students. What the specific requirements of implementing quality education for college English education and how to further strengthen and improve the university English education are hot topics and have been widely concerned. This paper analyzes the traditional mode of university English education quality, and set up a comprehensive evaluation mode of education model, so as to provide some help for the university English reform.

Keywords: university English education, character building education, analysis of advantages and disadvantages, comprehensive evaluation

1 Introduction

Knowledge economy means knowledge producing, processing, dissemination and application which is based on modern science and technology, and determines the cultivation and development of people's innovative ability. It is very important to cultivate people to control and apply the information and knowledge. The coming of knowledge economy will promote the internationalization of Education. National power competition is becoming increasingly fierce shows that the strength of national will depend on the quality of workers, and depend on the quality and quantity of the talents [1]. The college English plays a decisive role in the knowledge economy. It is a basic course for the university personnel training, which is the premise to decide personnel quality; therefore, more and more scholars focus their attention on the university English educational model. In the article of the Research of University Curriculum Factors of Autonomous English Promotion Learning [2], Gu Shimin, from Shanghai International Studies, conducted a systematic research on the involved curriculum elements of Requirements of Teaching University English Curriculum, and reflected on the teaching principle of Requirements of Teaching University English Curriculum. He also put forward some suggestions on how to meet the individualized learning needs in the university English teaching and effectively promote the occurrence of autonomous learning. In [3], Kong Deliang Luan Shuwen constructed the structure model and the practical model of cross culture teaching university English, and aimed at enhancing cross-cultural differences in sensitivity, tolerance and flexibility of the treatment of culture in College English teaching, and expanding university cultural influence, and enhancing the cultural soft power in Universities. Wang Shouren Reflected on College English Teaching [4]. In the new era, college English teaching should stick to the right reform direction, and continue improve and enhance college English teaching; furthermore, it is the mission and task for college to construct university English curriculum system, and perfect the university CET4 CET6 test, and create high-quality professional university English teachers, and improve teaching quality. In addition, create a new situation in university English teaching.

In this paper, the author made a detail analysis of the advantages and disadvantages of the traditional mode of university English Education, and put forward the specific college English teaching requirements, mapped out the reform direction to some extent, and established a comprehensive evaluation model of comprehensive and objective assessment of the English education mode.

2 Analysis of the traditional model of English education quality

The traditional English teaching experienced variability for several decades and gradually formed a set of intrinsic mode. The long existence of this kind of education mode must have its advantages, but with the innovative reform of teaching mode, its malpractice also revealed beyond doubt.

The advantages of traditional education:

1) Pay attention to the students' knowledge and skills training.
2) The student fixed grouping according to teaching plan, clear learning objectives, schedule orderly and controllable.

* ***Corresponding author's*** e-mail: byyingb@163.com

3) The course of teacher's teaching, easy to play, good interactivity.

The disadvantages of traditional education:

1) Monotonous education environment and learning environment, dull, formulaic cramming education process, spoon feeding education
2) Teacher centered, more explanation, less language practicing, little chance for communication.
3) Focus on reading, ignoring the oral practicing.
4) The language is scattered grammar, vocabulary, idioms and other language points, contrast analysis.

3 The specific requirements for college English education after implementing quality education

The Decision on Deepening Education Reform and Advancing Quality Education by The CPC Central Committee and State Council pointed out the direction for the construction of socialist education system in twenty-first Century has China characteristics, and put forward principle requirements on implementing quality education. According to the "decision" requirement, there are at least four points on implementing the requirements for the university education of English quality education, the four points are as the following:

1) For all students.

As college English is a basic course in university and also a public course, each undergraduate has to learn it for a long time. It has a broad coverage and, far-reaching impact, which other subjects are not comparable [5]. The key is whether the education is truly for all students; whether all students can receive education and all the students' English level can be really improved.

2) Adding the moral education in college education.

The quality education must include aspects of education such as moral education, intellectual education, physical education and aesthetic education, and integrated them in the different parts in education activities. All types of schools at all levels should pay more attention to moral education, and strengthen the ability of moral education. We also need to carry forward the Chinese excellent traditions and combine it with all outstanding civilization achievement of the world. In order to enhance the student's education of strengthening the correct world outlook, life outlook education and ethics and civilization, we must study the world outstanding civilization achievements, and only do this can the students have a broad horizon and have a deep understanding of our country condition, in addition, they can strengthen the patriotism, collectivism and socialism feelings.

3) Emphasize the cultivation of students' quality and ability.

The implementation of quality education requires colleges to cultivate the students' professional quality, scientific and cultural quality and ability [6]. The college English includes intensive reading, extensive reading, fast reading, listening, grammar and exercises. Each course should emphasize the cultivation of the student's quality and ability, such as the structure of intensive reading exercises can improve the students' ability of expression. The reading exercises can improve the students' reading skills, and the extensive reading covers a range of subjects, new topics and genre diversity. It can improve students' scientific and cultural quality because of its knowledge and interest. Grammar exercises enable students to apply the grammar knowledge into language practice, and improve the ability to use grammar knowledge on discourse level. Listening can enhance the understanding and proficiency at discourse level and improve students' analysis, induction, and comprehensive and inference ability. Fast reading can improve reading ability and speed. The above part not only has its own system but complement each other, which can lay a wide, thick and solid foundation for students and constitute the comprehensive ability and quality of students.

4) Emphasize the cultivation of scientific spirit and innovation spirit.

There are a large number of relevant scientific and cultural knowledge article, and the teacher should explain these articles, which can enhance the dialectical materialism and historical materialism education, and cultivate science philosophy and spirit of exploration and innovation. In addition, it can integrate the scientific knowledge, scientific thinking, scientific method and scientific spirit, and become a kind of strength power.

4 A fuzzy comprehensive evaluation model of college English

4.1 CONSTRUCTION OF THE EVALUATION INDEX SYSTEM OF COLLEGE ENGLISH EDUCATION

This paper selected the learning attitude, the degree of interaction, resource use and test scores as the four evaluation model of university English education dimension according to the characteristics of autonomous learning university English education and university English. It adopted the analytic hierarchy process to determine the weight and used the current learning assessment commonly used in excellent, good, pass, fail these five grade evaluation to establish evaluation index system. The specific content is as following (Table 1):

TABLE 1 Evaluation index system of English education in a university

First class indexUi	Second class indexUij	The evaluation factors
Learning attitude	Complete the task of learning according to a predetermined course	Browse Webpage learning; courseware learning; complete homework etc.
	Actively participated in Online teaching activity organized by the teacher	Put forward the question; comment; watch video; listen to the audio; Participate in answering; taking the test;
	Willing to interact with other learning partner	Put forward questions; answer the question; asynchronous communication; send E-mail etc.
	Able to finish challenging task	The proposed scheme; works; teachers' evaluations of student; provide resources etc.
	Complete autonomy learning notes and class record	Electronic notes; record lectures etc.
	To actively participate in the course of the necessary counseling	The time of participating in counseling, participating times
	To submit coursework in a timely manner, do not do things carelessly	Submission time and the operation results
	No cheating in the homework exam and quiz	Results students comment;
Cooperation	In answering the students often ask questions to the teacher	Participate in answering questions; times; browsing answering database etc.
	Published course related comments and opinions in the discussion areas	Answer questions, such as the number of published articles
	To find out the important theme in the other people's point of view, to extract the useful information, form their own point of view	Published articles; BBS statement; teachers and students evaluation
	Often exchange learning problems with teachers or students	Chat room floor, E-mail communication
	Often put forward construction suggestions to teaching education work	The times of giving advice
	Timely response to the teacher's questions	Respond time and respond times
	Actively finish the task with other learning partners	Participate in the discussion of the number of proposed solutions etc.
Resource utilization	Make full use of system platform	Log on to the system platform for learning time
	Upload valuable resources on the system platform	Upload resources and provide the link address
	Do some learning records in a notebook	Electronic notes, notes, tag, summarize the study, reflection
	Effectively choose all kinds of resources for learning	Webpage, document browsing; the use of video, audio resources; information search
	To be able to use the knowledge to solve practical problems	Work, works; participate in the competition; problem solving etc.
Papers	Examination	Multiple choice questions 25%
		Word and phrase dictation 10%
		Grammar and structure 10%
		Sentence repetition 10%
		Essay writing 15%
		Fast reading 10%
		Deep reading 20%
	Homework	Teachers score for the job
	Test	Test results

4.2 A FUZZY COMPREHENSIVE EVALUATION MODEL OF UNIVERSITY ENGLISH EDUCATION

1) The establishment of evaluation index factor set and evaluation set.

Set the evaluation index for U, $U=(u_1,u_2,...,u_n)$, u_i is the evaluation index.

Define the first class index set $U=(u_1,u_2,u_3,u_4)$, which is equal to the utilization of resources, cooperation, learning attitude, learning achievement.

The corresponding weight set $A=(a_1,a_2,a_3,a_4)$. a_i (i = 1,2,3,4) indicates that the index weight in the total scores of the students in U, $\sum a_i=1$.

Define the second level of index set $U=(u_{i1},u_{i2},.....)$, which is equal to (complete the learning task according to a predetermined course).

The corresponding weight set $a_i=(a_{i1},a_{i2}.....)$. a_{ij} (i = 1,2,3,4), indicates the weight index in total score of the students.

Define the fuzzy evaluation set $V=(v_1,v_2,v_3,v_4,v_5)$, each parameter of the fuzzy sets is evaluated by the score level assessment.

2) Determination of membership function of fuzzy judgment matrix.

Students ranked hypothesis relation and the actual level of meet the normal distribution, and this normal distribution has been widely applied in practical application of education evaluation. In view of the entire university English class ranking, the student's ranked sequence is

divided into five: (0%~10%], (10%~30%], (30%~60%], (60%~90%], (90%~100%]. We use fuzzy membership concept and expert experience value to determine a student's ranking in the (possibility of localization of 10%~30%] 0.6, which falls on (0%~10%] and (the possibility of 30%~60%] each of them is o.2. Thus, the students in the real class ranking level of the available vector can be expressed by r = [0.2 0.6 0.2 0 0], while the fuzzy evaluation index of matrix U $R=[r_1,r_2,...,r_n]^T$. N is the number of the U evaluation index.

5 Conclusions

At present, college English education should be further strengthened and improved, and make it play a greater role in implementing quality education. Firstly, we need to improve our consciousness, transfer the concept, and eliminate obstacles of university English education, and establish the confidence of improving college English teaching. Secondly, we should solve university English education problems emphatically, and take effective measures so as to ensure the university English education along the healthy track [7-9]. This paper started from the traditional mode of university English education quality analysis and combined China's current education reform content as reference index, thus clear the improvement direction of education pattern. In addition, it constructed the reference for the further improvement of future English teaching mode which is based on the fuzzy theory and promotes comprehensive construction evaluation model and the English education career.

References

[1] Zhang P 2010 Research on Collaborative Learning Evaluation Model under the Network Environment *Sci-Tech Information Development & Ec* (34) 102-3 (*in Chinese*)
[2] Gu S M 2013 The Promotion of Curriculum Study Factors in College English Autonomous Learning *Shanghai International Studies University* (*in Chinese*)
[3] Kong D, Luan S W 2012 Construction of Modes for Cross-Cultural College English Teaching-The Status quo and Theoretical Reflection *Foreign Language World* (2) 17-26 (*in Chinese*)
[4] Wang S R 2011 Some Thoughts on College English Teaching in China *Theory and Practice of Foreign Language Teaching* (1) 1-5 (*in Chinese*)
[5] Tang Z X 2010 Perspective on the Phenomenon of Inefficient English Teaching and Teaching Management (6) 35-6 (*in Chinese*)
[6] Li Y M 2010 Some Thoughts on Chinese Language Planning *Journal of Foreign Languages* (1) 2-8 (*in Chinese*)
[7] Yang X Y 2010 A Study on the Cultural Introduction in College English Teaching *Journal of Sichuan International Studies University* (4) 127-30 (*in Chinese*)
[8] Ma H M,Xu S C, Zhao K J 2013 Research on the Strategy of the Complementary Advantages of the Traditional Teaching Mode and Teaching Mode of Interactive Network University English Teaching *Journal of Qinghai University* 31(2) 85-8 (*in Chinese*)
[9] Zheng S T 2011 The Application in the Formation of College English System of Network Autonomous Learning *Journal of Southeast University* (6) 232-4 (*in Chinese*)

Author

Ying Bai, 1966, Gansu Province of China.

Current position, grades: associate professor.
University studies: Master's degree of English education, Northwest Normal University in 2003.
Scientific interest: English education and college English teaching.

Research of golden week's tourism flow based on tourism system model visual

Yan Zhang*

Shangluo University, Shaanxi, 726000, China

Received 1 October 2014, www.cmnt.lv

Abstract

Travel system is a set of various travel issues; some negative effect of tourism factors in the overall tourism system will affect the tourists' decision. Since the "golden week" holiday, the growing numbers of residents traveller, form a tourist flow for a period of time, while promote domestic economic growth, accompanied by some negative effects, such as poor quality of "golden week" travel, poor tourist satisfaction, cause certain environmental pollution and so on, all aspects of the social concern is caused. This article make Huangshan scenic spot and the ancient villages in southern Anhui province as an example, based on tourism system vision, on the basis of the tourism golden week to explore, set up and analyze the "golden week" tourism flow system model. Classified tourism flow mobility factors by using the model, analyzing influence factor of golden week tourism, according to affect the positive and negative effects of tourism flows, come up with some suggestions and improvement measures, in order to improve the management and planning of tourism, meet the needs of the national holiday.

Keywords: travel system, golden week, tourist flow, Huangshan, villages in southern Anhui province

1 Introduction

Tourist flows are an important part of tourism geography research and left tourist's space behavior in time and space research to a level, become the focus of tourism management and development planning. Tourism development is vulnerable to outside influence, certain festivals, golden week, and all kinds of incidents are likely to cause fluctuations of tourism development, form a special tourism development period [1,2]. Since implementation of the golden week holiday system, it is to stimulate domestic demand, promoting consumption and major economic development initiatives, as well as to provide short periods of holidays [3]. From the practical standpoint, the golden week not only meets the needs of citizens tend to holiday getaway, to some extent, it fuelled the economy development of the consumer market.

Residents prefer to travel in the week holiday, for relieve work pressure, proper rest, family reunion, close to nature, a form of relaxing form of recreation. However, in reality, only to suffer road traffic congestion, tourist overload, unbalance of supply and demand, resulting in reduced quality of tourism services and other awkward situation. Exposed the imperfection of our holiday system design, especially Golden Week Holiday travel supply and demand contradiction is particularly acute. This paper will proposed recommendation based on tourism system models for study basis, analyses the impact of tourists, tourist routes and destinations tourism mobility factors, on how to improve our country's Golden Week travel, complete citizens leave systems, promote civic holiday recommenddations for the quality level.

2 Tourist flow influence factors analysis

2.1 LEIBOER TOURISM MODEL

Tourism flow refers to the subject of passenger flow, material flow, information flow, energy flow and cultural flow, etc, is by the tourists' need thrust, destination attraction and related external driving force, under the joint action of between tourists and destination, or multiple destinations of agglomeration and diffusion between flow collection [4,5]. In order to better demonstrate the characteristics of tourist flow and the influence of different factors on the tourist flow. This paper uses LeiBoEr tourism system models (see Figure 1), combine the travel demand, supply and affecting factors these related factors, research golden week tourism flow between the relationship of the destination, destination of tourists and tourism channel [6].

LeiBoEr tourist flow model system includes three parts: tourism, destination of tourists and channels between tourists and tourism destinations. In the tourist destination of tourists to travel from the process of forming the travel channel, namely, tourist flow, including the flow of tourists and the flow of tourism industry. The formation of tourist flow depends not only on the thrust of the tourists and the tensile strength of the tourist destination, the tourist flow's path and direction is influenced by political, economic,

* ***Corresponding author's*** e-mail: zylsjsa@163.com

cultural, personal hobby, tourist facilities and so on. So the tourist flow should not only study the tourist flow within the system of thrust and tension mechanism, and to study the effect of external factors on the tourism flow.

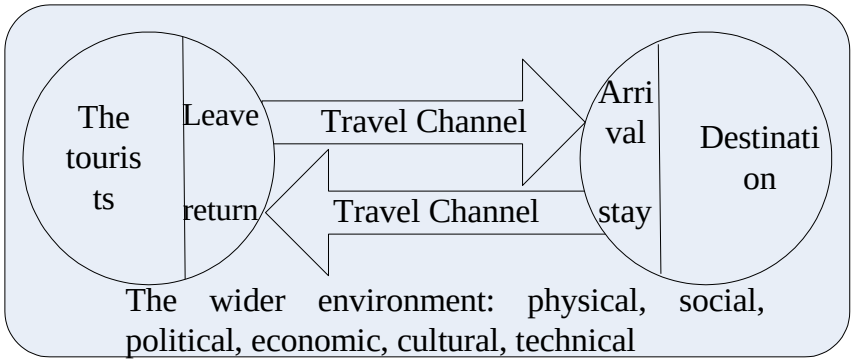


FIGURE 1 LeiBoEr tourism system models

2.2 TOURIST FLOW PATH AND ITS INFLUENCING FACTORS

According to Lei Boer tourist flow model, tourist flow path is a channel of tourist to destinations, from a different point of view, we can divided the tourist flow path into different types. From the economic needs, tourist flow is a research of supply and demand between tourist and tourist destination, calling it a "N-S", is an expression of needs are met by tourist destinations. Starting from the geo-spatial mobile is the study of space moves and layout of tourists and tourism industry, known as the "O-D".

From tourism demand factors, affecting people's Golden Week travel for many reasons and factors, including income, leisure time, and action and so on. Personal disposable income increased, tourism demand will increase. Travel spending is a time consuming, is the discretionary spending of tourists. Analyses pull factors of tourism demand, tourism flow influenced by product supply, pricing strategies, cultural similarities and differences between the source and destination and other factors "7" (see Figure 2)

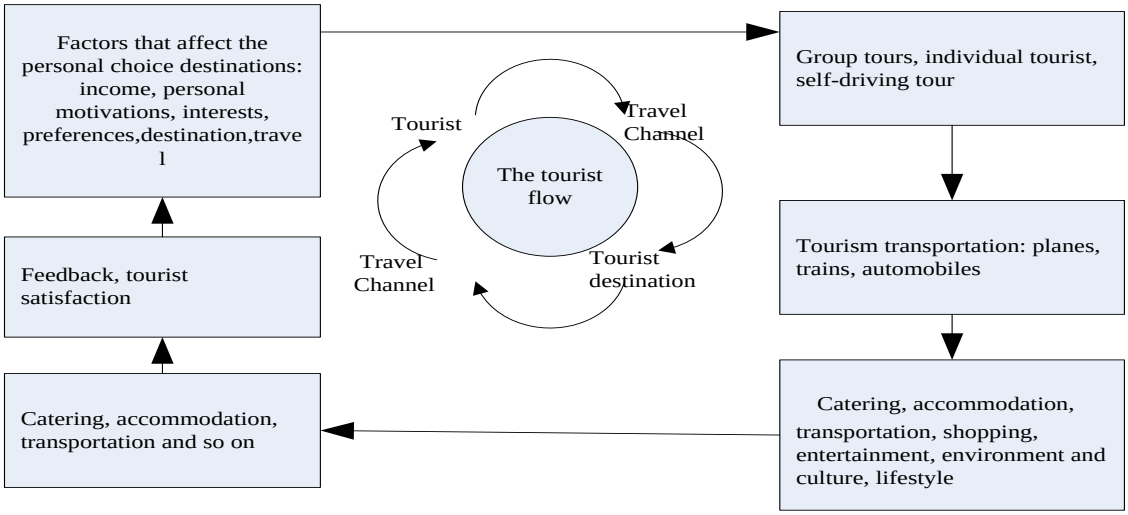


FIGURE2 Tourist flow path and its influencing factors

According to the research emphasis of this article, we analyzed the influence between tourism of tourists and tourist destination tourist flow and its influencing factors of tourist flow according to the economics point of view, that there is a travel demand supply (driver factor) and destination (pull factor), and the external factors influencing the tourist flow, such as government policy, social, economic and environmental status of destination is divided into four categories (see Table 1).

TABLE 1 Tourist influencing factors

First level categories	Second level categories	Influencing factors
Travel demand	Travel motivations	Willingness to travel
	Ability to pay	Disposable income, travel prices
	Leisure time	Length of holiday
Travel supply	Tourism resources	Types of cultural resources, natural resources
	Tourist facilities	Supply of food, lodging, travel shopping entertainment
	Tourist services	Ticket prices and traffic indicators, tourists streaming, information
	Tourism infrastructure	Infrastructure and public restrooms, visitor centers, hospitals
Policy and management	Policy support	Paid vacation, a small car-free tolls
	Government Regulation	Emergency plans, legislation, and other
	Scenic regulation	Swim smooth trails, parking and other
	Travel safety	Destination of the security situation
Destinations affection	economy	Focus consumer, price index
	Social cultural	Subject-object, custom merchandise
	environment	Eco-system

3 Golden week's effect on huangshan and ancient villages in Southern Anhui

3.1 TOURIST FLOW EVALUATION ON HUANGSHAN SCENIC AREA AND THE ANCIENT VILLAGE DURING GOLDEN WEEK

Huangshan scenic spots and ancient villages in southern Anhui are the world heritage site. Huangshan tourism system is top level tourism system in Huangshan city, systems of South Anhui ancient villages in Yixian County, is one of the five secondary tourist destination system of Huangshan [8]. The golden week period and different tourist tourists travel around the same time, prone to traffic concentration. According to tourist flows analysis from source to destination in the tourism model of flow paths and influence factors of tourist flow, this paper validated golden week affect tourists' satisfaction with the Huangshan scenic spots and ancient villages in southern Anhui (see Table 2). "+"expressed positive feedback, has no negative effect on the quality and experience, the government and the scenic no action, "-"indicates problems occurring in the tourism system, affect the satisfaction of tourists, have had some negative feedback, need ancient villages in southern Anhui and Huangshan scenic spots and local governments should take appropriate measures for improvement.

TABLE 2 Tourist flow evaluation table on Huangshan scenic area and the ancient village during golden week

First-class classification	Number	Second-class classification	Positive effects	Negative effects	Measures
Demand	1	Willingness to travel	+		
	2	Disposable income	+		
	3	Demand-oriented		-	Reasonable travel guide, control tourist flow
Supply	4	Humanities, natural resources	+		
	5	Food, lodging, travel shopping entertainment		-	Supply, increase the supply of temporary
	6	Tickets, of signs, tips, banners, etc		-	Appropriate increase in the ticket office, set the prompt
	7	Infrastructure such as hospitals, public toilets		-	Reasonable construction of temporary toilets and so on
Management	8	Small cars from tolls	+		
	9	Tourist trails, parking		-	Increasing the unimpeded protection counseling staff
	10	Crisis management		-	Timely and effective in resolving the crisis
Destination influence	11	Environmental impact		-	Strengthening education on environmental protection and appropriate penalties for the violators
	12	Travel experience			Increase the experiential tours, guided eco-tours
	13	Visitors and residents' consumption		-	Increase tourism supply, control prices
	14	Folk show		-	Strengthen management, proper guidance

3.2 ANALYSIS OF HUANGSHAN SCENIC SPOTS AND ANCIENT VILLAGES IN "GOLDEN WEEK" TRAVEL INFLUENCING FACTORS

Huangshan scenic area and ancient villages has geographical environment, rich in tourism resources. Among them, the Huangshan scenic area known as "the most fantastic mountain", with "pines, rocks, clouds, spring," four famous and ancient villages are outstanding representatives of China's Yangtze River residential villages in mountainous areas. Huangshan scenic spots on the surrounding area with strong radiative forcing [9], passénger flow in ancient villages in southern Anhui Huangshan obviously influence by the Huangshan passenger flow [10], due to the time and location of traffic restrictions, tourists are scheduled on the same day two scenic tours, tourists usually visit the resource levels and high visibility of Huangshan scenic area, followed by return, or continue down the Hill to visit ancient villages in southern Anhui.

Lei Boer tourism system models, travel demand is elastic, tourists can choose different tourist attractions according to their own needs,and tourism supply, especially the destination facility supply is rigid, maximum limits the pull of tourist destinations. Golden Week maximum release of tourists 'travel needs, increased tourism spending. But at the same time demand, appeared in Huangshan scenic area and the imbalances between supply and demand of the ancient villages, scenic overload, many negative consequences caused in Huangshan scenic spots and ancient villages. This paper analyses according to the main factor in Huangshan scenic spots and ancient villages, provide pertinent suggestions for scenic system improvement.

3.3 DEMAND FACTOR

Needs of tourists according to their capacity to pay, a form of willing and able to meet their own needs. Tourism demand has a point, driving tourists from source to destination and even tourism industry direction of the flow, accumulation. According to the research; we can find different tourist groups with social characteristics are different in selection of tourist attractions [11].

There are many factors that affect travel demand, including income, leisure time, and tourism and so on. When assessing item1-3 in evaluation table can find people's disposable income increases, travel demand is also growing; golden week was able to meet tourists on long-distance and trans-regional travel needs. Meanwhile, Huangshan scenic spots and ancient village with its famous attraction is so strong that tourism in the tourism system exerts a very strong cluster effect.

3.4 SUPPLY FACTORS

Implementation of the golden week, make the Huangshan tourist passenger demand and traffic congestion caused by the phenomenon, ancient village voice-belching, losing its origin color, does not know the atmosphere of the ancient village. Sparked conflicts between a tourist and scenic spots management, without tourists' safe and scenic spots of Huangshan scenic spots and ancient villages causing a negative impact.

From Table 2, item 4-7, one of the main factors that cause travellers dissatisfaction is area congestion, infrastructure is not complete. This scenic reception have been unable to meet the demand of tourist, demand and supply become imbalance. So although the Huangshan scenic area and the ancient village has the advantage of human, natural and landscape features, but due to inadequate supply, mismanagement, incomplete infrastructure during golden week, greatly reduced the tourists 'satisfaction.

3.4 MANAGEMENT FACTORS

In 2012, issued by the State Council on approving and departments such as the Ministry of transport on major holidays to exempt small passenger car tolls of notification of programmes should be implemented, greatly contributed to the drive travel and long distance tour. Meanwhile, caused heavy traffic, improper management and emergency measures of the problem in scenic. Travel safety is found to be the most fundamental requirements of tourist, destinations that security in tourist activities, prevent from physical and mental burden of injury or property damage, thereby increasing tourist satisfaction [12]. Huangshan scenic area to resolve risk management measures improper reduces the satisfaction of tourism.

3.5 DESTINATIONS AFFECTED

Passengers to destinations in their own needs can be met, have a pleasant and enjoyable travel. However, during the Golden Week travel, concentration of tourist information, as well as the demand of tourists, visitors make the Huangshan scenic area and the ancient village clashed with the management side, exposed the disadvantages of the reception capacity. Meanwhile, area admin area regardless of area of affordability, on the number of visitors without restrictions, results in crowded, ecological and environmental damage. Increased commercialization of ancient village tourism, loss of traditional culture of truth and simplicity, influences the attractiveness of destinations.

4 Suggestion

The "Golden Week" holidays in the form of travel, study it in the tourism system, are actually tourists demand and the impact of moving in space and time. "Golden Week" holidays of this centralized system, bringing" travel fashion" can easily lead to popular scenic spots heavily overloaded, causing similar problems of Huangshan scenic spots and ancient village tourist area. This paper asked on the study on ancient village tourism in Huangshan scenic area and analysis of influencing factors concerning the improvement of Golden week's tourism made the following recommendations:

1) Improve the holiday system and meet the diversifycation of tourism demand.

The rapid development of economy and technology, increasing residents' disposable income and time, enterprises can promote staggered holiday, tourism can be divided into time to travel, diversion of tourist flows play a role, alleviate the pressure of the golden week travel overcrowded, avoid crowded traffic, accommodation and tourism vacation problem such as poor quality.

2) Establish the perfect travel system and to divert divert visitors.

Travel system work, established vacation holiday travel forecast system based on prediction of golden week holiday tourism flows to help guide the management of scenic spots and providing visitors with useful travel information for their reference. Timely dissemination of early warning information to guide visitors through the broadcast or network, do an emergency plan to help tourist choice tourist attractions.

3) Limited tourist capacity, improving the quality of tourism.

Travel size must be in strict accordance with the scenic spot management system, develop scenic maximum carrying capacity of government violations of the scenic spots to take coercive measures, economic punishment as stipulated. Guaranteed quality of service, improve visitor satisfaction. Meanwhile, strengthening environmental education, increasing the visitor experience programs to prevent commercialization of scenic.

5 Conclusions

Passenger traffic is an important symbol of tourism development; passenger volume size can reflect the appeal to tourists. Implementation of the golden week, increasing passengers' chance to travel, has also led to generally lower quality of passenger travel experience. Implementation of the tourism law, area volume control to a certain extent, but the effect is not obvious. Hot scenic spot overload, poor tourism service quality, environment and resources has been disrupted, traffic and security issues highlighted, some people worried about congestion, safety and other factors delayed time to travel, that golden week can also delay a part of people's travel needs. This paper based on tourism system model, taking Huangshan scenic spots and ancient villages as case study, analysis of various factors of Golden Week affecting tourism flows and make recommendations, hope to implement new management measures, to solve problems arising during the golden week.

References

[1] Xie X M, Sun G N, Han C X 2010 Analysis of influence of tourism on the tourism industry in Xinjiang *Arid land geography* 33(3) 487-92 (*in Chinese*)
[2] Liu Z H, Zhang J 2010 Special study on structural response time distribution of tourism geography of tourism flow-In Beijing, Huangshan, Xian TDD Golden Week tourists as an example *Human geography* (1) 129-33 (*in Chinese*)
[3] Yu W B 2011 Constructing experimental teaching system of tourism management major *Overview of tourism (industry)* (1) 61-2 (*in Chinese*)
[4] Zhang Y Y, Gu J, Ma Y F 2013 Progress, evaluation and prospect of research on tourist flow *Journal of tourism* 28(6) 38-45 (*in Chinese*)
[5] Pang W, etc 2011 Tourist flow architecture evolution based on fractal theory *Journal of Shanxi University of technology (natural science Edition)* 27(2) 89-94 (*in Chinese*)
[6] He X H, Bai K 2011 Xian special study on fractal structure of tourist flow size-with "October 1st" golden week for example *Arid land geography* 34(5) 858-65 (*in Chinese*)
[7] Wang Z F, Zhang L 2010 Summary of the research on residents' perception of tourism impacts *Economic geography* 30(9) 1563-8 (*in Chinese*)
[8] Zhang S Z, Meng C 2011 Huangshan tourism economic development pattern transformation and upgrading *Journal of West Anhui University* 27(1) 78-83 (*in Chinese*)
[9] Zhu F 2010 Study on spatial structure of tourist area in southern Anhui based on aggregate fractal *Yunnan geography study* 22(1) 94-9 (*in Chinese*)
[10] Yu X Y, Sha R, etc 2013 Neighboring scenic seasonal comparison of passenger flow and its dynamical correlation *Geographical research* 32(1) 191-200 (*in Chinese*)
[11] Peng C, Lin B, etc 2014 Analysis of different groups of tourists to Hainan provincial tourism scenic area selects type *Jiaying University* 32(6) 50-1 (*in Chinese*)
[12] Zou Y G, Zheng X M 2012 Research of Tourists security formation mechanism and influence factors of promotion strategy *Human geography* 27(3) 103-8 (*in Chinese*)

Author

Yan Zhang, 1982, Shaanxi Province of China.

Current position, grades: lecturer.
University studies: Master's degree of human geography, Shaanxi Normal University in 2010.
Scientific interest: regional development and urban and rural planning.

Establishment and implementation of network congestion control algorithm based on real-time streaming transmission

Pingping Xiao*

Changchun Guanghua University, Changchun, China

Received 1 October 2014, www.cmnt.lv

Abstract

In order to improve the network performance of computer and avoid the occurrence of network congestion better, this paper analyzed the problems faced by the integration of real-time streaming transport and network and their solutions in the perspective of the characteristics of real-time streaming transport. As to the congestion control of streaming media, TFRC algorithm was analyzed emphatically. Based on TFRC, network congestion was improved, monitored and predicted with parameter of real-time cache length; sending rate was corrected based on cache length when the network was saturate, in order to avoid congestion in time and improve fairness; at last, the test comparison proceeded by network simulation platform NS2. The result indicated that, the improved TFRC was fairer; meanwhile, it showed more friendliness to TCP.

Keywords: real-time streaming transport, congestion control, TFRC, cache length

1 Introduction

Network congestion is caused by the mismatching of network resources and data transmission, for example, the unreasonable network cache space, bandwidth capacity, processor performance, network structure; they detailed present as information packet reception delay, discard probability increasing, upper application system performance decreasing, etc. In the current stage, network protocol plan is widely applied in stream transmission [1,2]. Therefore, the research on network congestion control mainly focus on network TCP protocol, including improving TCP protocol, developing new congestion control protocol by simulating AIMD algorithm of TCP, or regulating transmission speed of media stream according to the TCP throughput capacity under stable state [3,4]. Wang Wenliang et al proposed a TCP congestion control method based on LwIP, detailed stated the implementation method of slow start algorithm, congestion avoidance algorithm, fast retransmission algorithm and restoring method, and optimized TCP congestion control method of LwIP according to correction of fast recovery algorithm of TCP. Lv Guanqiao [6] et al proved the effectiveness of TCP improvement as compared in the perspective of channel utilization, communication stability and fairness. Besides the above methods, TCP friendly congestion control mechanism is the most discussed and promising mechanism in recent years [7]. This paper first introduced TCP-friendly rate control protocol, then analyzed TFRC algorithm emphatically, adjust transmission speed through packet loss probability, round-trip time with the current formula. Meanwhile, smooth packet loss internal algorithm was introduced to precede weight average calculation on the newest eight packet loss rate to ensure smoothness of transmission speed. And then, based on TFRC, these papers introduced additional web cache parameters, proposed MulRFRC algorithm, estimate the behaviors of various TFRC data flow modify transmission speed and improved TCP friendliness and finally made experimental verifycation with network simulation platform NS2.

2 Real-time streaming transmission

2.1 DEFINITION OF STREAMING TRANSMISSION

Streaming transmission is process to transmit multimedia data from transmitting end to destination in the form of continuous data streaming and then play multimedia while receiving data in destination. It can be divided into real-time streaming transmission and progressive streaming transmission. Progressive streaming transmission refers to that, user only can visit the multimedia data that has been downloaded, moreover, the data is downloaded according to certain order. Real-time streaming transmission refers to transmit data into user side through certain network protocol, which is suitable for live event [8].

2.2 CHARACTERISTICS OF REAL-TIME STREAMING TRANSMISSION AND PROBLEM OF ITS INTEGRATING WITH INTERNET

The characteristics of real-time streaming transmission includes high bandwidth, burstiness, the same speed, voice frequency and video sampling rate while receiving end playing multimedia data in transmission.

In practical network, communication network is characterized by limited link bandwidth, unpredictable data packet delay, which is inconsistent with high bandwidth, real time and burstiness of real-time multimedia data strea-

* ***Corresponding author's*** e-mail:xpp163xpp@163.com

ming. Transmitting data by dedicated links is expensive and unpractical. As a resources-shared network system, some characteristics of internet is not suitable for multimedia data transmission. Normal transmission of multimedia data flow in internet should take some factors into account: internet has enough bandwidth, data packet can be transmitted to destination address of the same group at the same time, thus to reduce to burden of multicast; internet has relative mechanism and can meet the demand of enough bandwidth resources; internet transmission technology is based on data packet switching, and should ensure real-time data packet to reach receiving end in time according to certain order, thus play continuously and synchronously.

3 Research on network congestion control

3.1 NETWORK CONGESTION ANALYSIS

Network congestion mainly refers to that, the demand on network resources of user exceed the current volume, thus cause overload condition, decrease the communication of network and network throughput capacity, increase packet loss rate, data packet delay, and even cause network collapse. The main reason is that, link bandwidth, data packet processing capacity of network node and node cache cannot meet the demand of user [9]. Limitation of link bandwidth is because the channel capacity of transmission media is limited and the exceeding data volume capacity can cause loss or error of data packet; the limitation of data at node refers to the congestion when processor cannot process data packet, update of routing table and retrainsmission of data packet; node cache is induced by overflow of data packet and change of performance in buffer memory space caused by limited link bandwidth and insufficient processor. Moreover, complex network structure and unreasonable routing algorithm principle are also possible to induce network congestion.

3.2 TFRC ALGORITHM STUDY AND PROTOCOL

TCP protocol based on AIMD algorithm is not suitable for all internet application. New transport layer protocol is needed to meet multimedia application in order to meet the requirements of stable bandwidth, low delay and small shake. TFRC ensures the service quality of network in the premise of ensuring friendliness of TCP. TFRC is a kind of congestion control algorithm suitable for traditional and TCP data flow-dominated network environment. It can ensure data flow to have smooth transmit speed, reduce transmit speed through theoretical formula and is suitable for real-time multimedia data flow. The relative function is:

$$T=\frac{s}{R\sqrt{\frac{2p}{3}}+t_{RTO}\left(3\sqrt{\frac{3p}{8}}\right)p\left(1+32p^2\right)}. \tag{1}$$

In the Equation (1), T is the maximum of transmit rate of data flow, s is the size of data packet, R is the round-trip time of data packet, p is the packet loss rate in stable state, t_{RTO} is the retransmission time of TCP packet loss. If compared the different data flow effect in the same network environment, it is better to choose same response function to control transmission speed and fairly use network resources. Therefore, good use of TCP response function can ensure fair competition of TFRC and TCP as well as smooth of transmit rate. The flow of TFRC congestion control mechanism is as shown in Figure 1:

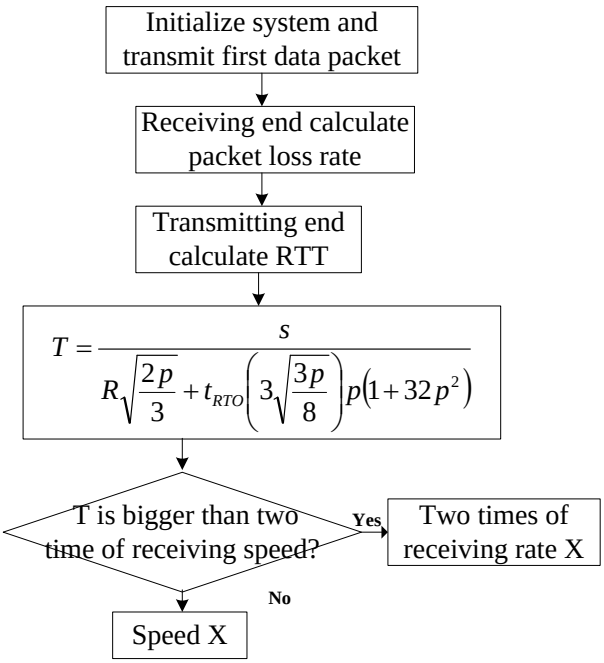


FIGURE 1 Flow chart of TFRC congestion control mechanism

Calculating packet loss rate is the most important part of TFRC algorithm. TFRC selects packet loss internal smooth algorithm to calculate the probability of packet loss event with weighted average, and can restrain some non-representative packet loss as well as reduce the shake of packet loss rate. Weighted average algorithm is as follows:

$$\hat{s}(1,n)=\frac{\sum_{i=1}^{n}w_i s_i}{\sum_{i=1}^{n}w_i}, \tag{2}$$

where is refers to the former number i packet loss event internal, which can be expressed as the number of data packets, $\hat{s}(1,n)$ refers to the smooth result of the former n packet loss event interval; the calculation of weight w_i is as follows:

$$\mathrm{w}_i=\begin{cases}1, & 1\le \mathrm{i}\le n/2\\ 1-\dfrac{i-n/2}{n/2+1} & n/2<i\le n\end{cases}. \tag{3}$$

When sample s_i take i=0, s_0 refers to the number of data packet received during the last packet loss event to the current time. The mean value of s_0, s_1,…,s_{n-1} is:

$$\hat{\mathrm{s}}(0,n-1)=\frac{\sum_{i=0}^{n-1}w_{i+1}s_i}{\sum_{i=0}^{n-1}w_{i+1}}. \tag{4}$$

Finally, it is obtained that the average interval $\hat{s} = \max(\hat{s}(1,n), \hat{s}(0,n-1))$, and the probability of packet loss event is:

$$p = \frac{1}{\hat{s}}. \tag{5}$$

The use of packet loss event probability can simulate "reduce transport protocol while coming across packet loss" [10]; meanwhile, use of TFRC algorithm can make transmit rate be smoother. However, TFRC responses slowly to packet loss, cannot be able to decrease transmit speed in time, thus occupy too much bandwidth. This paper aimed to improve based on this defect, thus enhance the fairness of TFRC.

3.3 TFRC IMPROVED ALGORITHM

Packet loss rate can be updates in a high speed through reduce sample number N, or increase weights of the newest samples, at the same time, shake of throughput rate increases and the redundant noise is introduced. This paper introduced extra network parameter - real-time cache length to correct packet loss interval, as shown in Equations (2)-(6):

$$s_m = \hat{s} \cdot f_{\text{RealTime}}, \tag{6}$$

where S_m is the packet loss interval after correction, f_{RealTime} is the introduced network parameter, which is not related to network condition, but can reflect and predict the congestion and saturation of network. In network with coexisting of TFRC and TCP, one-way delay of TFRC data packet can monotonically increase or decrease within a large range, emerging lots of peak value. Transformation process of network state: free network, transmit speed increases, use efficiency improves – data packet increases, one-way delay grows – network congestion – congestion control – free network, in dynamic balance. One-way delay of TFRC is monotonous. When the delay value increases, data packet is considered to be cached, and the amount of data packet is the network cache length; when the delay value remains the same, then network is free and the cache is 0; when the delay value decreases, then cache length decreases, as shown in Figure 2:

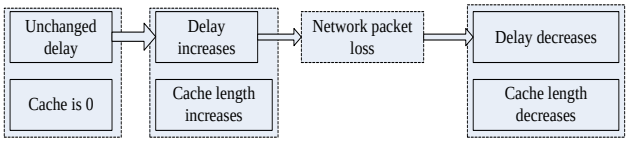


FIGURE 2 Change of cache length

When data packet reach receiving end, the one-way delay time of data packet is calculated. The equation is as follows:

$$\text{stt\_sample}_i = t\_\text{now-t\_send}_i. \tag{7}$$

In addition, instantaneous noise should be filtered, for example, the first sample takes $\text{stt}_{i-1} = 0$. The calculation method is:

$$\text{stt}_i = df * stt_{i-1} + (1-df) * stt\_sample_i \tag{8}$$

Smoothing factor can select the value same with RTT. When obtaining stt sequence, changes between samples are calculated, in order to deduce the current state of network. They are expressed by three parameters as follows:

Inc_count records the length of increased cache, indicating the network is in stage of gradual saturation. dec_count recorded the length of decreased cache, indicating the network is in stage of congestion control. equal_count records the amount of data packet when the cache is 0, indicating the network is free. Parameter zero_delay is used to filter the unnecessary delay shake, can be valued as 10-3ms. In addition, shake threshold value should be set to reduce shake and filter noise.

This paper mainly adopted RED [1] queue management mechanism, selected Minth as the shake threshold of parameter in gradual saturation state. For example, in free state, it is detected that inc_count is larger than Minth, then it is believed that network starts to enter gradual saturation state, that is, B_r =inc_count.

Real-time caches B_r can be used to correct transmit speed of TFRC. When B_r fluctuates near 0, it means network is free and do not need to correct speed; when B_r increase, network begins to saturate; before packet loss event and after some data, the bigger B_r is, the higher network saturation is, the more packet loss interval need to reduce. At this moment, a correction threshold value B_{th} should be set. The correction threshold value selected in this paper is the cache length when the first data packet is abandoned. When B_r is larder than B_{th}, algorithm can adjust packet loss interval. When B_r is larger than or equal to B_{th}, then the correction is not needed. The value of f_{RealTime} proceeds according to the following equation:

$$f_{\text{RealTime}} = \begin{cases} 1 & B_r < B_{th} \\ \sqrt{\dfrac{B_{th}}{B_r}} & B_r \geq B_{th} \end{cases} \tag{10}$$

Therefore, the Equation (10) for packet loss rate pm after improvement is as shown in Equations (2)-(11):

$$p_m = \frac{1}{s_m} = \frac{1}{\hat{s} * f_{\text{RealTime}}} \tag{11}$$

Detailed flow of improved TFRC is as shown in Figure 3:

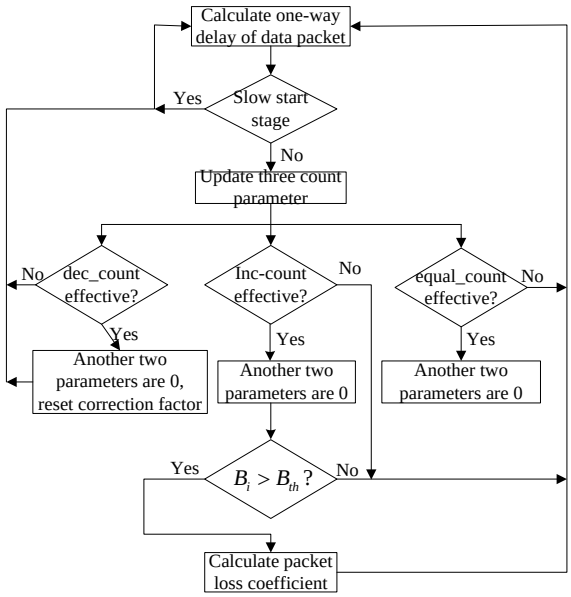


FIGURE 3 Algorithm flow figure

4 Performance analysis of algorithm

4.1 SET OF EXPERIMENTAL ENVIRONMENT

This paper selected NS-2 network simulation system as experimental environment. NS-2 is mainly used for solving problem of network research and simulates various network protocols such as TCP, routing and multicast in wireless or wire network. Network topological structure is this simulation is as shown in Figure 4:

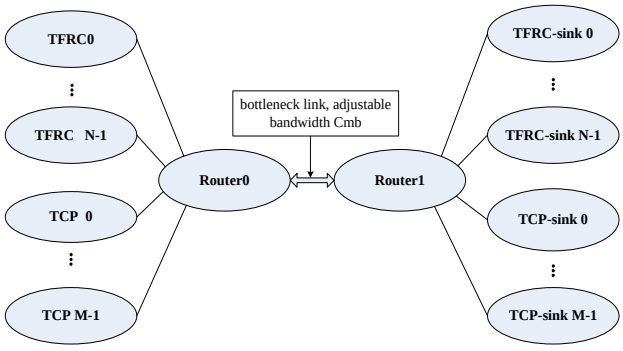


FIGURE 4 Simulation network protocol structures

R0 and R1 are two routers. Between them, it is the bottleneck of network. The delay is 15 ms. Queue scheduling adopts RED queue management. The minimum threshold of length is 5 packets and the maximum is 15 packets. Weight w is 0.002. The maximum temporary probability is 0.1. The link bandwidth between sending node and receiving node with router is 10Mb/s and the network delay is 0.

4.2 EXPERIMENTAL RESULT AND ANALYSIS

1) Analysis on fairness and throughput rate in network environment with different bottleneck bandwidth.

Set topology network according to the simulation scene of network environment with different bottleneck bandwidth. The fairness of simulation result is as shown in Figure 5 and the throughput rate of link is as shown in Figure 6. It can be found that, after the bottleneck bandwidth increases, packet loss event of TFRC data flow decreases, therefore, TFRC data flow tend to be more stable after bandwidth increasing. In addition, the improved TFRC can reduce resource occupation when congestion occurs, release more bandwidth for TCP data flow, improve the fairness; the average throughput rate of bottleneck is basically unchanged.

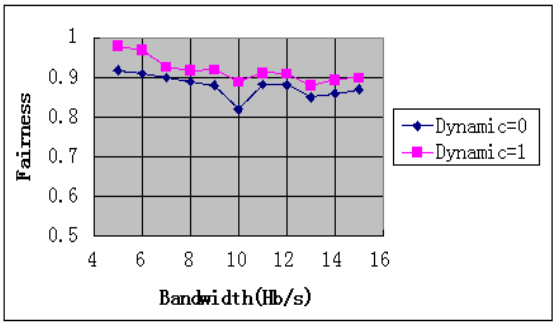


FIGURE 5 Fairness of TFRC in condition of different bandwidth

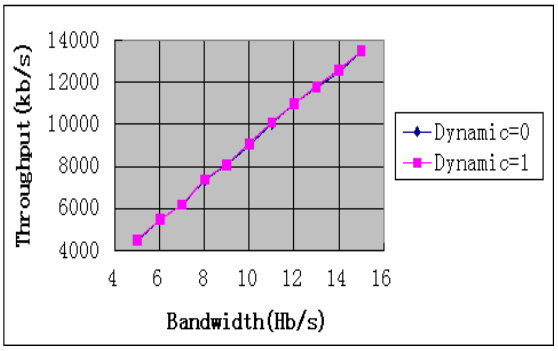


FIGURE 6 Throughput rate of TFRC in condition of different bandwidth

2) The analysis of fairness and throughput in network environment with different amount of TFRC data flow.

Set topology network according to the simulation scene of network environment with different amount of TFRC data flow. The fairness parameter of the simulation result is as shown in Figure 7 and the link throughput rate is as shown in Figure 8. It can be seen that the improved TFRC data flow cannot improve all sampling points because of the randomness of network; however, it can decrease sending rate in a high speed during network congestion, thus network resources can be released in time and its fairness also improves. With the increase of TFRC data flow, the improvement of TFRC algorithm improves fairness by sacrificing the average throughput rate of link.

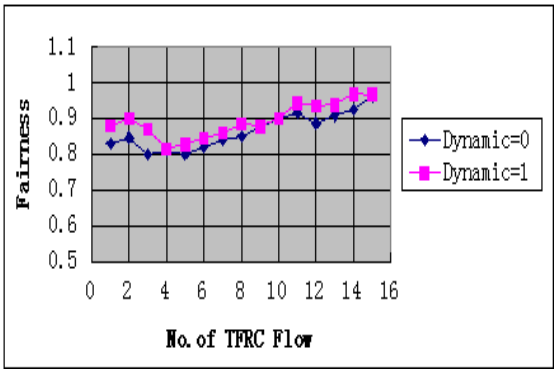


FIGURE 7 Fairness in the situation of different TFRC data flow

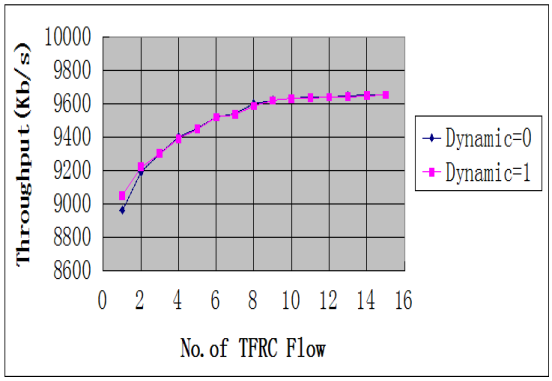


FIGURE 8 Throughput rates in the situation of different TFRC data flow

5 Conclusions

The key point of TFRC achieving good effect is to respond timely and accurately when the network occur congestion. TFRC algorithm can improve the quality of stream transmission service based on IP to some extent as well as the friendliness of TCP. This paper made improvement aiming at the insufficient fairness, introduced network parameter - real-time cache length to estimate network condition and improve the fairness of TFRC. In addition, it preceded test through network simulation platform, veryfying that the improved TFRC had higher fairness and showed good friendliness. Although real-time streaming transmission does not require the complete correction of data, proper packet loss retransmission can improve the transmission quality of multimedia. Therefore, the following research is to speculate network congestion degree according to real-time cache degree and precede retransmission of packet loss, in order to provide better transmission service for network.

References

[1] Lei X H, Jiang XH, Wang CH 2013 Principle and Application of Streaming Media Technology Based on RTMP *Journal of Communication University of China Science and Technology* **20**(6) 59-64 (*in Chinese*)
Adobe Systems Incorporated 2012 Adobe Flash Media Server4.5 Developer's Guide[EB/OL]. http://help.adobe.com/en_US/flashmediaserver/devguide/flashmediaserver_4.5_dev_guide.pdf
[2] Kong J S, Ren P Y 2014 Summary of TCP Network Congestion Control Research *Computer Technology and Development* **24**(1) 43-6 (*in Chinese*)
[3] Liu X Y 2011 Simulation of TCP Congestion Control Algorithm Based on NS *Experimental Technology and Management* **28**(9) 79-81 (*in Chinese*)
[4] Wang WL, Zhang XY, Li Y 2010 TCP Congestion Control Method Improvement Based On LwIP *Journal of Changchun Teachers College* **29**(1) 45-7 (*in Chinese*)
[5] Lv G Q, Liu H B 2014 The Design of Network Congestion Control Algorithm Based on TCP Protocol *Software Guide* **13**(1) 56-8 (*in Chinese*)
[6] Xiao F, Wang R C 2010 Research of TCP-friendly Congestion Control Protocol in Wireless Network *Computer Science* **37**(7) 50-3 (*in Chinese*)
[7] Wang F, Cai J J 2012 The Design of Real-time Video Transmission System Based on Wireless Network *Microelectronics & Computer* **29**(5) 58-61 (*in Chinese*)
[8] Chen L R, Kong J S 2012 The Research of Congestion Control on Internet *Computer Knowledge and Technology* **8**(7) 1502-5
[9] Xie J, Yu L, et al 2010 Congestion Control Based on Queuing Delay and Packet Loss Probability *Journal of Electronics & Information Technology* **32**(9) 2058-64 (*in Chinese*)
[10] Sun D D, Wang Y B 2010 Research and Simulation of Queue Management Mechanism RED Based on NS2 *Collected Paper of 2010 Academic Annual Conference of Computer Association of Guangxi*
[11] Liang P 2010 A Study into the Network Simulation Based on NS2 *Journal of Hechi University* **30**(2) 57-61 (*in Chinese*)

Author

Pingping Xiao, born in 1974, Jilin Province of China.

Current position, grades: associate professor.
University studies: PhD degree of control theory and control engineering, Jilin University in 2008.
Scientific interest: wireless sensor network (WSN), a distributed intelligent control, network congestion control.

Analysis evaluation and quantitative research on economic law behaviour cost and market efficiency

Jindong Chen*

Yuanpei College, Shaoxing University, Zhejiang, 312000, China

Received 1 October 2014, www.cmnt.lv

Abstract

Fiscal expenditure, as the material support of economic law behaviour, plays a critical function in regulating resource distribution, promoting economic development and guaranteeing social far equity. Analysis evaluation of fiscal expenditure is obtained through connotation of fiscal expenditure and investigation evaluation. This paper thought that, analysis evaluation of economic law behaviour cost referred to make comprehensive evaluation in aspects of economy, efficiency and benefits on fiscal expenditure process and result according to specific principles and using scientific and normative evaluation method and technology, with fiscal expenditure as management objectives. The results it obtained will be mostly reflected on market effect. Through the quantitative research on fiscal expenditure and economic benefits evaluation in our country, this paper made digitalization and objectification analysis on the comparison of economic law behaviour cost and market benefit, studied the relationship between fiscal expenditure scale and market economic, further estimated the optimal fiscal expenditure scale in our county, in order to provide basis and rule for the formulation of economic law behaviour cost based on quantitative analysis.

Keywords: economic law behaviour, fiscal expenditure, market benefits, quantitative analysis

1 Introduction

For a long time, the function of fiscal policy in economic growth is one of the core topics concerned by theoretical researchers and policy makers. Modern fiscal theory and practice proved that, fiscal expenditure and economic growth is closely related [1,2]. According to Keynesian fiscal theory, positive fiscal policy can promote economic growth by multiplier benefits, but meanwhile, it also can be decreased because of crowding-out effect [3]. At present, estimation of optimal fiscal expenditure scale are mainly based on two methods: one is to use natural efficiency condition – marginal output of public service is 1 under linear model framework to establish production function; the other one is to use the concave function relationship between economic growth and fiscal expenditure scale in the perspective of nonlinear to construct equation for estimation, taking Armey curve as example [4]. When Armey curve is expressed as small fiscal expenditure scale, the enhancement effect on output takes the leading role. At that moment, increasing fiscal expenditure can promote economic growth, when the scale reaches some critical point A (as shown in Figure 1), then the generated enhancement effect decreases progressively and further expansion of scale will inhibit output increase.

This paper analyzed the status of fiscal expenditure and market benefits in our country, made certain economic evaluation and confirmed the boundary of economic law behaviour based on quantitative analysis, thus to provide basis and rule for the relative research on economic law behaviour in our country.

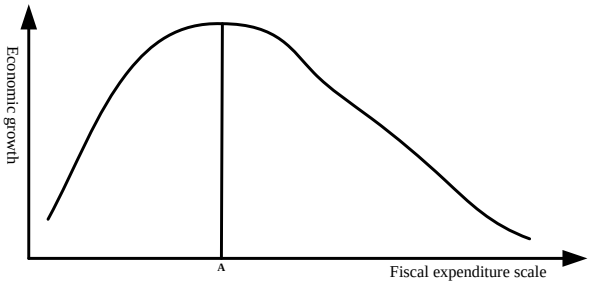


FIGURE 1 Armey curve

2 Analysis of fiscal expenditure status

2.1 FISCAL EXPENDITURE SCALE

Since the reform and opening-up policy, the scale of fiscal expenditure in our country showed change of down first and then rise. This change fully reflects the characteristics of economic transition period. In 1996, fiscal expenditure in our country decreased to the historical low point, accounted for 11.15% in GDP. It demonstrates that, market plays a leading role in resource distribution in the process of turning to market economic system. Year 1992 to 1996 is the stage of tight fiscal expenditure scale; year 1998 to 2002 is the stage of expanded fiscal expenditure scale; year

* ***Corresponding author's*** e-mail: cjdsxwl@163.com

2003 to 2007 is the stage of implementation of firm fiscal policy, and the fiscal expenditure scale remains between 17.8% and 18.7%; year 2008-2009 is the stage of expanded expenditure scale. In addition, calculated based on the current year's price, the total amount of Chinese fiscal expenditure from 1978 to 2012 grows as much as 111.03 times over the 34 years with an average annual growth of 3.27 times. We can found that, fiscal expenditure constantly increases with the development of modern society, and the expansion of fiscal expenditure scale is related to the fiscal policy and economic situation in that year [5]. For instance, in the second half of 2008, financial tsunami induced by subprime crisis spread around the world. As a result, the economic growth in our country slowed down and economic pressure became bigger. In order to resist the adverse effect of external environment on economy, our country proposed 4 trillion investment plan and a series of stimulus measures used for expanding domestic demand. The positive fiscal policy helped our country effectively resist financial crisis, made contribution for the early recovery of Chinese Economy but also aggravated fiscal deficit.

2.2 FISCAL EXPENDITURE RELATED TO ECONOMIC LAW BEHAVIOUR

2.2.1 Fiscal expenditure related to economic law behaviour in macroscopic view

Fiscal expenditure and fiscal revenue together constitute all steps of financial distribution, which detailed reflects on government's arrangement, supplement and utilization and management on the financial fund, indicating the scale, structure, flow direction and use of financial fund [6,7]. Fiscal expenditure can provide the financial resources needed by government, thus to ensure the normal operation of state machinery and realize function of the state; secondly, it can provide public goods and service for society; thirdly, fiscal expenditure can regulate the scale and structure of financial operation and make macro-adjustment on national economic policy, thus realize social equity and promote the healthy development of macro economy.

Public expenditure and government expenditure is also similar to the concept of fiscal expenditure. Fiscal expenditure contains all non-public expenditure used for investtment on competitive field such as economic construction as well as the public expenditure such as providing public products for non-competitive field; government expenditure is government-dominated and government duties-based financial expenditure activity. Public expenditure refers to in particular to the expenditure needed for providing public goods and service by public organization under market economy. It not only contains the main provider with government as core, but also contains lots of non-government public organization. The relationship between these three is as shown in Figure 2.

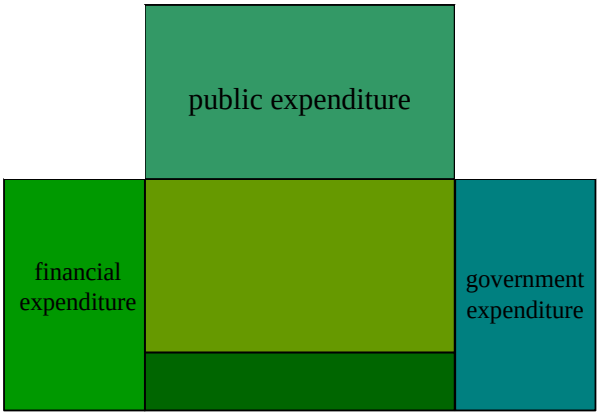


FIGURE 2 The relationship between financial expenditure, public expenditure and government expenditure

Fiscal expenditure project involves fields such as national defence, economy, education, social insurance. In recent years, the proportion of education expenditure in fiscal expenditure becomes higher and higher, increasing from 6.69% in 1978 to 15.10%. General public service and social insurance and employment take the second and third place, respectively. The development tendency of financial policy of our government is to carry out the people's livelihood finance and give way to the market. Therefore, the investment on people's livelihood projects such as education, social insurance and employment, medical care become more and more, while the expenditure proportion on grain and oil material management affairs and business service industry constantly decreases.

2.2.2 Fiscal expenditure related to economic law behaviour in microcosmic view

Analyzed from microcosmic view, the intervention of economic law behavior on economic life in aspect of fiscal expenditure roughly reflects on market control and economic services. Market control, in fact, is that, state use public power to correct and standardize deviation state in economic life from external market. Financial regulation accounts the most in market regulation. However, the foods and drugs regulatory affairs that affect the health of consumer has not greatly supported by finance. The regulatory efforts are not conforming to the impact strength needed for major food safety accidents such as swill-cocked dirty oil and red yolk eggs. In addition, the market regulatory strength on the retail price of goods cannot suppress the rapid growth of price. In the short term, the investment on market regulatory field is not perfect; meanwhile, missing merchant mortality and disordered economic operation also exist.

Economic service is the basis of normal economy operation, among which, the stand or fall of public transportation construction is related to the circulation and transaction of products in market. The normal supply of water and electricity is the security of normal operation of social

economy. For the past two years, our country mainly invested in transportation and infrastructure construction of small towns with large fund. For instance, in 2011, the expenditure on transportation is 749.78 billion, infrastructure of small towns 36.78 billion, total 786.56 billion, accounting for 7.2% in national fiscal expenditure.

3 Evaluations on the financial expenditure effect

3.1 FISCAL EXPENDITURE AND ECONOMIC GROWTH

Fiscal expenditure has multiplier effect and acceleration effect [8]. Fiscal expenditure is related to the national income to some extent. National income is increasing with the growth of fiscal expenditure. Then it feeds back to fiscal expenditure. As a result, fiscal expenditure increase promotes the increasing of national income, and then the income increasing stimulates the expenditure increasing.

Fiscal expenditure and GDP growth rate are all positive number (except 1980 and 1981), indicating GDP increases with the increase of fiscal expenditure. However, the fiscal expenditure growth rate does not have necessary correlation with GDP growth rate. Figure 3 shows the relationship between fiscal expenditure and GDP growth rate. It can be seen that, in 1979~1996, fiscal expenditure is basically consistent with the GDP growth rate curve, showing positive correlation. However, in 1996~2003, the curve moves towards the opposite direction, showing negative correlation. In 2003 to 2012, they become consistent again.

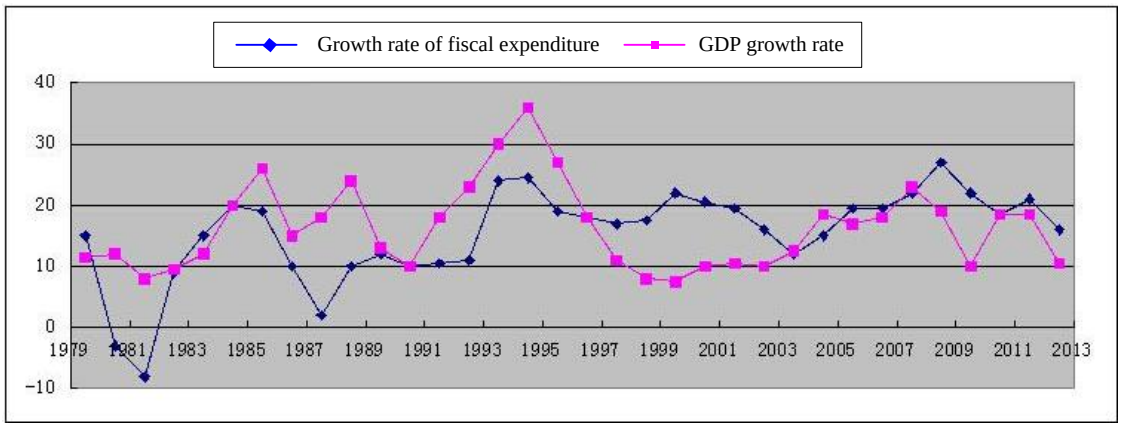


FIGURE 3 Growth rate of fiscal expenditure and GDP in 1979~ 2012(%)

Ratio of growth rate of fiscal expenditure and GDP fluctuates greatly over the 33 years. The ratio of growth rate of fiscal expenditure and GDP contains two information: the contribution rate of fiscal expenditure on GDP growth and the elastic coefficient of fiscal expenditure growth. As shown in Figure 3, 1996 is the turning point of the relationship between fiscal expenditure and GDP. 1996 can be regarded as the watershed of planned administrative intervention and market rules regulating. Negative financial policy and tight monetary policy is adopted to reply the increasingly serious inflation from 1993 to 1996; As the financial crisis broke out in 1997, our country began to carry out the positive financial policy that takes expanding domestic demand as basic approach, and enhance infrastructure construction and public investment. Before 1996, the promotion function of GDP comes from fiscal expenditure with pertinence and political orientation. After 1996, the state of GDP mainly relies on the development of market rule, and its relationship with fiscal expenditure decreeses.

3.2 ECONOMIC CONSTRUCTION EXPENDITURE AND ECONOMIC GROWTH

Figure 4 shows the curve relationship between economic construction cost and GDP growth rate. It can be seen that, from 1978 to 1997, the curve trend of economic construction expenditure and GDP growth rate is basically consistent. These two show obvious positive correlation while negative from 1997 to 2009. It demonstrates that, in the initial stage of market, the economic construction expenditure can significantly stimulate GDP. But with the economic development, economic construction cost is negatively correlated to economic growth. Therefore, in the meanwhile of economic development, economic construction expenditure should make structural adjustment.

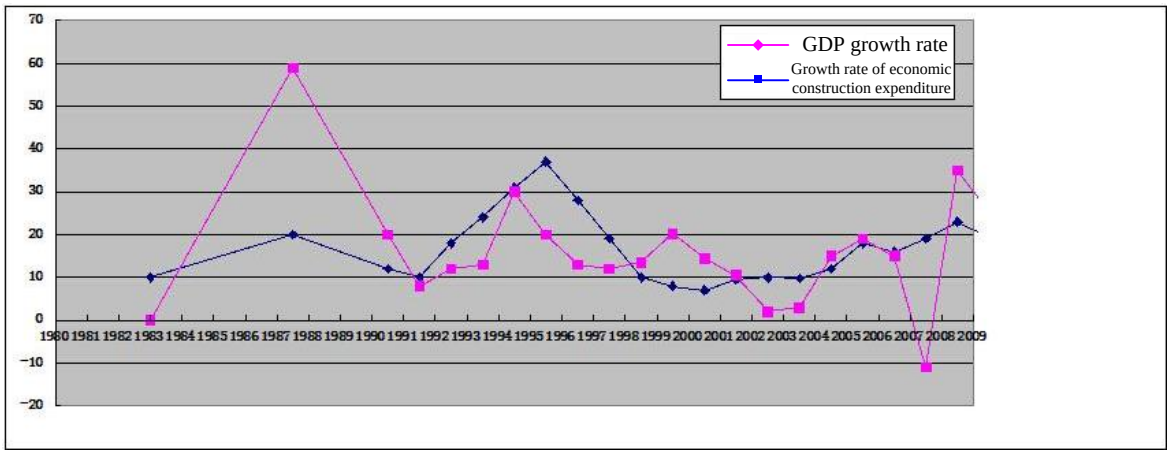


FIGURE 4 Growth rate of economic construction expenditure and GDP in our country from 1980 to 2009 (%)

4 Economic law behaviour based on quantitative analysis

It is of great significance of economic law behavioural norms to economic development. The boundary content of economic law behaviour contains intervention range, intervention object and intervention degree [9]. Accurate positioning of economic law behavioural norms can ensure market rule to have sufficient regulation space as well as providing probability for the defect intervention that is hard for market itself to overcome. In reality, macroscopic and idealized planning boundary has limited guidance on market intervention behaviour, and the objective and quantized reality boundary has stronger operability and adapts to market demand. According to quantitative information, government is regarded as rational economic man, and then the cost and benefits of governmental intervention behaviour is calculated, thus to be used to judge if the governmental intervention is necessary, feasible as well as how the effect is. To be specific, the confirmation of economic law behaviour norms can accord the following three points.

4.1 TIME AND SPACE DIMENSIONS

In practice of Chinese socio-economic transformation, concept of state intervention and market regulation both play functions, that is, market needs the economic intervention of state. Meanwhile, the concept and value orientation of state on economic intervention changes. That is, the state established intervention on the basis of respecting market rule. For instance, the state transforms the control method dominated in direct executive orders into the current control method dominated in market-oriented means; in different development stages, public power exertion will differ in range and strength as well as its functions and role positioning.

First, the national intervention boundary and economic development stage is closely related. In the initial stage of economy, government greatly increases the investment on environment and infrastructure in order to stimulate the rapid growth of economy. The total amount of governmental expenditure increases greatly, and the economic construction investment is dominated structurally; in medium term of economic development, the role of government slowly transforms to the extender of investment, the investment ratio may decreases and the direction transforms from economic construction to national education and health care. With the constant development of economy, the predominating direction of country will constantly improve to the development of market economy.

Secondly, the regional difference is also the content that should be considered by economic law behavioural norms. Regional difference not only reflects the disunion of economic development level and the different intervention function of government, but also requires economic law to adjust intervention range and degree according to different development.

4.2 TRANSFORMING FROM QUANTITATIVE CHANGE TO QUALITATIVE CHANGE

Market economy in our country from formation tom maturity experiences the process from administration dominating economy to full playing of market mechanism, meanwhile, reflects the transforming from quantity to quality of the intervention of economic law on economy. In the initial stage of market economy, government action shoulders productivity function, especially the fiscal expenditure even dominates the development of national economy; in economic mature stage, the dominant position of government should give way to market rule and retire after economic life, since too many government intervention will inhibit economic development rather than generate great function and significance on the regulation and development of market economy. In the real world, government should concern the transformation of the effect of fiscal expenditure on economic growth from quantity to quality, and then regulate economic law boundary norms relatively in order to make it adapt to the development of market economy.

4.3 OBSERVATION DATA AND FORECASTING DEVELOPMENT

Observation of cost of economic law behaviour is a key problem of economic law behaviour boundary. Economic law behaviour boundary, as one mean of state interfering market economy, affects the future economy, people's life and development and trend of social life. Its objective analysis and argument on economic performance is essential. Data can verify the stage that the economic development locates, market demand, situation of government demand and whether the intervention behaviour is conforming to expectation. Only by analyzing whether the norms of economic law behaviour is conforming to the development of market economy based on quantitative analysis can it be close to reality in accuracy.

Economy has certain predictability. In March, 2012, Prime Minister Wen announced that the economic growth objective in 2012 was 7.5%. Most analysts thought that, this conclusion was too cautious. However, it was proved when the growth of Chinese economy in the second quarter decreased from 8.1% in the first quarter to 7.6%. Economic prediction is made based on the full master of Chinese economic situation, economic rule and the current economic data. Through analyzing data and grasping rule, we predict the trend of economic development more accurately.

5 Conclusions

The research on the cost of economic law behavior is aimed at dynamics, those us, to look for the set factors of economic law behaviour norms in an absolute no balanced state. With the development of market economy, economical law behaviour norms with analysis of objective data constantly changes errors in it, which makes the market benefit and financial expenditure proceed in economic law behaviour that is conforming to market.

Acknowledgement

This paper belongs to the project supported by the planning projects of Zhejiang philosophy and social science (14SWH10YB)" Research on the legal security system in Zhejiang province's drinking water source protection under the background of the five water treatment", and is supported by the subject of philosophy and social science planning of Shaoxing City (125J113)" Research on Zhejiang province's drinking water source protection legal system in the context of the five water treatment".

References

[1] Zhang Y 2012 Empirical test on the effect of structure of fiscal expenditure on economic growth *Journal of Zhengzhou University (Philosophy and Social Science Edition)* **45**(1) 81-5 (*in Chinese*)
[2] Zheng S 2012 Empirical test on the effect of fiscal expenditure structure distortion on regional economic growth based on provincial panel data in eastern, middle and western regions *Review of Economy and Management* (4) 87-91 (*in Chinese*)
[3] Zhang Y 2010 Crowding-out effect of Chinese fiscal policy based on the empirical test on 1952~2008 Chinese annual data *Journal of Financial Research* (1) 58-65 (*in Chinese*)
[4] Li J, Zhang S 2011 Amery curve effect of government scale on economic growth based on the empirical test on Chinese provincial panel data *Shandong Economy* (6) 37-42 (*in Chinese*)
[5] Liu C 2011 The Impact of Macro-economic Policies on China's Pro-poor Growth *Research of Financial and Economic Issues* (4) 17-26 (*in Chinese*)
[6] Mo Y, Zhang Z 2011 The impact of China's fiscal expenditure on income since reform and opening-up *Reform of Economic System* (2) 122-6 (*in Chinese*)
[7] Bai Y, Yu H 2013 Research on allocation structure of fiscal expenditure on three -level education in China *Journal of Finance and Economics Theory* (2) 47-54 (*in Chinese*)
[8] Chen Y 2011 Multiplier effect of positive fiscal policy *Proceedings of International Conference on Engineering and Business Management (EBM2011)* 3530-3
[9] Zhang B 2010 The screening of economic law behavior boundary and its error correction – a kind of recognition in social conflict perspective *Productivity Research* 106-9 (*in Chinese*)

Author

Jindong Chen, born in 1978, Zhejiang Province of China.

Current position, grades: lecturer.
University studies: Master's degree of law, Shanxi University in 2007.
Scientific interest: economic law, environment and resources protection law.

Incentive system of skill-oriented dispatched employees based double moral hazard

Yanli Yang*, Xianyu Wang

School of business, Sichuan University, Chengdu, 610000, Sichuan, China

Received 1 July 2014, www.cmnt.lv

Abstract

As market competition gets increasingly fierce, the elastic way of employment arouses great concern among more and more enterprises, which results in rapid development of labour dispatch. Labour dispatch is different from the traditional way of employment and it involves the three parties of the employer, the accepting entity and the dispatched employee with the striking feature of asymmetric information. From the perspective of the employer, this paper analyses the incentives of the three parties in labour dispatch while taking the employer as the first principal and the accepting entity as the second principal in consideration of the factors that the intermediary agent's training will influence the quality of the employee. It verifies through models the sharing ratio paid by the accepting entity to the employer, the sharing ratio paid by the employer to the dispatched employee and the fixed wage paid by the accepting entity to the dispatched employee in expectation of offering concrete suggestions to the practice of enterprises.

Keywords: Labour Dispatch, Agent's Effort, Double Programming, Sharing Ratio

Introduction

With the development of global economy and the intensity of market competition, many enterprises endeavour to save business costs and enhance profit margins in order to keep and improve the core competitiveness. Different from the focus on the strict division of labour and centralization of functional authority under the background of industrialized production, the rapidly changing market requires that there should be more flexibility of the function and communication of the enterprise's various systems. Under such circumstances, flexible way of employment is widely adopted. On a global scale, although the elastic way of employment is of a small scale, it develops very rapidly.

Labour dispatch is that the dispatching unit signs the dispatching agreement with the accepting entity in accordance with the demands of the accepting entity and dispatch the employee with labour contract relations to the accepting entity so as to establish a special labour relation that the dispatched employee provides labour under the management of the accepting entity while the dispatching unit gets the dispatching payment from the accepting entity and pay the dispatched employees labour remuneration. Through the research on the situation of labour dispatch in Chicago, Peck and Theodore [1] find that this business is becoming divided, from which it derives the short-termed, low skilled and low-waged low-end service and the high-end service with a long-term cooperation and a close relation to the inner flow of the client's organization.

Labour dispatch is different from the traditional dual employment relationship, and it involves the three parties of the dispatching unit, the accepting entity and the dispatched employee. In general, the accepting entity does not know enough about the dispatching unit and the credit level and ability of the dispatched employee, which leads to the obvious asymmetric information. In the operation of labour dispatch, the training effort of the dispatching unit will influence the ability of the dispatched employee, and the ability and effort of the dispatched employee will directly influence the actual output of the accepting entity. Therefore, the double effort of the dispatching unit and the dispatched employee will influence the revenue of the accepting entity. This paper makes an analysis of the incentives of skill-oriented dispatched employees based on the bi-level program (i.e. when the work ability is stronger than the work effort) while taking the dispatching unit as the first principal and the accepting entity as the second principal under the circumstances of double moral hazard. The sharing revenues of the three parties are manifested through the models. The author wishes it a valuable reference to the practice of the enterprises.

Related literature

Labour dispatch involves the three parties of the dispatching unit, the accepting entity and the dispatched employee. The dispatching agreement is signed between the dispatching unit and the accepting entity and the employment contract is signed between the dispatching unit and the dispatched employee. Thus, the dispatching unit serves as the bridge in the business of labour dispatch. Literature from the perspective of the accepting entity mainly focuses on the motivation of choice of the accepting entity, such as literature [2] to [6] explaining the enterprise's motivation of the employment from the economic and social function. Literature from the perspective of the dispatched employee mainly focuses on the psychological contract and organization involvement, such as

* ***Corresponding author*** e-mail: yydouble@sohu.com

literature [7] to [9] discussing the relationship between psychological contract and attitude and action of the casual workers. Based on a study on 1091 dispatched workers and regular workers from 12 automobile factories in 6 provinces, Xie Yuhua et al. [10] analyses the difference of employee – organization relationship between the dispatched workers and regular workers. Based on a field survey of telecommunications in Jilin Province, Wang Hongyu [11] analyses the features of organizational justice of the dispatched employees in order to probe into the idea and effective ways to solve the organizational injustice. Most of the existing literature focuses on the practical experience of the labour dispatch unit instead of the theoretical research.

Moral hazard is a case of information asymmetry between the principal and the agent. Cooper [12] at first studies the problem of double moral hazard, proposing to encourage the two parties to reveal their actions by means of paying insurance money. Romano [13] makes a study on the double moral hazard model of linear contract. Al-Najjarz [14] expands the theory of linear contract in two-sided moral hazard to the situation of multiple agents. Under the background of the enterprise's incentives to the workers, considering that both the principal and the agent are risk averse, Agrawal [15] thinks that revenue-sharing contracts can make both parties' effort level reach the optimality. Tsoulouhas [16] applies the model of multiple agents' double moral hazard into the field of pricing the labour. De Janvry [17] makes a study on the sharing contract of labour revenue with moral hazard, manifesting those two-sided incentives of the employer and the employee can be obtained through making the sharing contract of labour revenue.

The existing research on double moral in the labour market focuses on the traditional dual relationship between the employer and the employee, while few focus on the study of the three parties of labour dispatch. Moreover, in the relationship of multiple agents, there is mainly paralleled competition or cooperation among the agents. However, there is no paralleled relationship among the multiple agents in labour dispatch. The effort of the dispatching unit will influence the working ability of the dispatched employees, and then the output level of the accepting entity. Therefore, the relationship among the multiple agents is the new principal-agent relationship.

3 Description of the models

This paper mainly discusses the incentives among the dispatching unit, the accepting entity and the dispatched employee in labour dispatch. The operation process in reality is that the accepting entity signs the agreement with the dispatching unit, and the dispatching unit signs the contract with the dispatched employee. The dispatching unit undertakes the training of the employees in expectation of improving their working ability, and then dispatches qualified employees to the accepting entity. The accepting entity pays some money to the employees and the dispatching unit according to the achievement of the dispatched employees. This paper mainly makes a study on the incentives of skill-oriented dispatched employees on the basis of double moral hazard (i.e. when the work ability is larger than the elastic coefficient work effort). For the convenience of research, following assumptions are made about the models:

1. Suppose that $f(\mathrm{e})$ is the output function of effort e, according to the actual situation, it should meet the requirement $f'(\mathrm{e}) \succ 0$, signifying that the more effort the dispatched employee and the dispatching unit make, the larger the output of the accepting entity. If the equation is $f''(\mathrm{e}) \le 0$, it signifies that the effort's marginal output is digressive (The equal sign signifies the marginal output is invariant). To simplify the assumption, here let the equation be $f_1(\mathrm{e}_1)=f_2(\mathrm{e}_2)=\mathrm{e}$.

2. Let $f_2(\mathrm{e}_2)$ be the function of the improvement of the dispatched employee's quality influenced by the effort e_2 of the dispatching unit. Let the coefficient of the dispatching unit's training ability be A_2, and its effort degree is e_2, both of which will directly influence the coefficient of the dispatched employee's working ability A_1. Moreover, A_1 will be influenced by the employee's original working ability d, that is, $A_1 = df_2(A_2, \mathrm{e}_2)$, thus $A_1 = dA_2 e_2$. The cost function of the dispatching unit's effort is $\mathrm{C}_2(\mathrm{e}_2) = \frac{b_2 \mathrm{e}_2^{\,2}}{2}$, and here b_2 is the cost coefficient of the dispatching unit's effort.

3. Suppose that the output function of the dispatched employee's effort for the accepting entity is $f_1(\mathrm{e}_1)$, and the cost function of the dispatched employee's effort is $\mathrm{C}_1(\mathrm{e}_1) = \frac{b_1 \mathrm{e}_1^{\,2}}{2}$, and here b_1 is the cost coefficient of the employee's effort.

4. The dispatching unit's training effort will influence the dispatched employee's working ability, and the working ability and working effort of the dispatched employee will directly influence the output level of the accepting entity. That is to say, the dispatching unit's training effort and the dispatched employee's working effort will directly influence the employee's output level. In this paper, let the output be based on the simplified Cobb-Douglas function, and the working ability and the working effort of the dispatched employee will exert different degrees of influence on the output. Let the dispatched employee's output be

$$Q = kA_1 e_1^{\alpha} + \varepsilon = kdA_2 \mathrm{e}_2 e_1^{\alpha} + \varepsilon . \quad (1)$$

Here, let the elastic coefficient of the dispatched employee's working ability be 1, and α is the elastic coefficient of the dispatched employee's effort for output, and also $0 < \alpha < 1$ (for details see the reduction of the

value range of α attached to this paper); ε is a random variable, and let it be accordant to the normal distribution $(0,\sigma^2)$, and suppose that the average revenue of the accepting entity's product is m, that is, the total revenue of the accepting entity is Qm;

5. In order to obtain the job opportunity provided by the dispatching unit, the dispatched employee needs to pay the fee v to the accepting entity while signing the employment contract with the dispatching unit.
6. The dispatching unit's revenue is directly connected with the dispatched employee's output level. The accepting entity pays the dispatching unit according to the proportion of the dispatched employee's output and income μ_1, therefore the dispatching unit's revenue is $\mu_1 Qm$. The income of the dispatched employee consists of regular wage and performance bonus. The accepting entity pays regular wage ω to the dispatched employee. The dispatching unit pays performance bonus to the dispatched employee according to its proportion of revenue μ_2. Thus the dispatched employee's income is $\omega+\mu_1\mu_2 Qm$.
7. Suppose that the accepting entity, the dispatching unit and the dispatched employee are all risk neutral.
8. This paper analyses the dispatching unit's function as the intermediary agent from the perspective of the dispatching unit, consequently taking the dispatching unit as the first principal and the accepting entity as the second principal.

Based on the above assumptions, the utility functions of the accepting entity, the dispatching unit and the dispatched employee can be concluded respectively as follows:

$$\pi_{the\ emplyer} = (1-\mu_1)kdA_2e_2e_1^{\alpha}m-\omega, \tag{2}$$

$$\pi_{the\ accepting} = \mu_1(1-\mu_2)\ kdA_2e_2e_1^{\alpha}m-\frac{1}{2}b_2e_2^{\ 2}+v, \tag{3}$$

$$\pi_{the\ dispatched} = \mu_1\mu_2 kdA_2e_2e_1^{\alpha}m+\omega-\frac{1}{2}b_1e_1^{\ 2}-v. \tag{4}$$

In the relationship of the three parties in labour dispatch, the accepting entity is the principal, the dispatching unit's training effort will influence the dispatched employee's working ability, and the working ability and effort of the dispatched employee will directly influence the accepting entity's actual output. Therefore, attention should be paid to the accepting entity's pursuit of maximizing its self-interest, the dispatched employee's incentive compatibility constraint and participation constraint, and also the maximizing interest of the dispatching unit. This paper carries on the research from the perspective of the dispatching unit, consequently taking the dispatching unit as the first principal and the accepting entity as the second principal. Thus, the following bi-level programming model is concluded:

$$\max_{\{\mu_2,e_2\}}\ \pi_{the\ employer}=\mu_1(1-\mu_2)kdA_2e_2e_1^{\alpha}m-\frac{1}{2}b_2e_2^{\ 2}+v, \tag{5}$$

$$s.t.\ \max_{\{\mu_1\}}\ \pi_{the\ accepting}=(1-\mu_1)kdA_2e_2e_1^{\alpha}m-\omega, \tag{6}$$

$$s.t.\ \ e_1\in \arg\max\ \mu_1\mu_2 kdA_2e_2e_1^{\alpha}m+\omega-\frac{1}{2}b_1e_1^{\ 2}-v, \tag{7}$$

$$\mu_1\mu_2 kdA_2e_2e_1^{\alpha}m+\omega-\frac{1}{2}b_1e_1^{\ 2}-v\geq s_1. \tag{8}$$

Let $R=kdA_2m$.

From equation (3) we get:

$$e_1^{o}=\left(\frac{\alpha\mu_1\mu_2\,\mathrm{Re}_2}{b_1}\right)^{\frac{1}{2-\alpha}}. \tag{9}$$

Substitute Eq. (9) and Eq. (8) into Eq. (6), we obtain:

$$\pi_{the\ accepting}=\frac{\mathrm{R}^{\frac{2}{2-\alpha}}e_2^{\frac{2}{2-\alpha}}\alpha^{\frac{\alpha}{2-\alpha}}\mu_1^{\frac{\alpha}{2-\alpha}}\mu_2^{\frac{\alpha}{2-\alpha}}}{2b_1^{\frac{\alpha}{2-\alpha}}}(2-2\mu_1+2\mu_1\mu_2-\alpha\mu_1\mu_2)-s_1-v. \tag{10}$$

From (10) we get

$$\frac{\partial\pi_{the\ accepting}}{\partial\mu_1}=\frac{\mathrm{R}^{\frac{2}{2-\alpha}}e_2^{\frac{2}{2-\alpha}}\alpha^{\frac{\alpha}{2-\alpha}}\mu_2^{\frac{\alpha}{2-\alpha}}}{2b_1^{\frac{\alpha}{2-\alpha}}}\cdot$$

$$\left[\frac{\alpha}{2-\alpha}\mu_1^{\frac{2\alpha-2}{2-\alpha}}(2-2\mu_1+2\mu_1\mu_2-\alpha\mu_1\mu_2)+\mu_1^{\frac{\alpha}{2-\alpha}}(2\mu_2-2-\alpha\mu_2)\right]$$

Let $\frac{\partial\pi_{the\ accepting}}{\partial\mu_1}=0$, then:

$$\mu_1^{o}=\frac{\alpha}{2+\alpha\mu_2-2\mu_2}. \tag{11}$$

Substitute Eq. (9) and Eq. (11) into Eq. (5), we obtain:

$$\pi_{the\ employer}=\frac{\alpha^{\frac{2+\alpha}{2-\alpha}}\mathrm{R}^{\frac{2}{2-\alpha}}e_2^{\frac{2}{2-\alpha}}(\mu_2^{\frac{\alpha}{2-\alpha}}-\mu_2^{\frac{2}{2-\alpha}})}{\mathrm{b}_1^{\frac{\alpha}{2-\alpha}}(2+\alpha\mu_2-2\mu_2)^{\frac{2}{2-\alpha}}}-\frac{1}{2}b_2e_2^2+v. \tag{12}$$

From (12) we get

$$\frac{\partial\pi_{the\ employer}}{\partial\mu_2}=\frac{\alpha^{\frac{4}{2-\alpha}}\mathrm{R}^{\frac{2}{2-\alpha}}e_2^{\frac{2}{2-\alpha}}\mu_2^{\frac{2\alpha-2}{2-\alpha}}}{\mathrm{b}_1^{\frac{\alpha}{2-\alpha}}(2+\alpha\mu_2-2\mu_2)^{\frac{4-\alpha}{2-\alpha}}(2-\alpha)}(2+\alpha\mu_2-4\mu_2). \tag{13}$$

Let $\frac{\partial\pi_{the\ employer}}{\partial\mu_2}=0$, then: $\mu_2^{o}=\frac{2}{4-\alpha}$. (14)

From (13) we get

$$\frac{\partial^2 \pi_{the\ employer}}{\partial \mu_2^2} = \frac{\alpha^{\frac{4}{2-\alpha}} \mathrm{R}^{\frac{2}{2-\alpha}} e_2^{\frac{2}{2-\alpha}} \mu_2^{\frac{\alpha}{2-\alpha}}}{\mathrm{b}_1^{\frac{\alpha}{2-\alpha}} (2-\alpha)^2 (2+\alpha\mu_2-2\mu_2)^{\frac{4}{2-\alpha}}}$$
$$[(\frac{4\alpha-4}{\mu_2}+\alpha^2-4\alpha)(2+\alpha\mu_2-2\mu_2) \quad . \quad (15)$$
$$+(2+\alpha\mu_2-4\mu_2)(8-6\alpha+\alpha^2)]$$

From (14) and (15), we get

$$\frac{\partial^2 \pi_{the\ employer}}{\partial \mu_2^2}\Big|_{\mu_2=\mu_2^o} = -\frac{\alpha^{\frac{4}{2-\alpha}} \mathrm{R}^{\frac{2}{2-\alpha}} e_2^{\frac{2}{2-\alpha}} (\frac{2}{4-\alpha})^{\frac{\alpha}{2-\alpha}}}{\mathrm{b}_1^{\frac{\alpha}{2-\alpha}} (2-\alpha)^2 (\frac{4}{4-\alpha})^{\frac{2+\alpha}{2-\alpha}}} (\alpha-2)(\alpha-4) \quad . \quad (16)$$

From (12) we get

$$\frac{\partial \pi_{the\ employer}}{\partial e_2} = \frac{\alpha^{\frac{2+\alpha}{2-\alpha}} R^{\frac{2}{2-\alpha}} \mu_2^{\frac{\alpha}{2-\alpha}} (1-\mu_2)}{b_1^{\frac{\alpha}{2-\alpha}} (2+\alpha\mu_2-2\mu_2)^{\frac{2}{2-\alpha}}} \cdot \frac{2}{2-\alpha} e_2^{\frac{\alpha}{2-\alpha}} - b_2 e_2 \quad . \quad (17)$$

Let $\frac{\partial \pi_{the\ employer}}{\partial e_2}=0$, then

$$e_2^o = \frac{\alpha^{\frac{2+\alpha}{2-2\alpha}} R^{\frac{2}{2-2\alpha}} \mu_2^{\frac{\alpha}{2-2\alpha}} (1-\mu_2)^{\frac{2-\alpha}{2-2\alpha}} (\frac{2}{2-\alpha})^{\frac{2-\alpha}{2-2\alpha}}}{b_2^{\frac{2-\alpha}{2-2\alpha}} b_1^{\frac{\alpha}{2-2\alpha}} (2+\alpha\mu_2-2\mu_2)^{\frac{2}{2-2\alpha}}} \quad . \quad (18)$$

From (17) we get

$$\frac{\partial^2 \pi_{the\ employer}}{\partial e_2^2} = \frac{\alpha^{\frac{2+\alpha}{2-\alpha}} R^{\frac{2}{2-\alpha}} \mu_2^{\frac{\alpha}{2-\alpha}} (1-\mu_2)}{b_1^{\frac{\alpha}{2-\alpha}} (2+\alpha\mu_2-2\mu_2)^{\frac{2}{2-\alpha}}} \cdot \frac{2}{2-\alpha} \cdot \frac{\alpha}{2-\alpha} e_2^{\frac{2\alpha-2}{2-\alpha}} - b_2 \quad . \quad (19)$$

From (18) and (19), we get

$$\frac{\partial^2 \pi_{the\ employer}}{\partial e_2^2}\Big|_{e_2=e_2^o} = b_2 \frac{2\alpha-2}{2-\alpha} \quad . \quad (20)$$

From (13) we get

$$\frac{\partial^2 \pi_{the\ employer}}{\partial \mu_2 \partial e_2} = \frac{\alpha^{\frac{4}{2-\alpha}} \mathrm{R}^{\frac{2}{2-\alpha}} \mu_2^{\frac{2\alpha-2}{2-\alpha}} (2+\alpha\mu_2-4\mu_2)}{\mathrm{b}_1^{\frac{\alpha}{2-\alpha}} (2+\alpha\mu_2-2\mu_2)^{\frac{4-\alpha}{2-\alpha}} (2-\alpha)} \cdot \frac{2}{2-\alpha} e_2^{\frac{\alpha}{2-\alpha}} \quad . \quad (21)$$

From equation (14), (18) and (21) we get:

$$\frac{\partial^2 \pi_{the\ employer}}{\partial \mu_2 \partial e_2}\Big|_{\mu_2=\mu_2^o, e_2=e_2^o} = 0 \quad . \quad (22)$$

From above we get: $A = \frac{\partial^2 \pi_{the\ employer}}{\partial e_2^2}\Big|_{e_2=e_2^o} = b_2 \frac{2\alpha-2}{2-\alpha}$;

$$B = \frac{\partial^2 \pi_{the\ employer}}{\partial \mu_2 \partial e_2}\Big|_{\mu_2=\mu_2^o, e_2=e_2^o} = 0 \quad .$$

$$C = \frac{\partial^2 \pi_{the\ employer}}{\partial \mu_2^2}\Big|_{\mu_2=\mu_2^o} = -\frac{\alpha^{\frac{4}{2-\alpha}} \mathrm{R}^{\frac{2}{2-\alpha}} e_2^{\frac{2}{2-\alpha}} (\frac{2}{4-\alpha})^{\frac{\alpha}{2-\alpha}}}{\mathrm{b}_1^{\frac{\alpha}{2-\alpha}} (2-\alpha)^2 (\frac{4}{4-\alpha})^{\frac{2+\alpha}{2-\alpha}}} (\alpha-2)(\alpha-4) \quad . \quad (23)$$

According to the necessary and sufficient condition to determine the binary function's extremum, only when $AC-B^2>0$, the binary function has the extreme point; only when $A<0$, it has the maximum value

From $A<0$ we get: $\begin{cases}2\alpha-2>0\\2-\alpha<0\end{cases}$ or $\begin{cases}2\alpha-2<0\\2-\alpha>0\end{cases}$ that is:

$\alpha>2$ or $\alpha<1$. (24)

From $C<0$, we get: $\begin{cases}\alpha-2>0\\\alpha-4>0\end{cases}$ or $\begin{cases}\alpha-2<0\\\alpha-4<0\end{cases}$ that is:

$\alpha>4$ or $\alpha<2$. (25)

Simultaneous equations (24) and (25), then:

$\alpha>4$ or $\alpha<1$. (26)

Since μ_1 and μ_2 are the sharing ratios, then the ranging scope is $[0,1]$.

From (12) we get: $\mu_2^o = \frac{2}{4-\alpha}$, thus $\alpha<2$. (27)

From equations (9) and (12) we get:

$$\mu_1^{\ o} = \frac{\alpha}{2+\alpha(\frac{2}{4-\alpha})-2(\frac{2}{4-\alpha})} = \frac{\alpha(4-\alpha)}{4} \quad . \quad (28)$$

Thus $0<\alpha<4$. (29)

Consider equations (26), (27) and (29) at the same time we get:

If and only if $0<\alpha<1$, this model achieve it optimum.

Proposition 1:

Under the double moral hazard, the sharing ratios, optimal effort degree and revenues of the dispatching

unit, the accepting entity and the dispatched employee are respectively as follows:

$$\mu_2^o = \frac{2}{4-\alpha};\ \mu_1^o = \frac{\alpha(4-\alpha)}{4};\ e_2^o = \frac{\alpha^{\frac{2+\alpha}{2-2\alpha}} R^{\frac{2}{2-2\alpha}}}{2^{\frac{2}{2-2\alpha}} b_2^{\frac{2-\alpha}{2-2\alpha}} b_1^{\frac{\alpha}{2-2\alpha}}};$$

$$e_1^o = \frac{\alpha^{\frac{3}{2-2\alpha}} R^{\frac{2}{2-2\alpha}}}{2^{\frac{2}{2-2\alpha}} b_2^{\frac{1}{2-2\alpha}} b_1^{\frac{1}{2-2\alpha}}};$$

$$\omega^o = s_1 + v + \frac{\alpha^{\frac{4+2\alpha}{2-2\alpha}} R^{\frac{4}{2-2\alpha}}}{2^{\frac{6-2\alpha}{2-2\alpha}} b_2^{\frac{2}{2-2\alpha}} b_1^{\frac{2\alpha}{2-2\alpha}}}(\alpha-2);$$

$$\pi_{\text{the employer}} = \frac{\alpha^{\frac{4+2\alpha}{2-2\alpha}} R^{\frac{4}{2-2\alpha}}(1-\alpha)}{2^{\frac{6-2\alpha}{2-2\alpha}} b_2^{\frac{2}{2-2\alpha}} b_1^{\frac{2\alpha}{2-2\alpha}}} + v;$$

$$\pi_{\text{the accepting}} = \frac{\alpha^{\frac{4+2\alpha}{2-2\alpha}} R^{\frac{4}{2-2\alpha}}}{2^{\frac{6-2\alpha}{2-2\alpha}} b_2^{\frac{2}{2-2\alpha}} b_1^{\frac{2\alpha}{2-2\alpha}}} \cdot \frac{2(2-\alpha)}{\alpha} - s_1 - v \qquad (30)$$

Also, from the above we get:

$$\frac{\partial \mu_2^o}{\partial \alpha} = \frac{2}{(4-\alpha)^2} > 0\ ;\ \frac{\partial \mu_1^o}{\partial \alpha} = \frac{2-\alpha}{2} > 0\,. \qquad (31)$$

Proposition 2:

The sharing ratios paid by the accepting entity to the dispatching unit and paid by the dispatching unit to the dispatched employee will increase with the increase of the dispatched employee's effort degree.

4 Conclusion

This paper makes an analysis of the incentives of skill-oriented dispatched employees based on the bi-level program (i.e. when the work ability is stronger than the work effort) while taking the dispatching unit as the first principal and the accepting entity as the second principal under the circumstances of double moral hazard. The sharing revenues of the three parties are manifested through the models. The author wishes it a valuable reference to the practice of the enterprises.

References

[1] Peck J, Theodore N 2001 Contingent Chicago: Restructuring the Spaces of Temporary Labour *International Journal of Urban and Regional Research* **25**(3) 471-95
[2] Rubery J, Earnshaw J, Marchington M, Fang Lee Cooke, Vincent S 2002 Changing organizational forms and the employment relationship *Journal of Management Studies* **39**(5) 645-72
[3] Kalleberg A L, Reynolds J, Marsden P V 2003 Externalizing Employment: Flexible Staffing Arrangement in US Organizations *Social Science Research* **32** 535-52
[4] Bas Koene, Jaap Paauwe, Groenewegen J 2004 Understanding the development of Temporary Agency Work in Europe *Human Resource Management Journal* **14**(3) 53-73
[5] Uzzi B, Barsness Z I 1998 Contingent Employment in British Establishments: Organizational Determinants of the Use of Fixed-term Hires and Part-time Workers *Social Forces* **76**(3) 967-1005
[6] Lautsch B A 2002 Uncovering and Explaining Variance in the Features and Outcomes of Contingent Work *Industrial and Labour Relations Review* **56**(1) 23-43
[7] Budd J W 2006 *Employment with a Human Face: Balancing Efficiency, Equity and Voice* ILR Press :New York
[8] Hirschman A O 1970 *Exit, Voice and Loyalty* Harvard University Press: Cambridge, MA
[9] Guest D 2004 Flexible Employment Contracts, the Psychological Contract and Employee Outcomes: an Analysis and Review of the Evidence *International Journal of Management Reviews* **5**(1) 1-19
[10] Xie Yuhua, Xiao Qiaoling, Anita Chan, Zhao Wei, Yiu Por Chen 2013 A Comparative Study on the Organization Relations of the Dispatched Employees and the Regular Employees *Comparative Economic & Social Systems* **169**(5) 183-92
[11] Wang Hongyu 2012 On the Structure Characteristics and Countermeasures of the Dispatched Employee's Organizational Justice *Social Science Front* **5** 267-8
[12] Cooper R, Ross T W 1985 Product Warranties and Double Moral Hazard *The RAND Journal of Economics* **16**(1) 103-13
[13] Romano R E 1994 Double Moral Hazard and Resale Price Maintenance *The RAND Journal of Economics* **25**(3) 455-66
[14] Al-najjar N I 1997 Incentive Contracts in Two-sided Moral Hazards with Multiple Agents *Journal of Economic Theory* **74**(1) 174-95
[15] Agrawal P 2002 Double Moral Hazard, Monitoring and the Nature of Contracts *Journal of Economics* **75**(1) 33-61
[16] Tsoulouhas T 1999 Do Tournaments Solve the Two-sided Moral Hazard Problem? *Journal of Economic Behaviour & Organization* **40**(3) 275-94
[17] De Janvry, Sadoulet A E 2007 Optimal Share Contracts with Moral Hazard on Effort and in Output Reporting: Managing the Double Laffer Curve Effect *Oxford Economic Papers* **59**(2) 253-74

Authors

Yanli Yang, born on December 19, 1982, China

Current position, grades: doctor student at Sichuan university, China.
University studies: master degree in management science and engineering from Hebei engineering university, China in 2007.
Scientific interest: human resource and incentive theory.

Xianyu Wang, born in May, 1947.China

Current position, grades: researcher at Sichuan university, China.
Scientific interest: incentive theory and supply chain management.

Comparative study of personalizing recommender systems based on shopping system

Zhihang Tang, Zhonghua Wen

School of Computer and Communication, Hunan Institute of Engineering Xiangtan 411104, China

Abstract

Making choices is an integral part of everyday life; Recommender systems facilitate decision-making processes through informed assistance and enhanced user experience. To aid in the decision-making process, recommender systems use the available data on the items themselves, Personalized recommender systems subsequently use this input data, and convert it to an output in the form of ordered lists or scores of items in which a user might be interested. These lists or scores are the final result the user will be presented with, and their goal is to assist the user in the decision-making process. Recommender systems facilitate making choices, improve user experience, and increase revenue, therefore should be easily accessible for deployment to interested parties. The implementation of recommender systems in RapidMiner has been additionally simplified through the Recommender Extension.

Keywords: shopping system, personalized recommender systems, RapidMiner, decision-making

1 Introduction

Recommender systems have become extremely common in recent years, and are applied in a variety of applications. The most popular ones are probably movies, music, news, books, research articles, search queries, social tags, and products in general. However, there are also recommender systems for experts, jokes, restaurants, financial services, life insurance, persons (online dating), and twitter followers [1].

Recommender systems typically produce a list of recommendations in one of two ways – through collaborative or content-based filtering [2]. Collaborative filtering approaches build a model from a user's past behavior (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users; then use that model to predict items (or ratings for items) that the user may have an interest in [3].

Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties. These approaches are often combined (see Hybrid Recommender Systems).

Each type of system has its own strengths and weaknesses. In the above example, Last.fm requires a large amount of information on a user in order to make accurate recommendations. This is an example of the cold start problem, and is common in collaborative filtering systems. While Pandora needs very little information to get started, it is far more limited in scope (for example, it can only make recommendations that are similar to the original seed). Recommender systems are a useful alternative to search algorithms since they help users discover items they might not have found by themselves. Interestingly enough, recommender systems are often implemented using search engines indexing non-traditional data. Montaner provides the first overview of recommender systems, from an intelligent agents perspective [4]. Adomavicius provides a new overview of recommender systems [5]. Herlocker provides an additional overview of evaluation techniques for recommender systems [6]. They also provide a literature survey on research paper recommender systems [7].

Recommender system is an active research area in the data mining and machine learning areas. Some conferences such as RecSys, SIGIR, KDD have it as a topic.

2 Basic conception

2.1 RECOMMENDATION OPERATORS

The Recommender Extension has a total 26 recommendation operators. These operators are grouped in the following categories: Item Recommendation, Item Rating Prediction, and Recommender Performance.

Item recommendation operators operate over large matrices that contain information about which user consumed which item. These input matrices often contain large numbers of empty entries, and can thus be used in a more space-efficient way. We describe the appropriate format used for efficient data loading and storage in the next subsection.

2.2 DATA

Typical recommender systems, operating on user usage data, are built on top of large matrices called utility matrices. These matrices usually contain elements from a limited set of numbers, for example from 0, 1, 2, 3, 4, 5, where 0 typically denotes that the user had no interaction with the item, and the rest describes the level of that

interaction in a form of rating. Due to a large number of both users and items, these matrices are typically very large. In addition, since users mostly consume a very small portion of items out of the total number of items, these matrices tend to contain a lot of zeros-they tend to be very sparsely populated. This is why special data structures for handling sparse data need to be implemented. In RapidMiner, we can use AML reader operators to read such datasets. Input datasets used to learn a recommender system model must be formatted in two columns; for example, the first column can contain user IDs, while the second can contain item IDs. Attributes names, and their positioning can be arbitrary. Prior to applying recommenddation operators to input datasets, proper roles have to be set for these attributes. Any additional attributes will not be considered. An example of an AML, and a related DAT file for item recommendation operators is given in Fig. 1.

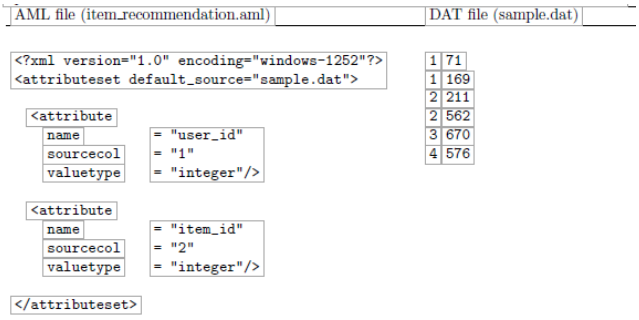


FIGURE 1 An example of an AML and a related DAT file for item recommendation operators

The recommender system datasets used throughout this paper consists of content and collaborative data. Content data was taken from the shopping system.

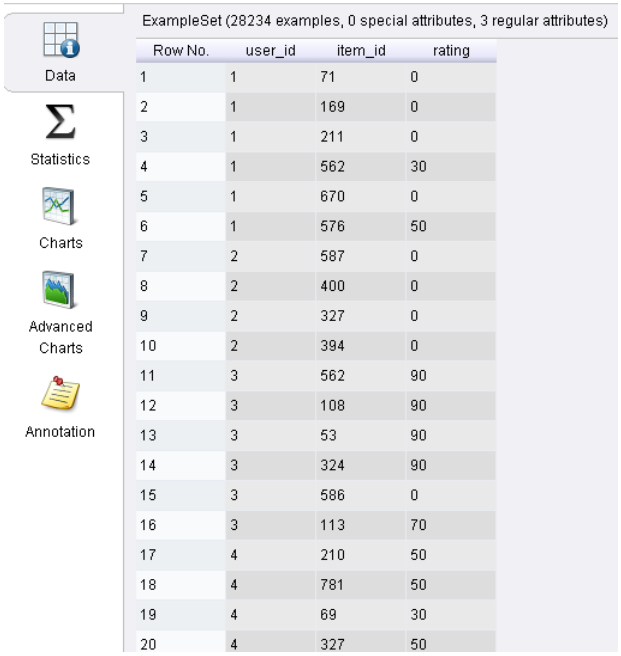
The content dataset described contains the following content attributes for each item:

brand ID: a unique integer that represents a lecture;

brand name: a text string containing a name of a particular lecture;

brand description: a text string denoting a description of a particular lecture.

A particular item identifier is denoted by the small letter i and the set of all items is denoted by the capital letter I. Collaborative data contains synthetic click streams of users, where each click stream is a sequence of items viewed by a particular user in some time interval. In the following text, we refer to the synthetic users as users. A particular user identifier is denoted by the small letter u and the set of all users is denoted by the capital letter U. Click streams are transformed into the sparse matrix A, which is called the usage matrix. The non-zero elements of the usage matrix (A (i, u)) tell us that the item i was consumed by the user u. Using this dataset, we construct collaborative and content recommender systems in the following sections. The collaborative recommender systems rely on the usage matrix A while the content recommender systems rely on items textual descriptions. We can get Fig. 2.



Row No.	user_id	item_id	rating
1	1	71	0
2	1	169	0
3	1	211	0
4	1	562	30
5	1	670	0
6	1	576	50
7	2	587	0
8	2	400	0
9	2	327	0
10	2	394	0
11	3	562	90
12	3	108	90
13	3	53	90
14	3	324	90
15	3	586	0
16	3	113	70
17	4	210	50
18	4	781	50
19	4	69	30
20	4	327	50

FIGURE 2 Initial data from shopping system

2.3 COLLABORATIVE-BASED SYSTEMS

The main idea of collaborative recommendation approaches is to use information about the past behavior of existing users of the system for predicting which item the current user will most probably like and thus might consume. Collaborative approaches take a matrix of given user-item ratings or viewings as an input and produce a numerical prediction indicating to what degree the current user will like or dislike a certain item, or a list of n recommended items. The created list should not contain items the current user has already consumed.

Neighborhood-based recommender systems work by counting common items two users have viewed for every pair of users in the system, or the number of common users that viewed the same pair of items. Using this count, similarity between two users or items is calculated. Neighborhood systems use intuition that two users who have viewed a large number of common items have similar tastes. That information can be used to recommend items that one user consumed and the other one did not. We are interested in finding pairs of users having the most similar taste, or pairs of items having the most users that viewed both items. Those pairs of users/items are called "the closest neighbors". We describe two main approaches of the neighborhood-based recommender systems: user and item-based nearest neighbor recommendation.

2.3.1 User-based nearest neighbor recommendation

Given a user-item viewing matrix and the ID of the current user as input, identify other users having similar past prefe-

rences to those of the active user. Subsequently, for every product the active user has not yet consumed, compute prediction based on the product usage of the selected user subset. These methods assume that users, who have previously shared similar tastes, will share similar tastes in the future, and that user preferences remain stable and constant over time.

To calculate similarity between users, two typical similarity measures are used: the Pear- son correlation and the Cosine correlation [6]. In our item recommendation problem we used cosine correlation as a similarity measure. Typically, we do not consider all users in the database when calculating user similarity, rather the k most similar ones.

2.3.2 Item-based nearest neighbor recommendation

When dealing with large problems, consisting of millions of users, user-based collaborative filtering approaches lead to increased memory usage and execution time. Since the system is required to calculate a large volume of potential neighbors, it becomes impossible to compute predictions in real time. In some cases, the number of users dominates the number of items in the system so it would be natural to try to use items for making recommendations. That is the reason for creating a second neighborhood-based recommender system based on items instead of users.

As opposed to the user-based approach, the item-based recommendation approach computes prediction using the similarity between items. We use a cosine similarity measure, as we did in the user-based approach. Likewise, as in the user-based approach, we use k-nearest neighbors, i.e. the k most similar items for prediction.

3 Personalizing Recommender Systems

Collaborative recommender operators use the user-item matrix to build a recommendation model. This user-item matrix is presented as an example set of user-item pairs describing user consumption history. The recommendation model built with this matrix is used to recommend items to users from a query set. The query set is an example set containing identification numbers of users for which we want to make recommendations. For each user in the query set we recommend only the items not consumed by this user. Figure 3 depicts a basic collaborative recommender operator workflow.

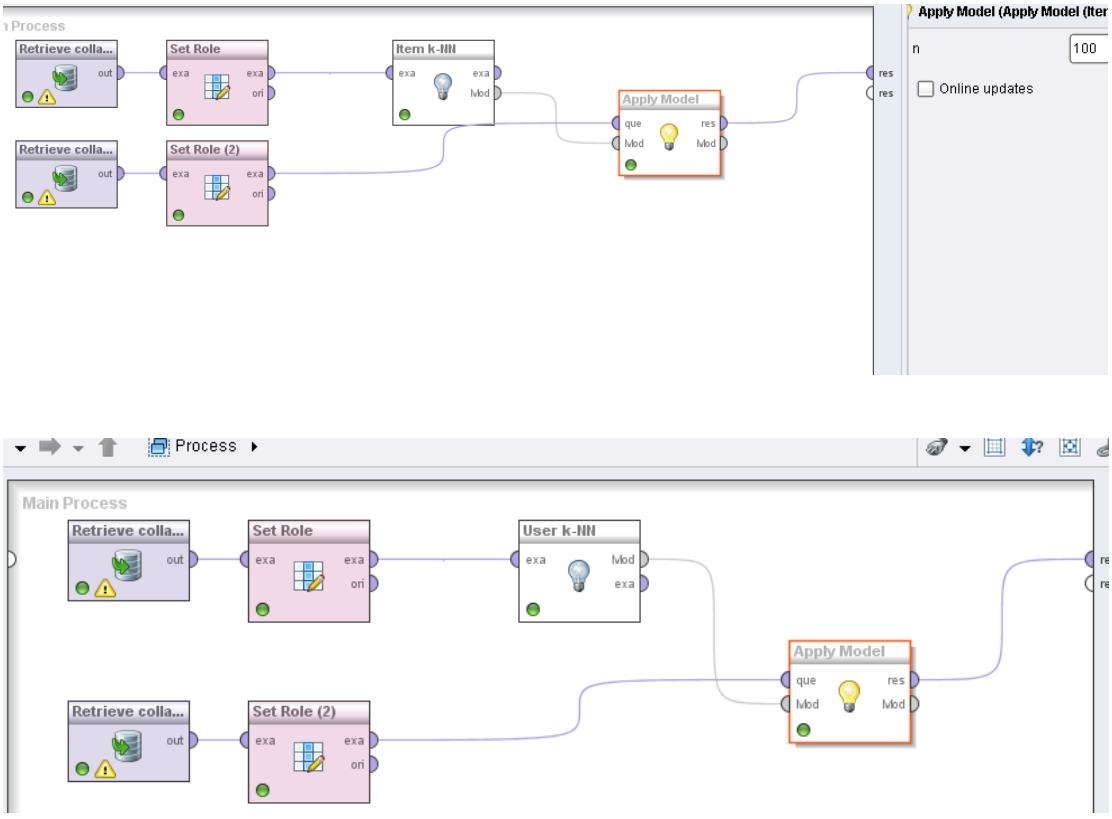


FIGURE 3 Two examples of an item recommendation workflow

The Recommended results shown in Figure 4.

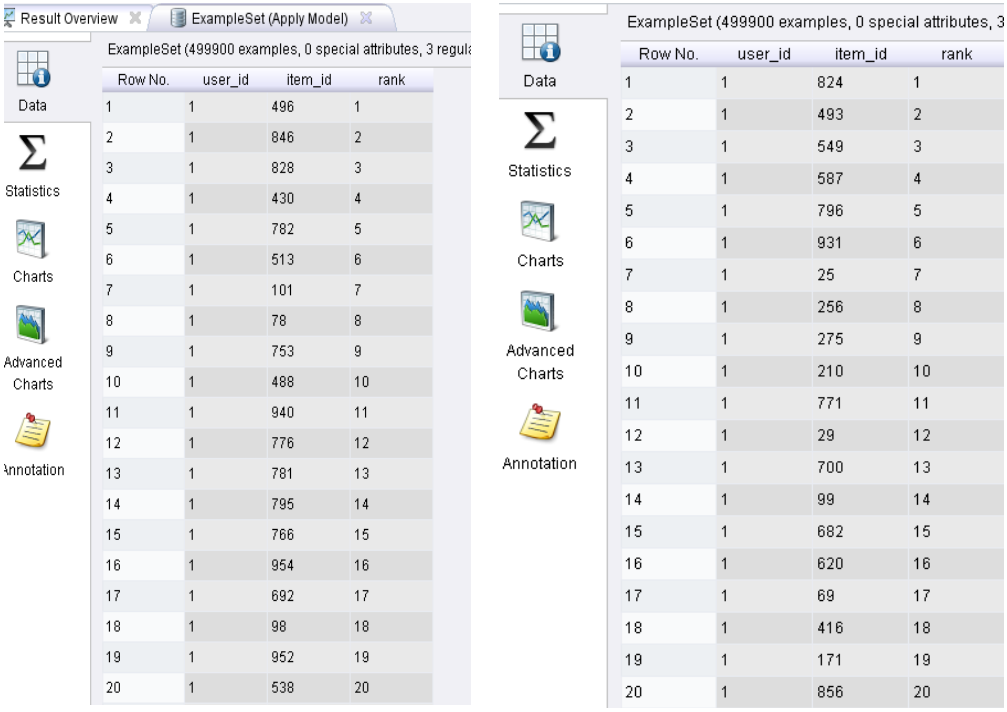


FIGURE 4 Comparative study of the recommended results

In the item recommendation workflow, the first two operators read the train and the query example sets using the Read AML operators (1,4). Following, the appropriate roles are set to attributes using the Set Role operator (2). The user identification role was set to user id attribute and item identification role to item id attribute. Data attributes can have arbitrary names but roles for those attributes must be set. Next, we use the train data with the appropriately set roles to train an Item k-NN model (3). At this point we can use our trained model to recommend new items to users in the query set using the Apply Model operator (6). Prior to model application, the user identification role was set for the query set (5). The Apply Model operator (6) returns an example set containing the first n ranked recommendations for every user in a query set. In Figure 3 we have seen how to make recommendations for particular users. In the following figure, Figure 4, we show how to measure performance of a recommendation model.

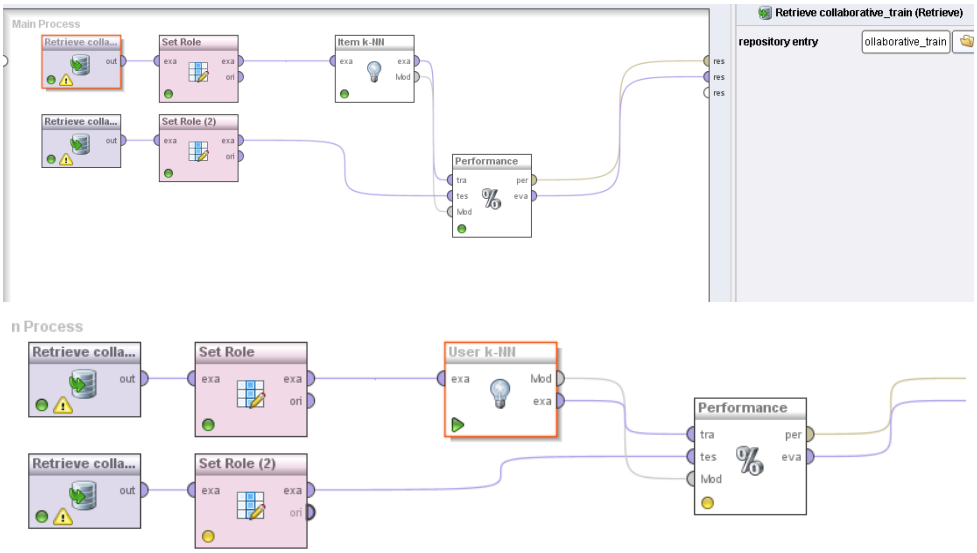


FIGURE 5 Comparative study of measuring performance of a recommendation model

The data management part of the workflow for measuring recommender model performance in Figure 5 is the same as in Figure 3. We use the Read AML operators (1,4) to load the data input, and the Set Role operators (2,5) to set the appropriate roles. In this workflow we use the test data (4) containing two attributes, the user id and the item id attribute and we set user identification and item identification roles to those at-tributes, respectively. The difference from the previous workflow is the need to calculate the performance of our built recommendation model (3). We use the Performance operator (6) to measure standard recommendation error measures we previously defined: AUC, Prec@k, NDCG, and MAP. The Performance operator (6) returns a performance vector and an example set containing performance measures. This enables a user to choose which format suits his or her needs. We can get Figure 6.

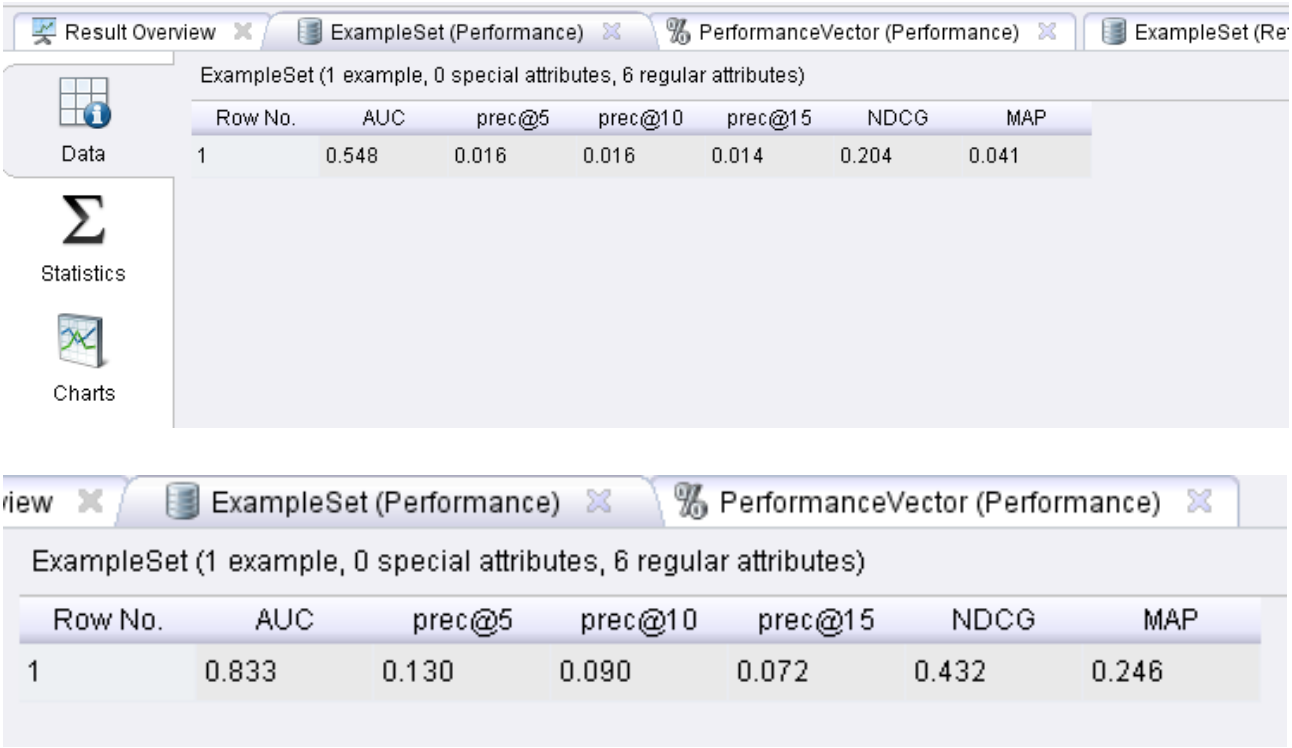


FIGURE 6 Comparative study of the performance of recommender systems

4 Conclusions

Recommender systems became essential in an information- and decision-overloaded world. They changed the way users make decisions, and helped their creators to increase revenue at the same time. Bringing recommender systems to a broader audience is essential in order to popularize them beyond the limits of scientific research and high technology entrepreneurship. The goal of the Recommender Extension for RapidMiner and this paper was to bring recommenders to a broad audience, in a theoretical, practical, and above all, application way.

In this paper we presented recommender systems and their different techniques: collaborative filtering, content-based recommender systems, and hybrid systems. We presented the advantages and disadvantages of each of those systems and demonstrated how they could be implemented easily in RapidMiner. The application of recommender systems outlined was just a small introduction to the possibilities of the extension. We hope you will use the knowledge obtained through this paper in your own applications, problems, and businesses, and that recommender systems will assist you in reaching quality, informed decisions.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61272295) and 2014 science and technology plan of Hunan province (No. 2014GK 3157).

References

[18] Gupta P, Goel A, Lin J, Sharma A, Wang D, Zadeh R 2013 WTF: The Who to Follow Service at Twitter *Proceedings of the 22th International World Wide Web Conference (WWW 2013)* 505-14
[19] Jafarkarimi H, Sim A T H, Saadatdoost R 2012 A Naïve Recommendation Model for Large Databases *International Journal of Information and Education Technology* **2**(3) 216-9
[20] Melville P, Sindhwani V 2010 Recommender Systems *Encyclopedia of Machine Learning*
[21] Montaner M, Lopez B, de la Rosa J L 2003 A Taxonomy of Recommender Agents on the Internet *Artificial Intelligence Review* **19**(4) 285-330
[22] Adomavicius G, TuzhilinA 2005 Toward the Next Generation of Rec-ommender Systems: *IEEE Transactions on Knowledge and Data Engineering* **17**(6) 734-49
[23] Beel J, Langer S, Genzmehr M, Gipp B 2013 A Comparative Analysis of Offline and Online Evaluations and Discussion of Research Paper Recommender System Evaluation *Proceedings of the Workshop on Reproducibility and Replication in Recommender Systems Evaluation (RepSys)* 7-14
[24] Gunawardana A, Shani G 2009 A survey of accuracy evaluation metrics of recommendation tasks *Journal of Machine Learning Research* **10** 2935-62

Authors

Zhihang Tang, born in August 1974, Hunan, China.

Current position, grades: teacher at the department of computer and communication, Hunan Institute of Engineering (Xiangtan, China) since 2003.
University studies: PhD degree at Donghua University China in 2009.
Scientific interests: intelligent decision and knowledge management
Publications: 30 papers

Zhonghua Wen, born in May 1966, Hunan, China.

Current position, grades: professor at the Hunan Institute of Engineering.
University studies: PhD degree in computer engineering at Zhongshan University, China.
Scientific interests: intelligent decision and knowledge management.
Publications: 50 papers.

Analysis on Ideological and Political Teaching Curriculum Design and Information Transfer Based on Complex Adaptive System Theory

Caihong Zhang*

Xi'an Physical Education University, Xi'an, 710068, China

Received x xx 2014, www.cmnt.lv

Abstract

In the field of technology, based on scholar's research results of instructional system design complexity, this paper discussed the development of instructional system design thought. It also analyzed the basic framework of ideological and political teaching curriculum design based on complex adaptive system.

Keywords: system, complexity, instructional design

1 Introduction

With the development of modern science and technology, informatization and networkization have become the trend of Chinese development. At the same time, with the development of Chinese higher education from elite to mass, the increased enrollment and building of colleges and universities, and running a school of multi-campus, teachers of ideological and political lessons have the reality of "mobile-mode of teaching" in teaching practices. The extracurricular contact time between political and ideological lesson teachers and students is lesser. Then students' consolidation and review of knowledge are total by themselves. However, such kind of way will cause lack of deep understanding on ideological and political lessons. In addition, students have strong material gaining attitude on learning, which causes students are to learn for examination. Therefore, as for how to utilize after class time in high efficiency has become another key point of teaching methods in ideological and political course of colleges and universities. Fro example, to strengthen the coaching of students to ideological and political course; to deepen students' understanding on knowledge of ideological and political course; to develop student's ability of logical thinking and to enhance the instruction effect of ideological and political course. With the help of researches of scholars in educational technology field on the complexity of instructional system design, this paper established ideological and political teaching curriculum design based on complex adaptive system.

2 Complex Adaptive System Theory

Complex adaptive system (CAS) theory put forward by Holland in 1994. CAS refers to the complex dynamical system of subjects improving their own behaviors through study in the process of evolution and development, and they are inter-coordinated, mutual adapted and inter active [1].

Its basic idea is that system is made up of subjects that have their own finality and initiative. Adaptive subject has connected micro and macro world together. In microcosmic aspect, subject can constantly learning in the interaction with environment and other subjects, and changing their own structure and behavioral pattern according to their learned experience, so as to better survival in objective environment, that is the subjective adaptive. In macrocosmic aspect, in interaction between subject and environment, the whole system shows complex evolutionary process. The initiative and adaptation expressed among subject, environment and other subjects are the basic agents of system development and evolution. The evolutions of whole system are all derived on this basis, including the generation, differentiation and appearance of diversity of new hierarchies, and the appearance of larger subjects that generated by new polymerization. This is the core idea of CAS theory-"adaptation creates complexity" [2].

3 Teaching System and Its Design

A. TEACHING SYSTEM AND ITS COMPLEXITY

According to the basic idea of system theory, we regard various educations and teaching organizational form that with the purpose of achieving certain education and teaching objectives and realizing certain education and teaching function as education system or teaching system. Teaching system is the subsystem of education system, which contains four basic elements: teacher, students, textbook and teaching media. It is the premise of system operation and forms the spatial structure of system. The mutual effect, interdependence and interaction among sys-

* ***Corresponding author's*** e-mail: zchongrainbow@163.com

tem elements have constituted the complex operational process between system input and output. American famous educational technical expert, B. H, Banathy thought that teaching system is the hierarchical system complexus. Teaching system is generally divided into four hierarchical systems, which are respectively system in institutional level, management level, teaching level and learning level from high to low [3].

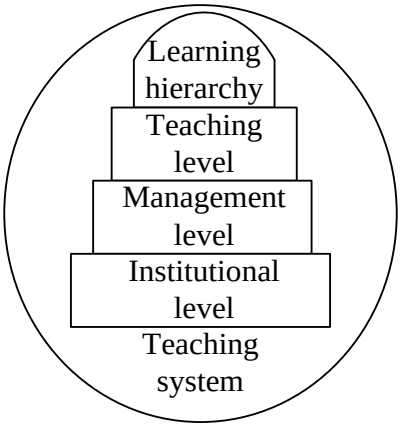


FIGURE.1 hierarchical structure of teaching system

Therefore, seeing from hierarchical structure and function, teaching system has the property and status of "complex".

B. INSTRUCTIONAL SYSTEM DESIGN AND THE COMPLEXITY OF ITS PROCEDURE SCHEMA

Instructional system design (ISD, which is also called teaching design), seeing from the history of the formation of its subject, is a subject that developed with the development of technology based on the integration of multiple theories. As for the definition of the concept of teaching design, different scholars have different opinions. Professor He Kekang [3] thought that teaching design mainly use system approach to convert the principle of learning theory and teaching theory into teaching objectives, teaching content, teaching method and strategy and teaching evaluation, so as to conduct specific planning and creation on the "process" or "procedure" of teaching and learning system. However, the fundamental objective of creating teaching and learning system is to promote students' learning. ISD has the following characteristics: the research object of ISD is learning and teaching system of different hierarchies, which includes the content, condition, resource, method and activity that promote students' learning; ISD refers to research and explore the essential contact among various elements, and element and whole using system method. It also comprehensively consider and coordinate their relations, which organically integrate elements so as to finish the function of teaching system; ISD aims to apply basic theories like learning theory and teaching theory to solve the practical problems of teaching, thus to form the teaching and learning system that has experienced verification and can realize intended functions.

The procedure schema of ISD is a set of programmed steps. The general procedure schema of ISD contains four basic elements: learner, objective, strategy and evaluation (Wu Meina, 1994). Currently, procedure schema of ISD has hundreds of types, of which the ISD schema with operability already has hundreds of types. It development is generally experienced four generations of evolutionary process (ISD1, ISD2, ISD3, ISD4). From the perspective of science of system, its evolutionary process also expresses as "linear cycle–flow chart–cycle in phase and linear out of phase–dynamic interaction network" [4]. We can induce the previous three phases of ISD procedure schema as linear teaching design. However, procedure schema of teaching design at the fourth generation is non-linear teaching design schema that based on the thought of complex adaptive system. There exist apparent differences between the two. We can compare the characteristics between them (as shown in Table 1[5]).

TABLE 1 Comparison of the main characteristics between linear ISD and nonlinear ISD

linear ISD	**nonlinear ISD**
linear, closed, static system	nonlinear, open, dynamic system
Teaching event presents linear causal relationship.	Teaching event presents interactive causal relationship.
Whole is the simple sum of each part.	Whole is not the simple sum of each part.
It can predict the change of learner's cognition and skill.	It can not conduct long-term prediction (butterfly effect). a tiny disturbance may produce unpredictable consequence. Integrate the unpredictable teaching event into the dynamic relation so as to increase its flexibility.
By decomposing the whole into the independent part to conduct analysis, and to simplify the complex teaching system	Part-whole-part…, constant cyclic and timely analysis, especially stress to use anthropological studies to understand the overall characteristics of complex teaching system.
Provide the predesigned teaching objectives and teaching sequence, until teaching is finished. There are seldom other alternative learning paths	Teaching objective is diversified and hierarchical with multiple alternatives of learning paths.
Factors like mistake and disturbance are corrected through negative feedback, which is an important means for maintaining the equilibrium state of teaching system.	Factors like mistake and disturbance are queries, explored and introspected in each level through positive feedback, which promote the constant evolution and development of teaching system and learners.

4 Development of ISD Thought

In long term of exploration, educational technology workers have confirmed the system theory thought as the guiding thought of ISD, considered teaching as system for study and defined system approach as the core method of teaching design. With the development of system theory and the constant deepening of cognition of teaching system, system thought and method used by ISD have successively experienced three phases: definitive linear system, definitive nonlinear system and indefinitive nonlinear sys-

tem. Currently it is developing in the direction of complex adaptive system [6].

The above three system thoughts have possessed the characteristics of some complex thoughts, but it cannot fundamentally solve the problems like the complexity of system, initiative adaptation of system elements and learning ability. Until the put forward of complex adaptive system (CAS), can these problems obtained better reply. CAS thinks that all subjects are in a common political and economic environment, but each part conducts adaptive learning and evolution concurrently and independently according to its surrounding partial microenvironment. Chen Liezun, et al tried to apply CAS thought for carrying out teaching design research and practice based on CAS theory, and also put forward the basic key points of ideological and political teaching curriculum design based on CAS thought [7], as shown in Table 2:

TABLE 2 Key points of ideological and political teaching curriculum design based on CAS

thought framework	**complex theory and CAS thought**
external relationship	It focuses on the social distant imagination. The designation is come forth with the purpose of conscious pursuing and adapting the uncertainty of future society. Its main aim is to culture future social citizens.
knowledge opinion	Knowledge is not only the reflection of objective and dynamic world. It also implicates the passion, faith, conscience and responsibility of human intelligence, which includes explicit knowledge and implicit knowledge. Among knowledge of different types and different subjects, they realize the mutual conversion, sharing, using and creation of knowledge though SECI* process.
Design focus	The creation strengthening of resource, rule, knowledge conversion (teaching platform) and mechanism, and the launch of activity. It attaches importance to practical teaching, extracurricular scientific activity and social practical activity
adaptivity	Subject automatically adapt to environment of indefinitive changes. It has strong environmental adaptability and forms mutual symbiosis with environment. It mainly relies on internal monitoring, automatic regulation and internal incentive.
internal relationship	System design experts provide design resource and conduct common design, interaction and collaborative operation with teachers. It will be much better if teaches are also system design experts. Cyclic, multi-causal relationship, nonlinear relationship and the coexistence of positive and negative feedback require to reform of system itself. The coexistence of self organization with other organization, and the harmonious coexistence of competition and cooperation.
procedure evolution	Teachers and students create together. Subject behavior or system behavior is unknown and cannot be totally predicted or controlled. All students design together and determine the orientation and process of teaching or spontaneously emerged. Keep open attitude to change, and guide by dynamic opinion of environment coordination evolution. Plan interacts with action.
Results assessment	individualized, polybasic, procedural and developed evaluation mode

Teaching system is considered as hierarchical complexus, instead of just staying in decomposing teaching system into teacher, student, teaching content and teaching media. This cognition has reached consensus in field of teaching design. Seeing from China's practical situation, there exist two design orientations: the ISD that defines teaching level as basic level and ISD that considers learning level as basic level. In China's teaching system of class teaching, teaching design with teaching level as basic level occupies leading role. However, in family education, social education, web education and school teaching that centered on student activities, teaching design with learning level as basic level has become the mainstream. At the same time, in other teaching forms, we can also consider institution level and management level as basic level for conducting teaching design.

According to the key points of ideological and political teaching curriculum design of CAS, this paper established the basic framework of ideological and political teaching design based on CAS, as shown in Figure 2.

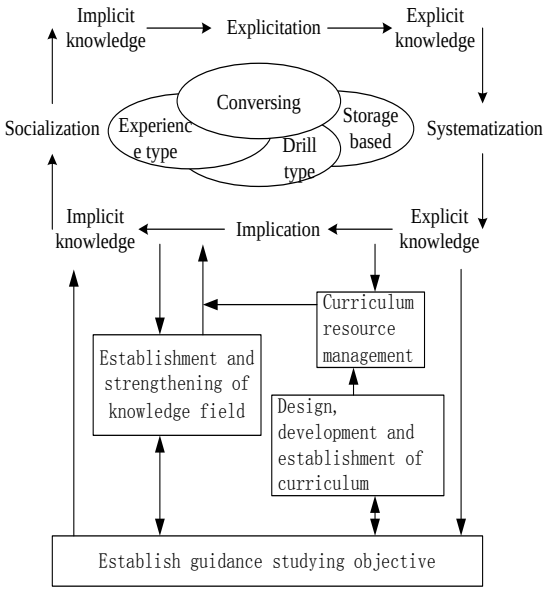


FIGURE. 2 basic framework of ideological and political teaching design based on CAS

Basic framework of ideological and political teaching design based on CAS has the following characteristics:

(1) Stimulate and intensify students' subject adaptation, which makes the inter coordination of demand between system objective and individual development;

(2) Respect the generative of teaching activity, advocate creative learning and culture student's innovation consciousness and ability;

(3) Competition and cooperation are harmoniously coexisted;

(4) Attach importance to the inter conversion and spiral upward of implicit knowledge and explicit knowledge.

5 Conclusion

Seeing from the development of ISD thought, the ideological and political teaching curriculum design based on CAS thought will certainly become the mainstream of ISD. Regard teaching system as hierarchical and stereo complexus, and ISD that respectively with teaching level as basic level and with learning level as basic level are the mainstream in the field of teaching design practice. It is believed that it can play tremendous effect in the future teaching practice, so as to improve the education quality of ideological and political curriculum. In addition, it also provides well reference for other subject educations.

References

[1] [America] John. H. Holland, Zhou Xiaomu, Han Huiyi 2000 Implicate Order – Adaptation Creates Complexity. *Shanghai: Shanghai Education of Science and Technology Press.*
[2] Lu Weilin 2005 Basic Principle of Ideological and Political Education under the Theoretical Perspective of CAS. *Exploration and Free Views*, (5).
[3] He Kekang, Zheng Yongbo, Xie Youru 2002 Teaching System Design. *Beijing: Beijing Normal University Press*, 2-17.
[4] Wu Meina 1994 Teaching Design. *Beijing: Higher Education Press*, 10-11.
[5] Wen Xiaojiao 2010 Teaching Design Based on CAS Theory. *Yunnan University.*
[6] Huang Lu 2006 Complex Science Theory and Expanding New Methods in Ideological and Political Education of College Students. *Journal of South-Central University for Nationalities (Humanities and Social Sciences)*, (1) 182-184.
[7] Zhang Tian, Li Ziyun 2013 Reconstitution of Teaching Design Framework in the View of Complex Thought. *Electrified Education Research*, **02** 23-26.

Authors

Caihong Zhang, born in 1978, Hebei Province of China

Current position, grades: associate professor.
University studies: PHD degree was earned in major of sports humanities and sociology, Beijing Sport University in 2011.
Scientific interest: ideological and political education, sport law.

The Attractiveness and Motivation of China's International Investment Based on Positive Investment Inertia

Liming Xiao*, Rui Jing

School of Economics and Management, Shanxi Normal University, Shanxi,, 041004, China

Received x xx 2014, www.cmnt.lv

Abstract

Nowadays, economic globalization is the main trend of world economic development. Each country takes the participation in international investment as a strategy. This paper analyzes China's international investment from the view of investment inertia. Though the attractiveness and motivation of China's investment keeps increasing, the integral development level is low which stays in the stage of zero inertia. The total amount of China's investment is increasing. But the invitation of investment is unreasonable; the short-term investment in foreign country prevails. We seek the strategic asset of developed country eagerly, and prefer to seek the market and resource of developing country.

Keywords: international investment; inertia; attractiveness; motivation

1 Introduction

International investment is the capital movement across the world. It is the investment in foreign country of enterprise, personage or government [1]. Since 21st century, the total amount of investment of each country is increasing rapidly as economic globalization enhancing. And it has become an important strategy to take part in international investment. By now, our international investment can be divided into invite investment and investment abroad. So, it means a lot to our international investment in the future to analyze the investment situation of our country based on the attractiveness, motivation and inertia of international investment which is already existed.

2 Relevant Content of International Investment

2.1 INTERNATIONAL INVESTMENT ATTRACTIVENESS

The international investment attractiveness is the host country's ability to attract and adsorb the capital of investment country. It is an objective economic power exists in the host country to regulate and dominate the orientation of the international capital [2]. The international investment attractiveness varies with the different GNP of each country. The higher GNP, the more powerful the attractiveness is [3]. Besides the GNP, there are some other factors that decide the international investment attractiveness of a country, such as, the country's capital demands condition, the accommodation power to foreign capital and the return on investment in the country.

2. 2 INTERNATIONLA INVESTMENT MOTIVATION

International investment motivation is the investment motivation and ability of one country or area to another [4]. Previous economist thought that investment abroad has direct relation with the situation of the country's economic development. That is to say, the motivation of a country depends directly on GNP, and GNP is the main factor that decides the motivation and attractiveness of investment abroad.

In China, someone put forward the Foreign Direct Investment theory – the theory of meeting each other's demand. They think the reason why international investment appears is both sides has different demands. An actual investment can be done only when both side that taking part in international investment must consider each other's demand [5].

2.3 INTERNATIONAL INVESTMENT INERTIA

International investment inertia is an objective economic phenomenon of international investment. The investment scale and speed of the country or area keeps the same while the international investment motivation, international investment attractiveness and the other economic and politic situation stays unchanged. It is divided into zero international investment inertia and positive international investment inertia. It can also be called the inertia of investment abroad [6]. The capital to invest abroad and continuous investment abroad are the two preconditions.

The international investment inertia theory depends mainly on the PLC (Product Life Cycle) model [7]. It has two basic points. Firstly, it is the original intention of international investment inertia to reduce the cost. Mostly, the investment country has the advantage of technology

* ***Corresponding author's*** e-mail: xlmxiaolm@163.com

and supervision while the host country has location advantage such as low price resource and labor. Secondly, transnational corporation eke out monopoly interest to bring out inertia from investment. The purpose of investment abroad is, to some extent, to prolong the monopoly advantage of the innovation country with the low cost of host country, and meanwhile to strike pirate.

3 Analysis of China's Investment Attractiveness and Motivation

3.1 ANALYSIS OF CHINA'S INVESTMENT ATTRACTIVENESS

We analyze international investment attractiveness with the model of international investment attractiveness, specifically as follow

$$E(a/b) = F(a/b) \cdot \frac{I(a/b) \cdot R(a/b)}{H(a/b)} \quad . \qquad (1)$$

Here: E(a/b) means China's general investment attractiveness coefficient to foreign country; F(a/b) means the ratio of national income per capital between China and target country; R(a/b) means the ratio of bank rate between China and foreign country; I(a/b)means the ratio of return on invest of certain industry between China and foreign country; H(a/b) means the ratio of investment density of certain industry between China and foreign country; Formula (1) shows that China's international investment attractiveness is proportional to the ratio of national income per capita, the return on invest of certain industry and invest density between China and foreign country.

When E (a/b)>1, capital flows into China; When E (a/b) <1, capital flows flow out from China. According to formula (1), we calculated China's international investment attractiveness during 1990 to 2009; the result is shown in Table1.

TABLE 1 Attrativeness coefficient of China's international investment

Year	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**
Coefficient	-17.44	NA	5.25	-17.78	-11.25	3.81	-8.38	1.33	-14.84	-13.94
Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Coefficient	-5.80	1.72	2.95	1.50	NA	4.45	NA	4.78	4.97	NA

Comment: NA means the data is lacked or the calculated result does not exist.

Table1 shows that in most years before 2001 the attractiveness coefficient is less than 1, which suggests China's lack of international investment attractiveness in those years. That is because, planned economic, the main part of China at that time, low opening degree and defective laws and regulations and policy and measure made economically advanced country not tend to favor China's investment situation. After 2001, the coefficient is more than 1. It suggests China's international investment attractiveness was gradually increasing. Mainly because China accelerated the market economy reform and opening degree, enacted lots of preferential policy to attract foreign capital, perfected investment climate further. That made foreign country looks good of the potential of China's market gradually and increased investment in China.

3.2 THE PROBLEM OF FOREIGN CAPITAL INTRODUCTION

1). The excessive speed of foreign capital introduction; the total amount of invite investment is 17.143billion during 1970-1996.After 1996, we make use of more than 10 billion each year. And we made use of more than 50 billion foreign capital each year after 2007; the invited projects were mostly foreign direct investment [8]. We estimated the relation of the foreign capital we used and GDP, using the ordinary least square method. During 1990-2008, the ratio of invited foreign capital and China's economic growth is 0.57.Every Yuan we invited increased 0.57yuan for GDP. The promotion of invited capital to the economic development was at average level. It suggests the large amount of foreign capital we utilized contributed a little to the economic growth.

2). The industrial structure of invited capital is unreasonable. **(1)** Foreign chose the unreasonable industry and the industry structure was imbalanced. Local administration department emphasized the scale of capital introduction, but they did not take China's industrial structure into consideration .Local government favored short-term benefit. Foreign capital was mostly invested to secondary industry, especially industry which needs less investment, has quick effect, brings high profit, and earns foreign exchange easily, such as manufacturing industry. As to the weak industry whose production cycle is long, such as science and technology, infrastructure, agriculture, etc., got poor investment. **(2)** The area selection of foreign investment is unreasonable either. The west area lags far behind the east area in making use of foreign capital. According to the statistics, at the end of 2008, the east and middle area absorbed 91.47% foreign capital of that year while his west area absorbed only 8.52%. And the development gap of east and west area is about to increase.

3). Capital invitation is in conflict with the development if local enterprise. Foreign investment values China's

preferential policy which make them in the advantageous position. But the local enterprise with less capital and backward technology do not have the preferential policy. They need equal preferential policy. So, the government has to considerate both sides and work out with suitable preferential policy.

4). The investment climate has to be improved; to attract foreign capital, we must have the laws to protect their rights and interests. So they have laws to abide during their operating. But China's laws and regulations are imperfect. And this make the transnational enterprise can't protect them when they get infringed illegally during their operation in China. What's more, though we have accelerated the infrastructure constructor and begin to pay attention to the environmental problem, we still have to perfect ourselves in many aspects.

3.3 ANALYSIS ON CHINA'S INTERNATIONAL INVESTMENT MOTIVATION

International investment motivation is a country's own motivation and ability to invest abroad. The coefficient the international investment motivation is the more investment abroad. So, the motivation coefficient is proportional to the height of the scale of investment abroad. We still use Wang Dongjin's international investment motivation model:

$$D't = \frac{H't(R't + I't)}{2} \quad . \tag{2}$$

Here: D, t stands for the coefficient of China's investment abroad, the bigger it is, the stronger investment motivation; H, t is the coefficient of China's civil investment density variation; R, t is the deposit interest rate of Chinese bank and variation index in a given period; I, t is the variation index of China's civil return of investment in a given period. Formula 2 shows that in a given period, China's motivation of investment abroad is proportional to investment density variation coefficient in the same term, and it is inversely proportional to the average index of return of investment and deposit interest rate. When D, t=1, the international investment motivation stays the same; When D, t>1, the motivation increases; When D, t<1, the motivation decreases. According to formula 2, w calculated the motivation coefficient of China's international investment during 1990 to 2009.See the result in Table 2.

TABLE 2 Motivation coefficient of China's international investment

Year	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**
Coefficient	39.40	4.95	0.51	-0.30	-0.59	-0.89	0.70	2.40	-3.60	1.25
Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Coefficient	0.93	0.08	2.01	0.21	0.04	0.36	0.68	1.04	1.17	2.35

The analysis of China's international investment motivation and inertia suggests that, though China's international investment motivation is gradually increasing, the motivation coefficient is low and stays at zero investment inertia period .That means China should increase the investment abroad.

3.4 THE PROBLEM OF CHINA'S INVESTMENT ABROAD

1). China's scale of investment abroad is small and scientific and technological investment is weak. It has a lot to do with the backward technology and small capital scale of most Chinese enterprise. What's more, the government attaches little importance to investment abroad. In fact, Chinese enterprise has realized the importance of investment abroad and expending of investment business. Though they have strong motivation to invest abroad, their investment scale was restrained by the limited capital and backward technology. Besides, China has few research institution or industry with advanced technology. Though Chinese university has some advantage on research and development, the research achievement can hardly be turned to productivity.

2). The problem of investment continuity. Abundant capital is needed to invest abroad. And the feasibility and profitability should consider. We should avoid blindness. In 2008, the acquisition of Gibson Iron Mine was strongly opposed by Australian and thus brought great financial lost to Shougang Corporation. In addition, to invest abroad, we should stick to the principal of win-win. Try to benefit our country's development and the host country in the same time. As the earning of investment abroad is not evident in the earlier stage, we should mind the chronicity.

3). The structure of the investment abroad is unreasonable. The investment in advanced country is based on service, as to developing country, it is resource. China's investment is indirect investment mostly. According to the experience of Multinational Corporation, we should transform our indirect investment to direct investment gradually. That will form stronger productive force, increase enterprise income and, as a whole, will promote our country's economic development.

4 Conclusion and Suggestion

The above shows, China is short of international investment attractiveness and motivation, international investment inertia is at zero investment inertia stage. To solve the problems that exist we should grounded on the actual situation and pay attention to the following problems.

1). To balance investment attraction and investment abroad. Investment attraction and investment abroad are the two main ways of China to take part in international investment. To confront the shock of economic globalization, we must develop introducing of foreign capital and investment equally. First, invited foreign capital can help to develop our economy and drive the innovation and development of enterprise, thus, laid foundation for investment abroad. Secondly, it can create condition for larger scale of capital invitation. If we simply import and output, we will probably rely on foreign capital and technology, lower the civil capital's rate of operation and the ability of technology innovation.

2). To create good investment climate. In order to get the initiative in the worldwide compete of using international investment, China should use the experience and lessons of other country's international investment for inference to perfect the rules and regulations and to better the investment climate to attract foreign investors.

3). Strive to develop domestic enterprise, enhance the competitiveness and encourage the investment abroad. To encourage those enterprises with advantages for development to invest abroad, to participate in world-wide competitiveness thus entirely elevate competitive force of China's investment abroad.

4). Maintain the appropriate scale of international investment. **(1)** Utilize foreign capital rationally and keep the reasonable scale of utilizing foreign capital. In practice, due to the limitation of reality, the theoretic best scale can not be achieved; therefore we should stick to the second-best principle [9]. And our foreign capital utilization is unreasonable in construction and source and area orientation. In view of this, we should, in the future, synthetically consider the speed and direction of utilizing foreign capital. **(2)** Maintain the appropriate scale of investment abroad. Properly balance the out flowed capital and net export of commodity.

Acknowledgement

This work was financially supported by the Humanities and social science Project in Colleges and Universities of Shanxi Province about "A Research on the interaction mechanism of Shanx's industrial restructuring and technological innovation" (No. 2013322) and the soft science project of Shanxi Province about "FDI technology spillover effect and Shanxi's Innovation: A Comparative Study with the Chinese eastern" (No. 2013041037-02).

References

[1] Pei Changhong, Zheng Wen 2011 Country-specific Advantages: Supplementary Explanation of International Investment Theory. *Economic Research Journal*. 22-23.
[2] Liu Haiping, Song Yihong ,WEI Wei 2013 Factor Endowments, Institutional Characteristics and FDI Flows : An Empirical Study Based on Investment-Gravity Model. *International Business*, 45-46.
[3] DUNING JH 1981 International production and the multinational enterprise. London: George Allen&Unwin.
[4] Fang Zhi 2006 Research of industrial cluster oriented transregional group investment motivation mechanism. *Journal of ZheJiang University of Technology*, 12-13.
[5] Ai Lingyu 2010 Ecological investment theory and the practise research: analysis of cases in Shan Xi province. *Journal of Shanxi University*. 5-6.
[6] Deng Dan 2013 Trade growth inertia and the lagged-effects of trade policy:research based on export tax rebate policy. *Ji'nan University*, 4-6.
[7] She Biao 2013 Product lifecycle modeling and configuration optimization for mass customization. China Economic Quarterly (7), 14-16.
[8] Huang Haide, Hu Zhi 2006 Research on the Attractiveness, Motivations and Inertia of China's International Investment. Finance and Trade Research, 8.
[9] Bai Mei 2010 Research on Foreign Investment Attractiveness of China's Western Area in the Perspective of Sovereign Monetary Theory. *Shanxi Normal University*, 48.

Authors

Liming Xiao, Shanxi Province of China

Current position, grades: associate professor
University studies: PHD degree was earned in major of World Economy, Nankai University in 2007.
Scientific interest: multinational companies and foreign direct investment

Rui Jing, born 06.02. 1976, Shanxi Province of China

Current position, grades: master
University studies: Master degree is reading in major of international trade, Shanxi Normal University.
Scientific interest: the multinational corporation management and administration

Operation Path and Strategy of Logistics Enterprise Supply Chain Logistics Management under Horizontal Integration

Yue Zhang*

Shanxi Youth Vocational College, Taiyuan, 030032, China

Received x xx 2014, www.cmnt.lv

Abstract

This paper put forward the significance of horizontal integration service provided for society by logistics enterprise, and discussed in detail about how to realize horizontal integration service and what kind of operation model can be adopted by relevant sub-service. Researches on these contents not only have certain effect on correlated theoretical research, but also can provide certain guidance for logistics enterprise. This research aimed to promote the compound integrated service ability and international competiveness through horizontal integration management mode.

Keywords: Logistics, horizontal integration, operation and management mode.

1 Introduction

Logistics is an integrated service industry which involves numerous sub logistics business. With the increasing of market competition and refining of social division of labor, an in creasing number of clients well preparing core business for concentrating superior resources, outsourcing logistics business to professional logistics enterprises and wishing to obtain the "integration", "one-stop" or "point contact" comprehensive logistics service. However, most of logistics enterprises in China still rest on low side logistics market and provide single and low profit subsection logistics service, which lack of logistics resource integration competence. Logistics enterprises are obviously insufficient in aspect of providing horizontal integration service competence, for they center on providing subsection logistics service with single function. It will severely impede logistics enterprise to become bigger and stronger, and constantly widen its gap with international logistics tycoon. Therefore, the current logistics enterprise is badly in need of constantly promoting logistics comprehensive service competence through acquisition and reorganization and resource integration.

2 Overview of Logistics Enterprise Horizontal Integration Operation

Logistics enterprise horizontal integration strategy generally refers to the expansion aiming at same market or relevant market. In logistics industry, horizontal integration refers to that in order to realize providing comprehensive service for clients, logistics enterprise conduct joint operation on correlated business area of logistics business. The advantages of horizontal integration lie in that it can obtain faster enterprise development with lesser cost, decrease competitor in market, rapidly enhance market share and improve scale economy. Its main disadvantages are that enterprises will undertake risk of operation from larger scale, and the coordination difficulty within enterprises increase. The implementation of horizontal integration strategy mainly is "strong-weak" alliance. Large-scale enterprise is matrix of combined enterprises, and professional enterprises or middle and small-sized enterprises are joint enterprises. With the increasing of market competition and refining of social division of labor, an in creasing number of clients well preparing core business for concentrating superior resources, outsourcing logistics business to professional logistics enterprises and wishing to obtain the "integration", "one-stop-shopping" or "point contact" comprehensive logistics service. Customer's requirements are increasing, thus the third party logistics enterprise should change development conception, especially large-scale logistics enterprise, change the relatively single logistics business operation mode, increase logistics business type, expand network coverage and increase value-added cervices so as to provide comprehensive one-stop service for clients. Therefore, when providing logistics service for clients, logistics enterprises should not only provide basic transport service, warehouse service for clients, but also provide multimodal transport service, customs declaration and commodity inspection service, distribution processing service, distribution service, express service, changing or refunding service and order decision service for clients. Efficient and comprehensive one-stop logistics service models like logistics project formulation, supply chain fund settlement, etc are logistics enterprise horizontal integration operation service model. As shown in Figure 1.

* ***Corresponding author's*** e-mail: zyyuez@163.com

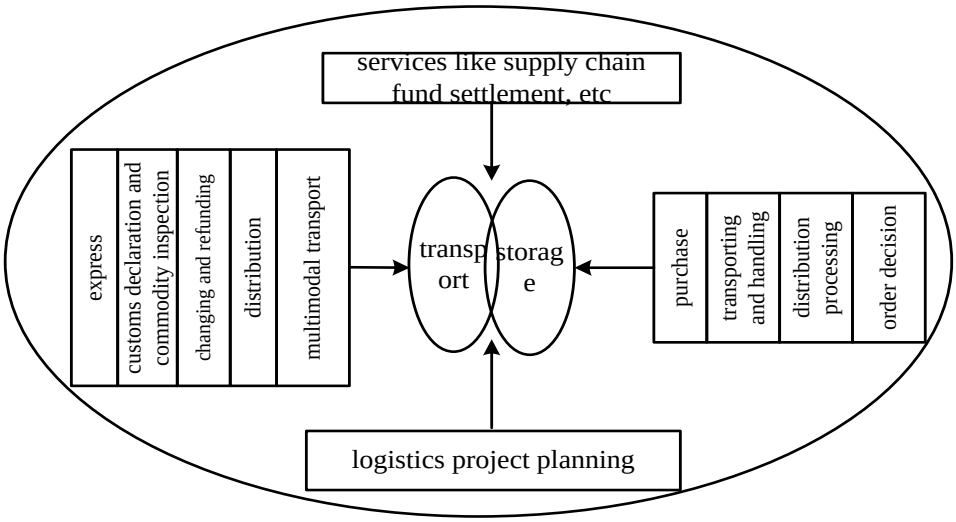


FIGURE 1: Operation model of logistics enterprise horizontal integration

3 Operation Path of Logistics Enterprise Horizontal Integration: Progressive Merger and Acquisition

When providing integration comprehensive service, as for how to obtain relevant resources that enterprise itself dose not possess is the primary problem enterprise will confronted, which also relate to the important decision of enterprise service quality. Besides the currently common merger and acquisition, investigation and reconstruction and outsourcing, paths of obtaining resources also include the newly emerged model-strategies alliance in recent years. Starting from bounded rationality and opportunism, the contemporary western economist Williamson [1] used the three properties (assets specificity, uncertainty and transaction frequency) that influence trade level and feature under incomplete contract to conduct further study on the nature of enterprise. He believed that there was a kind of governance structure in intermediate state based on transaction model between market and enterprise-alliance. Organization model of transaction is shown in Figure 2.

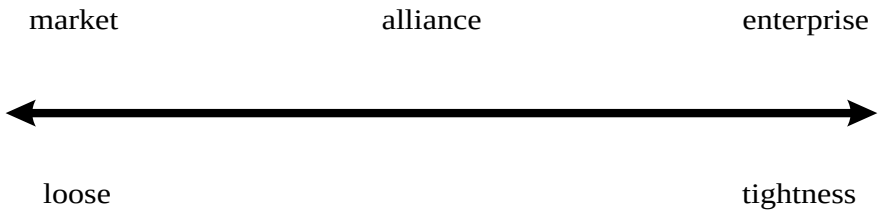


FIGURE 2: Williamson's alliance transaction organizational model

Alliance structure has both advantages of market and enterprise, is a kind of innovation and transition of transaction institution. This institutional arrangement can promote the mutual condition of market and enterprise under the limitation of protocol or property right, thus to play a role together. It also can overcome the malpractices of market and enterprise institution (information asymmetry in bureaucracy and enormous cost of one-time transaction in market) [2]. Based on Williamson's theory, enterprise integration operation can adopt market mode, enterprise mode and alliance mode. These three organizational modes have their own advantages and disadvantages in the operation process of enterprise integration.

"Progressive" is a kind of process gradually and slowly approach to objectives, which represent a kind of thought of quantitative change. Yi Chunguang [3] (2007) thought one of basic mode of expanding integrated logistics service is "break through", another is "progressive". He initially proposed to adopt progressive resources access path when providing integrated logistics service for clients. Progressive merger and acquisition on sub-business which logistics enterprise does not possess is an effective path for logistics enterprise to realize horizontal integration operation. Progressive merger and acquisition is divided in detail into six operation models, that is market-based outsourcing, customized service outsourcing, internal insourcing, contact type strategic alliances, share strategies alliances and merger and acquisition self-governance model. As shown in Figure 3, the advantages of applying progressive merger and acquisition are analyzed from two perspectives: small and medium-sized logistics enterprises and large-scale logistics enterprise.

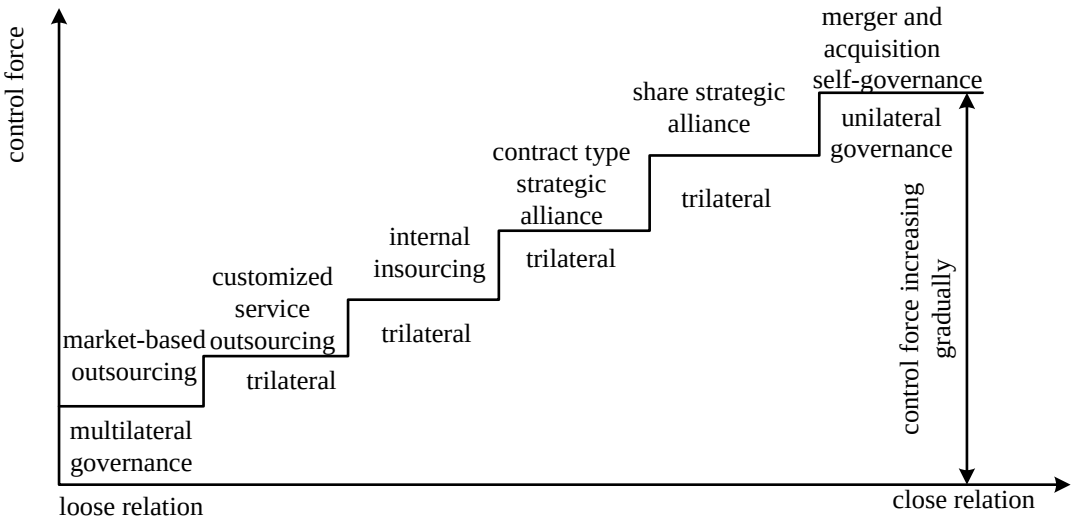


FIGURE 3: Six development stages of progressive horizontal merger and acquisition

Logistics enterprises can accumulate experience in process of constant try of integration operation by progressive horizontal merger and acquisition, for they adopt path of progressive resource access. As for current logistics development stage and developmental level, the accumulation is a kind of suitable development strategy for both small and medium-sized logistics enterprises and large-scale logistics enterprises. As for small and medium-sized logistics enterprises with limited capital capacity and in smaller scale, the adaptation of progressive merger and acquisition model is good for promoting the attempt of small and medium-sized enterprise providing comprehensive logistics service. Logistics industry is in developing phase with many features of "many, small, loose and weak". As for small and medium-sized logistics enterprises with limited resources, they start from outsourcing integration, and gradually grope the experience of integrated service, which avoid the overcrowding of many small and medium-sized enterprises in low-end market. It also guide middle and small-sized enterprises to attempt medium and high-end logistics market, thus to make them bigger and stronger until they have the strength to conduct horizontal mergers and acquisition. As for some large-scale enterprises with relatively abundant capital, progressive mergers and acquisition is good for avoiding enterprises make offensive mistakes. If not taking the practice of their self-development into consideration, and blindly study behaviors of international logistics tycoon and constantly conduct plenty of merger and acquisition, some enterprises especially those lack of integrated service experiences will have poor digestion. Moreover, they will suffer a lot because of lacking related sub-business operation experiences and merger and acquisition experiences [4-5]. As for logistics sub-business that has not yet involved in that, they should set down from low-end whether they are large or not. After they possess operation experiences, they can take the opportunity and implement merger and acquisition with self-support [6].

4 Logistics Enterprise Problems under Horizontal Integrated Operation Model and the Coping Strategy

There are three misunderstandings of logistics enterprises in the process of integrated operation, that is lack of horizontal integrated service awareness, excessively rely on outsourcing, blindly merger and acquisition and over expand.

4.1 LACK OF HORIZONTAL INTEGRATED SERVICE AWARENESS

Facing increasing number of market logistics comprehensive service needs, some logistics enterprises lack courage of enlarging range of logistics business, content with the status quo, fear to encounter risks and only provide simplified and segmented logistics service in low-end logistics market. They cannot meet customers needs, thus they lose middle and high-end market. The lack of horizontal integrated service awareness is one of vital reasons for logistics enterprise causing the fall behind of logistics horizontal integrated operation.

4.2 EXCESSIVELY RELY ON OUTSOURCING

Logistics enterprise horizontal operation model is to outsource sub-business. Although outsourcing can rapidly access resources, it has limitations. After logistics enterprises outsource sub-business, there exist not only transacttion cost, but the control on outsourcing business is also reduced, which lead to the decrease of client's satisfaction. Therefore, logistics enterprises should select the operation model of logistics sub-business according to the three-dimensional model established in this paper. Sub-business dose not suit for outsourcing can not outsource. Only in this way can logistics enterprises constantly enhance service quality in the process of integrated service, thus to realize the bigger and stronger of enterprises.

4.3 BLIND MERGER AND ACQUISITION AND EXPANSION

Some logistics enterprises adopt integrated operation model of over risk of expansion. In order to become bigger and stronger and reach to effect of horizontal integrated operation in short time, some logistics enterprises blindly implement merger and acquisition, of which the first bad results is rapture of capital chain and worse of financial condition. The second bad results are poor effect of merger and acquisition, and a series of troubles in the launce of business. Therefore, when implementing merger and acquisition, logistics enterprises should adopt progressive expansion style according to the enterprise strength and experience. As sown in Figure 4.

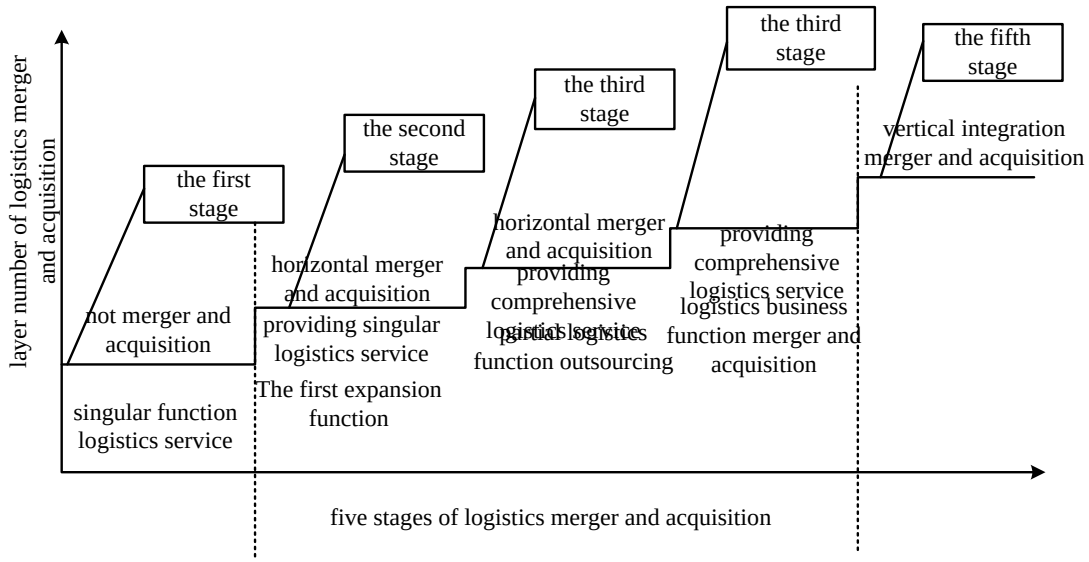


FIGURE 4: Progressive integrated merger and acquisition process of large-scale logistics enterprise

4. 4COPING STRATEGIES ON HORIZONTAL INTEGRATION OF VARIOUS LOGISTICS ENTERPRISES

Production enterprises' logistics department: the core business of production enterprises is not logistics business, thus it is difficult to realize large-scale. Logistics department also can meet the integration logistics service requirement of production enterprises thorough ways of outsourcing or alliance. Or they can cooperatively manage with the third part of logistics enterprise, and take advantages of each other so as to realize large-scale operation. It can both fully use the available resources and meet integration operation requirements [7].

Middle and small-sized logistics enterprise: outsourcing and alliance is very suitable operation model for vast middle and small-sized logistics enterprises in process of horizontal integrated operation, since their scale and resources are limited. However, in order to not reduce the effect of integration service, they should strengthen the control and management on outsourcing and alliance segment. In addition, middle and small-sized logistics enterprises also can adopt market-nicher strategy and depth appreciation model.

Large-scale logistics enterprise: enhance comprehensive degree of logistics service; adopt merger and acquisition means in good time; set up unified information platform; adopt horizontal networked organization structure.

5 Conclusion

Aiming at development status of logistics, logistics enterprises especially large-scale logistics enterprise should adopt horizontal integration operation model, which provide "one-stop" and "one-point-contact" comprehensive logistics service for clients so as to enhance their competiveness and shorten their gap with world logistics tycoon. The adoption of horizontal integrated operation model by logistics enterprises is good for improving the financial performance of logistics enterprises. Progressive merger and acquisition is an effective path to realize logistics enterprise horizontal integrated operation. Progressive merger and acquisition can be divided into six operation paths, that is market-based outsourcing, customized service, internal insourcing, contract type strategy alliance, share strategy alliance and merger and acquisition self governance model. According to the current situation of logistics enterprises in comprehensive integration service, there exist several misunderstanding of logistics enterprises in operation. The first is logistics enterprise lacks horizontal integration service awareness. The second is logistics enterprises are excessively rely on outsourcing operation model. The third is large-scale logistics enterprises are tending to blindly expand in process of horizontal integration. The participation or implementation of

logistics enterprise in supply chain logistics management is both serve for short-term objective and devote themselves to medium and long term objective. Logistics enterprise should orientate in customer value analysis and assessment, wholly enhance logistics management ability and service ability from supply chain level and height and timely and accurately find and recognize customer need and value point. At the same time, they should optimally configure the internal fine resources and external customer resources of supply chain, gather them in the business activity that can culture enterprise core competiveness and comprehensively optimize competiveness of supply chain logistics management of logistics enterprise.

Acknowledgement

Science research project of Shanxi Province: A Study on the construction of e-commerce logistics system of Shanxi province. 2012041081-04.

References

[1] Y.W.Guo 2010 The Review of Williamson's Theory of Economic Organization. *Reformation & Strategy*, (11) 1-2.
[2] M.X.Wang 2011 Research on Knowledge Management of Supply Chain Alliance. *Southwest Jiaotong University*.
[3] C.G.Yi, S.D.Ju 2007 Expand Comprehensive Logistics Service of Railway Freight Station. *Railway Freight Traansport*, (3), 1.
[4] Q.Guo 2011 Study on Post-acquistion Network Collaborative System of Multi-modal Logistics Enterprises. Doctoral Dissertation, *Beijing Jiaotong University*.
[5] X.w.Lin 2011 The Research on Synergy for System of Logistics Enterprise Resource Based on Integrating of M & A. Doctoral Dissertation, *Jiangxi University of Finance and Economics*.
[6] J.p.Zhou 2005 Study of Ipo and M & A of Chinese Logistics Enterprise. *Pioneering With Science & Technology Monthly*, (10) 9-11.
[7] X.F.Shi 2008 Third Party Logistics Enterprise Implements Integration Logistics Service Strategy. *Transportation Enterprise Management*, (9) 55-56.

Author

Yue Zhang, born in 1973, Shanxi Province of China

Current position, grades: associate professor,
University studies: PHD's degree was earned in major of e-commerce, Xi'an Jiaotong University.
Scientific interest: e-commerce

An entropy method-based index system for the competiveness of industrial cluster – a case study on the typical clusters in Zhejiang province in China

Weidong Wang[1,2*]

[1]*College of Public Administration, Zhejiang University, Hangzhou, Zhejiang Province, China*

[2]*College of Humanities and Social Sciences, China Jiliang University, Hangzhou, Zhejiang Province, China*

Received 1 October 2014, www.cmnt.lv

Abstract

Basing on the previous research results, this study constructed an index system for the competiveness of the industrial cluster using analytic network process (ANP) method. Moreover, it employed the entropy method in objective assignment method to assign weights to the indexes and conducted empirical analysis by exampling the typical clusters in Zhejiang province. The results showed that the industrial concentration degree, specialization degree (location quotient), and Herfindahl-Hirschman index (HHI) took relatively high proportions in the indexes concerning the competitiveness of industrial clusters. This study also drawn an important conclusion, namely, high industrial concentration degree was conducive to improve cluster competitiveness and reduce cluster risk, while lower industrial concentration degree facilitated the formation of high overall cluster competitiveness.

Keywords: ANP, entropy method, Industrial cluster competitiveness

1 Introduction

In the context of economic globalization, industrial clusters have become remarkable regional trends, which have had a huge role in promoting economic development in both developed and developing countries. However, industrial clusters will likely lead to a recession or even crash of the entire regional economy if there are problems. So the correct evaluation of the competitiveness of industrial clusters has become an important issue. To comprehensively evaluate the competiveness of industrial cluster, it is needed to construct a scientific evaluation index system by treating industrial cluster as a network and comprehendsively analysing the competitiveness and relationships of all network layers.

2 The theoretical framework of the indexes for evaluating the industrial cluster competiveness

2.1 THE BASIS FOR CONSTRUCTING EVALUATION INDEXES FOR INDUSTRIAL CLUSTER COMPETIVENESS

This study built an evaluation index system for the industrial cluster competitiveness from the following aspects in reference with previous research results:

1) Using ANP method, index system was divided into the target layer, criterion layer, and index layer in turn. Meanwhile, in reference with related study results concerning industrial cluster network structure, the criterion layer of the index system was further decomposed into enterprise network, support network, and social network [1].
2) The core index for the competitiveness of enterprise network included five aggregation indexes, namely, industrial concentration, spatial Gini coefficient, EG index, location entropy, and two indexes reflecting the dynamic variations of cluster enterprises, namely, the derivation degree and innovation degree of enterprises.
3) The indexes for support network employed three indexes that were closely correlated with the cluster development, namely, public service platform construction, professional market perfection, and industry association service. The three indexes were scientifically evaluable using corresponding quantitative indexes through conversion [2].
4) Social network index was mainly selected according to the cluster embed ability of the social capital and social network in the social network structure of industrial cluster. The core indexes in the two factors, including entrepreneurship and social capital, were used for qualitative analysis. In real application, the two indexes could be subjectively analysed and evaluated using expert advice method and Delphi method [3].

2.2 THE INDEX SYSTEM FOR EVALUATING THE COMPETITIVENESS OF INDUSTRIAL CLUSTERS

According to the construction principle and basis above and the network structure characteristics of industrial cluster, this study constructed the index system for evaluating the competitiveness of industrial clusters, as shown in Table 1.

[*] ***Corresponding author's*** e-mail: *wwdn2002@aliyun.com*

TABLE 1 The index system for evaluating the competitiveness of industrial clusters

Target layer A	Criterion layer B	Index layer X		
		Index code	Index	Index connotation
The competiveness of industrial cluster A	The competiveness of enterprise network B1	X1	Industrial concentration	The total output of a cluster / the output of a industry in national level or in a high-level regional range
		X2	Spatial Gini coefficient	The distribution equilibrium degree of industrial space of cluster: G=∑N(si-xi) 2
		X3	EG index	Cluster scale and the cluster aggregation degree bearing regional differences γ=[Gi-(1-∑rxj2)Hi]/[(1-∑rxj2)(1- Hi)]
		X4	Cluster specialization degree (location quotient)	The ratio of cluster output in regional total output/ the ratio of provincial industrial output in total provincial output
		X5	The enterprise aggregation in a cluster(HHI)	The quadratic sum of ratio of the output of the leading enterprises in the cluster to the total output of the cluster
		X6	Derivation degree of cluster enterprises	The investment of fixed assets /the total output of cluster
		X7	The creative degree of cluster enterprises	The output of new products/ total output of cluster
	B2 The competiveness of supporting network B2	X8	Supporting degree of public service platform and	Total assets of service platform/the number of service enterprises
		X9	The perfection degree of professional market	Total market turnover/ number of market stands
		X10	The service level of industrial associations	Number of member enterprises/ number of industrial associations
	The competiveness of social network B3	X11	Entrepreneurship	The innovation consciousness of entrepreneur
		X12	Social relationship	The monopoly of social assets

3 The entropy method-based calculation method for index weight

At present, index weigh is mainly determined using subjective and objective assignment methods. In light of that the indexes of the competiveness of industrial cluster were constructed using the ANP in subjective assignment method, the weight of these indexes were determined using the entropy method in objective assignment method. Moreover, five typical clusters in Zhejiang Province was employed for empirical analysis.

3.1 INTRODUCTION OF THE PRINCIPLE OF ENTROPY METHOD

"Entropy" was sourced from the Greek and represents the variant capacity. Clausius pointed out that, since entropy was similar with energy and were both functions of state; the two concepts should also be similar in word form. In reference with the German expression of energy, namely, energy, Clausius noted the entropy as entropie. Chinese physicist Hu Gangfu firstly translated the entrpoie into Chinese word as "熵" by adding the word "商" with a "火" on the left in 1923 in view of the ratio of entropy energy variations to temperature [4].

In mathematics, entropy signifies the uncertainty of a situation or problem. In 1950s, American proposed the information theory. In this theory, entropy was directly treated as the uncertainty of the state of the signals sent from an information source and signified the information-sending ability of non-information source. Since then, entropy spread to non thermodynamic field in a new look. Subsequently, entropy was widely and rapidly applied into the fields of economy, city planning, decision analysis, artificial intelligence, and philosophy etc. In project evaluation, it is always needed to take account of the importance of each evaluation index. One of most direct and convenient methods to represent the importance is assigning weight to each index (weight coefficient). The research conclusion regarding the entropy theory above suggests that the accuracy and reliability of decision are dependent on the information amount grasped by decision maker in the decision process. Meanwhile, entropy serves as an ideal scale for the evaluation on different decision processes and case-handling effects [4,5].

3.2 THE BASIC STEPS FOR CALCULATING THE INDEX WEIGHTS USING ENTROPY METHOD

First, selecting positive or negative indexes as the evaluation indexes according to the research object and calculating y_{ij} according to standard equation, where, the treatment equation for positive index is calculated by (the larger the index, the better)

$$y_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} . \quad (1)$$

The treatment equation for the negative index is given by (the smaller the index, the better)

$$y_{ij} = \frac{\max x_{ij} - x_{ij}}{\max x_{ij} - \min x_{ij}}. \quad (2)$$

Secondly, calculating the entropy of evaluation indexes: in an evaluation issue containing m evaluation indexes and n evaluation objects, the entropy H_j of the *j*-th evaluation index is defined as:

$$H_j = -k\sum_{i=1}^{n} f_{ij} \ln f_{ij} \quad (j=1,2,...,m), \quad (3)$$

where, $f_{ij}=y_{ij}/\sum iy_{ij}$ (i=1,2,…,n), k=1/lnn. Moreover, it is assumed that when $f_{ij}=0$, $f_{ij}\ln f_{ij}=0$. The smaller the entropy of an index, the more useful the information provided by this index to the decision maker.

Thirdly, calculating the entropy weights of evaluation indexes: in the evaluation issue of (m,n), the entropy weight of the *j*-th index is defined as:

$$w(j) = \frac{1-H(j)}{m-\sum_{j=1}^{m} H(j)}. \quad (4)$$

4 An empirical analysis on the weights of the competiveness indexes of five typical clusters in Zhejiang Province

4.1 STANDARDIZATION OF THE ORIGINAL INDEXES FOR THE COMPETIVENESS OF INDUSTRIAL CLUSTERS

In this study, the original data of 5 typical industrial clusters in Zhejiang Province in 2011 were obtained from the data submitting platform for the key industrial clusters in Zhejiang province. Moreover, according to the data in the statistical yearbook of Zhejiang Province, the competitiveness indexes of the industrial clusters were calculated. In view of the competitiveness of support network showed little influences and was hard to be quantitatively analyzed as well, this study merely analyzed the weight of enterprise network competiveness of the industrial clusters using entropy method. The X1 to X5 of competiveness of enterprise network were calculated using the original data of the five industrial clusters on the "data submitting platform for the key industrial clusters in Zhejiang province" of the Zhejiang Province Economic and Information Commission in 2011. The calculation process was neglected in this study [6]. Table 2 shows the calculation results. Table 3 show the calculation results in Table 2 after being treated by the standardization equation of entropy method.

TABLE 2 The original data of the enterprise network competiveness of five industrial clusters in Zhejiang province

Criterion layer	Index code	Index	Typical industrial clusters				
			Shaoxing textile cluster	Haining leather cluster	Cixi household appliances cluster	Yongkang hardware cluster	Yueqing industrial electrical cluster
The competiveness enterprise network B1	X1	Industrial concentration	0.290	0.146	0.065	0.089	0.101
	X2	Spatial Gini coefficien	0.022	0.016	0.001	0.006	0.007
	X3	EG index	0.0210	0.0049	-0.0035	0.0036	0.0046
	X4	Location quotient	2.13	7.47	3.32	14.78	5.39
	X5	The aggregation degree of cluster enterprises (HHI)	0.0014	0.0112	0.0045	0.0024	0.0024
	X6	The derivation degree of cluster enterprises	0.038	0.022	0.050	0.067	0.124
	X7	The innovation degree of cluster enterprises	0.206	0.308	0.197	0.178	0.258

Data source: The "data submitting platform for the key industrial clusters in Zhejiang province" of the Zhejiang Province Economic and Information Commission.

TABLE 3 The standard data of the enterprise network competiveness of five industrial clusters in Zhejiang Province

Criterion layer	Index code	Index	Typical industrial clusters				
			Shaoxing textile cluster	Haining leather cluster	Cixi household appliances cluster	Yongkang hardware cluster	Yueqing industrial electrical cluster
The competiveness enterprise network B1	X1	Industrial concentration	1	0.36	0	0.107	0.16
	X2	Spatial Gini coefficien	1	0.762	0	0.238	0.286
	X3	EG index	1	0.343	0	0.290	0.331
	X4	Location quotient	0	0.422	0.094	1	0.258
	X5	The aggregation degree of cluster enterprises (HHI)	0	1	0.316	0.102	0.102
	X6	The derivation degree of cluster enterprises	0.157	0	0.275	0.441	1
	X7	The innovation degree of cluster enterprises	0.215	1	0.146	0	0.615

4.2 THE CALCULATION AND ANALYSIS ON THE COMPETIVENESS OF INDUSTRIAL CLUSTERS

Index weights were calculated according to the basic steps of entropy method, as shown in Table 4 (the calculation process is omitted).

TABLE 4 The index weight of the competiveness of the typical clusters in Zhejiang Provinces

Warning index X_J		Weight(W_J)
X_1	Industrial concentration	0.167
X_2	Spatial Gini coefficient	0.114
X_3	EG index	0.111
X_4	Location quotient	0.150
X_5	The aggregation degree of cluster enterprises (HHI)	0.190
X_6	The derivation degree of cluster enterprises	0.131
X_7	The innovation degree of cluster enterprises	0.138

According to weights, the indexes were sequenced as X5> X1> X4> X7> X6> X2> X3. This result suggests that the industrial concentration degree, specialization degree (location quotient), and HHI index take relatively high proportions in indexes of enterprise network competiveness. Moreover, industrial concentration, spatial Gini coefficient, and EG index are internally correlated. Gini coefficient and EG index take small proportions since they are used to detect the regional competiveness degree of a cluster.

By comparing the economic benefit indexes of the five typical clusters in Zhejiang Province in Table 5, it is found that Shaoxing textile cluster and Yueqing industrial cluster show higher economic benefits. Comparison on the network competiveness of the five clusters reveals that the two clusters above have high industrial concentration degree but low specification level and enterprise concentration degree. This result suggests that high industrial concentration degree is conductive to improve the cluster competiveness, while low industrial concentration degree means the low competiveness among enterprises and the high overall cluster competiveness formed thereby [7].

TABLE 5 The economic benefit indexes of the five typical clusters in Zhejiang Province

Indexes / Clusters	Labor Productivity(ten thousand Yuan per person)	Per capita profit rate (%)
Shaoxing textile cluster	76.38	2.96
Haining leather cluster	29.29	1.30
Cixi household appliances cluster	30.38	1.05
Yongkang hardware cluster	27.06	2.16
Yueqing industrial electrical cluster	42.1	2.89

Data source: The data submitting platform for the key industrial clusters in Zhejiang province

5 Conclusion and discussion

This study preliminarily attempts to quantitatively evaluate the competitiveness of industrial clusters. As for the large and complex industrial cluster network system, some factors are hard to be quantified and dynamic data are difficult to be collected [8]; limited by time, paper length, and objective conditions, this study merely empirically analyzed the typical industry clusters in Zhejiang Province. This determines the limitation of this study.

Acknowledgement

The work was funded by China's national social science fund program 2011 (No. 11BGL096).

References

[1] Saaty T L 2004 Fundamentals of the Analytic Network Process-Dependence and Feedback in Decision-Making with a Single Network *Journal of Systems Science and Systems Engineering* **6**
[2] Li C 2005 The Method of Selecting Early Warning Index of Real Estate Based On Nuclear System and Coritivity Theory *Mathematics in Practice and Theory* **35**(11)
[3] Li W 2009 Evolution for the Competiveness of the Industrial Cluster based on Analytic Network Process *Science & Technology Progress and Policy* **7**
[4] Peng Y, Shen X 2011 The Commonly used Quantitative Methods in Economic Management *Economic management press Beijing*
[5] Duan S, Gong X 2012 Comprehensive Evaluation of Regional Circular Economy Development based on Entropy Method Taking Xinjiang Province as an Example *Science and Technology Forum of China* **11**
[6] Lu J 2008 Theoretical Foundation and Model Construction for Competitiveness Evaluation of Industrial Clusters *Statistics and Decision* **19**
[7] Lee L, Yu J 2010 Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects *Journal of Econometrics* **154**(2)
[8] Miao C, Wei Y, Lv L 2011 *New Economic Geography Science press, Beijing*

Author

Weidong Wang, China.

Current position, grades: associate professor of Zhejiang University, Post Doctor; China Jiliang University, China.
University study: PhD degree in public administration, Huazhong University of Science and Technology, 2010.
Research activities: urban and regional management, crisis management, industrial economy.
Professional activities and memberships: Director, China Economic Geography Research Association, from 2008.

Wireless vehicle detection node based on tunnelling magneto resistance sensor

Wei Huang*

School of Automation, Hangzhou Dianzi University, Hangzhou, China

Received 15 September 2014, www.cmnt.lv

Abstract

A wireless node based on tunnelling magneto resistance sensor was designed for large scale vehicle detection in Intelligent Transport System. With regard to the sensor's characteristic of high resistance, rapid response and high linearity, the signal acquisition and regulation circuits were designed to meet the requirement of geomagnetic measure for three dimension axis of the sensor. A vehicle noise pre-detection unit was implemented to wake up the microcontroller from sleep state before vehicle enter the detection area. Low power chips were considered and the all sensor units were power supplied by MCU, cooperated with the improved power-efficient ATA algorithm, the power consumption was minimized. Experimental results showed that the designed node was capable of capturing the magnetic feature of different vehicle types and on line vehicle flow detecting for long time.

Keywords: tunnelling magneto resistance, geomagnetic induce, vehicle detection, ATA, WSN

1 Introduction

Vehicle detection and recognition is one of key technologies in Intelligent Transport System (ITS). From vehicle detection sensors, road traffic information are gathered, such as the vehicle occupying, driving in and departing, the vehicle speed and traffic flow. Currently, Surveillance technologies can be divided into two types, the Intrusive and Non-Intrusive [1]. Sensor used in the former are often embed under the surface of the pavement. Examples include inductive loop, pneumatic road tube, piezoelectric cable, and weigh-in-motion system. Drawbacks include the disruption of traffic for installation and repair, failures induced by poor road conditions, and system reinstallation caused by road repairs or resurfaces [2]. Non-Intrusive technology include microwave radar, infrared, Video Image Processing, etc. They are installed above the pavement or on roadside so that the installation and repair of such a system can be done without disrupting the traffic. However, the performance is greatly affected by the environment: confusing signal from sunlight, IR energy is absorbed or scattered by atmospheric particulates, fog, rain and snow. Besides, the devices listed above need auxiliary power supply circuit, which increase the costs and not suitable for distributed in a large scale space.

The increasing traffic congestion is a growing problem in many countries. Motivation for using Wireless Sensor Networks (WSN) to solve the problem needs searching for reliable and cost-effective devices, which can provide traffic data. Magneto Resistance Sensors have received lot of attentions. They are small size, flexibility in deployment configuration, and rarely affected by meteorological environment [3-5].

However, most of proposed research of magneto resistance were based on GMR or AMR, whose performance are limited by the Low-Frequency noise, nonlinear saturation field and low sensitivity for weak magnetic field measurement [6, 7]. There's serious magnetic hysteresis exists in the AMR, so that degaussing circuit are required in the sensor device, which restrict the technique applying widely in WSN apply [8, 9].

This paper presents the design and implementation of a brand new wireless sensor nodes for vehicle detection based on the Tunnelling Magneto Resistance (TMR), manufactured from CMOS-MEMS. The node was designed as an embed system with a microcontroller. The TMRs were used as the sensitive elements in three directions to measure the geomagnetic field of each dimension. Signal conditioning circuit were designed to ensure the analogue signal acquitted accurately. Secondly, An simple aural signal measuring circuit was realized to pre-detecting the vehicle from remote distance by the sound it emitted, so that the main system could in sleep mode for energy saving. Thirdly, an improved ATA detection algorithm was established. By the feature distance between the data series in sample window defined to observe magnetic field change, the algorithm counts the vehicles and schedules the state machine of the system. Outdoor experiments were conducted, the on road testing result shows the node have good performance and suitable for WSN application in ITS.

2 The measuring principle and TMR

2.1 GEOMAGNETIC DETECTION PRINCIPLE

The geomagnetic field is weak, ranged from 0.3 to 0.7Oe, but the magnitude is relative uniform and stable in a square

***Corresponding author** e-mail: hw@hdu.edu.cn

area. Since the Magneto conductivity of Iron is far larger than that of the air, ferromagnetic substance would cause a change in a magnetic field, distort the magnetic lines of flux. Since almost all vehicles have significant amounts of ferrous metals in their chassis the magnetic field disturbance created by a vehicle is sufficient to be detected by a magnetic sensor, which makes it a good candidate for detecting vehicles.

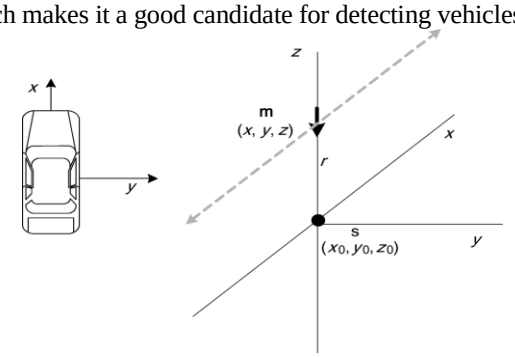


FIGURE 1 magnetic dipole of vehicle moves in x axis

The simplest mathematical model to describe the magnetic signature of a vehicle is a magnetic point dipole with a magnetic moment *m* cantered in the vehicle and parallel to the Earth's field (Figure 1). The field components B_x, B_y, B_z are

$$B_x = \frac{\mu_0}{4\pi}\frac{m_x(2x^2-y^2-z^2)+3m_y xy+3m_z xz}{r^5}, \quad (1)$$

$$B_y = \frac{\mu_0}{4\pi}\frac{m_y(2y^2-x^2-z^2)+3m_x xy+3m_z xz}{r^5}, \quad (2)$$

$$B_z = \frac{\mu_0}{4\pi}\frac{m_z(2z^2-x^2-y^2)+3m_x xy+3m_y xz}{r^5}, \quad (3)$$

where μ_0 is the permeability of free space, m_x, m_y, and m_z are the magnetic dipole moments in *X*, *Y* and *Z* direction, $s(x_0, y_0, z_0)$ and is the distance from the dipole to the observation point.

Suppose a vehicle moves in the x direction, m_x=m_y=0, the magnetic induction of magnetic dipole B_z is determin-ed by m_z and E_z, *where* E_z is the component of Earth's field in Z direction.

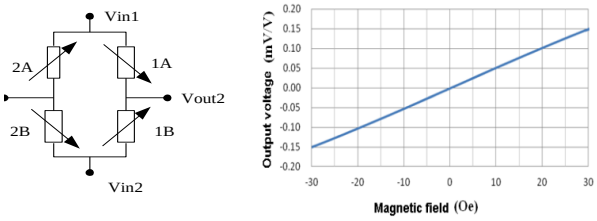
$$B_z = Z_E + \frac{\mu_0}{4\pi}\frac{m_z(2(z-z_0)^2-(x-x_0)^2)}{\left[(z-z_0)^2-(x-x_0)^2\right]^{\frac{5}{2}}}, \quad (4)$$

where B_Z can be used to judge whether a vehicle pass though the detection area. For each type vehicle may have their own ferrous metal distribution, distinct magnetic signature with multiple peaks they yield is benefit to vehicle recognition.

2.2 TMR SENSOR

TMR was first used in hard disk technology. With the change of external magnetic field, the Magnetic Tunnel Junction (MTJ) generates a change in resistance. Under the applied magnetic field, the magnetization direction of free layer will change, while the pinning layer keep still. So the relative orientation between the two layers is changed, resistance of the MTJ across the insulating layer is observed. Since this physical effect is based on the electron tunnelling effect in insulating layer, it is called as the Tunnelling magneto resistance effect [10].

This used the TMR chip MMLP57F with SOP8 packaging and a Wheatstone bridge inside consists of four pieces of TMR elements. The internal structure and output specific are shown in Figure 2.



a) b)

FIGURE 2 MMLP57F specification: a) Internal structure, b) Output feature in±30Oe

While the external magnetic field is changing, the resistance changing trend of 1A and 1B are opposite to that of 2A and 2B. Assume the four elements have equal resistance R when the detecting field is at zero. Output voltage is:

$$V_{out} = V_{-out1} - V_{-out2} = (V_{in1} - V_{in2})\frac{\Delta R}{R}, \quad (5)$$

where V_{in1} and V_{in2} are the bias, ΔR is the absolute value of changes in resistance. The output feature is shown in Figure 2b. Here, bias is 5V, from –30Oe to 30Oe, the sensitivity is 4.8 mV/V/Oe, the nonlinearity less than 1% and the hysteresis error is about 0.1%. That is sufficient to meet the requirement for geomagnetic field detection.

3 Hardwire design of the sensor

3.1 THE STRUCTURE OF NODE

The node is consists of five main units: the acoustic signal detection unit(ASD), the three axis magnetic detection and signal conditioning(MDSC), the wireless transport unit(WTU) and a microcontroller (MCU). The block diagram of the system is presented in Figure 3. The MCU is EFM32G, specially designed for low-power applying [11]. The core of chip is Cortex M3, highest working frequency 32MHz, supports five mode from EM0-EM4 and capable to wake up in 2 μS from sleep mode. Peripherals has the autonomous working ability in CPU sleep duration. The ADC is 8 channel, 12bit, rate at 500K/S with operating current 200μA, which is helpful for high speed data acquisition.

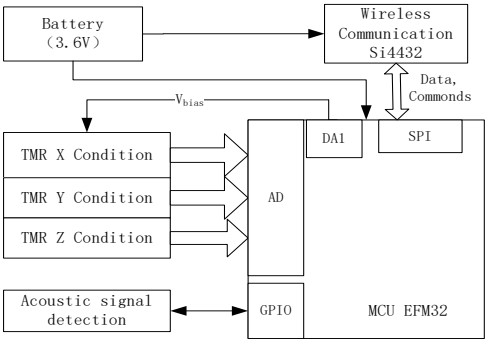


FIGURE 3 The block diagram of the system

Wireless chip Si4432 is used for communication, the bias 3.6V, transmission frequency 900 MHZ, the maximum data transfer rate 256 bps. The diversity antenna, power amplifier, digital modem and a 64B FIFO are internal integrated. External components are one 30MHZ crystal oscillator and a transceiver. The EFM32G communicates to the Si4432 through the SPI interface, the work mode settings and the data are transferred in packets.

3.2 THE MAGNETIC SIGNAL ACQUISITION AND CONDITION

The magnetic induction by vehicles is about -200uT to 200uT, and the responding output range of a single TMR is -32mV to 32mV. So, a power amplifier circuit is considered to conditioning the signal to 2.5V level to satisfy the AD sampling requirement. Since the equivalent internal resistance of TMR bridge is greater, and the magnetic signal to measure is weak, which require the amplifier has high CMRR. The Instrument amplifier chip INA332 can exercise its restraints upon the CMRR and reduce the disturbance from temperature drift and null shift. The chip has low power consumption that meets the system energy management. The conditioning circuit is shown in Figure 4.

The amplifier gain *G* is calculated from:

$$G = 5 + 5 \cdot \left(\frac{R5}{R6} \right), \tag{6}$$

where, R5=147K, R6=10K, G≈80. Vref=1.25 V. The common mode voltage(CMV) of TMR is 3.3V/2=1.65V. To ensure the input signal have maximized dynamic range, the ADC bias is set to 2.5V.

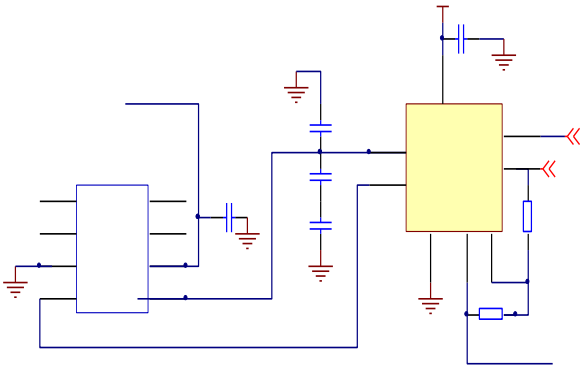


FIGURE 4 The circuit of a single TMR Sensor

Not like other MR chip that should designed with a special power circuit, the three dimension TMR chips are biased by the DA channel of the MCU directly, V_{bias}=3.3V. This is benefit from TMR extremely high input impedance, almost at MΩ level. Meanwhile, MCU provides the power in a intermittent way. The duty time span including the signal setup time and the analogue digital conversion time. For INA332, the AD time is 1μS. The designed circuit limited the amplifier gain under 100, so that operation bandwidth is about to 1MHZ, that minimizes the signal setup time.

3.3 THE ACOUSTIC SIGNAL DETECTION UNIT

To promote energy usage rate, the MCU and magnetic sensor should in speed mode or low power condition in spare time. The acoustic signal detection is used for pre-detecting the audio signals from vehicles, when the threshold is reached, an input is trigged to the MCU, and the system shift into vehicle detecting mode.

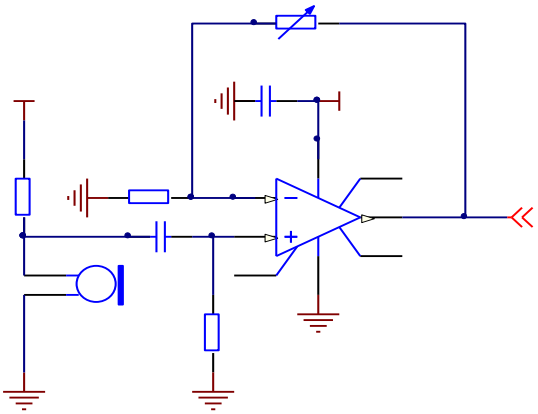


FIGURE 5 The acoustic signal detection circuit

In Figure 5, a two order active amplifier circuit is designed with high pass filtering. The GPIO of MCU have a high driving preferment of 20mA, 3V. External interrupt is supported to wake up the MCU from low power mode (EM1-EM3). The WM-034, an Omni directional electrets microphone has sensitivity of -44dB-36dB. With the in-phase proportion amplifying structure, Op07 magnification is A_{vf}=R4/R2. The range of sensitivity is adjusted by R4, the magnification changed in same way. Generally, the audio frequency the road vehicles emitted between 500HZ to 5KHZ, and that of the background noise is lower than 500HZ. Thus, let the cut-off frequency f_L=500HZ, C1=01.uF, R3=1/2πf_LC1=3.2KΩ.

4 Software design of the sensor

4.1 THE VEHICLE DETECTION ALGORITHM

Jiagen proposed the model of process sensor data in vehicle detection, and based on simple fix threshold value, the Adaptive threshold algorithm (ATA) is applied. But the ATA is not consider of the characteristics of vehicle magnetic signal (VMs), mathematics fault detection may lies in

the algorithm. Second, the algorithm require processing on each sample data, which would keep the MCU in busy mode with large emerge consumption [12].

Here, an improved ATA algorithm (IATA) is using in the sensor, the new algorithm can calculate the signal feature between the background window and the current window by separating vehicle coming and leaving events to realize vehicle detection. Based on the energy saving principle in WSN, Duty-cycling policy is applied to decrease energy consuming [13].

The signal characteristic function *I(K)* is used to distinguish the detecting signal from the background signal. Let:

$$I(k) = be^{ak} + c\,, \tag{7}$$

where $ab>0$. The factors may different for different vehicle type at different speed. For two real time signal sequences $X=\{x(1), x(2), \ldots, x(\mathrm{L})\}$, $Y=\{y(1), y(2), \ldots, y(\mathrm{L})\}$; the normalized power series for *I(K)* is:

$$\delta_f(k) = \frac{I(k)}{\sqrt{\sum_{i=1}^{L} I(K)^2}}\,. \tag{8}$$

The feature distance is defined as:

$$D_L(X,Y) = \sum_{i=1}^{L} \delta_f \cdot \left[y(k) - x(k) \right]^2. \tag{9}$$

4.2 THE IMPROVED ATA MODEL

The algorithm flow chart is shown in Figure 6. There are four basic status in IATA: No Vehicle sleep (NVS), High threshold enter (HTE), Middle threshold hold (MTH) and Low threshold leave (LTL). The feature distance used to detect the end point of VMS for HTE and LTL, while Duty-cycling for NVS and MTH.

Let $s(k)$ as current sample series, S_b is the average value of background, M is the background windows size.

$$S_b = \frac{S(k) + S(k-1) + \ldots + S(k-M+1)}{M}\,. \tag{10}$$

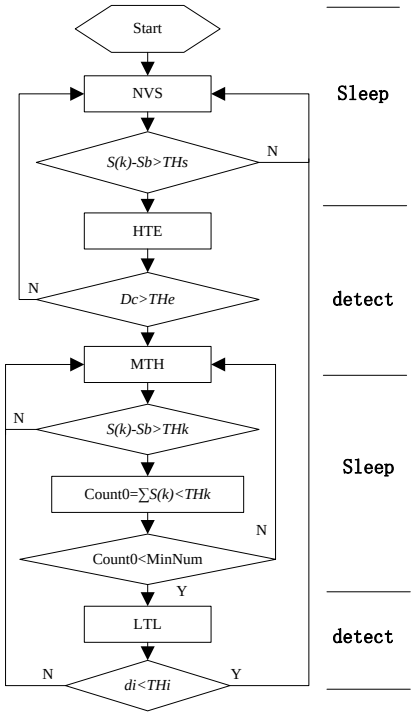


FIGURE 6 The IATA algorithm flow chart

5 Experimental Results and Discussion

The sensor node is in a compact structure, the size is 50mm*50mm, shown in Figure 7. A tiny WSN is built by a Ether2USB device connected to the laptop computer, act as the AP. When the SN is power on, the AP commands are processed in prior, ADC sample rate is set to 100HZ. SN sends a batch of data per second.

5.1 MAGNETIC SIGNATURE OF DIFFERENT PARKING CARS

In this experiment, the Acoustic Unit is Closed. We send AP commands to start the Recording function of SN, by which the SN will upload the raw sample data directly.

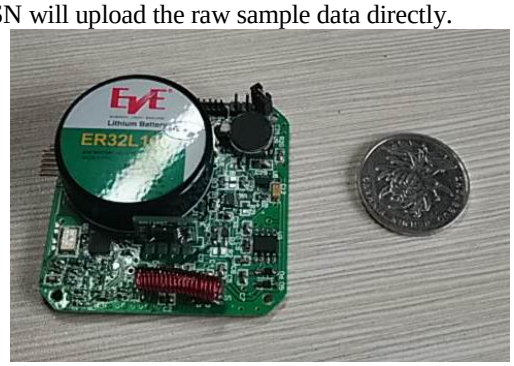


FIGURE 7 The picture of a real SN

We measured on parked vehicles and during parking maneuvers in an outdoor parking area. A compact vehicle (skoda Octavia), a subcompact compact (JinLong) and a truck (GreatWall) were involved. The sensor measured each vehicle on eleven positions along a line centred in the floor

of the vehicle (Figure 8d). D1, D2 are the effective detection area from the head and lap of vehicle respective.

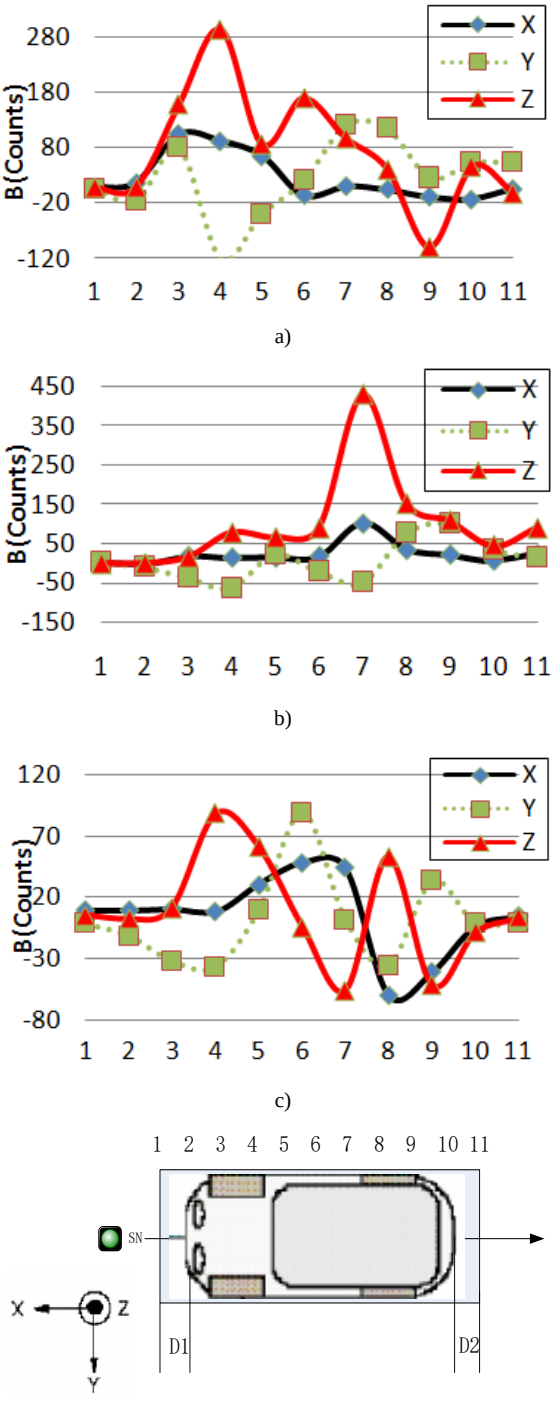


FIGURE 8 Magnetic signature of different parking cars: a) compact vehicle (D1=0.8m, D2=0.6m), b) subcompact vehicle (D1=0.4m, D2=0.3m), c) trunk (D1=0.7mm,D2=0.3mm), d) direction and position

Figure 8(a-c) shows the magnetic induction simultaneously measured by the sensor. Each value is averaged by ten raw data, and the resolution achieved was 36.7nT (12bit), the measurement time was 88mS.

The signatures suggest that compact cars can be roughly described by three magnetic dipoles: two larger dipoles for the engine and wheel axes and a smaller dipole between them. The trunk may have more dipoles with more wheel axes. As to B_Z, the peak counts, peak values and frequencies are distinguished from each type. The B_x and B_y also have distinct features to used for vehicle detecting and movement direction, i.e. the magnetic declination [14].

5.2 ON LINE DETECTING FOR MOVING CARS

The on line test was carried on in a campus lane and a main road to a logistic park. In Table 1, the false and missed count are produced by the IATA detection algorithm and the acoustic detection algorithm respectively. Experiments results shows the average accuracy is 95.8%.

TABLE 1 On line test results

Place	Actual counts	Detected counts	Wakeup times	False	Missed
Lane	86	83	84	1	6
Road	387	394	513	10	3

The magnetic background in campus lane is more stable than the main road, the IATA detection algorithm can had an accurate measurement output. But the missed count in campus is greater than the road, since there the vehicles are most compact type and low speed, if the acoustic threshold not set properly, the MCU would not be trigged when a car passed by slightly. The main road has large flow and through output, and vehicle move at high speed. The acoustic unit trigged about two times than the actual vehicle count because of the noise by adjacent vehicle.

5.3 CONTINUOUS WORKING TIME ESTIMATION

The power consumption of a node is:

$$P = \sum I_{ak} V_{DDK} , \tag{11}$$

were, I_{ak} and V_{DDk} is the average current and supply voltage of the unit *k*.

$$I_{ak} = \frac{t_k}{T} I_k + \left(1 - \frac{t_k}{T}\right) I_s , \tag{12}$$

I_s is current consumption in sleep mode, I_k is current consumption in active mode, T is the clock period, t_k is duty time on each cycle.

TABLE 2 Energy consumption of node units

Unit	Mode	V_{DD}	I_k(mA)
MCU	Active	3.6	0.18
	Sleep		9×10^{-4}
MDSC	Active	3.3	0.02
	Stop	0	0
ASD	Active	3	0.3
	Stop		0
WTU	Rx	3.6	18.5
	Tx		30
	adjust		8.5
	Sleep		1×10^{-3}

Assume the road lane vehicle flow is 220 per hour, and the node works autonomously, the nodes are distributed a long distance between each other, where communication channel collision is low probability. Duty cycle of ASD is 1%, miss trigger rate is 15%, a handshake package sent to AP every 5min and listen to the AP broadcasting per seconds. By Table 2, power consumption in 1 day is about 10.8mW. So the battery (9000mAh) would last about 1.5 years with the consideration of other battery loss of 30%.

6 Conclusion

A magnetic sensor node is proposed to detect the vehicles by measuring then changes of the geomagnetic field. When vehicles arrive or depart, the acoustic sensor detects the changes in illumination and wakes up the sensor node. The TMR sensor measure the magnetic field and IATA algorithm processes the data with energy saving policy of Duty-cycling is building in the control software. Experimental results showed the node can provide fast detection. The Node is compact, low-cost, and very low-power vehicle detector for WSN applications, and that is also reliable and easy to install and maintain.

References

[1] Gupte S, Masoud O, Martin R K, Papanikolopoulos N P 2002 *IEEE Transactions on Intelligent Transportation Systems* **3**(1) 37-7

[2] Su D, Wang L, Ma S 2007 Vehicle Detection Method Based on Magnetoresistive Sensors *Computer and Communications* **25**(136) 9-13

[3] Gao H, Heuer T, Dimitropoulos K 2009 Safe airport operation based on innovative magnetic detector system *Intelligent Transport Systems* **3**(2) 236-44

[4] Lan J, Shi Y 2009 Vehicle detection and recognition based on a MEMS magnetic sensor *Proceedings 4th IEEE International Conference NEMS* 404-8

[5] Feng Z, Wang M, Ni F 2010 Passing vehicle classification algorithm based on AMR sensors *Application Research of Computers* **127**(7) 2533-5 2555

[6] Edelstein A S, Fischer G A, Pedersen M, Nowak E R, Cheng S F, Nordman C A 2006 Progress Toward a Thousand Fold Reduction in 1/f Noise in Magnetic Sensors Using an Ac Micro electromechanical System Flux Concentrator *Journal of Applied Physics* **99**(08B317) 1-6

[7] Wei W, Jizhong S, Lida D 2013 Self-Powered Magnetic Sensor Node for Vehicle Detection *Chinese Journal of Sensors and Actuators* **26**(12) 1734-9

[8] Jianzhong Q, Yan L, Fen W 2009 Application of an isotropic magnetoresistive sensor in vehicle detection *Transducer and Micro system Technologies* **28**(5) 106-8

[9] Sifuentes E, Casas O, Pallas-Areny R 2011 *IEEE Sensors Journal* **11**(8) 1669-79

[10] Sonehara M, Kamada H, Iida S, Sato T 2013 *IEEE Transactions On Magnetics* **49**(7) 3854-7

[11] Silicon LABS 2013 *EFM32_product_brief* http://www.silabs.com

[12] Bohui H, Hang H, Kefu C, Zheng C, Fang T, Liu H 2007 Power Efficient Vehicle Detection Algorithm Using Wireless Magnetic Sensor Node *Journal of Data Acquisition and Processing* **22**(3) 336-41

[13] Dutta P, Grimmer M, Arora A, Bibyk S, Culler D 2005 Design of a wireless sensor network platform for detecting rare, random and ephemeral events *Fourth International symposium on Information Processing in Sensor Networks* LA USA 497-502

[14] Taghvaeeyan S, Rajamani R 2014 *IEEE Transactions Systems* **15**(1) 73-8

Author

Wei Huang, born in February, 1974, Guangyuan, Zhejiang Province, China

Current position, grade: lecturer at the school of automation, HDU, China. Member of the Intelligent robot control institute.
University studies: Computer application.
Scientific interests: wireless network, sensors and testing, research and development of motion control, robotic applications, ITS, industrial automation.
Publications: 8 papers.

Mechanical characteristics of multi-span concrete continuous girder oblique bridge based on ANSYS

Wan Ning*

Southwest Jiaotong University-Emei, Sichuan, China, 614202

Received 6 October 2014, www.cmnt.lv

Abstract

In order to adapt to environment and reduce the damage of construction for environment, the principle that environment adapts to construction is updated for bridge design, therefore lots special-shape bridges including oblique bridge emerge. However, there is no accurate and developed theory for computing the mechanical characteristics of oblique bridge. Generally, equivalent straight bridge is applied for computing. By performing finite element modelling, the internal force and displacement of oblique bridge and equivalent straight bridge with load were compared. The results showed that due to the special structure of oblique bridge, torque is generated under load, accompanying with coupling effect of bend-twist. Therefore, the mechanical characteristics analysis of straight bridge can not accurately explain that of oblique bridge.

Keywords: ANSYS, oblique bridge, computer simulations, finite element modelling

1 Introduction

Oblique bridge is a kind of special-shape bridges that different with straight bridge. Its normal line of bridge axis is not parallel to the support sides, between which the acute angle is called slope, which reflects the lean degree of the bridge. Benefited from high-grade highway, high-speed railway, urban three-dimensional traffic, etc. [1], oblique bridge develops continuously. With the development of economy, road network construction accelerates as well and new traffic lines are not always vertical to the existing lines and waterways in connection and crossing; considering the coordination with surrounding constructions and environment, the new traffic lines have to bring as little damage as possible to the environment; additionally, based on green building principle, the requirements for reducing water resistance ability of water piers and the navigation, and the change of design concept, the idea that traffic lines submit to bridges is not applicable any more. All these factors cause the existence of oblique bridge and its increasing proportion. To ensure quality of traffic lines, oblique bridges are mainly constructed in high-grade highways, such as urban three-dimensional traffic, expressway, and high-speed railway [2-5]. Regarding mechanical characteristics, there are large differences between oblique bridge and straight one. Compared with straight bridge, oblique bridge generates torque and complex internal forces with bend-twist coupling under loads. There are many factors influencing the internal forces of oblique bridge, including supporting stiffness, stiffness ratio of bending and torsion, and the angle between bearing line and axis line, that is, the slope. At present, research on oblique bridge in China and abroad is in the initial stage and there is lack of perfect design theory and method. Moreover, there are no definite stipulations for mechanical characters, construction features, and construction points of oblique bridge in related specifications [6].

2 Static finite element analysis principle

Figure 1 shows a skew girder element in horizontal plane, and there are three degrees of freedom for the node [7-11]. Slope is the angle between the oblique section of end point and *y*-axis (symbol description: it is positive when *y*-axis clockwise turns to the oblique section; otherwise, it is negative).

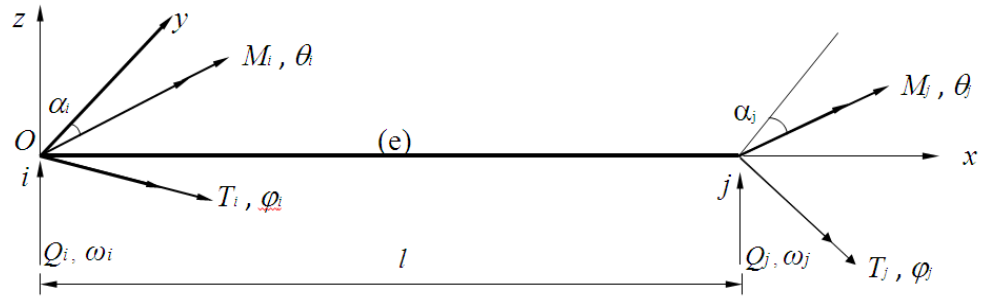


FIGURE 1 Skew girder element

***Corresponding author** e-mail: ningwan@163.com

Suppose the *i* end of the element (*e*) is fixed and the *j* end is free, when there are forces Q_j, M_j and T_j on the *j* end, each component of the internal force of *x* section are:

$$\left.\begin{aligned} Q(x) &= -Q_j \\ M(x) &= Q_j(l-x) - M_j\cos\alpha_j + T_j\sin\alpha_j \\ T(x) &= -\left(M_j\cos\alpha_j + T_j\sin\alpha_j\right) \end{aligned}\right\}. \tag{1}$$

The strain energy of the element is as follows without considering shear deformation:

$$U = \frac{1}{2}\int_0^l \frac{M^2(x)}{EI}dx + \frac{1}{2}\int_0^l \frac{T^2(x)}{GJ}dx. \tag{2}$$

According to Cass Tino theorem, we have:

$$\omega_j = \frac{\partial U}{\partial Q_j}, \theta_j = \frac{\partial U}{\partial M_j}, \varphi_j = \frac{\partial U}{\partial T_j}. \tag{3}$$

Therefore, the relationship of the node displacement and the nodal force of *j* end of the element is expressed by

$$\begin{bmatrix} \omega_j \\ \theta_j \\ \varphi_j \end{bmatrix} = \begin{bmatrix} \frac{l^3}{3EI} & -\frac{l^2\cos\alpha_j}{2EI} & \frac{l^2\sin\alpha_j}{2EI} \\ -\frac{l^2\cos\alpha_j}{2EI} & \frac{l\cos^2\alpha_j}{EI} + \frac{l\sin^2\alpha_j}{GJ} & \frac{l\sin 2\alpha_j}{2GJ} - \frac{l\sin 2\alpha_j}{2EI} \\ \frac{l^2\sin\alpha_j}{2EI} & \frac{l\sin 2\alpha_j}{2GJ} - \frac{l\sin 2\alpha_j}{2EI} & \frac{l\sin^2\alpha_j}{EI} + \frac{lcos^2\alpha_j}{GJ} \end{bmatrix} \begin{bmatrix} Q_j \\ M_j \\ T_j \end{bmatrix}. \tag{4}$$

That is:

$$\boldsymbol{\delta}_j = \boldsymbol{D}_{jj}\boldsymbol{F}_j. \tag{5}$$

The coefficient matrix $\boldsymbol{D}_{jj}$ is the flexibility matrix of the *j* end of the skew girder element. By performing inversion for the matrix, the stiffness matrix $\boldsymbol{K}_{jj}$ is:

$$\boldsymbol{K}_{jj} = \boldsymbol{D}_{jj}^{-1}, \tag{6}$$

where $\boldsymbol{D}_{jj}^{-1}$ is stiffness ratio of bending and torsion, namely:

$$\lambda = \frac{EI}{GJ}.$$

According to the equilibrium condition of the element, the relationship between force of *i* end and *j* end is:

$$\boldsymbol{F}_i = \boldsymbol{H}\boldsymbol{F}_j, \tag{7}$$

where $\boldsymbol{H}$ is stiffness transformation matrix.

The distal stiffness $\boldsymbol{K}_{ij}$ and proximal stiffness $\boldsymbol{K}_{ii}$ of *i* end of the element are:

$$\boldsymbol{K}_{ij} = \boldsymbol{H}\boldsymbol{K}_{jj}, \tag{8}$$

$$\boldsymbol{K}_{ii} = \boldsymbol{H}\boldsymbol{K}_{jj}\boldsymbol{H}^T. \tag{9}$$

Therefore, the stiffness matrix of the skew girder element is:

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{H}\boldsymbol{K}_{jj}\boldsymbol{H}^T & \boldsymbol{H}\boldsymbol{K}_{jj} \\ \boldsymbol{K}_{jj}\boldsymbol{H}^T & \boldsymbol{K}_{jj} \end{bmatrix}. \tag{10}$$

Suppose that the torsion angle function and the vertical displacement function of the element are cubic linear functions, the displacement function of the element is:

$$\boldsymbol{\delta}(x) = \boldsymbol{N}(x)\boldsymbol{\Delta}, \tag{11}$$

where $\boldsymbol{\delta}(x)$ is the vector of the displacement function of the element.

$$\begin{aligned} \boldsymbol{\delta}(x) &= \left[\omega(x) \quad \theta(x) \quad \varphi(x)\right]^T, \\ \boldsymbol{\Delta} &= \left[\omega_i \quad \theta_i \quad \varphi_i \quad \omega_j \quad \theta_j \quad \varphi_j\right]^T. \end{aligned} \tag{12}$$

Furthermore, the shape function of the element is calculated. After obtaining the shape function matrix, the equivalent nodal forces of each load are calculated based on energy principle.

3 Experimental case

The example in the research is a concrete double-line railway bridge with continuous rigid frame and three spans. The slope of the bridge is 45°, and large general finite element software ANSYS14.0 is used [12, 13]. The span of the equivalent straight bridge is the diagonal of the oblique bridge. No. shell181 plate element is applied in the construction of the upper structure of the bridge and the bridge pier model. Solid modelling is constructed by defining nodes first, through which establishing elements, and then building the overall model based on the elements (Figure 2). The rigid connection of pier and girder is simulated by coupling all degrees of freedom using nodes, and the pier bottom is restrained by consolidation. Static load simulation is conducted when two trains cross the bridge at the same time and produce maximum load.

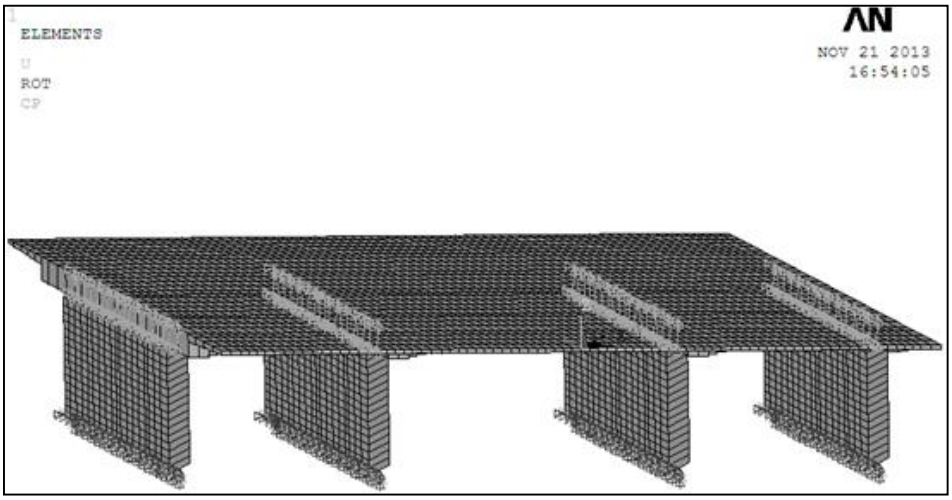


FIGURE 2 Whole oblique bridge model before the trains are loaded

3.1 DISPLACEMENT CALCULATION RESULTS

As there are lots nodes in the model, 21 key points in the oblique bridge and the straight bridge correspondingly are investigated to analyse their displacement, static force, and bending moment. The arrangement of the 21 points is demonstrated in Figure 3.

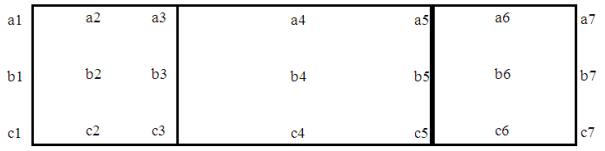


FIGURE 3 Distribution of 21 key points in the research

The static force calculation for vertical displacement of the 21 points of the oblique bridge and the straight bridge is illustrated in Figure 4. By turning the calculation results around y-axis (along the transverse direction), the results are displayed in Figure 5.

Figure 4 shows the calculation results of vertical displacement. It indicates that under common medium-live load of bi-direction, the vertical displacement of the straight bridge is regular. The midpoints of each span are the maximum vertical displacements of each span (with maximum deflection). When the bridge is divided by the central point and there are equal loads along the both sides, the vertical displacement of the bridge is symmetric for the central point of the span. The vertical displacement at flange plate is close to 0 without obvious warping deformation. While under common medium-live load of bi-direction, the oblique bridge shows significant irregular vertical deformation. The calculation results display that the maximum vertical displacements of each span are also at the midpoints of each span. While, besides this, different with the straight bridge, distinct vertical deformations are observed at flange plate and the angular points (a1, c7) of the long diagonal, which causes irregular deformation of the whole bridge.

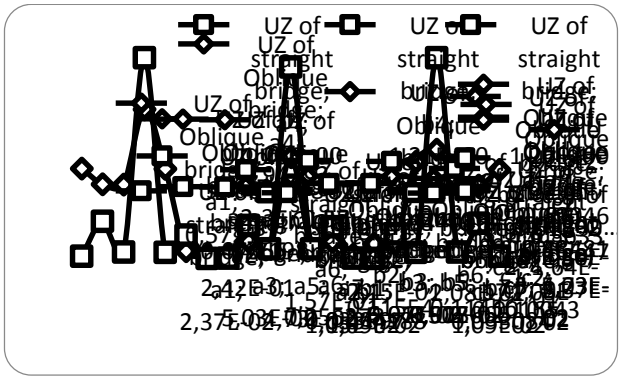


FIGURE 4 Calculation results of vertical displacements of the 21 points of oblique bridge and straight bridge (unit: mm)

The calculation results (Figure 5) demonstrates that the angular displacement of the straight bridge maintains dynamic equilibrium and fluctuates above and below zero alternatively. The displacement is regular without apparent change. Meanwhile, the angular displacement of the oblique bridge is irregular and shows peak values. For examples, there are large angular displacements at the angular point, bridge pier, and the centres of spans at both sides. Therefore, there is obvious twist deformation of the surface of the oblique bridge.

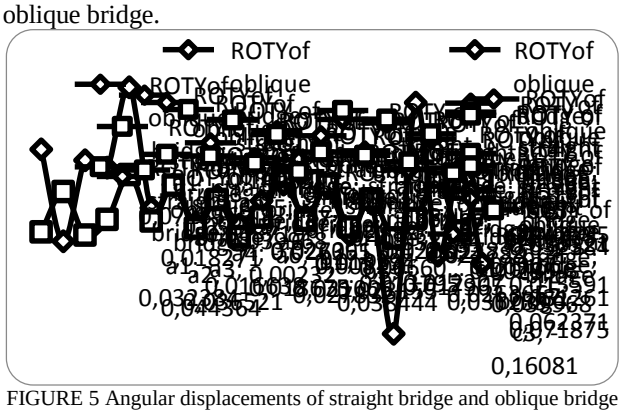


FIGURE 5 Angular displacements of straight bridge and oblique bridge by turning around *y*-axis (transverse direction) (unit: rad)

Merely specific points on each pier were analysed in the calculation of static internal forces, and the concrete positions of the points are demonstrated in the Figure below.

3.2 INTERNAL FORCE CALCULATION RESULTS

The static forces of the oblique bridge and the straight bridge are analysed and the comparison of calculation results of the axial force of the bridges is illustrated in Figure 6.

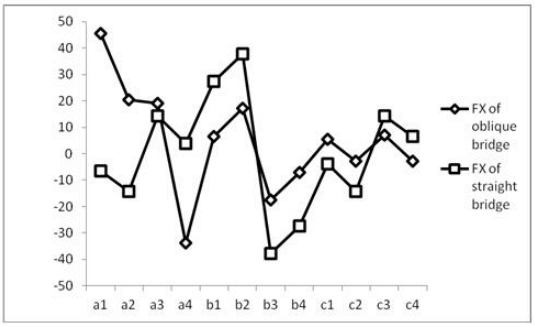


FIGURE 6 Calculation results of horizontal forces along the oblique bridge and the straight bridge (unit: kN)

It is observed in Figure 6 that among the selected points, the maximum axial horizontal force of the straight bridge is at the midpoint of the left central pier. The force is also positive. Maximum axial horizontal force difference (about 80 kN) is found between the midpoints of two central piers, where the directions of the two forces are opposite. For the oblique bridge, the variation of the axial horizontal force is mainly observed at two ends of the bridge, with a maximum difference of 90 kN around. Obviously, the mechanical character is different with that of the straight bridge.

Figure 7 shows that the maximum positive bending moment of the straight bridge is at the midpoint of the bridge surface and the maximum negative bending moment is at the left central pier. Considering the middle span, all the bending moments are large and positive and there is distinct variation of the bending moment along the axis. For the oblique bridge, the maximum positive bending moment is at point a3, which is near to the flange and the maximum negative bending moment is at the left bottom point a1. Meanwhile, the variation of bending moment along the axis is less than that of the straight bridge, with a maximum difference about 3 times. All these reveal that, for safety, it is better to calculate the bending moment along the axis using the data of straight bridge; while as the bending moments at side span and angular points of bridge surface are obviously different between the two bridges, the calculation of bending moments of oblique bridge using data of straight bridge can not ensure safety.

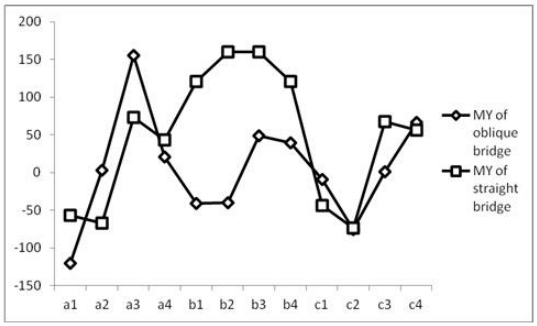


FIGURE 7 Bending moments of the oblique bridge and the straight bridge by turning around y-axis (unit: kN·m)

4 Conclusions and analysis

1) It is safe to calculate vertical displacement of oblique bridge using equivalent straight bridge at the midpoints of each span, where the vertical displacement of straight bridge are larger than that of oblique bridge. However, the vertical displacement of straight bridge is almost zero at flange and the oblique crossings of two ends; while oblique bridge generates coupling effect of bend-twist with load due to its special structure, and therefore leads to a larger internal force and large vertical displacements at the above positions. For example, obvious difference can be observed at the angular points (a1. c7), where there is no vertical displacement for straight bridge while about 0.6 mm of vertical displacement for oblique bridge. So, straight bridge cannot be applied in calculating vertical displacement of oblique bridge; otherwise, the safety of the oblique bridge cannot be guaranteed in usage.

2) The calculation of angular displacement by turning around y-axis shows that the angular displacement of straight bridge is stable without obvious variation; while owing to not completely symmetric structure, the oblique bridge generates torque under loads, accompanying with bend-twist coupling. This intensifies the local torsional deformation, for examples, the deformations at a5, c3, and c4. But the straight bridge is not influenced by the factor.

3) In the calculation of axial horizontal force, there is significant difference between the results of oblique bridge and the straight bridge. At some positions, the axial horizontal forces are in opposite directions, such as a1 and a4. There is about 50 Kn difference between the horizontal forces in the two computations, and the forces are in opposite indirections. Furthermore, the difference increases in dynamic calculation due to the movement of the loads. The inaccurate computation of horizontal force fails to accurately calculate and judge the stress and deformation of the bridge piers, and cannot provide valid data and reference for the design of lower structure. All these possibly cause transfinite deformation, even collapse of the bridge piers, therefore cannot guarantee the safety of the oblique bridge structure in usage.

4) Figure 7 demonstrates that, except several points are close to the straight bridge, the bending moments of oblique bridge around y-axis are significantly different with that of the straight bridge. This is clearly observed on the axis and angular points. On the axis, the maximum My difference reaches 200 Kn·m with opposite directions. When calculate bending moment using equivalent straight bridge, the tension and compression of girder and pier concrete of the oblique bridge cannot be accurately estimated. This probably causes that the tension of certain side of the concrete exceeds its tensile limit and therefore cracks. This inevitably threatens the safety of the bridge.

In summary, actual mechanical condition of oblique bridge cannot be accurately obtained in the computation of internal force and deformation using equivalent straight bridge, as the method fails to accurately consider or simulate the additional bending moment, coupling effect of bend-twist, non-uniform internal force and complex stress-strain

of the actual structure of oblique bridge. Presently, many scholars suggest that to improve the efficiency of the equivalent computing method by adding certain correction coefficient. However, with limited experience, specifications of each country have not provided accurate correction coefficients for each factor. Therefore, for safety, oblique bridges are generally designed with excessive bearing and cost vast manpower and material resources. The method for computing mechanical characteristics of oblique bridge using equivalent straight is not scientific and applicable. Further research is needed to provide more accurate computing method or correction coefficients for each factor apart from finite element simulation.

References

[1] Najafi F T, Nassar F E 1996 Comparison of high-speed rail and maglev systems *Journal of transportation engineering* **122**(4) 276-81

[2] Zheng N, Aboudolas K, Geroliminis N 2013 Investigation of a city-scale three-dimensional Macroscopic Fundamental Diagram for bi-modal urban traffic *Proceedings of the IEEE Conference on Intelligent Transportation Systems ITSC* 2013 1029-34

[3] Liu S, Wu Q 2009 Theory of "Compensating differential" for land expropriation in construction of expressway in China *Proceedings of the 2nd International Conference on Transportation Engineering ICTE* 2009 735-40

[4] Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103

[5] Chen Y-S, Li B, Xiao R-M 2011 Transportation status of Chinese expressway network in 2010 *Jiaotong Yunshu Gongcheng Xuebao/Journal of Traffic and Transportation Engineering* **11**(6) 68-73 *(in Chinese)*

[6] Nie Y, Qian C 2011 Constitutive relation of sulfate attacked concrete based on uniaxial loading *Advanced Materials Research* **168-170** 50-6

[7] Wang Z-C, Ren W-X 2011 Dynamic analysis of prestressed concrete box-girder bridges by using the beam segment finite element method *International Journal of Structural Stability and Dynamics* **11**(2) 379-99

[8] Zhang C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34

[9] Yang S-R, Zeng Q-Y 2011 Analysis of steel truss girder bridge vibrations by finite truss element method *Shenzhen Daxue Xuebao (Ligong Ban)/Journal of Shenzhen University Science and Engineering* **28**(3) 195-9 *(in Chinese)*

[10] Zhang Y, Li L, Lin L, Sun X 2013 Beam-segment finite element analysis on shear lag effect of thin-walled box girder adopting additional deflection as generalized displacement *Tumu Gongcheng Xuebao/China Civil Engineering Journal* **46**(10) 100-7 *(in Chinese)*

[11] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81

[12] Zeng Y 2013 Three-dimensional rock slope stability analysis of the longjiang river bridge *Proceedings of the 2013 5th Conference on Measuring Technology and Mechatronics Automation ICMTMA 2013* 1260-3

[13] Zhao M-H, Yin P-B, Zhang Y-J, Yang M-H 2013 The design and calculation method of pile-column bridge pier foundation in high and steep slope *Lixue/Engineering Mechanics* **30**(3) 106-11 *(in Chinese)*

Author

Ning Wan, born in November, 1978, Sichuan Province, P.R. China

Current position, grades: master at the School of Southwest Jiaotong University, China.
University studies: MSc in Electrical Engineering and Automation in 2008 at the University of Southwest, Jiaotong University in China.
Scientific interests: structural dynamics, earthquake bridge.
Publications: more than 5 papers.
Experience: teaching experience of 12 years, 3 scientific research projects.

Design of new type biomass pellet forming machine with plunger roller and ring-die at room temperature

Xiangyue Yuan, Zhongjia Chen*

School of Technology, Beijing Forestry University, Beijing, 100083 China

Received 6 October 2014, www.cmnt.lv

Abstract

Bio-energy has become the fourth major energy resource after oil, coal, natural gas energy on account of the properties of green, clean, and renewable. And biomass densification is an important technology which makes the biomass resource be low cost and high value. But for current biomass forming equipment, there exist common shortcomings of high energy consumption, low efficiency, serious wear of forming parts, and etc. In this paper, a forming mechanism of a new type biomass pellet machine with plunger roller and ring-die was proposed, which using the mesh extrusion pressure between plungers and forming cavities (or called holes) to make the loose biomass material densification. It could avoid direct contact and intense friction of forming parts and biomass materials out of the holes. So the energy consumption of the materials in ineffective forming process was reduced and the wear of forming parts was effective alleviated. Then the new type biomass pellet forming machine was designed according to the requirements by Solidworks after force analysis and theoretical calculation, and a prototype was built with the whole power 22 kW and production capacity 450 kg/h. Compression experiments were done finally on the prototype. It was shown that the forming quality of solid pellet fuel tended to be stable after 30 minutes' running and the pellet density could reach more than 0.8 g/cm^3. So it represented that the designed pellet machine had a good practical operation and met the design demands.

Keywords: forming technology of biomass densification, pellet forming mechanism with plunger roller and ring-die, pellet fuel

1 Introduction

Energy shortage and environmental pollution are common serious problems currently restricting the development of world economy and the society. So it is the necessary way for the future to develop renewable and green energy, which means a low-carbon, sustainable and scientific development [1]. Although biomass materials are widely distributed with a good renewability and richness in natural resources, its greatest inadequacy lies in the scattered distribution, the low original density, and low energy density [2], which severely limits the direct utilization of biomass energy. So technologies should be developed to convert bio-materials into a new form fuel with high unit energy density.

Biomass densification is an important kind of technologies to achieve the target of the biomass resource utilization with low cost, high value and efficiency at present. The process of biomass densification is that compress biomass material (mainly residues of forest felling, thinning, processing and agriculture) physically at heating or room temperature into solid fuels shaped as columnar, cube, or pellet with high unit density after the pretreatment of drying, crushing and etc. [3]. The unit density and combustion value of the solid fuels after processing could respectively reach 0.8~1.4t/m^3 and 16~21MJ/kg in terms of kinds of biomass materials such as agricultural straws, grasses, woody resources like wood chips and sawdust. Combustion efficiency could be as high as 90%, and the conversion cost is low as well [4].

So far, there are many kinds of biomass forming machine, but an objective reality that the high power consumption (≥100kw/h per ton) and quick wear of main forming parts (≤1000h) still exists, which restricting the development of biomass solid fuels. To overcome shortcomings in traditional forming machines, a completely new type of biomass pellet forming mechanism with plunger roller and ring-die was designed after deeply studied the densification of biomass material in this paper, which effectively avoided the extrusion and friction of biomass material out of forming cavities. It would greatly reduce the power consumption of densification, and prolong the service life of forming molds.

2 Design of biomass pellet forming machine with plunger roller and ring-die at room temperature

2.1 THE WORKING PRINCIPLE OF THE FORMING MECHANISM

A new forming mechanism with plunger roller and ring-die was presented after deep study on compression and shaping process of many kinds of biomass forming machines, shown in Figure 1, which was similar to a pair of inner meshing gears. It depended on the mesh extrusion pressure between plungers circumferentially on the roller and forming cavities uniformly distributed in the ring-die to make the loose biomaterials densification. Compared with the traditional rolling ring-die forming mechanism, shown in Figure 2, it avoided direct contact between the ring-die and the roller,

***Corresponding author** e-mail: chenzhongjia@bjfu.edu.cn

reduced mutual extrusion and friction between the roller and bio-materials out of the forming cavities and alleviated the force of forming parts. Therefore, it could effectively reduce the power consumption of biomass forming and wear of forming parts to improve the service life [5].

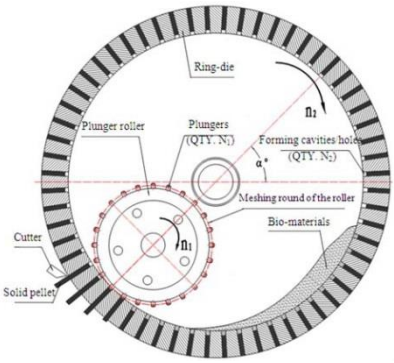


FIGURE 1 The working principle of biomass pellet molding mechanism plunger roller ring-die

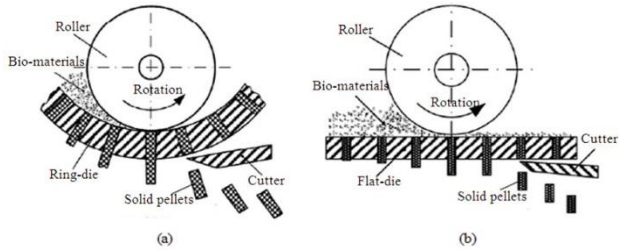
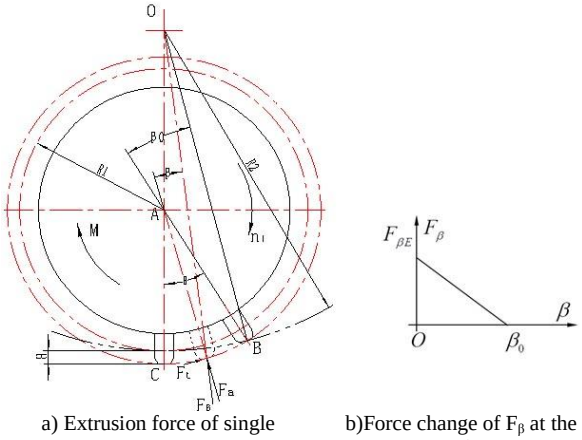


FIGURE 2 The traditional rolling ring-die forming mechanism

All of the plungers and forming cavities were uniformly distributed along circumference in the Figure 1. Let the quantity of plungers and forming cavities of each row separately be N_1 and N_2, the revolving speed of plunger roller and ring-die be n_1 and n_2 in turn. Then when the four parameters matched the following equation, the movement between the plungers and forming cavities was similar to the meshing motion of the gear: $\frac{N_2}{N_1}=\frac{n_2}{n_1}=\left(\frac{Z_2}{Z_1}:gear\right)$.

2.2 FORCE ANALYSIS OF PLUNGER RING-DIE FORMING PARTS

Figure 3 was the force schematic of plunger roller. In the Figure 3a, the plunger roller was rotated at speed of n_1 driven by torque (M) after bio-materials filling into the forming holes in the ring-die. The plunger began to mesh with forming hole at the position of B, that is to say, bio-materials in the hole were squeezed by the plunger. When rotated to the position C, the plunger reached a maximum depth of engagement with H mm, and it would gradually withdraw from the hole with the continual rotation of roller. It was thought that the top end of the plunger was subject to the action of force F_β to the centre of the ring-die O when the materials was squeezed because of the larger radius of ring-die compared with roller. The angle between F_β and the centre line of the roller was β. F_β could be decomposed into the radial force of the roller F_a and the tangential force F_t, so F_t would be the resistance to the rotation of roller. It could be simplified to a static force process due to the short residence time in the meshing zone BC of the plunger and higher rotating speed of the roller.



a) Extrusion force of single plunger

b)Force change of F_β at the top end of plunger

FIGURE 3 The force schematic of the plunger roller

O and A were the rotation centre of the ring-die and roller respectively. Point B and C were the entry and exit points of engagement separately. R_1 and R_2 were radius of roller circle and ring-die inner circle in turn. M was the driving torque, and n_1 was the rotating speed of the roller. θ was the central angle of the roller at the mesh zone of plunger and forming hole. β was the angle between axes of plunger and hole and the maximum was β_0 (i.e. at the position of B). F_β was the force of the plunger, and F_a and F_t were the radial and tangential force of the roller respectively, while the maximum was $F_{\beta E}$ at the position of C. H was the maximum insert depth of plunger.)

For ease of calculation and analysis, F_β was simplified as a linear growth in the Figure 3b, and the following Equations could be obtained.

$$F_{\beta E}=P_0\frac{\pi D^2}{4},\ \beta_0=\cos^{-1}\frac{R_2^2+(R_1+H)^2-OA^2}{2R_2(R_1+H)}, \quad (1)$$

where, P_0 was the maximum extrusion pressure per unit (MPa), D was the diameter of the forming hole.

Equation (2) can be calculated from Equation (1).

$$F_\beta=F_{\beta E}-\frac{F_{\beta E}}{\beta_0}\beta. \quad (2)$$

Decompose F_β into radial force F_a and tangential force F_t, and Equation (3) can be received:

$$F_t=\sin\beta F_\beta,\ F_a=\cos\beta F_\beta, \quad (3)$$

where resistance torque produced by the tangential force to the roller is:

$$T=\sum_{\beta=\beta_0}^{0}F_t\cdot(R_1+H)=(R_1+H)\sum_{\beta=\beta_0}^{0}\sin\beta\left(F_{\beta E}-\frac{F_{\beta E}}{\beta_0}\beta\right)= \quad (4)$$
$$(R_1+H)F_{\beta E}\left(\frac{\sin\beta_0}{\beta_0}-1\right).$$

The resultant of all radial force to the roller was R_a:

$$R_a = \sum_{\beta=\beta_0}^{0} F_a = \sum_{\beta=\beta_0}^{0} \cos\beta F_\beta = F_{\beta E}\left(\frac{\cos\beta_0 - 1}{\beta_0}\right). \quad (5)$$

Plungers would squeeze the bio-materials in the range of angle θ, and θ was received through the geometric relationship in the Figure 3a:

$$\theta = \cos^{-1}\frac{OA^2 + (R_1 + H)^2 - R_2^2}{2OA\cdot(R_1 + H)}. \quad (6)$$

Bio-materials were extruded by more than one plunger when the roller rotating because several ones were circumferential distributed per row on the roller. So let the number of meshing pair between plunger and forming hole which took part in squeezing simultaneously be ω:

$$\omega = \frac{\theta}{2\pi / N_1}, \quad (7)$$

where, N_1 was the number of plungers per row on the roller and ω should be an integer.

So the resistance torque and the resultant radial force are:

$$T^{total} = \omega Ts,\ R_a^{Total} = \omega R_a. \quad (8)$$

Equation (9) can be obtained integrating Equations (4), (5) and (8).

$$\begin{aligned} T^{Total} &= \frac{\theta}{2\pi / N_1}(R_1 + H)P_0\frac{\pi D^2}{4}\left(\frac{\sin\beta_0}{\beta_0} - 1\right), \\ R_a^{Total} &= \frac{\theta}{2\pi / N_1}P_0\frac{\pi D^2}{4}(\cos\beta_0 - 1)\frac{1}{\beta_0}. \end{aligned} \quad (9)$$

Bio-materials in each forming hole were extruded by appropriate plunger from the position *B*, and it reached the maximum depth *H* when at the position *C*. During the process, extrusion force increased gradually from 0 to maximum P_0 with the degree of the materials compression. Curve 2 in the Figure 4 showed the actual force change. To measure the complicated friction resistance between bio-materials and hole wall in compression process, relationship of extrusion force and deformation was simplified as a positive correlation, shown in curve 1 of the Figure 4.

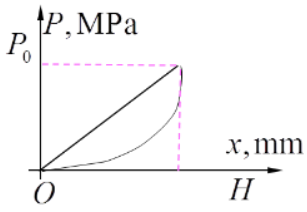


FIGURE 4 The relationship between deformation and pressure of bio-materials in one forming hole

Curve 1 and 2 were the linear simplified and actual relationship respectively between deformation and pressure of bio-materials in straight forming hole.

According to the forming theory of open straight hole [6, 7], it was known that the radial force generated by bio-materials in the mould, friction per unit length of straight hole wall, and the total friction are:

$$P_r = \varepsilon P, \quad (10)$$

$$dF_f = fP_r ds = 2\pi Df\varepsilon P dx, \quad (11)$$

$$F_f = \int_0^H dF_f = 2\pi f\varepsilon D\int_0^H P dx = \pi Df\varepsilon P_0 H. \quad (12)$$

The value of the total friction F_f based on curve 1 in Figure 4 was slightly larger than the actual friction of the straight hole for more safe and reliable.

Similarly, the total friction of the ring-die was expressed in the Equation (13) because there were several (the number was ω) forming holes working for extrusion simultaneously at forming zone *BC*:

$$F_f^{total} = \omega F_f = \omega\pi Df\varepsilon P_0 H. \quad (13)$$

In Equations (10)-(13), ε is a side pressure coefficient and *f* was friction coefficient of materials and forming hole (usually to be 0.1-0.7) [8].

Equations (1)-(13) are the approximate static calculation results for pressure of plunger and forming hole per row. If the row number of plunger and forming hole was *N*, the force of forming mould should be *N* times.

2.3 GENERAL DESIGN OF THE FORMING MECHANISM WITH PLUNGER ROLLER AND RING-DIE

The pellet forming machine was mainly composed of transmission system, forming parts, enclosure bodies and frames, and so on, where, the key points of the design were the transmission system and size of the roller and the ring-die.

2.3.1 Design of the transmission system

Figure 5 was the design of the transmission system. It was known that the ring-die spur gear 2 and the roller spur gear 1 with their shafts were in rigid connections respectively. The ring-die and the roller were mounted on the frame by the bearings, and the idle gear with its shaft was mounted on the frame as well.

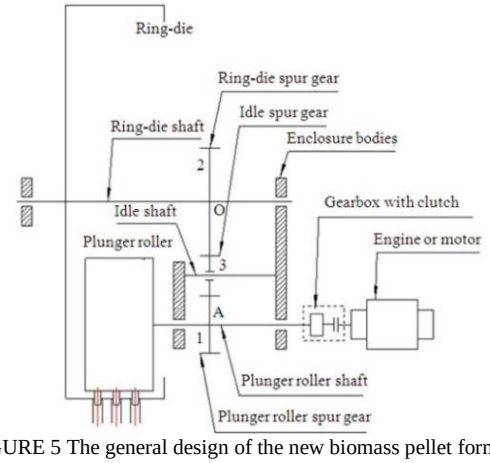


FIGURE 5 The general design of the new biomass pellet forming mechanism

The engine power was transported to the roller shaft through the clutch, which driving the roller rotation. Then the ring-die parts rotated by the gear transmission system 1-3-2, and finally turned round together with the plunger roller at a certain ratio. The running speed of the forming mould could be changed by gearbox in real time according to the actual forming conditions of the materials' characteristics to improve the forming fuel quality and productivity.

Considering the transmission system and design demands, spur gears parameters of the roller and the ring-die could be determined finally referred to size of the same type of pellet forming equipment existing on the market, shown in Table 1.

TABLE 1 The spur gears' design parameters of the plunger roller and the ring-die

Z_1	Z_2	Z_3	m/mm	i	R_1/mm	R_2/mm	OA/mm
24	60	24	3	2.5	132	330	198

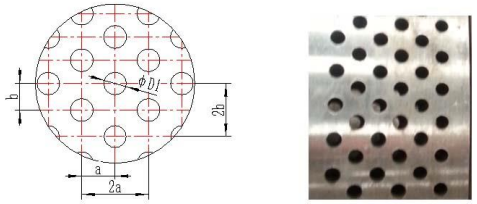
2.3.2 The design of the key forming parts

1) Design of the ring-die parts: according to the design requirements, the productivity (Q) of the new type forming machine was greater than or equal to 450 kg/h. The rotation speed of the ring-die was elected to be n_2, which was equal to 240r/min, to ensure the forming mould working under lower shock load. The layout of the forming straight hole in the ring-die was shown in Figure 6. In the figure, two adjacent rows of straight holes were symmetrical and staggered distribution, which could increase the hole wall thickness and its strength. And the pressure of the roller and the ring-die could be relatively more dispersion at the same time when the bio-materials in the forming holes were squeezed in turn by each row of plungers on the roller. So it was benefit to the smooth operation of the mould. As the diameter of pellet fuel was already determined (D was 8mm according to the design demands), the diameter of the forming hole (D_1) should be 8.5 mm in order to ensure the enough buckling strength of the extrusion plunger body in the same size, and offset the transmission error caused by gear backlash and machining error.

The number of holes per row would be determined based on the productivity of design requirement:

$$N_2 = \frac{4\times10^6 Q}{60\pi N D^2 H \rho n_2 \zeta} \approx 114.9\,, \tag{14}$$

where, Q=500kg/h, N=6, D=8mm, H=6mm, ρ=0.9g/cm^3 (density of pellet), n_2=240r/min, ζ=0.167.



a) Layout of forming holes in the ring-die b)Actual forming holes

FIGURE 6 The straight forming hole layout of the ring-die

Let the number of holes per row (N_2) be 120, and the design parameters of the ring-die was listed in Table 2.

TABLE 2 The structure parameters and dimensions of the ring-die

Inside Diameter/mm	Outside Diameter/mm	Number of Rows/N	Number of holes per row/N2	Hole Diameter D1/mm	Space in rows/mm
φ660	φ760	6	120	φ8.5	12.5

2) Design of the plunger roller parts: the number ratio of plungers and holes per row must be equal to the speed ratio of the roller and the ring-die because all the plungers on the roller and the corresponding straight holes in the ring-die needed precision engagement. Based on the data in the Tables 1 and 2, the number of plungers (N_1) per row on the roller should be 48 (N_1=N_2/i). Like the layout of forming hole in the ring-die, the plungers were circumferential uniform distributed. The position and space of the two adjacent rows were same as the forming holes. The design parameters of the plunger roller parts were expressed in Table 3 and the structure was in Figure 7.

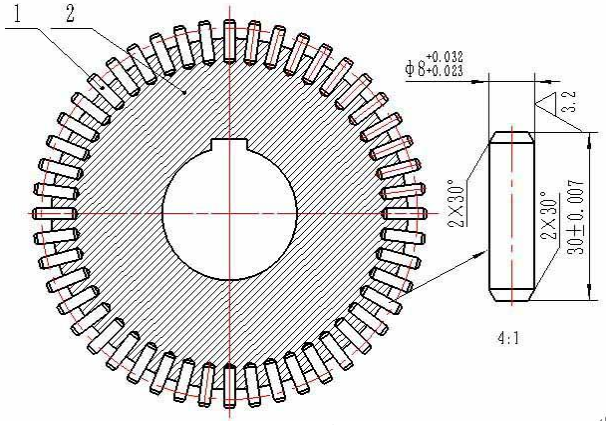
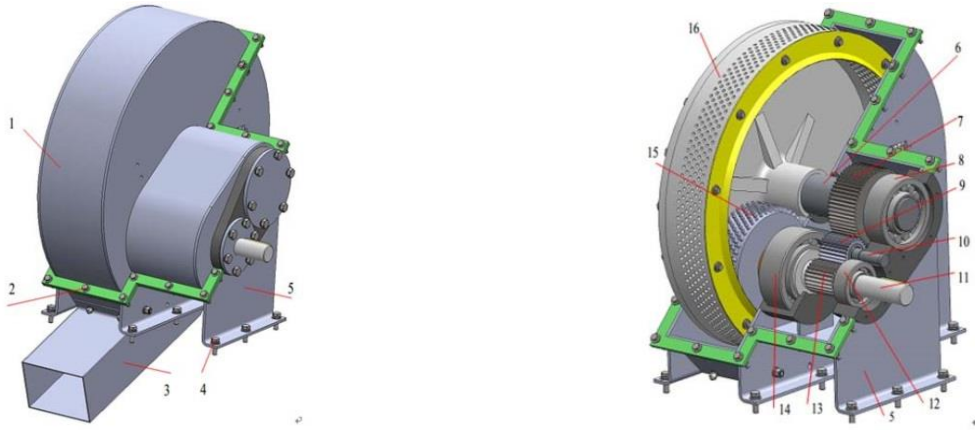


FIGURE 7 The structure schematic of plunger roller. 1) The plunger body 2) The roller body

TABLE 3 The structure parameters and dimensions of the plunger roller parts

Mesh Circle Diameter of the Roller /mm	Mesh Depth of the Plunger /H/mm	Row of the Plunger /N	Number per Row/N_1	Diameter of Plunger/D/mm
φ264	6	6	48	φ8

According to the above design of several parts and reference of similar equipment on the market, the new biomass pellet forming machine with plunger roller and ring-die was designed completely, shown in Figure 8.



a) Appearance of the forming machine b) The internal structure of the forming machine

FIGURE 8 Complete design of the new type biomass pellet forming machine 1) Upper enclosure body. 2) Enclosure body connecting bolts. 3) Outlet hopper. 4)Forming machine fixing bolts. 5) Lower enclosure body. 6) Ring-die shaft. 7) Ring-die spur gear. 8) Ring-die bearings. 9) Idle spur gear. 10) Idle shaft. 11) Plunger roller shaft. 12) Right plunger roller bearings. 13) Plunger roller spur gear. 14) Left plunger roller bearings. 15)Plunger roller 16) Ring-die

2.4 POWER CALCULATION FOR THE FORMING MACHINE

The other parameters of angle and force could be calculated based on the Equations (1)-(13) and the design parameters of the forming parts. And the total weight of the plunger roller parts and the ring-die parts could be obtained as well on the foundation of design, listed in Table 4.

TABLE 4 The other parameters of angle and force for forming parts

θ/°	β_0/°	T^{total}/N·m	R_a^{total}/N	F_f^{total}/N	G_P/N	G_R/N
22.05°	13.02	119.7533	11258.133	21191.5	592.7435	1977.64

Note: β was the angle between axes of plunger and hole and the maximum was β_0.

Table 5 shows the bearing types for both side support of ring-die and plunger roller respectively.

TABLE 5 Bearing types

Right side of ring-die	Left side of ring-die	Right side of plunger roller	Left side of plunger roller
NJ2316E	NJ2316E	NJ2310E	NJ2319E

Let frictional coefficient μ and friction arm d_m of the rolling bearing to be 0.006 and $(d_i+D_0)/2$ apart, then the rolling friction torque of the couple forming parts would be get in the Equation (15), and the power consumption of the friction torque in Equation (16) below:

$$T_1 = \mu(G_1 + R_a^{total})d_{m1},\ T_2 = \mu(G_2 + F_f^{total})d_{m2}, \tag{15}$$

$$P = \frac{Tn}{9550}, \tag{16}$$

where, P – power consumption (kw); T - torque (N·m); n - rotation speed (r/min); d_i - inner ring diameter of bearing (mm), D_0 - outer ring diameter of bearing (mm).

The calculations of resistant torque and power consumption for the couple forming parts were listed in Table 6.

For probably fieldwork, a chipping equipment (power was P_2=3 kW) was installed at the feed inlet to improve the application performance of the forming machine. So the total power consumption of the whole machine was calculated below.

TABLE 6 The resistant torque value and power consumption of forming parts

	Resistant torque / $N \cdot m$			n/rpm	P/kW
	T^{total}	$T_{1,2}$	$\sum T$		
Roller part	119.753	10.488	130.241	600	8.1827
Idler part	–	–	–	240	–
Ring-die part	–	17.376	17.377	240	0.4367
				$P_1=\sum P$	8.6194

Note: Speed of forming mould n was calculated according to the theoretical productivity Q=450 kg/h

$$P_0 = k\frac{P_1 + P_2}{\eta} = 1.3\frac{3 + 8.6149}{0.85} = 17.8 \text{ kW}, \tag{17}$$

where, P_0 was the power of the whole forming system with unit kw, k was the power factor (k=1~1.1 when $P_0 \leq 5$kW, k=1.1~1.2 when P_0=5~10kW and k =1.2~1.4 when P_0>10kW), and was selected to be 1.3 according to the power calculation. η was the total efficiency of the transmission system (η=0.80~0.90), and was selected to be 0.85.

It could be known from Equation (17) that total power consumption of the new pellet machine with productivity 450 kg/h was about 39.6 kW/t theoretically, much less than that of traditional forming equipment on the market (about 80~100 kW/t) [9].

A multi-cylinder diesel engine with type of ChangChai ZN3850Q, was chosen finally in the light of lighter weight at the same power to make the machine with enough power reserve. And the main parameters were listed in the Table 7.

TABLE 7 The main parameters of ChangChai ZN385Q diesel engine

Number of cylinders	Engine displacement	Rated power	Idle speed	Maximum torque	Fuel consumption	Cooling	Shape size	Weight
Unit	L	kw/ rpm	r/min	N.m /rpm	g/kw.h	–	mm	kg
3	1.617	22/2400	800	91 /1800	≤220	Water cooling	614×492×628	185

3 Forming test on the prototype of the new biomass pellet forming machine

A prototype of the new biomass pellet forming machine was assembled completely after design and machining, shown in the Figure 9, which included a drum chipper directly mounted on the side panel of lower enclosure body for integration fieldwork operation. And the machine and engine were fixed on the frame by bolts.

FIGURE 9 The biomass forming system with plunger roller and ring-die at room temperature 1) Frame. 2) Forming mechanism. 3) Drum chipper. 4) Clutch lever. 5) Gearbox lever. 6) Changchai ZN845Q diesel engine. 7) NJ130 Gearbox. 8) Clutch

3.1 UNLOAD TEST

Unload test must to be done before the actual operation to keep normal meshing of each couple of plungers and forming holes. The test was conducted according to the following steps.

Firstly, started the forming machine by hand. The transmission must be in "neutral", and turned the roller shaft slowly by manually to run the forming mechanism. Any sound like metal friction and knocking must be noticed during the rotation. After several times of manual rotation, friction between plunger and entrance to forming hole would be well fit.

Then, started the diesel engine, and unload test with multi-speed running would be done.

1) The forming machine was idle run driven by the engine at the speed of about 1000 r/min with the first gear engagement, and maintained the speed of 10 minutes to make the forming pairs mesh fully at high speed.

2) Adjusted the engine to the speed of 1800 r/min with maximum torque and the second gear engagement, the ring-die parts would be run at fastest speed of the design about 240 r/min. Kept the state for 10 minutes to observe the actual operation situation, and the results showed well run.

3) Increased slowly the speed of the engine to the maximum 2400 r/min, namely, the ring-die parts would be run at speed of about 320 r/min. Kept the state for 10 minutes to observe the actual operation situation, and the results showed normal working.

It was known from the above unload test results that the designed forming machine could run well at different speeds after low speed running.

3.2 COMPRESSION TEST

At the beginning of work, the state of the new and traditional forming machine was similar. Loose bio-materials entering into the forming parts would directly drop out with the centrifugal force of rotation due to the open end of the forming hole. With the continuous supplement, bio-materials were gradually compacted to densification under the combined action between the friction produced by hole wall and the extrusion pressure of plunger. The larger density of the pellet, the greater of friction and extrusion pressure. After stable run for a period of time, steady-state balance was developed of the friction and extrusion, when the forming density of bio-materials tended to be stable, and reached the maximum. Then the solid pellets were constantly squeezed out.

Compression tests were done with sawdust and paper scraps at the condition of 12% moisture content and ≤2mm granularity as test materials [10, 11]. Adjusted the speed of engine at about 1600 r/min with the first gear engagement, that is, the ring-die part was run at 100 r/min. After 10 minutes feeding, bio-materials (sawdust) in forming holes would be dense gradually and begin to be squeezed from the holes, seen in Figures 10a and 10b. The pellet density was lower (about 0.4-0.5g/cm^3) for the poor particle binding force at the moment in Figure 10c. With the growth of work time, the pellet production quality had been improved greatly and tended to be stable, when the pellets density could reach 0.8-0.9 g/cm^3 finally shown in Figures 10d and 10e. Figure 10f was the paper pellets produced at the engine speed of 1000 r/min, the density of which was measured to be 0.85 g/cm^3.

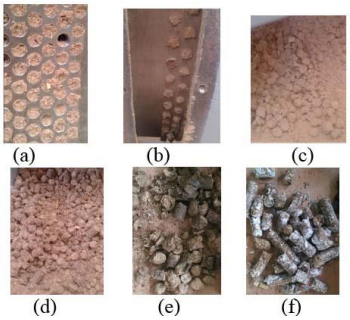


FIGURE 10 The compression experiments of the new biomass pellet forming machine a) Material extrusion condition at the entrance end of forming holes after 10 minute sawdust feed. b) Discharge condition at the outlet end of forming holes after 10 minute sawdust feed. c) Biomass pellet product after 10 minute sawdust feed. d) Biomass pellet product after 20 minute sawdust feed. e) Pellet fuel made by sawdust after 30 minute operation. f) Pellet fuel made by paper scraps after 40 minute operation

4 Conclusion

At present, there exist common shortcomings of low efficiency, high energy consumption, serious wear of forming mould, and etc., general existing in current biomass forming equipment's. In this paper a forming mechanism of a new type pellet machine with plunger roller and ring-die was proposed, which using the extrusion force of mesh between plungers circumferentially on the roller and forming cavities uniformly distributed in the ring-die to make the loose biomaterials densification. It could avoid direct contact and intense friction of forming parts and materials out of the holes. So the energy consumption of the materials in ineffective forming process was reduced and the wear of forming parts was effective alleviated. Then the new type biomass pellet forming machine was designed according to the requirements by CAD/CAE system (SolidWorks), and a prototype was built with the power 22 kW and production capacity 450 kg/h. Unload and compression test were done finally on the prototype. It was shown that the machine could run well at different speeds when idle load, the forming quality of solid pellet fuel would tend to be stable after 30 minutes' continuous operation and the pellet density could reach more than 0.8 g/cm^3. So it represented that the new pellet machine had a good practical operation and met the design demands. The test results ensured feasibility and practicability on the mechanism of plunger roller and ring-die forming machine.

Acknowledgement

The financial supports of the National Science & Technology Pillar Program during the 12th Five-year Plan Period (No. 2012BAD30B00) and the Fundamental Research Funds for the Central Universities (No. YX2013-26) are gratefully acknowledged.

References

[1] Tumuluru J S, Wright C T, Hess J R 2011 Biomass densification systems to develop uniform feedstock commodities for bio-energy application *Landtechnik* **54**(1) 28-35
[2] Junginger M, Bolkesj T, Bradley D 2008 Developments in international bioenergy trade *Biomass and Bioenergy* **3**(2) 717-29
[3] Zhang C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34
[4] Mielenz J R 2001 Ethanol production from biomass technology and commercialization status *Current Opinion Inmicrobiology* **1**(4) 324-9
[5] Yuan D L, Yuan X Y, Chen Z J 2014 Design of enclosure assembly for Q450 pellet molding machine based of Solidworks *Advanced Materials Research* **971-973** 881-4
[6] Song X, Yu G, Jiang C, Du K, Guo X 2011 Parameter Research of Bio-mass Shaping with Open Lineal Mold in Natural *Temperature Heilongjiang Agricultural Sciences* (11) 36-8 *(in Chinese)*
[7] Xiong K, Zhang Y, Zhang Z, Wang S, Zhong Z 2014 PA-NEMO: Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103
[8] Fang G 2011 *Design and research of suspended forest small-biomass pellet fuel forming machine* Northeast Forestry University Master Thesis *(in Chinese)*
[9] Pang L, Tian Y, Hou S, Zhao L, Meng H 2012 Development status and prospects of biomass pellet forming device *Journal of Agricultural Mechanization Research* **5**(9) 237-41 *(in Chinese)*
[10] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization: The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81
[11] Yan W 2011 Research on Biomass Open Compaction at Normal Temperature Beijing Forestry University *(in Chinese)*

Authors

Xiangyue Yuan, born in September, 1976, Beijing, P.R. China

Current position, grades: lecture at the School of Technology, Beijing Forestry University, China.
University studies: Doctor degree in Engineering at Beijing Forestry University in China.
Scientific interests: development of the bio-energy equipment and forestry machinery.
Publications: more than 20 papers.
Experience: teaching experience of 7 years, 10 scientific research projects.

Zhongjia Chen, born in January, 1981, Beijing, P.R. China

Current position, grades: lecture at the School of Technology, Beijing Forestry University, China.
University studies: Bachelor and Doctor degrees of Engineering in Machine Design and Theory at Beijing Forestry University in China.
Scientific interests: development of bio-energy equipment and forestry machinery.
Publications: more than 10 papers.
Experience: teaching experience of 5 years, 10 scientific research projects.

Reaction mechanism of carboxymethyl starch-based wood adhesive

Yong Zhang, Lixia Li*

Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming, China

Received 5 November 2014, www.cmnt.lv

Abstract

The technology involved in manufacturing carboxymethyl starch-based adhesive was studied with the aim of enhancing its water-resistance. The optimum formula was based on previous research using a quadratic orthogonal rotation combination design to synthesise the carboxymethyl starch-based wood adhesive. Compared with the original starch, carboxymethyl starch improves the performance of starch adhesive due to its advantageous properties such as hydrophilicity and stability during freezing and thawing etc. Infrared spectroscopy and scanning electron microscopy were used to analyse the synthesis of the carboxymethyl starch-based adhesive. Many functional groups could be identified by their characteristic Fourier transformation infrared vibration frequency and this was used to explain the structural changes. At the same time, the microstructure of the adhesive was analysed by scanning electron microscopy (SEM), and porous, round or elliptical granules in the carboxymethyl starch were observed. An SEM of the main starch agent revealed a network structure; the granules did not aggregate together to form large groups due to the polyvinyl alcohol that prevented granule agglomeration. However, the SEM of the starch adhesive showed a smoother and more compact network structure, due to the large number of strong chemical bonds that were formed by the association of isocyanate with the main starch agent.

Keywords: carboxymethyl starch, wood adhesive, optimisation, mechanism

1 Introduction

In recent years, sustainability has gained greater attention due to the world's heavy dependence on petroleum and natural gas for energy and various products; the development of alternatives to petrochemical-based products from renewable materials improves sustainability [1]. At present, urea-formaldehyde resin, phenolic resin and melamine formaldehyde resin are still widely used to synthesise wood adhesives; however, these are not sustainable in the long run. Furthermore, the emission of carcinogenic formaldehyde during their production has been a serious concern in recent years [2-4]. Natural resources have unique properties and advantages. Starch is an abundant natural polymer that is excellent for use in the manufacture of wood adhesives. For a long time, starch has been used as an additive in conventional adhesives [5-7] because it is easily obtained, the process is simple, it is easy to operate, it is low cost, gives good adhesion and is environmentally friendly; it is therefore potentially a useful product to investigate [8]. Li Lixia [9] achieved the best operating parameters of wet strength adhesive after bonding wood materials and analysed the effects of the formula and synthesis technology using a quadratic orthogonal rotation combination design. Shi Junyou [10] investigated the bonding mechanism of the starch-based aqueous polymer isocyanate and birch wood using Fourier transformation infrared (FTIR) and electron spectroscopy for chemical analysis (ESCA). Wang Zhenjiong [11] analysed starch-based wood adhesive by FTIR, nuclear magnetic resonance (NMR) and SEM.

These studies on starch wood adhesive led to significant progress, but there are still some issues such as problems of low starch content and poor water resistance. Compared with the original starch, carboxymethyl starch can improve the performance of starch adhesive due to its advantages, including hydrophilicity and good stability during freezing and thawing etc. Therefore, in this paper we investigated the optimal formula for the preparation of carboxymethyl starch-based wood adhesive [9]. We aimed to reveal the reaction mechanism of the starch-based adhesive through comparative analysis using infrared spectroscopy and SEM to analyse the starch, the main agents of starch and the starch-based adhesive. The ability to enhance the water resistance of starch-based adhesives is of great importance.

2 Material

Material: carboxymethyl starch (Heilongjiang Longfeng Modified Corn Starch Factory); polyvinyl alcohol (1788, Tianjin Kemiou Chemical Reagent Development Centre); two-component polyisocyanate emulsion adhesiv Prefere. 6150 (Taier Chemical (Shanghai) Co., Ltd.); carboxylic butadiene-styrene latex Nipol LX473D (Shanghai Lishen Chemical Co., Ltd, Shanghai, China).

Instruments: HH.SY11-Ni electric-heated thermostatic water bath (Beijing Changfeng Instrument Company), JJ-1 type power basic stirrer (Jintan City, Jiangsu Province Medical Instrument Factory), three-necked flask (250 ml; Sichuan Shubo Co., Ltd.), thermometer (100 °C), measuring cylinder (100 ml), pipetting device (20~200 L; Shanghai Dragon Medical Equipment Co., Ltd.), JM1102 electronic

***Corresponding author** e-mail: lilixia_2013@126.com

balance (Yuyao Jiming Weighing Electronic Balance and Check Equipment Co., Ltd.), VECTOR22 Fourier transform infrared spectrometer (Swiss Bruker company), Scanning electron microscope and E-1010 ion sputtering apparatus (Japan Hitachi company).

3 Methods

3.1 PREPARATION OF THE ADHESIVE

The optimum formula for carboxymethyl starch adhesive was based on previous research using a quadratic orthogonal rotation combination design to synthesise the starch-based wood adhesive [9].

3.2 FOURIER TRANSFORMATION INFRARED (FTIR)

The starch and the samples of carboxymethyl starch adhesive were fully milled in an agate mortar, and samples were prepared by grinding the finely powdered starch with potassium bromide. The mixture was then pressed into tablets before measurement. The IR spectra were obtained using a VECTOR22 IR Spectra Scanner, and they were recorded over a wave range of 400–4000 cm^{-1}.

3.3 SCANNING ELECTRON MICROSCOPY (SEM)

The carboxymethyl starch-based wood adhesive was synthesised using the optimal parameters from the above-mentioned experiment. A sample was placed on the carrier using a pipette, then it was solidified in a drying vessel. The dehydrated samples were coated with gold powder to avoid charging the sample under the electron beam. Finally, the microstructure was observed using a scanning electron microscope.

4 Results and discussion

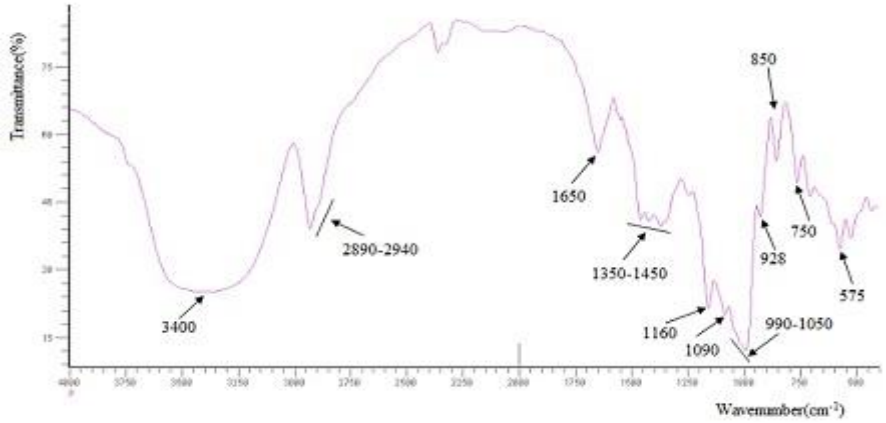
4.1 ANALYSIS OF FTIR

Infrared spectroscopy was used to analyse carboxymethyl starch, carboxymethyl starch main agent and carboxymethyl starch-based adhesive; the results of IR spectra are shown in Figure 1 (Figure 1a, Figure 1b and Figure 1c).

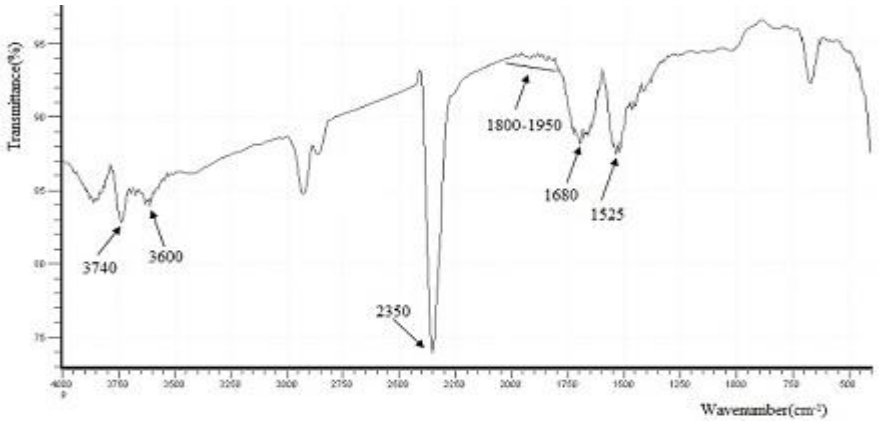
The information on chemical structure from FTIR spectroscopy in Figure 1a shows a broad band at 3400 cm^{-1} attributed to O-H stretching [12-14]. This is due to the hydroxyl groups present in the starch molecules, and indicates that these hydroxyl groups are not free; they form large amounts of intramolecular and intermolecular hydrogen bonds [5]. Meanwhile, the C-H stretching vibration gave a strong signal at 2890–2940 cm^{-1} [15-17]. The peak at 1650 cm^{-1} was attributed to C=O stretching [18-20], and the peaks at 1350–1450 cm^{-1} were related to the C–H bending of CH_2 [21-23], this indicated that the hydroxyl groups of starch molecules were carboxymethylated. The results were consistent with previous reports concerning different kinds of carboxymethyl starch. The peaks at around 1160 cm^{-1} and 1090 cm^{-1} were characteristic of C–O–H in starch [24-26], while the peak between 990 cm^{-1} and 1050 cm^{-1} was attributed to C-O bond stretching in the C-O-C group [27-29]. A similar phenomenon was observed in the absorption peaks present at 928 cm^{-1} for skeletal mode vibrations of glycosidic linkages (C–O–C) [30]. Bands at 850 cm^{-1} were due to the C(1)-H deformation of CH_2 in the glucose unit, and the bands at 575 cm^{-1} and 750 cm^{-1} related to the C–C stretching of skeletal modes in the glucose unit, this observation is similar to the results of Joshi et al. [23, 30].

Compared with Figure 1a, the information of chemical structure from FTIR spectroscopy in Figure 1b shows a smaller peak at 3740 cm^{-1}; this was attributed to the free O-H stretching. The band, which was displaced at 3600 cm^{-1} as a sharper and smaller peak in carboxymethyl starch main agent was the characteristic absorption peak for the stretching vibration of O-H groups because of the complex vibrational stretching associated with intramolecular and intermolecular bound hydroxyl groups. In addition, this peak had a high transmittance in the spectrum of carboxymethyl starch main agent attributed to the introduction of polyvinyl alcohol and carboxylic butadiene-styrene latex. The results indicated that the reaction of the hydroxyl and carboxyl weakened the peak intensity and wave number of hydroxyl absorption, and hydroxyl groups were blocked to proper lever in the carboxymethyl starch main agent. The band occurring at 2350 cm^{-1} was most probably due to the presence of carbon dioxide in test samples. The absorption bands at 1680 cm^{-1} were attributed to C=O stretching and this is consistent with the results of Machová et al. [31, 32], indicating the presence of carboxyl in the samples. It was also observed that the absorption peak present at 1525 cm^{-1} was characteristic for benzene [33-35], and the peaks at 1800–1950 cm^{-1} indicated the vibrating property of mono-substituted benzene derivate, suggesting the presence of carboxylic butadiene-styrene latex. The presence of the carboxyl group and phenyls made significant improvements in the viscosity and adhesive strength of the polymer and thus improved the property of the adhesive.

The absorption bands between 2310 and 2350 cm^{-1} were ascribed to - NCO antisymmetric stretching vibration (see, Figure 1c) and a large number of strong chemical bonds were formed by the association of –NCO groups and carboxymethyl starch main agent, which improved the wet strength of the adhesive. It was also observed that absorption peaks were present at 1550 cm^{-1} for N-H bending deformation and C-N antisymmetric stretching vibration [36, 37]. The peaks at 1700 cm-1 and 1650 cm^{-1} were attributed to the C=O stretching vibration [38, 39]. Moreover, the broad band at 1250 cm^{-1} was attributed to N-CO-O asymmetric stretching vibration, indicating the introduction of isocyanate. This significantly increased the viscosity and wet strength of the adhesive.



a)



b)

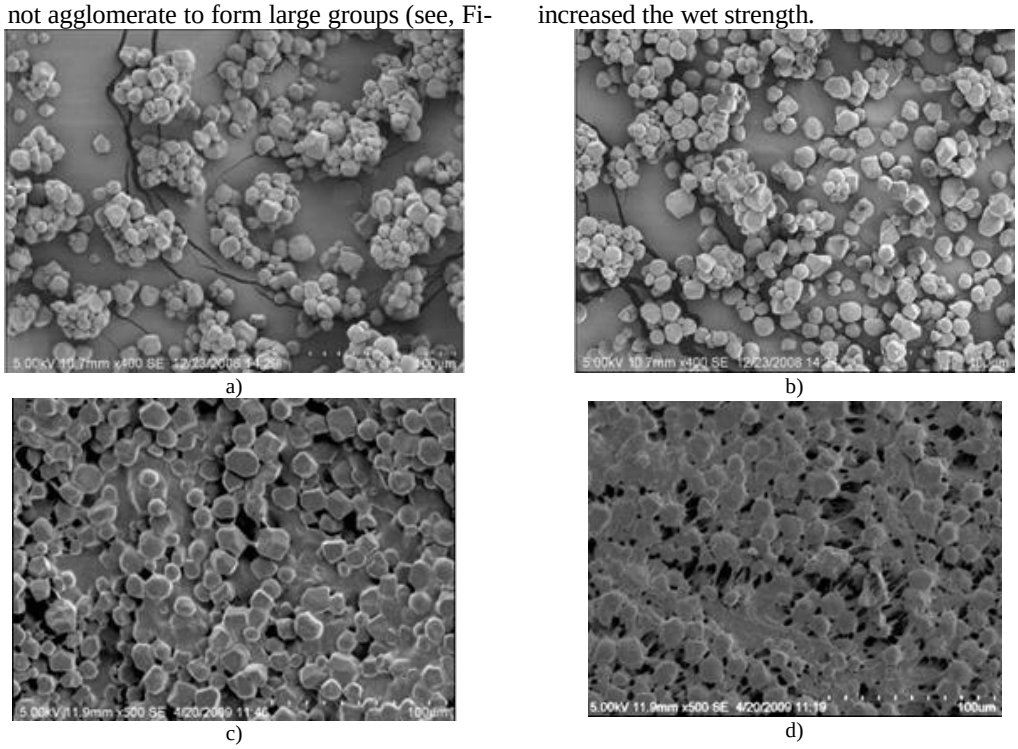
c)

FIGURE 1 FTIR: a) IR spectra of carboxymethyl starch, b) IR spectra of carboxymethyl starch main agent, c) IR spectra of carboxymethyl starch adhesive

4.2 ANALYSIS OF SEM

The microstructure of starch-based adhesive that was analysed by scanning electron microscope is shown in Figure 2. As can be seen in Figure 2a, round or elliptic shaped granules were present in the cornstarch [40-43] and the surface of the cornstarch had a smooth, full and compact structure [44-46]. The starch granules appeared to agglomerate due to the strong interaction of intermolecular and intramolecular hydrogen bonds in the starch molecules [47]. Compared with Figure 2a, carboxymethyl starch (see, Figure 2b) was still present in a granular form and it was similar in shape to the corn starch. However, holes were apparent on the surface of granules that increased the surface area thus improving solubility and adsorption. This observation suggests that the starch crystallinity was altered, thus allowing the etherifying

agents to have more access to the starch molecules for the carboxymethylation processes [48, 49]. The changes observed on the modified starch granules were also due to the breakage of chemical bonds in starch molecules affected by the strong alkaline environment together with heat treatment, this means that the carboxymethyl reaction took place both on the surface and within the starch granules [50]. The network structure following the introduction of a carboxyl group, benzene and polyvinyl alcohol can be seen in Figure 2c. This significantly improved the viscosity and adhesive strength of the polymer and improved adhesive properties. Each granule still maintained its individual entity but the granules did not agglomerate to form large groups (see, Figure 2c). This was due to the polyvinyl alcohol, which prevented granules agglomerating and accelerated the emulsifying process. When compared with Figure 2c, the much smoother and more compact network structure can be seen in Figure 2d. The starch adhesive was highly compatible because the uptake of isocyanate greatly improved the adhesive strength. The isocyanate was uniformly distributed in the water-soluble solvent of emulsifying by polyvinyl alcohol as long as it was reacting with the main agent and then forming solid amine fat keys, urea key chemical bonds and so on. Subsequently, these chemical bonds reacted with hydrogen, forming a network of structural molecules that increased the wet strength.



a) b) c) d)

FIGURE 2 Scanning electron micrograph: a) Starch, b) Carboxymethyl starch, c) Carboxymethyl starch main agent, d) Carboxymethyl starch adhesive

5 Conclusions

1) FTIR of carboxymethyl starch revealed a broader band at 3400 cm^{-1} attributed to the large number of hydrogen bonds that were formed by the association of hydroxyl and carboxyl. The C-H stretching vibration gave a strong signal at 2890-2940 cm^{-1} and the primary and secondary alcohols of the bands at 990-1160 cm^{-1} indicated C-O stretching. The FTIR of carboxymethyl starch main agent showed absorption bands at 1680 cm^{-1} attributed to C=O stretching, thereby indicating the introduction of carboxyls. It was also observed that the absorption peaks present at 1525 cm^{-1} were characteristic of benzene, and the peaks at 1800–1950 cm^{-1} had the vibrating property of monosubstituted benzene derivate; this indicated the introduction of phenyl of carboxylic butadiene-styrene latex. The introduction of polyvinyl alcohol and carboxylic butadiene-styrene latex made the viscosity and adhesive strength of the polymer increase significantly and thus improved the adhesive property. The FTIR of carboxymethyl starch adhesive absorption bands at 2310 -2350 cm^{-1} were ascribed to the - NCO antisymmetric stretching vibration, and isocyanate was introduced into the carboxymethyl starch adhesive.

2) Round or elliptic shaped granules appeared in the carboxymethyl starch and the granule surfaces became pocked with a number of holes. The carboxymethyl starch granules appeared to become agglomerated due to the strong interaction of the intermolecular and intramolecular hydrogen bonds of the starch molecules. The SEM of carboxymethyl starch main agent presented a network structure; however, each granule maintained its individual entity while the granules did not bind together to form large groups as compared with Figure 1b. This was due to the polyvinyl alcohol that prevented granule agglomeration and improved stability. In addition, the SEM of the carboxymethyl starch adhesive displayed a smoother and more compact network structure, due to the large number of strong chemical bonds that were formed by the association of isocyanate and the starch main agent.

Acknowledgements

The authors gratefully acknowledge all the scientific members of the research team. This study was supported by research project of the Heilongjiang (GB06B501-2-2). Also we are grateful to the Heilongjiang Longfeng Modified Corn Starch Factory for kindly supplying the carboxymethyl starch, and the Shanghai Lishen Chemical Co., Ltd for the carboxylic butadiene-styrene latex.

References

[1] Jang Y, Huang J, Li K 2011 A new formaldehyde-free wood adhesive from renewable materials *International Journal of Adhesion & Adhesives* **31**(7) 754-9
[2] Wang Z, Gu Z, Hong Y, Cheng L, Li Z 2011 Bonding strength and water resistance of starch-based wood adhesive improved by silica nanoparticles *Carbohydrate Polymers* **86**(1) 72-6
[3] Moubarik A, Charrier B, Allal A, Charrier F, Pizzi A 2010 Development and optimization of a new formaldehyde-free cornstarch and tannin wood adhesive *European Journal of Wood and Wood Products* **68**(2) 167-7
[4] Xu H, Shen Q, Ouyang X, Yang L 2012 Wetting of soy protein adhesives modified by urea on wood surfaces *European Journal of Wood and Wood Products* **70**(1-3) 11-6
[5] Nie Y, Tian X, Liu Y, Wu K, Wang J 2013 Research on starch-g-polyvinyl acetate and epoxy resin-modified corn starch adhesive *Polymer Composites* **34**(1) 77-87
[6] Wu G, Sun E, Huang H, Chang Z, Xu Y 2013 Preparation and properties of biodegradable planting containers made with straw and starch adhesive *BioResources* **8**(4) 5358-69
[7] Li Z, Wang J, Cheng L, Gu Z, Hong Y, Kowalczyk A 2014 Improving the performance of starch-based wood adhesive by using sodium dodecyl sulphate *Carbohydrate Polymers* **99**(2) 579-83
[8] Tan H, Zhang Y, Weng X 2011 Preparation of the plywood using starch-based adhesives modified with blocked isocyanates *Procedia Engineering* **15** 1171-5
[9] Li Lixia, Lia Fuguo 2009 Study on synthesis factors of carboxymethyl starch adhesive on bonding strength *Journal of Northeast Agricultural University* **20**(8) 100-4 (*in Chinese*)
[10] Shi J, Gu J 2009 Description of bonding mechanism of starch-based AIPI and birch by FTIR and ESCA *Journal of Northeast Forestry University* **37**(3) 55-57 (*in Chinese*)
[11] Wang Z, Li Z, Gu Z, Hong Y, Cheng L 2012 Preparation, characterization and properties of starch-based wood adhesive *Carbohydrate Polymers* **88**(2) 699-706
[12] Wang Y, Ju B, Zhang S2012 Viscosity properties of acetylated carboxymethyl starch *Carbohydrate Polymers* **90**(1) 696-702
[13] Luo Z, Xu Z 2011 Characteristics and application of enzyme-modified carboxymethyl starch in sausages *LWT-Food Science and Technology* **44**(10) 1993-8
[14] Lipsa R, Tudorachi N, Vasile C, Chiriac A, Grigoras A 2013 Novel Environmentally Friendly Copolymers Carboxymethyl Starch Grafted Poly(Lactic Acid) *Journal of Polymers and the Environment* **21**(2) 461-71
[15] Zou W, Yu L, Liu X, Chen L, Zhang X, Qiao D, Zhang R 2012 Effects of amylose/amylopectin ratio on starch-based superabsorbent polymers *Carbohydrate polymers* **87**(2) 1583-8
[16] Zhou G, Luo Z, Fu X 2014 Preparation and characterization of starch nanoparticles in ionic liquid-in-oil microemulsions system *Industrial Crops and Products* **52** 105-10
[17] Rashid I, Al Omari M H, Leharne S A, Chowdhry B Z, Badwan A 2012 Starch gelatinization using sodium silicate: FTIR, DSC, XRD and NMR studies *Starch-Stärke* **64**(9) 713-28
[18] Selamat M E, Sulaiman O, Hashim R, Hiziroglu S, Aidawati Wan Nadhari W N, Sulaiman N S, Razali M Z 2014 Measurement of some particleboard properties bonded with modified carboxymethyl starch of oil palm trunk *Measurement* 53,251-9
[19] Balsamo B, López-Carrasquero F, Laredo E, Conto K, Contreras J, Feijoo J L 2011 Preparation and thermal stability of carboxymethyl starch/quaternary ammonium salts complexes *Carbohydrate Polymers* **83**(4), 1680-89
[20] Anirudhan T S, Parvathy J 2014 Novel semi-IPN based on crosslinked carboxymethyl starch and clay for the in vitro release of theophylline *International Journal of Biological Macromolecules* **67** 238-45
[21] Muscat D, Adhikari B, Adhikari, Chaudhary D S 2012 Comparative study of film forming behaviour of low and high amylose starches using glycerol and xylitol as plasticizers *Journal of Food Engineering* **109**(2) 189-201
[22] Spagnol C, Rodrigues F H A, Pereira A G B, Fajardo A R, Tubira A F, Muniz E C 2012 Superabsorbent hydrogel nanocomposites basedon starch-g-poly(sodium acrylate)matrix filled with cellulose nanowhiskers *Cellulose* **19**(4) 1225-37
[23] Joshi M, Aldred P, McKnight S, Panozzo J F, Kasapis S, Adhikari R, Adhikari B 2013 Physicochemical and functional characteristics of lentil starch *Carbohydrate Polymers* **92**(2) 1484-96
[24] Çokaygil Z, Banar M, Seyhan A T 2014 Orange peel-derived pectin jelly and corn starch-based biocomposite film with layered silicates *Journal of Applied Polymer Science* **131**(16) 40654-65
[25] Klein B, Vanier N L, Moomand K, Pinto V Z, Colussi R, da Rosa Zavareze E, Dias A R G 2014 Ozone oxidation of cassava starch in aqueous solution at different pH *Food Chemistry* **155**(15) 167-73
[26] Xie Y, Chang P R, Wang S, Yu J, Ma X 2011 Preparation and properties of halloysite nanotubes/plasticized Dioscorea opposita Thunb starch composites *Carbohydrate Polymers* **83**(1) 186-91
[27] Liu H, Adhikari R, Huo Q, Adhikari B 2013 Preparation and characterization of glycerol plasticized (high-amylose)starch-chitosan films *Journal of Food Engineering* **1116**(2) 588-97
[28] Mendes A C, Boesel L F, Reis R L 2012 Degradation studies of hydrophilic, partially degradable and bioactive cements (HDBCs) incorporating chemically modified starch *Journal of Materials Science: materials in medicine* **23**(3) 667-76
[29] Teoh K H, Lim C-S, Ramesh S 2014 Lithium ion conduction in corn starch based solid polymer electrolytes *Measurement* **8** 87-95
[30] Fan D, Ma W, Wang L, Huang J, Zhao J, Zhang H, Chen W 2012 Determination of structural changes in microwaved rice starch using fourier transform infrared and Raman spectroscopy *Starch-Stärke* **64**(8) 598-606
[31] Machová E, Bystrický P, Malovíková A, Bystrický S 2014 Preparation and characterization of carboxymethyl derivatives of yeast mannans in aqueous solutions *Carbohydrate Polymers* **110**(22) 219-23
[32] Zhang B, Gong H, Lü S, Ni B, Liu M, Gao C, Huang Y, Han F 2012 Synthesis and characterization of carboxymethyl potato starch and its application in reactive dye printing *International Journal of Biological Macromolecules* **51**(4) 668-74
[33] Ouyang Z, Wu L, Yi D 2013 Preparation and characterization of polystyrene-SiO2/NiFe2O4 magnetic microspheres *New Chemical Materials* **41**(4) 123-5 (*in Chinese*)
[34] Guo Y, Zheng Y, Long H, Long H, Ge L, Li F 2012 Intercalation of polystyrene in kaolin via in-situ polymerization *Spectroscopy and Spectral Analysis* **32**(1) 99-102 (*in Chinese*)
[35] Balkan T, Sarac A S 2011 Synthesis and characterization of electrically conductive composite films of polypyrrole/poly(acrylonitrile-co-styrene) *Fibers and Polymers* **12**(5) 565-71
[36] Lei H, Du G, Wu Z, Xi X, Dong Z 2014 Cross-linked soy-based wood adhesives for plywood *International Journal of Adhesion and Adhesives* **50** 199-203
[37] Ahmed H E, Kolisis F N 2011 An investigation into the removal of starch paste adhesives from historical textiles by using the enzyme α-amylase *Journal of Cultural Heritage* **12**(2) 169-79
[38] Nistor M T, Vasile C 2013 TG/FTIR/MS study on the influence of nanoparticles content upon the thermal decomposition of starch/poly (vinyl alcohol) montmorillonite nanocomposites *Iranian polymer journal* **22**(7) 519-36

[39] Abd El-Mohdy H L, Hegazy E A, El-Nesr E M, Abd El-Wahab M 2011 Control release of some pesticides from starch/(ethylene glycol-co-methacrylic acid) copolymers prepared by γ-irradiation *Journal of Applied Polymer Science* **122**(3) 1500-9
[40] Zhang B, Chen L, Zhao Y, Li X 2013 Structure and enzymatic resistivity of debranched high temperature-pressure treated high-amylose corn starch *Journal of Cerealscience* **57**(3) 348-55
[41] Sujka M, Jamroz J 2013 Ultrasound-treated starch: SEM and TEM imaging, and functional behaviour Food Hydrocolloids **31**(2) 413-9
[42] Koo S H, Lee K Y, Lee H G 2010 Effect of cross-linking on the physicochemical and physiological properties of corn starch *Food Hydrocolloids* **24**(6-7) 619-25
[43] Qiu L, Hu F, Peng Y 2013 Structural and mechanical characteristics of film using modified corn starch by the same two chemical processes used in different sequences *Carbohydrate polymers* **91**(2) 590-6
[44] Chen Y, Huang S, Tang Z 2011 Structural changes of cassava starch granules hydrolyzed by a mixture of α-amylase and glucoamylase *Carbohydrate Polymers* **85**(1), 272-5
[45] Utrilla-Coello R G, Hernández-Jaimes C, Carrillo-Navas H 2014 Acid hydrolysis of native corn starch: Morphology, crystallinity, rheological and thermal properties *Carbohydrate polymers* **103**(15) 596-602
[46] Jiménez A, Fabra M J, Talens P 2012 Influence of hydroxypropylmethylcellulose addition and homogenization conditions on properties and ageing of corn starch based films *Carbohydrate Polymers* **89**(2) 676-86
[47] Qiao X, Tang Z, Sun K 2011 Plasticization of corn starch by polyol mixtures *Carbohydrate Polymers* **83**(2) 659-64
[48] Li X, Gao W, Huang L, Wang Y, Huang L, Liu C 2010 Preparation and physicochemical properties of carboxymethyl Fritillaria ussuriensis Maxim. Starches *Carbohydrate Polymers* **80**(3) 768-73
[49] Gao W, Lin X, Lin X, Ding J, Huang X, Wu H 2011 Preparation of nano-sized flake carboxymethyl cassava starch under ultrasonic irradiation *Carbohydrate Polymers* **84**(4) 1413-8
[50] Liu J, Ming J, Li W, Zhao G 2012 Synthesis, characterisation and in vitro digestibility of carboxymethyl potato starch rapidly prepared with microwave-assistance *Food Chemistry* **133**(4) 1196-205

Authors

Yong Zhang, born in May, 1987, Jining, Shandong Province, China

University studies: Engineering degree in Mechanical design manufacturing and automation at Qingdao university of science and technology institute of Piano Island in China.
Scientific interests: agricultural machinery.
Experience: 1 scientific research projects.

Lixia Li, born in January, 1983, Harbin, Heilongjiang Province, China

Current position, grades: lecturer of Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, China.
University studies: master degree at Northeast Agricultural University in China, doctor degree at Northeast Agricultural University in China.
Scientific interests: agricultural machinery, agricultural products processing, biomass materials.
Publications: more than 14 papers.
Experience: 7 scientific research projects.

Experimental study on heat-preservation wall materials made from waste foam

Lipeng Cai*

Department of Civil Engineering, Luoyang Institute of Science and Technology, Luoyang City, Henan Province, China, 471023

Received 28 October 2014, www.cmnt.lv

Abstract

In this paper, a new energy-saving-and-environmental-protection heat-preservation wall material is studied. This material is mainly made by waste foam, cement and sand in the structure of "sandwich" - cement foamed plastic is in the middle of cement mortar. Experimental study reveals that the heat-preservation block of this material has lighter self-weight and higher compressive strength and better performance in heat preservation, sound insulation and frost resistance, that is to say, its overall property is obviously superior to that of others such as aerated concrete block. This block consumes a large amount of waste foam that is difficult to dispose, which conducive to environmental protection and energy is saving. It can bring favourable social, economic and environmental benefit. Besides, less energy is consumed in the production of this material and it is easy to realize volume production, so this material is a new wall material of energy saving and environment protection.

Keywords: waste foam, heat preservation, wall material, new technologies

1 Introduction

With heavy self-weight and high energy consumption, the traditional wall materials in China waste lots of land resources, so it is to the disadvantage of environment protection and energy saving. Heat preservation property of the building envelope is a major factor influencing energy consumption of a building and different envelope materials need different energy costs. In China the insulating layer is generally not built in the outer wall of house by tradition, resulting in poor heat preservation and deafening effect as well as more energy consumption [1]. There is one important aspect in the architectural idea of "sustainable development": laying emphasis on energy conservation [2]. Currently the development priority in wall materials is to develop the new wall material, which needs low energy consumption and causes less pollution local resources by utilizing local resources [3]. Meanwhile, the waste foam is difficult to resolve in natural environment, causing serious pollution in modern times. With the development of social economy and rapid improvement of living standard, the demand for foamed plastic is increasing and waste foam is bringing more serious pollution and some experts have listed this pollution on the top three public pollution following water pollution and air pollution. As the aggregate of concrete heat presservation block, recycled polyphony l granule cannot only avoid "white pollution", but has many advantages such as good particle size gradation, being easy to be coated by size, good property and low price. The concrete heat preservation block in this paper is produced through simple technique, consumes less energy and has many advantages including light self-weight, good heat preservation property and high strength. This block consumes a large amount of waste polystyrene foam, opening up a way of removing white pollution and helping to environmental protection and energy saving, so it is a new green wall material with broad prospect.

2 Experiment materials

Cement: PO 32.5, PO 42.5 and PO 52.5 ordinary Portland cement from Luoyang Yellow River Cement Plant.

Foamed plastics particles: waste polystyrene foam is broken into particles of 0.4 mm~5 mm.

Sand: medium sand with the fineness modulus of 2.6 and qualified grain size gradation.

Tackifier: new environment-friendly II glue from Luoyang Kelong Coating Co. Limited.

Water: drinking water.

3 Experimental study

3.1 MAKING BUILDING BLOCK

1) To make the test mould in size of $150mm \times 150mm \times 150mm$ and its size and structure are shown in Figure 1. Steel plates with the thickness of 5mm are selected to make the test mould in the net size of $150mm \times 150mm \times 150mm$; the plates can be disassembly and place together flexibly and no gap is allowed after the assembly. 2 scale marks which are 20mm to the upper and lower surfaces are drawn on the side wall of the test mould in order to control the thickness of screed and that of cement foam layer. Waste engine oil is painted on inwalls for easy mold release.

***Corresponding author** e-mail: cai-lipeng@163.com

2) Cement:Sand:Water = 1 : 2 : 0.6 and cement mortar is mixed mechanically. Put the cement mortar on the bottom of test mould as thick as 10mm.

3) Cement:Sand:Tackifier=1:10:0.10 to make the cement paste; fully mix the paste with recycled polyphony l granule; after the paste attaché evenly to polyphony l granules, to put the mixture in the test mould about 5mm to the top edge of the block.

4) To spread the cement mortar to the top edge and smooth it over with a spatula.

5) To cover the surface of the test mould immediately once it shapes up to prevent evaporation of water. After 24 hours standing at the temperature of 20°C±5°C, to mark serial numbers and remove the block from the mould.

6) To put the test block in a standard curing room with the temperature of 20°C±1°C and working humidity above 95%. After the test block is maintained for 28d, the cement heat preservation block can be formed.

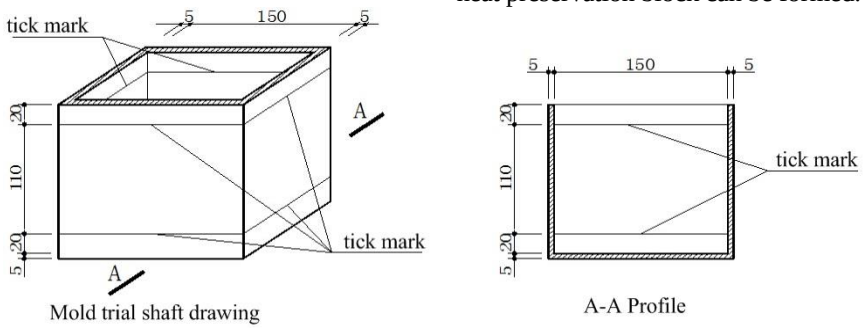


FIGURE 1 Test mould dimension drawing

3.2 PERFORMANCE DENSITY TEST

There are 18 test blocks in 6 groups for this test. The performance density is tested in the following steps:

1) To bake the test block in a 105~110°C baking oven until it has constant weight and to cool it to room temperature in a dryer.

2) To measure the size of the block with a straightedge (mm). Since it is a cube, the arithmetic mean value of 3 extents on all surfaces is used as the test value. Note down the value and calculate the volume (V_0).

3) To weight the mass with a balance and calculate its performance density ρ_0 according to the Equation - $\rho_0 = m/V_0$. Table 1 displays the test data.

TABLE 1 Performance density test data

Groups	No	Test block size/ cm			Test block volume V_0/cm³	Test block mass m/ g	ρ_0/(g/cm³)	Mean value/(g/cm³)
		Side length	*Side length*	*Side length*				
Group 1	1	15.0	15.1	15.2	3442.80	2065.7	0.600	0.610
	2	15.0	14.9	15.1	3441.90	2109.9	0.613	
	3	14.9	15.0	14.9	3330.15	2054.7	0.617	
Group2	1	15.0	15.0	15.2	3420.00	2072.5	0.606	0.607
	2	14.9	14.9	15.0	3330.15	2034.8	0.611	
	3	15.0	15.1	15.2	3442.80	2082.9	0.605	
Group3	1	14.9	15.0	15.1	3374.85	2058.3	0.610	0.606
	2	15.0	15.1	15.2	3442.80	2093.2	0.608	
	3	14.9	15.1	15.2	3419.85	2047.8	0.599	
Group4	1	15.0	14.9	15.0	3452.50	2106.0	0.610	0.612
	2	15.1	15.0	15.2	3442.80	2107.3	0.612	
	3	15.0	14.9	15.0	3352.50	2061.2	0.615	
Group5	1	15.1	15.0	15.2	3442.80	2105.4	0.612	0.611
	2	15.0	14.9	14.8	3307.80	2040.5	0.617	
	3	15.0	15.0	15.1	3374.85	2042.3	0.605	
Group6	1	15.0	15.0	14.9	3352.50	2038.3	0.608	0.613
	2	15.1	15.0	15.1	3442.80	2110.3	0.613	
	3	15.0	15.0	14.9	3352.50	2071.4	0.618	

According to Table 1, the performance density is: $\frac{0.610+0.607+0.606+0.612+0.611+0.613}{6} = 0.61$.

The performance density is 0.61g/cm³, that is 610 kg/m³.

3.3 COMPRESSIVE STRENTH TEST

There are 18 test blocks (used in the above test) in 6 groups for the compressive strength test. The compressive strength is tested in the following steps:

1) To wipe and measure the size of the test block (1mm) as soon as it is taken out from the curing room and to calculate its compression area A (mm2).

2) To place the test block on the bearing plate according to Illustration 2: the pressure-bearing surface of the block should be perpendicular to the top surface when it shapes up and its centre should align with the centre of the tester. To start the tester and adjust the ball seat for balanced contact between the top and bottom platens and the top and bottom surfaces of the test block.

3) In this test, loading should be increased constantly and uniformly with the speed of 0.3MPa/s. When the test block begins to deform, stop to adjust the accelerator of the tester until the block is destroyed. Note down the collapse load $P(N)$.

4) To calculate the compressive strength according to $f_{cu} = P/A$. 3 arithmetic mean value of the strength is used as the compressive strength value of their group (0.1MPa). If $\min imun\ value - mediam > 15\%$ median or $maximun\ value - mediam > 15\%$ median, the minimum value and the maximum value should be removed in the calculation; that is the median is the compressive strength value. If $\min imun\ value - mediam > 15\%$ median and $maximun\ value - mediam > 15\%$ median, the test results are invalid.

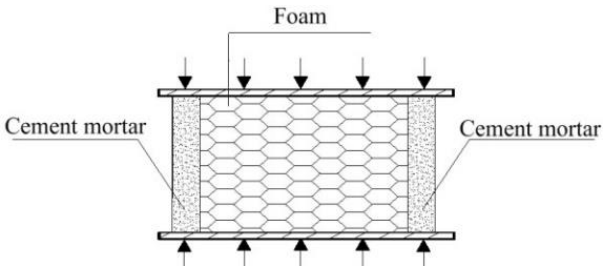


FIGURE 2 Compressive Strength Test

Table 2 presents the compressive strength test data.

TABLE 2 Compressive Strength Test Data

Groups	NO.	Size(mm)		A(mm^2)	P (N)	CS(MPa)	
		a	*b*				
	1	150	151	22650	90250	3.98	
1	2	150	149	22350	85250	3.81	3.81
	3	149	150	22350	80850	3.62	
	1	150	150	22500	78400	3.48	
2	2	149	149	22201	72550	3.27	3.41
	3	150	151	22650	78750	3.48	
	1	149	150	22350	84000	3.76	
3	2	150	151	22650	81200	3.58	3.63
	3	149	151	22499	79550	3.54	
	1	150	149	22350	85500	3.83	
4	2	151	150	22650	91500	4.04	3.84
	3	150	149	22350	81550	3.65	
	1	151	150	22650	82500	3.64	
5	2	150	149	22350	77800	3.48	3.57
	3	150	150	22500	80550	3.58	
	1	150	150	22500	84000	3.73	
6	2	151	150	22650	89500	3.95	3.75
	3	150	150	22500	80500	3.58	

According to Table 2, the compressive strength is: $\frac{3.81+3.41+3.63+3.84+3.57+3.75}{6} = 3.67$.

The average value of compressive strength is 3.67MPa and the minimum value is 3.27MPa.

3.4 HEAT PRESERVATION PROPERTY TEST

This test is a comparison test. The property of heat preservation is compared between heat-preservation concrete block, aerated concrete block and common brick to conclude the heat preservation property and heat conductivity of this block.

3 sealed experiment rooms of the same size are built up with common bricks, aerated concrete block and the concrete block studied in this paper [7]. Place an electric stove (power: 2500w) in each room to heat the experiment rooms. Stop the heating when the temperature reaches 95°C and observe how the temperature changes in each room at the interval of 30m for 9 times. Note down the temperatures observed.

Do this test for 3 times. The temperature-time variation curves in the 3 experiment rooms are shown in Figures 3, 4 and 5.

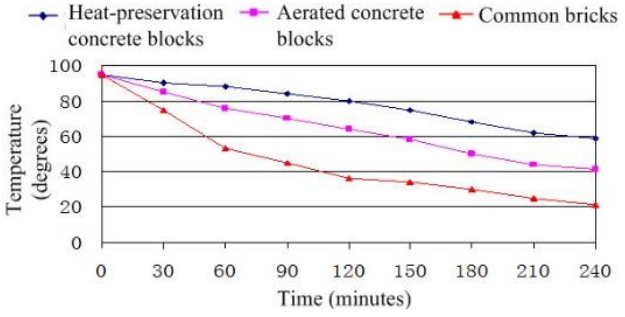


FIGURE 3 Temperature variation curve in rooms (1st time)

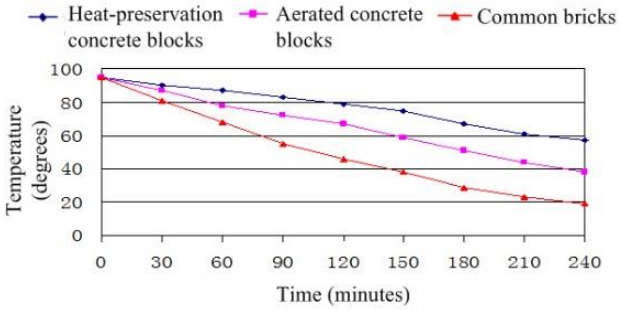


FIGURE 4 Temperature variation curve in rooms (2nd time)

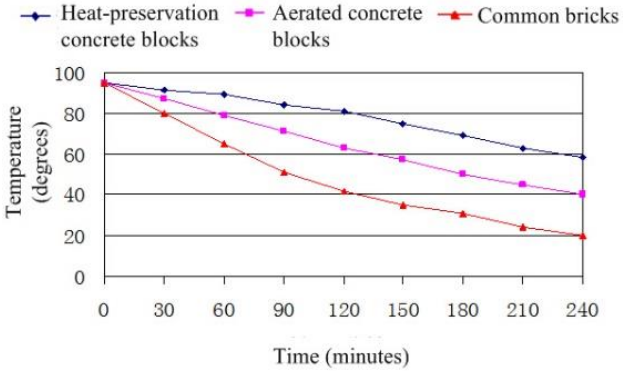


FIGURE 5 Temperature variation curve in rooms (3rd time)

Test results show that the temperature drops slowest in the room made by heat-preservation concrete blocks followed by the room made by aerated concrete blocks and common bricks. Thus, it can be inferred that the heat preservation property of the block studied in this paper is obviously superior to aerated concrete block and common brick, and that its heat conductivity is smaller than that of aerated concrete block, that is its heat conductivity $\lambda < 0.12 \ \ w/m.k$.

4 Conclusions

1) The concrete block studied in this paper can fully meet the requirement of new wall materials in properties such as performance density, compressive strength and heat presservation. Compared with the autoclaved aerated concrete blocks, this block has obvious advantages.

2) In engineering practice, the cement mortar layer on the outer coat of block is the main framework and the polystyrene granule in the block can bears little external force, so cement mortar of high strength can produce the heat-preservation concrete block of higher strength.

3) The technique in this paper is quite simple and easy and consumes less energy. Large equipment is not needed; small investment requires; desired results are produced rapidly and it is easy to realize volume production.

4) Blocks of various specifications can be produced according to the thickness of wall for different environment. For example, the $190*190*H$ (the mortar layer is as thick as 15mm) and $240*240*H$ (the mortar layer is as thick as 20mm).

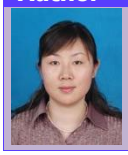
5) This concrete blocks have digested a lot of waste polystyrene plastics, protecting the environment and steering the building materials in the direction of saving resources, utilizing wastes and protecting environment.

6) To carry out the studies on heat-preservation concrete blocks made from recycled polyphony l granules can acelerate the progress of wall material innovation in China and the greenization of building materials industry and is conducive to the popularization of energy saving construction. To build houses with high-quality new wall materials has obvious building energy efficiency and improves comfort level of living and can meet the requirement of economic and social development and upturn living standarts. To launch the studies plays a positive role in propelling the progress of wall material innovation and the greenization of building materials industry and the popularization of energy saving building in China.

References

[1] China Building Materials Academy 2003 Green building materials and green building materials *Chemical Industry Press* **01** 47-9 (*in Chinese*)
[2] Li B 2007 An introduction to green building *Chemical Industry Press* **01** 57-9 (*in Chinese*)
[3] Ding D 2005 The wall reform and sustainable development *Machinery Industry Press* **01** 167-82 (*in Chinese*)
[4] Zhang, C, Huang L, Zhao Z 2013 Research on combination forecast of port cargo throughput based on time series and causality analysis *Journal of Industrial Engineering and Management* **6**(1) 124-34 (*in Chinese*)
[5] Li C, Cai L2003 Testing research of heat-insulated bonding mortar for environmental coordination *Architecture Technology* **10** 760-4 (*in Chinese*)
[6] Xiong S, Zhang K, Zhang Y, Wang Z, Zhong S Z 2014 PA-NEMO:Proxy mobile IPv6-aided network mobility management scheme for 6LoWPAN *Elektronika ir Elektrotechnika* **20**(3) 98-103 (*in Chinese*)
[7] Qi Y, Tang M, Zhang M 2014 Mass customization in flat organization:The mediating role of supply chain planning and corporation coordination *Journal of Applied Research and Technology* **12**(2) 171-81 (*in Chinese*)
[8] Cai L 2014 Study on the preparation of new wall material with recycled polystyrene pellets *New Building Materials* **03** 78-80 (*in Chinese*)

Author



Cai Lipeng, born in March, 1971, Luoyang, Henan Province, P.R. China

Current position, grades: the associate professor of civil engineering, Luoyang Institute of Science and Technology, Henan Province, China.
Scientific interest: energy-saving and environment friendly construction materials.
Publications: 30 papers.
Experience: teaching experience of 20 years, 10 scientific research projects.

The analysis of vibrations induced by variation section vortex in tension leg platform for a floating wind turbine

H F Wang[1]*, Y H Fan[1], L Yang[2]

[1]*School of Natural Sciences and Humanities, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, 518055, China*

[2]*Shanghai Electric Group Company Limited, Shanghai, 200000, China*

Received 26 July 2014, www.cmnt.lv

Abstract

During the previous decade, several offshore wind-farms were constructed for offshore wind energy generation showed promise as a source of green energy. However, there are several challenges to be met in the design and construction of the foundations for offshore wind turbines. The fatigue load plays an important and crucial role in the design of the supporting structures. In this paper, the vortex-induced vibrations of the tension leg platform were studied. Two types of structures namely; cylindrical Tension Leg Platform (TLP) and variable cross-section TLP were designed and studies on them were conducted to compare the advantages and drawbacks resulting from vortex-induced vibrations. Both uniform and shear flow were considered to simulate water flow through the structures. The variable cross-section TLP, which possesses outstanding mechanical properties, gave lower vortex shedding frequency compared to the cylindrical TLP for the same velocity. This is the objective desired in vortex-induced vibration.

Keywords: vortex-induced, tension leg platform, variation section

1 Introduction

With increasing population and economy, many countries began to exploit the natural ocean resources such as offshore oil and wind energy. The foundation of the wind turbine is the most important component. For its operation at increasing depths the traditional foundation fails to meet the necessary requirements. At 60m-100m, the TLP is not only economical but also stable [1].

The TLP is the vertical basis of mooring and compliant structure, similar to the combination of rigid and flexible systems. It is a complex nonlinear dynamic system. Its structural principle is producing a greater weight than force of buoyancy for offsetting the gravity and preload. Besides, the flexibility of TLP can realize the plane movement conforming to the external load. As a result the structural internal load is not needed [2, 3].

The TLP works in a complex environment. One of its serious operational problems is vortex-induced vibration. When there is an ocean current flow over a flexible cylinder, there will be a shedding vortex behind the structure because of the non-uniform pressure distribution. The vibrations in the structure will be produced by the shedding vortex because of the cross-flow and in-line flow pulse pressure. The vibration is the main factor for the fatigue of the structure [4].

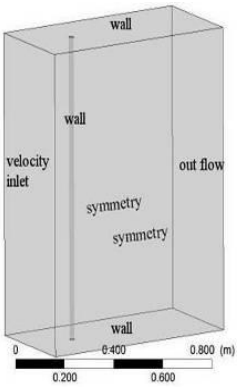
Three methods are employed to study the vortex-induced vibrations. These are experimental empirical modelling and numerical simulation. Feng [5] conducted experiments on the vibrations of one-degree-of-freedom of a circular cylinder in air flow. In his study, the vibration of the cylinder was confined to the cross-flow direction where only typical lock-in phenomenon was presented. In general, when a cylinder is subjected to a flow at a high mass ratio, only two amplitude response branches exist. These are initial and lower branches. If the cylinder is placed in a flow at low mass ratios, the third branch (i.e., upper branch) is also observed. The mostly known empirical model includes the VIVA, VIVANA, SHEAR7 and VICoMo. Along with the advances in computers, the science of Computational Fluid Dynamics (CFD) developed rapidly. Anagnostopoulos and Iliadis [6] carried the out numerical studies on the 1 DOF VIV of a circular cylinder in the stream direction and found that the response of the cylinder was amplified significantly, if the oscillatory flow frequency was close to the natural frequency of the cylinder. Zhao and Cheng [7], Zhao et al [8] simulated 1 DOF VIV of a cylinder in oscillatory flows at KC=10 and 20 and found that the response of the cylinder included more than one frequency as the reduced velocity exceeded KC=8. Ding [9] studied two-dimensional vortex-induced vibration with numerical simulation method and found that the lift coefficient and drag coefficient transformation laws were functions of the Reynolds number. Willden [10] researched on the vortex-induced vibration under the shear flow and observed a new locus for the movement of the structure.

The entire published research focused on the vortex-induced vibration of cylinder. The effect of varying section did not receive the attention of the researchers. Yanqiu Dong proposed the structure of non-uniform cross section for the oil platform, without analysing the mechanical characteristics for the structure. In this paper wake pattern lift and drag were compared for the cylindrical and cylindrical-conical structures at different stream velocities.

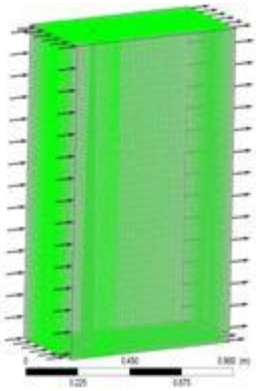
***Corresponding author** e-mail: phdwhf@163.com

2 Numerical methods

The material of the TLP was steel, with ρ=7850 kg/m^3, E=2.06E+11, and Poisson ratio of 0.3. The dimensions of the TLP were: 100min length, and 2m in diameter. For the variable cross-section of TLP, the diameter was 1m-3m. There was a preload of 100N at top of the TLP. The top of TLP was fixed to the floating structure with hinged joint and the bottom was fixed with the seabed foundation. The software ICEM was used to generate the mesh. The dimensions of the domain were 80m×40m×100m. The distance from the inlet and side face to the TLP was 20m. The grid elements around the TLP were encrypted because the subdomain was complicated and not only included fluid breakdown, but also vortex shed. All of the grid elements were in form of a structured grid. The geometrical and mesh models are shown in Figure 1a and Figure 1b respectively.



a) The geometric model



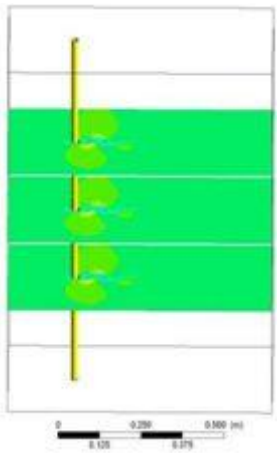
b) The mesh model

FIGURE 1 The geometric and mesh models

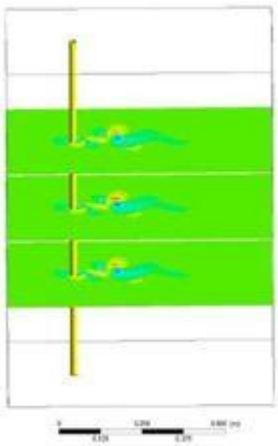
The initial and boundary conditions were set in CFX-pro, the TLP was non-slipping boundary, the top and bottom sides were walls, the left and right sides were in symmetry. In this paper, discretization format was Second-order central difference and k-ε was chosen as the turbulent model. All of the residuals were limited to 10E-4, the time step was chosen as 0.05s and the terminating time was 40s.

3 Results

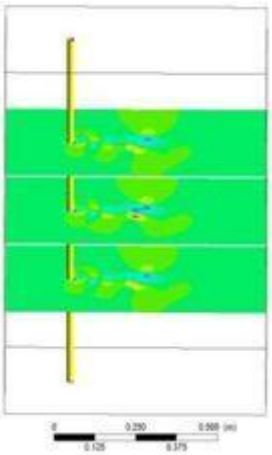
In this section, three velocities 0.2m/s, 0.4m/s and 0.2y+0.2 [0<y<1] (m/s) were chosen. The results for cylinder section and the cone section were obtained. Figure 2 shows the process of cylinder section vortex shedding. The corresponding times are 1.5s, 3s and 5s, the left column of figures is for the velocity 0.2m/s and the right column of figures is for 0.4 m/s. The results obtained from the velocity contour are that Reynolds number has a significant effect on the wake. Higher the Reynolds number sooner is the vortex generation and higher is the amount of energy is stored in the vortex. Besides, the linear distance between adjacent vortexes is less at 0.4m/s than that at 0.2m/s.



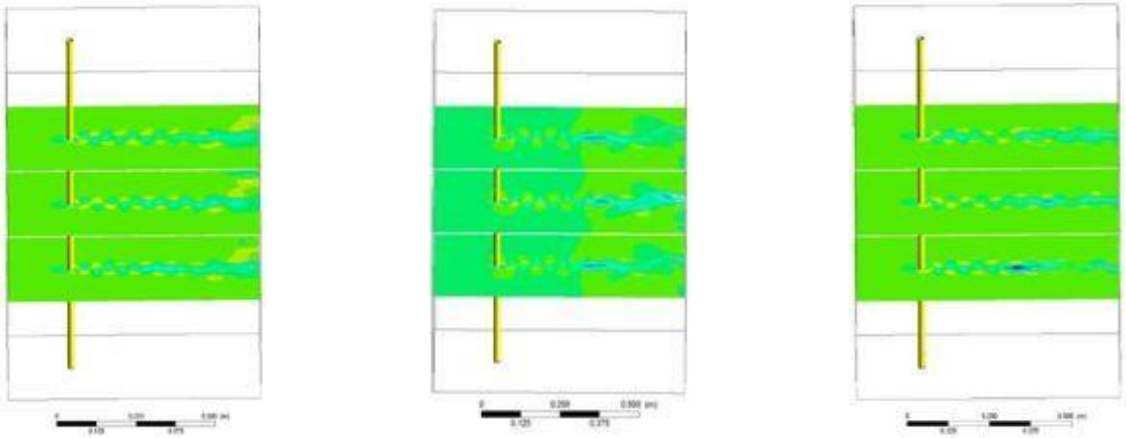
a) 1.5s at 0.2m/s



b) 1.5s at 0.4m/s



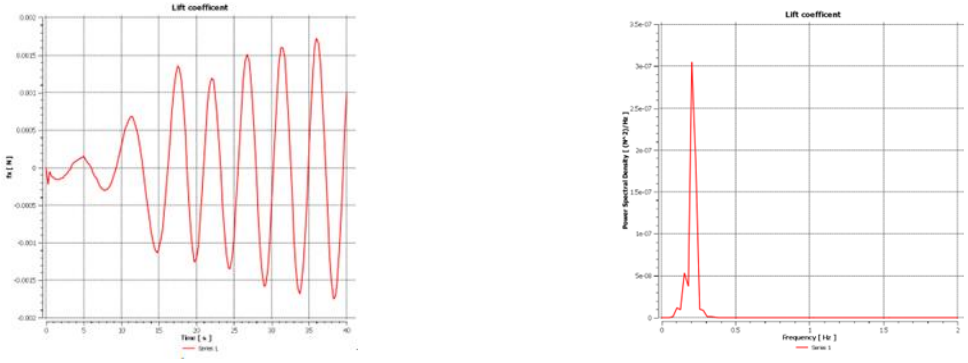
c) 3s at 0.2m/s



d) 3s at 0.4m/s e) 5s at 0.2m/s f) 5s at 0.4m/s

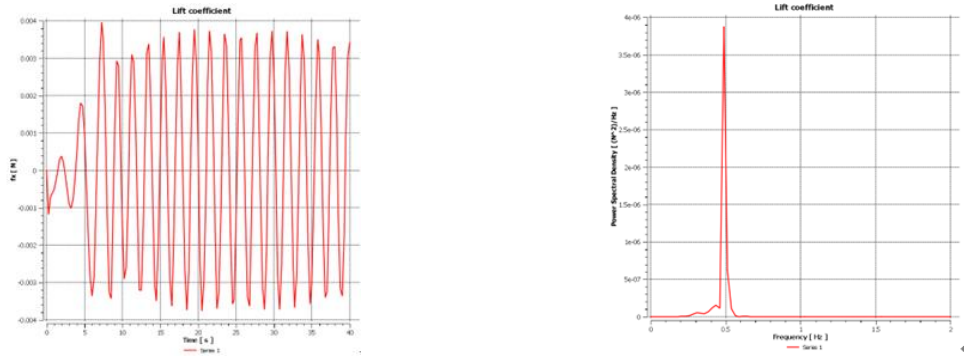
FIGURE 2 The velocity contours of cylinder TLP at different instants of time

The lift force, drag force time history curve and frequency spectrum for cylinder TLP and variable cross section TLP result are shown in Figure 3 to 10. It can be seen from these figures that when the velocity increases, the amplitude of lift force and the frequency increases. This means that the frequency of vortex shedding is accelerated. The lift force frequency is 0.251Hz at 0.2 m/s and the corresponding Strouhal number S_t is 0.23. When the velocity is 0.4 m/s, the lift force frequency is 0.498Hz and the S_t is 0.21. The above simulation results are in agreement with those in literature in which S_t was reported as about 0.2 when $300 < R_e < 1.5\times10^5$.



a) The time history curve of lift force b) The spectrum

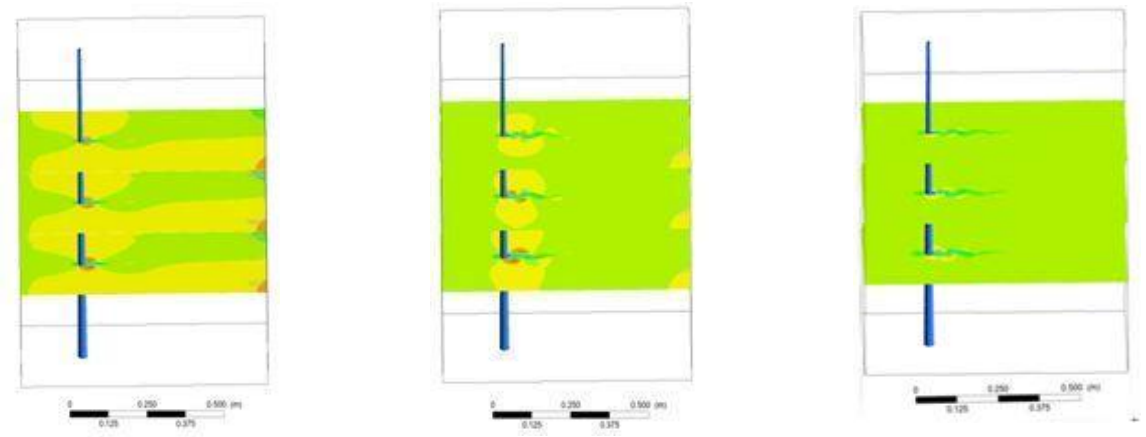
FIGURE 3 The time history curve of lift force and its spectrum at 0.2m/s for cylinder TLP



a) The time history curve of lift force b) The spectrum

FIGURE 4 The time history curve of lift force and its spectrum at 0.4m/s for cylinder TLP

The contours shown in Figure 5 below illustrate the variation of vortex at different times and positions. The effect of velocity is similar for the cylinder TLP, but vortex changes with the altitude and size of the section also has influence on the vortex. The diameter of the vortex is larger and the distance in-line between adjacent vertexes is more at bottom than at the top.

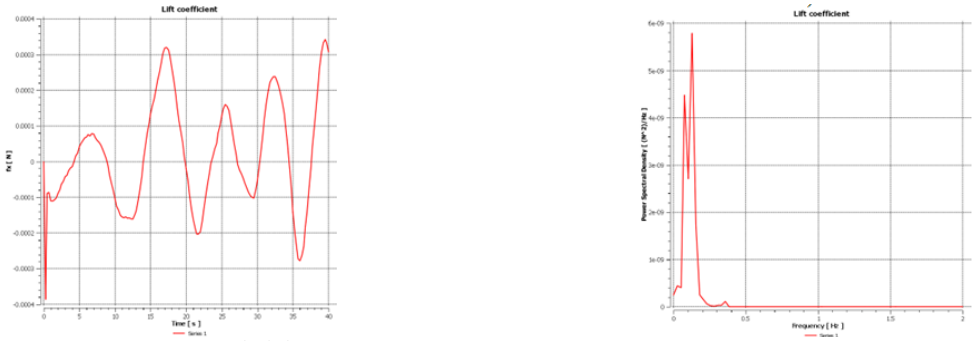


a) 1.5s at 0.2m/s, b) 1.5s at 0.4m/s c) 3s at 0.2m/s

d) 3s at 0.4m/s e) 5s at 0.2m/s f) 5s at 0.4m/s

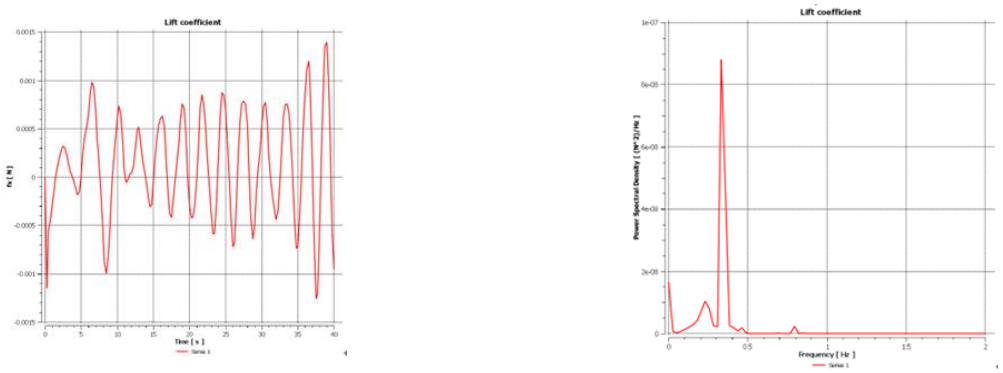
FIGURE 5The velocity contours at different instants of time for variable cross-section TLP

The time history curve of lift force in Figure 6 below shows the period of variable cross-section TLP is larger than of the cylinder TLP for the same velocity. Their frequencies are 0.11-0.198Hz at 0.2 m/s and 0.352Hz at 0.4 m/s. The stiffness of variable cross-section TLP is larger than that of the cylinder TLP. In fact, the stiffness has significant effect on the frequency of the vortex shedding. The trend of the time history curve for variable cross-section TLP is similar to that of cylinder TLP at different velocities.



a) The time history curve of lift force b) The spectrum

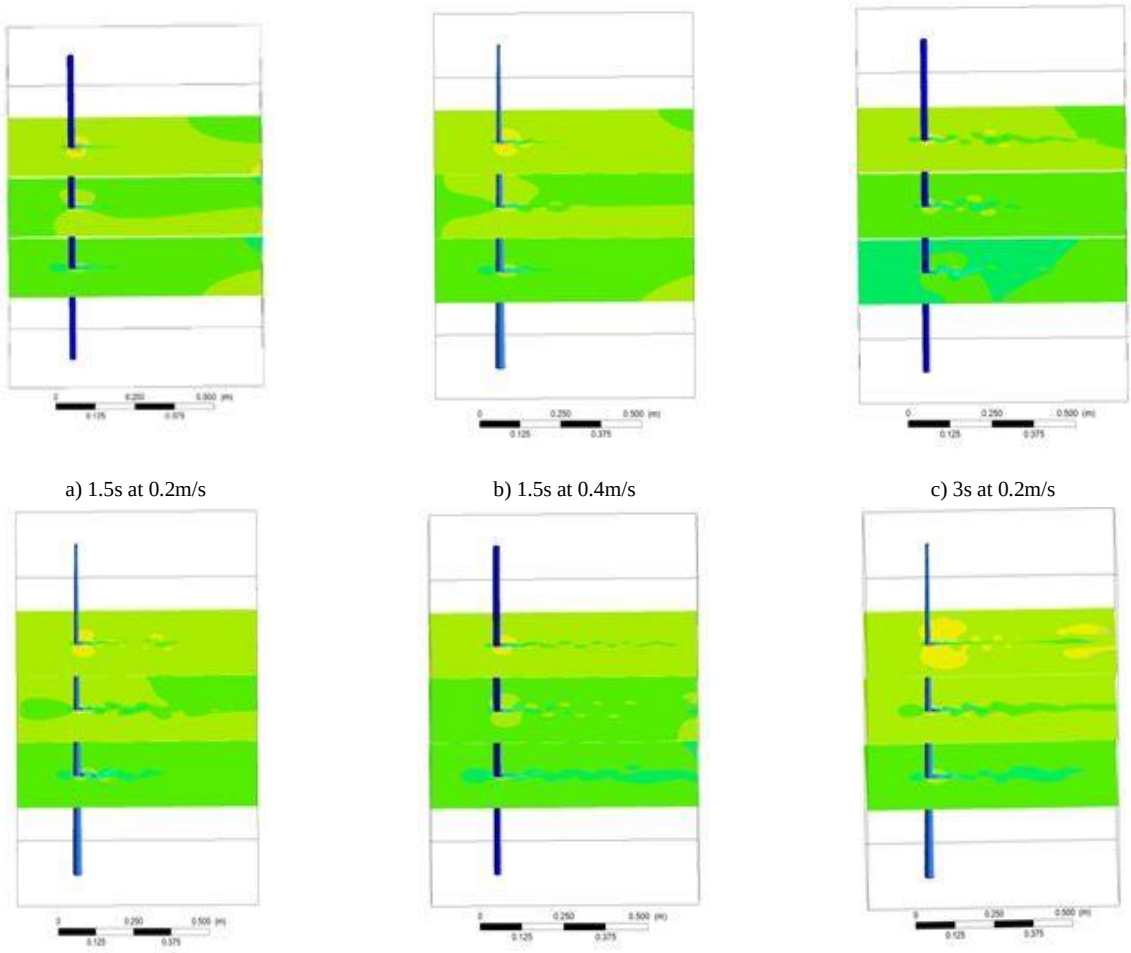
FIGURE 6 The time history curve of lift force and its spectrum at 0.2m/s for variable cross-section TLP



a) The time history curve of lift force b) The spectrum

FIGURE 7 The time history curve of lift force and its spectrum at 0.4m/s for variable cross-section TLP

As uniform flow cannot simulate the true environment of the TLP, shear flow is considered in this section. Firstly, the velocity contours of the variable cross-section TLP (Figure 8 b), d), f)) and the cylinder TLP (Figure 8 a), c), e)) are displayed below. For the cylinder TLP, the vortexes are different at different positions. The bottom of variable cross-section TLP does not show visible vortex, but it does not affect the top. There is vortex shedding at the top.



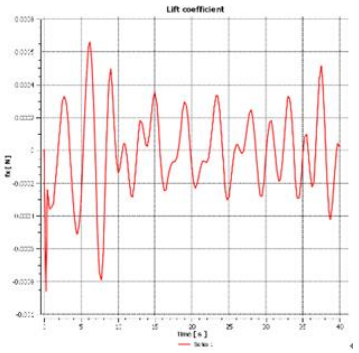
a) 1.5s at 0.2m/s b) 1.5s at 0.4m/s c) 3s at 0.2m/s

d) 3s at 0.4m/s e) 5s at 0.2m/s f) 5s at 0.4m/s

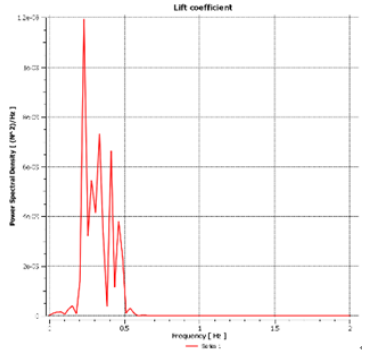
FIGURE 8 The velocity contours at different instants of time for cylinder TLP and variable cross-section TLP

In comparison to the uniform flow, for shear flow the time history curves of lift force for cylinder TLP and variable cross-section TLP change appreciably. For cylinder TLP, small oscillations are noticeable in every

period. The interval of frequency for vortex shedding is wider than before. The vortex shedding is in the range of 0.2Hz-0.5Hz. For the variable cross-section TLP, there are two types of oscillations. The trend of large oscillations is the same as that of the cylinder TLP, but the frequency of every small oscillation is larger than that of cylinder TLP. Though the frequency of vortex shedding for variable cross section TLP is smaller than that of the cylinder TLP, the vortex shedding begins from 0.157Hz. The shedding is also found at a high frequency of 0.697Hz.

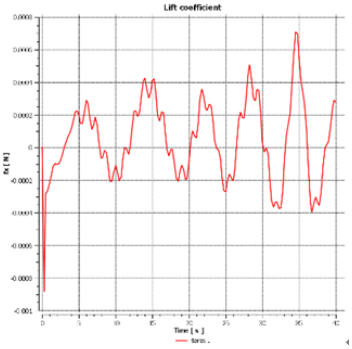


a) The time history curve of lift force

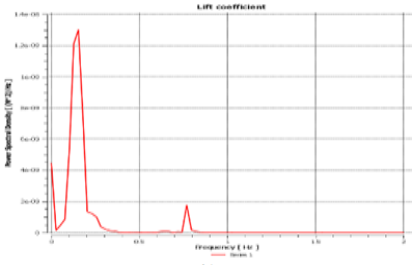


b) The spectrum

FIGURE 9 The time history curve of lift force and its spectrum for cylinder TLP at 0.2y+0.2[0<y<1] (m/s)



a) The time history curve of lift force



b) The spectrum

FIGURE 10 The time history curve of lift force and its spectrum for variable

From the Table 1 below, it can be seen that the frequency of variable cross-section TLP is always smaller than that of the cylinder TLP, although their average cross sections are equal and the volume of variable cross-section TLP is smaller than that of the cylinder TLP.

TABLE 1 The frequencies of TLP and variable cross section TLP at different velocities

	0.2(m/s)	**0.4(m/s)**	**0.2y+0.2[0<y<1] (m/s)**
Uniform cross-section (*D*=2m)	0.251Hz	0.498Hz	0.2Hz, 0.31Hz, 0.42Hz, 0.5Hz
Variable cross-section (*D*=1m-3m)	0.11Hz, 0.198Hz	0.352Hz	0.157Hz

4 Conclusions

This paper reports design of two types of TLP structures for simulation with uniform flow and shear flow. The velocity contours, lift force and drag force are illustrated to explain the results obtained. Following is the summary of final results:

1) For the two types of TLP structures studied, higher Reynolds number resulted in earlier vortex generation and higher energy storage in the vortex. Besides, the distance in-line between the adjacent vortexes is smaller at 0.4m/s than that at 0.2m/s. when velocity increases, the amplitude of lift force and also the frequency increase, which means that the frequency of vortex shedding is accelerated.

2) There are also several differences between the effects of the two types of TLP. The vortex is influenced by size of the section, that is the diameter of the vortex at the bottom is larger and the distance in-line between adjacent vortexes is more for the variable cross-section TLP than that for the cylinder TLP. Besides, a wider wake of variable cross-section is shown for the TLP. The time history curve of lift force shows that the frequency of variable cross-section TLP is smaller than that of the cylinder TLP for the same velocity though their average diameters are the same. In addition, low frequency vibration is observed in variable cross-section TLP.

3) When both types of TLP are subjected to shear flow, wake is different from the top to the bottom, and the vortex is generated earlier at the top. There are four peaks for cylinder TLP. The distribution of the peak value is amongst the uniform flow of 0.2m/s and 0.4m/s. Small oscillations arise in every period when variable cross-section TLP is subjected to shear flow. In the frequency spectrum curve, high frequency of lift force is found, but the main peak value is smaller than that of cylinder TLP.

This paper demonstrated that performance of variable cross-section TLP is better than the cylinder TLP with regard to the vortex-induced vibration. On the basis of results obtained in this study the variable cross-section TLP is recommended in engineering practice.

Reference

[1] Haifeng W, Youhua F, Yang L 2014 Dynamic analysis of a Tesion Leg Platform for offshore wind turbines *Journal of Power Technologies* **94**(1) 42-9

[2] Singh B, Mistri B, Patel R 2007 Comparison of foundation system for offshore wind turbine installtion *Coastal Engineering* **36** 13-6

[3] Leen D V, Peter F, Julie D R 2007 Wave run -up on cylindrical and cone shaped foundations for offshore wind turbines *Coastal Engineering* **54** 17-29

[4] Sandy B, Walt M, Jason J 2007 Engineering challenges for floating offshore wind turbines *National Renewable Energy Laboratory*

[5] Feng C C 1968 The measurement of vortex-induced effect in flow past a stationary and oscillating circular and D-section cylinder *Master's thesis: Canada*

[6] Anagnostoppoulos P, Iliadis G 1998 Numerical study of the flow pattern and the in-line resonse of a flexible cylinder in an oscillating stream *Journal of Fluids and Structures* **12** 225-58

[7] Ming Z, Liang C, Hongwei A 2012 Numerical investigation of vortex-induced vibration of a circular cylinder in transverse direction in oscillatory flow *Ocean Engineering* **41** 39-52

[8] Ming Z, Kalyani K, Yang X, Guirong Y 2013 Vortex-induced vibration(VIV) of a circular cylider in combined steady and oscillatory flow *Ocean Engineering* 73 83-95

[9] Ding DW 2010 Two-dimensional numberical simulation of flow around a cylinder and vortex-induced vibration *Master's thesis: China*

[10] Willden R H J, Graham J M R 2001 Numerical prediction of VIV of long flexible circular cylinder *Journal of Fluids and Structures* **15**(3-4) 659-69

Authors

H F Wang, born in May, 1982, Guang Dong Province, P.R. China

Current position, grades: PhD in Harbin Institute of Technology Shenzhen Graduate School, China.
University studies: MSc from Harbin Institute of Technology Shenzhen Graduate School in China.
Scientific interest: floating structure and simulate.
Publications: 5 papers.

Y H Fan, born in September, 1976, Guang Dong Province, P.R. China

Current position, grades: the professor Harbin Institute of Technology Shenzhen Graduate School, China.
University studies: M.Sc. and PH.D from Harbin Institute of Technology University in China.
Scientific interest: structural dynamics (vibration and control of wind turbine; vibration and control of aircraft structure).
Publications: 19 papers.
Experience: teaching experience of 11 years, 4 scientific research projects.

L Yang, born in May, 1987, Guang Dong Province, P.R. China

Current position, grades: engineer in Shang Hai Electric Group Company Limited, China.
University studies: MSc from Harbin Institute of Technology Shenzhen Graduate School in China.
Scientific interest: floating structure and simulate.

Study on the device of multipath household heat metering

Zhongguo Yang*

Collage of Engineering, Hei Longjiang Bayi Agricultural University, 2 Xin Yang Road, Daqing, China

Received 30 August 2014, www.cmnt.lv

Abstract

One of the most important contents of reforming heating system is that more attention will be paid to residential energy conservation and improvement of heating facilities in order to charge heating fee by heat consumption. But the out-dated technology prohibits the widespread use of multipath household heat metering. Since the heating system in China is still confronted by the substantial energy waste, the jumble of pipelines and difficult management of the system etc, a company in Liao Ning has developed a new kind of water supply system on multipath household heat metering. The new system blazes new trails on household heat metering, solving the problem of traditional pipelines. The paper uses the computational method of hydrodynamics and provides the experimental platform of pipe resistance measurement and its theoretical parameter for the system. The study shows that resistance coefficient of the experimental pipeline corresponds with the theory, and multipath household heat metering has favourable prospects.

Keywords: converter multipath, household heat metering, pipe resistance measurement

1 Introduction

The consumption of energy resources is one of the core problems for the "Scientific Concept of Development". How to effectively save energy and reduce consumption has become an important strategic task for the future of a country. According to statistics, heating energy consumption on the building accounts for about 30% of total energy consumption of the national economy and 52% of China's total building energy. Therefore, the heating control is the biggest and most potential way of energy conservation in China, a top priority in energy-saving work. [1] Urban central heating is mainly used in the northern area of China, an important urban infrastructure meeting the requirements of basic living quality of urban residents. At present, China's heating fee is charged according to the area, traditional, unreasonable, more importantly, wasting energy; so many experts put forward the concept of household heat metering.

2 System introduction

Household heat metering not only saves energy, but also charges fairly and manages easily [2]. It changes the heating pipes from the initial series or parallel operation into a set of system of each family, solving the problem that now if individual users do not pay, other users would be involved in real trouble. At the same time, it truly realizes the heating commercialization since it charges by the actual use of heating instead of the housing area. In view of that, a company in Province Liaoning developed a new type of water supply system of multipath household heat metering. The centralized control of the household heat metering spreads among seven floors, and it is shown in Figure 1. For the system, porous pipes and accessories of pipelines are mainly made of polyvinyl chloride resin (PVC-U) and polypropylene (PP-R) respectively, lastly centralized control system on water supply system of the household metering comes into being. Floor one and floor two share pipes with seven channels, and the interior of the pipe is divided evenly into small seven fan-shaped paths, which correspondingly suply one house. And water separator with seven holes turns into those with five holes between the second and third floor, so floor three and floor four use pipes with five channels. In Figure 2, since the interior of the pipe was divided into small five fan-shaped paths, water separator with five holes turns into those with three holes between floor four and floor five. So pipes with three channels are used between floor five and floor six, and between floor six and floor seven is three steering connector. The pipes between the floors uses flange to connect together, sealed by the sealing rubber ring. This system adopts mechanical circulatory system of single up-supply and down-circle pipe in practice, and the main water supply pipe for heating reaches the top of floor seven. Put the main water supply pipe and water separator of seven holes together and meanwhile water separator of seven holes connect pipes of seven channels through output of seven paths. Heating each floor, output of one path among output of seven hot water paths connects with heating equipment, and blocking pad with seven channels will seal firmly the hole of the bottom, so this part becomes empty, preparing backwater of heating for the user. The heating backwater of floor seven is directed into the original pipes of seven empty holes, finally achieving a heating return around floor seven. By this way, the heating shunt realizes household heating of floor seven to floor one by different paths. Completing the seven loop heating, the system extends to the underground and then into heating well or heating box through right-angle connector of seven holes, so in such condition is heating

***Corresponding author** e-mail: yzg111402@163.com

system of household metering developed through multi-channel self circulation. It charges heating fee by central heating metering, suitable for every house of new buildings, and also applicable to reconstruction of the old building. There is profound basis for heating household metering. Central heating, the infrastructure of the city, saves energy and reduces environment pollution, improving people's quality of life, and so on.

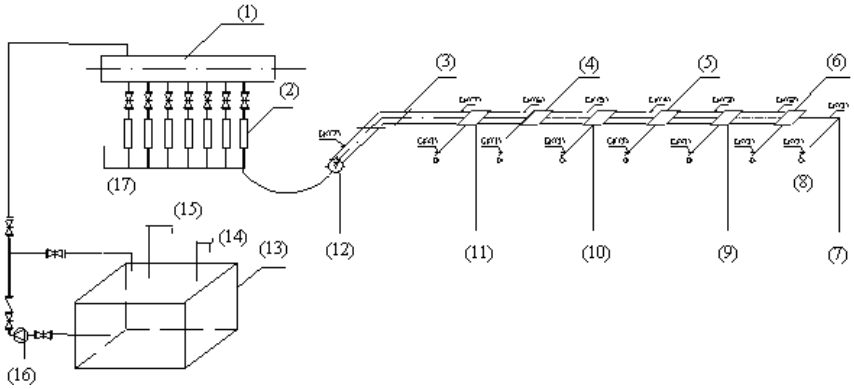


FIGURE 1 Centralized control system on water supply system of the household metering

(1) water separator (2) flow meter (3) right-angle connector of seven holes (4) water separator with five holes changed from those with seven holes

(5) water separator with three holes changed from those with five holes (6) head cover of three channels (7) 20 connectors (8) connecting with the water tank

(9) water separator with three holes (10) water separator with five holes (11) water separator with seven holes (12) merging device of seven holes

(13) water tank (14) connecting circulation of the sink (15) water replenishing (16) water pump (17) merging device

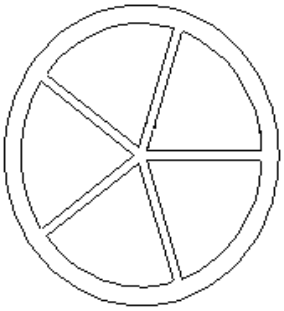


FIGURE 2 The diagram of the pipe with five channels

3 Calculation on heat transfer between different pipes

Now China is gradually charging fees by household metering of central heating. On the one hand, the method saves energy; on the other hand, it brings a lot of trouble. First, different users are located in the different places of the buildings, so each room has different face with various external and internal walls, and each user has different heating load caused by different circumstances of transferring the heat. Secondly, when temperature of one room is higher than its adjacent room, the room would transfer heat to the adjacent room, increasing heat load of the other room and meanwhile users do not need genuinely this part of increased load. Although this system is newly developed, there still exist some problems, such as heat transferring between the pipes. Take the pipe of seven channels as an example, temperature of indoor water is 85°C and backwater temperature 60°C in this system. Since the water still goes back to the duct, six hot channels and one cold path appear among the pipes of floor six (it is only relatively said that 85°C is hot, 60°C cold), and in the pipes of five floor occur five hot channels and two cold, and so on. Therefore, we carried out the calculation and measurement, thermal conductivity of PVC tube was 0.16w/m2 K, return pipe of the heating system in floor seven have six hot channels and one cold, and that of floor six five, hot channels and two cold.

Return pipe of the second floor is six cold channels and one hot, so conductive heating during the passage between the return water in the second floor to inlet supply pipe in the first floor accounts for the largest amount of heat conduction, and heat conduction is:

$$q = \frac{\Delta t}{\delta} = 666.67 w / m^2 . \quad (1)$$

The most unfavourable heat loss:

$$Q = 2.5 \times 0.25 \times q = 416.66875 w . \quad (2)$$

Heat loss rate:

$$416.66875/7200=5.787\%. \quad (3)$$

Heat transfer in every hole does exist, and the heat loss accounts for about 6% of total calories. To meet the needs

of users, the provision in calculation of the heating is that calculation of heat load increase to 6%-10%. Although this only takes pipes of seven channels as an example, in other cases, the heat loss and the heat load of the room is not different for the same housing.

4 The principle of resistance measurement

Since water resistance in multipath household metering has both pipe frictional drag and local resistance, while measuring the resistance, pipes with seven channels, those of five paths, those of three channels or their accessories are respectively supposed to measure both frictional drag and local resistance.

4.1 A SUBSECTION THE PRINCIPLE OF PIPE FRICIONAL DRAG

According to Bernoulli equation [4]

$$z_1+\frac{p_1}{\gamma}+\frac{v_1}{2g}=z_2+\frac{p_2}{\gamma}+\frac{v_2}{2g}+h_f. \tag{4}$$

By continuity Equation, when the diameter is constant, flow velocity is [5]

$$v_1=v_2=v=\frac{Q}{A}, \tag{5}$$

frictional resistance is [6]

$$h_f=(z_1+\frac{p_1}{\gamma})-(z_2+\frac{p_2}{\gamma}). \tag{6}$$

Then by the Equation, [7]

$$h_f=\lambda\cdot\frac{l}{d_e}\cdot\frac{v^2}{2g}. \tag{7}$$

friction coefficient λ can be calculated, meanwhile, also according to the Equation [8]

$$R_e=\frac{vd_e}{v}. \tag{8}$$

Reynolds number R_e is obtained and the diagram $R_e-\lambda$ with different Reynolds numbers drawn.

4.2 THE DETERMINATION PRINCIPLE OF LOCAL RESISTANCE

From the Equation

$$h_\xi=\xi\frac{v^2}{2g}, \tag{9}$$

$$\xi=\frac{2gh_\xi}{v^2} \tag{10}$$

is available.

Then the measurement of local resistance loss can be read by the piezometric tube, that is, the actual local resistance loss should be equal to the measurement of the local resistance loss minus loss of the two measurement points along the path.

5 The result of measurement

According to requirements of actual heating velocity, design velocity for the experimental pipeline is 0.8~1.2m/s and the corresponding Reynolds number about 14500~21000. In this range, regardless pipes with seven channels, those of five channels or those of three channels, the corresponding region in diagram Mo Di are smooth region of turbulent flow and transitional rough area of turbulent flow. Roughness of the pipe has been determined (about 0.0008 mm), so the curve $\mathrm{Re}-\lambda$ presents the downtrend. By using the method of curve regression, the paper shows the working system under the non-design condition. Figure 3 is a diagram $\mathrm{Re}-\lambda$ of pipes with seven channels.

In order to verify accuracy of the experimental results, in the same diagram can be drawn and compared theoretical curve of the regression for pipes with seven channels, those of five paths, those of three channels and cast iron pipe with the same roughness, as is shown in Figure 4. In Table 1 is local resistance coefficient.

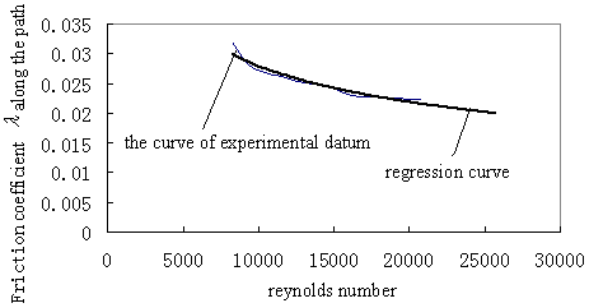


FIGURE 3 The diagram of pipes with seven channels

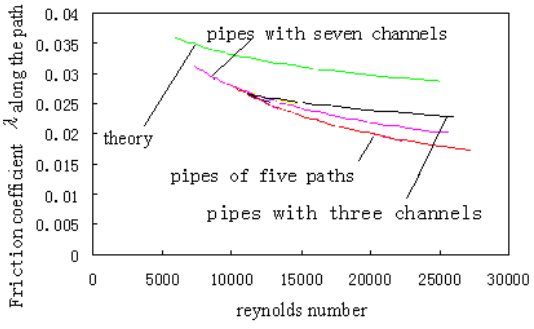


FIGURE 4 Regression curve $\mathrm{Re}-\lambda$

TABLE 1 The local resistance coefficient

Project	system of floor7	system of floor6	system of loor5	system of loor4	system of loor3	system of loor2	system of loor1
resistance coefficient of merging device of seven holes							1.54363
resistance coefficient of connector of seven holes	0.949072	1.700191	1.09999	1.41039	1.16111	1.88151	1.35818
resistance coefficient of water separator with seven holes	1.037447	1.127317	1.61171	1.06657	1.82916	1.56194	1.65768
resistance coefficient of water separator with five holes changed by those with seven holes	1.271711	1.226959	1.33825	1.16806	1.85618	1.72526	
resistance coefficient of water separator with five holes	1.569386	1.11457	1.1921	1.06456	1.88297		
resistance coefficient of water separator with three holes changed by those with five holes	1.496716	1.634158	1.49962	1.61476			
resistance coefficient of water separator with three holes	1.147256	1.511206	1.48419				
resistance coefficient of head cover of three channels	0.798686	4.812194					
20 connectors	0.441452						
mat of seven apertures							1.03224
mat of seven apertures							0.72685
mat of three apertures							1.35128

6 Conclusion

Seen from Figure 3, resistance coefficient of pipes with three channels is biggest, followed by that of pipes with seven channels, then is that of pipes with five channels, and resistance coefficient of cast iron pipe with the same roughness is obviously higher than the other three pipelines. For pipes with seven channels, those with five paths and those with three paths, is the same the friction coefficient of Reynolds number before about 11000, which is located in the smooth region of turbulent flow. Afterwards, is different the roughness of each kind of pipe, and so the turbulent transition region appears, which correspond to actual situation very well.

Error of friction coefficient is smaller, because there is no "instability" in the experimental process, and the graph $R_e - \lambda$ and theory also match very well, so the results can be directly applied to engineering practice. But the local resistance coefficient has relatively poor reliability, there are several reasons:

1) Unavoidable errors read by the man.

2) For flange mat of seven apertures, that of five apertures and that of three apertures, deformation exits greatly. When installing the system, tightening bolts leads directly to the flange deformation. Then large deformation appears in the channels, which deviates directly the test result of the local resistance coefficient greatly from the actual results. This is the reason why only water separator with seven channels on floor one is much higher than that of the other floors.

3) Do house-installed pipe, water separator with five holes changed by those with seven holes, water separator with three holes changed by those with five holes and head cover of three channels fuse poorly. The main problem is that house-installed pipe is not in the same height as the central line of the system, and there exits an angel, which directly causes the local loss of this place.

4) Water pump work longer, water tank is smaller, and the cooling effect is relatively not good. Consequently vapor contained in the water increases and easily collects in the plastic hose connecting with piezometric tube, which leads the water column astray.

References

[1] Yi J 2005 Current building energy consumption in China and effective energy efficiency measures *Journal of HV&AC* **35**(5) 30-41
[2] Liu D 2004 Brief introduction of the household heat metering *Journal of Oil-Gasfield Surface Engineering* **23**(8) 29
[3] Zhao L 2002 Revised study on energy conservation of the household heat metering *Journal of Heating and Ventilating Air-conditioner* **32**(1) 9-11
[4] Long T, Cai Z 2001 Fluid Mechanics *Press: China Architecture & Building Press* (*in Chinese*)
[5] Lamb S H 2010 Hydrodynamics *Press:Cambridge at the University*
[6] Lamb H 2012 Hydrodynamics *Press:Hardpress Publishing*
[7] Pope S B 2010 Turbulent Flows *Press: World Book Inc*
[8] Lamb H 2009 Hydrodynamics *Press:Dover Publications Inc; 6th Revised edition*

Author

Yang Zhongguo, born in January, 1979, Boli County, Heilongjiang Province, P.R. China

Current position, grades: the lecturer of Collage of Engineering of Hei Longjiang Bayi Agricultural University, China.
University studies: M.Sc. from LiaoNing University of Science and Technology in China.
Scientific interest: the fluid flow and heat transfer.
Publications: 21 papers.
Experience: teaching experience of 12 years, 10 scientific research projects.

Research on the relationship between moisture content and the dielectric constant of the tree trunk by the radar wave

Jingxia Lv, Lin Gao*, Jian Wen

School of Technology Beijing Forestry University, Beijing, China, 100083

Received 6 September 2014, www.cmtn.lv

Abstract

Radar wave technology offers a noninvasive, quick determination technique and has potential for the applications of the non-destructive detection (NDT) for the trees trunk and wood based materials. The precision of NDT determination by radar wave is influenced the wood dielectric constant which is closely related to the moisture content of the trees trunk. For our investigations we used TRU trees radar wave to detect the trees trunk. Four typical trees trunks were selected: polar, willow, pine, eucalyptus. Different trees trunks moisture content and dielectric constant were tested under the radar wave frequency respectively. Models of the relationship between moisture content and dielectric constant of the trees trunks were established for improving the accurate of radar wave NDT.

Keywords: radar wave, dielectric constant, moisture content, tree trunk, non-destructive detection

1 Introduction

As a new detection technology, Radar wave NDT has been widely applied in geological exploration. Electromagnetic wave signal was first used in detection of underground metal body by German Hulsmeyer. In 1910, Leimback and Lowy proposed radar's principles to ground-penetrating in the form of patents, it was the first time that the concept of ground penetrating radar (GPR) was put forward formally [1, 2]. In the early 1970 s, Domestic began the study of ground penetrating radar (GPR) and gradually applied to engineering application [2, 3]. In the early 1990 s, the technology was gradually widely used in many fields such as engineering geology exploration, quality inspection, mine detection, disaster geology survey, archaeological investigation, exploration, non-destructive evaluation [4]. In 2009, J R Butnor [5] etc. applied GPR technology to pinaceae standing timber, and the detection result was compared with hardwood. The TRU trees radar was developed by Tree Radar Company in the USA for wood internal defects detection. It was the first time the Radar wave was applied in the tree trunk NDT commercial application. Due to the fixed electric constant value of the TRU radar system, limited precision and poor suitability have limited its development and application in tree trunk NDT. Many experiment results showed that the detect result of wood internal defects changed by selected dielectric constants. And the moisture content has a great influence on the dielectric constant [6, 7]. Thus, it can be seen that the detecting precision of the defects is closely related to the dielectric constant of the measured wood, and wood moisture content is the key factor of dielectric constant of the wood. Therefore, in this paper, the different relationship of four typical woods such as poplar, willow, pine and eucalyptus between dielectric constant and moisture content were studied respectively.

2 Material and methods

The four kinds of typical wood adopted in this study are newly felled. It is reported that the dielectric properties of wood are related to wood structure, grain direction, moisture content, temperature, magnetic field frequency [8-10]. So, the radar detection system with a constant frequency of 900 MHz was used to investigate the dielectric properties at room temperature. Wood samples were in the form of cylinders with a diameter of 32 mm and a height of 3.0-3.5mm, in addition, the two main directions of wood samples were considered – parallel to grain and perpendicular to grain.

The dielectric constants were carried out by a Precision Impedance Analyser (Agilent 4294A) shown in Figure 1. Tree Radar Company's TRU radar equipment was used in this study shown in Figure 2.

FIGURE 1 Precision Impedance Analyser

* ***Corresponding author*** e-mail: gaolin0215@163.com

FIGURE 2 Field data manager and the radar antenna

Samples were placed under different moisture conditions for 96 hours.

The weight of the samples were measured respectively before and after each measurement of dielectric constant, denoted as w1 and w2. Dielectric constant values were measured at room temperature under the frequency of 900 MHz, denoted as ε.

After the test, samples were placed in the air.

Repeated step 1~2 until the weight had no significant change.

Finally the oven dried weight of the specimens was taken by drying in an electronic oven at 100 ± 3°C for 24 hours.

PASW Statistics software was carried out on the test data processing [11].

3 Results and discussion

First of all, the significance level of four groups' data was analysed respectively, the all result: sig. = 0.000. It shows there is a significant difference in moisture content and dielectric constant. Therefore, the data could be processed through curve fitting method.

R^2 is the curve fitting degree. That is how many percent of your sample curve similar. *F* value indicates a significant degree of R^2. *Sig*. is significant, the closer to 0.000, the better. Parameter estimation of b_0, b_1, b_2, b_3 is equation coefficient respectively:

$$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 . \quad (1)$$

Equation's inspection level α is 0.001, the value of sig. was 0.000. Four equations of poplar are significant. The highest fitting (R^2) is cubic curve, R^2 is 0.973. Therefore, cubic curve's fitting level is the best. The relationship between the dielectric constant and moisture content of poplar equation is:

$$\varepsilon = -0.351 + 14.994\omega - 19.475\omega^2 + 9.543\omega^3 , \quad (2)$$

where ε is the dielectric constant, ω is the moisture content.

The cubic curve in the Figure 3 indicates that the dielectric constant increases as the moisture content increases from 0 to 25%. The moisture content from 25% to 100% does not have much effect on the dielectric constant. The abrupt change of the dielectric properties is observed at very high moisture content and the curve becomes concave upward.

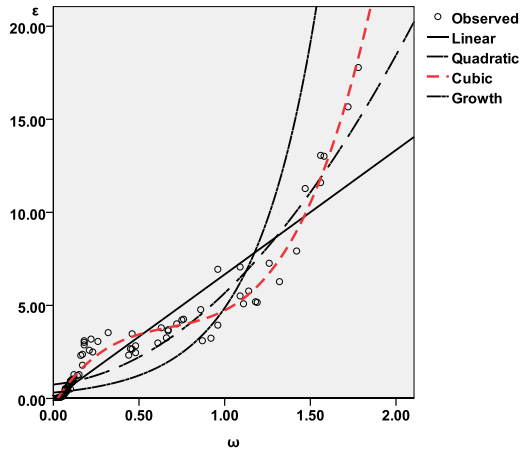


FIGURE 3 Curve fitting poplar moisture and dielectric constant

The various parameters in Tables 2 and 1 are the same. Equation's inspection level α is 0.001, the value of *Sig*. is 0.000. Four equations of willow are significant. The highest fitting (R^2) is cubic curve, R^2 is 0.964. Therefore, cubic curve's fitting level is the best. The relationship equation between the dielectric constant and moisture content of willow is:

$$\varepsilon = -0.132 + 13.738\omega - 22.307\omega^2 + 11.638\omega^3 , \quad (3)$$

where ε is the dielectric constant, ω is the moisture content.

TABLE 1 Popular model and parameter estimation

Equation	The Model Summary					Parameter Estimation			
	R^2	F	df_1	df_2	$Sig.$	b_0	b_1	b_2	b_3
linear	.868	597.189	1	91	.000	.029	6.755		
secondary	.913	474.960	2	90	.000	.755	1.412	3.650	
cubic	.973	1066.237	3	89	.000	–.351	14.994	–19.475	9.543
growth	.525	100.572	1	91	.000	–1.065	2.680		

TABLE 2 Willow model and parameter estimation

equation	The model summary					Parameter estimation			
	R^2	F	df_1	df_2	$Sig.$	b_0	b_1	b_2	b_3
linear	.711	201.997	1	82	.000	.386	4.345		
secondary	.817	180.651	2	81	.000	1.111	–.681	3.885	
cubic	.964	705.922	3	80	.000	–.132	13.738	–22.307	11.638
growth	.428	61.370	1	82	.000	–.715	2.228		

ω is the independent variables, ε is the dependent variable.

From cubic curve in the Figure 4 we can see, the dielectric constant increases as the moisture content increases from 0 to 20% the moisture content from 20% to 100% does not have much effect on the dielectric constant. The abrupt change of the dielectric properties are observed at very high moisture content and the curve becomes concave upward.

The various parameters in Tables 3 and 1 are the same. Equation's inspection level α is 0.001, the value of *Sig.* is 0.000. Four equations of pine are significant. But the curve fitting is not very good. R^2 is only 0.803, according to the law of the scatter distribution, piecewise fitting, depending on the moisture content of pine, is divided into below 25% and above 25%.

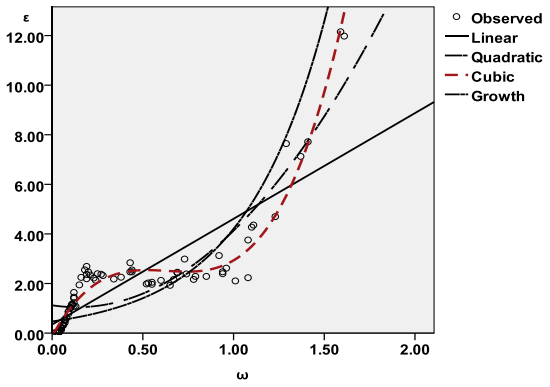


FIGURE 4 Curve fitting willow moisture and dielectric constant

TABLE 3 Pine model and parameter estimation

equation	The model summary					parameter estimation			
	R^2	F	df_1	df_2	$Sig.$	b_0	b_1	b_2	b_3
linear	.803	374.050	1	92	.000	.168	8.269		
secondary	.803	185.196	2	91	.000	.114	8.641	–.312	
cubic	.803	122.365	3	90	.000	.219	7.394	2.174	–1.223
growth	.528	102.908	1	92	.000	–.907	3.316		

ω is the independent variables, ε is the dependent variable.

TABLE 4 Pine model and parameter estimation(below 25%)

equation	The model summary					parameter estimation			
	R^2	F	df_1	df_2	$Sig.$	b_0	b_1	b_2	b_3
linear	.926	597.056	1	48	.000	–1.012	20.508		
secondary	.957	528.419	2	47	.000	–.346	6.297	58.053	
cubic	.967	451.443	3	46	.000	.101	–10.791	225.962	–455.280
growth	.823	222.439	1	48	.000	–3.408	25.786		

ω is the independent variables, ε is the dependent variable

The various parameters in Tables 4 and 1 are the same. Equation's inspection level α is 0.001, the value of *Sig.* is 0.000. Four equations of pine are significant. The highest fitting (R^2) is cubic curve, R^2 is 0.967. Therefore, cubic curve's fitting level is the best. The relationship equation between the dielectric constant and moisture content of pine (Figure 5 below 25%) is:

$$\varepsilon = 0.101 - 10.791\omega + 225.962\omega^2 - 455.280\omega^3 . \quad (4)$$

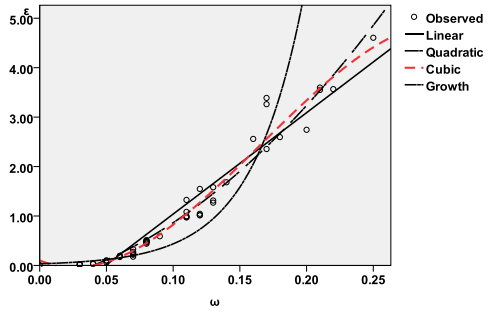


FIGURE 5 Curve fitting pine moisture and dielectric constant

The various parameters in Tables 5 and 1 are the same. Equation's inspection level α is 0.001, the value of *Sig*. is 0.000. Four equations of pine are significant. The highest fitting (R^2) is cubic curve, R^2 is 0.950. Therefore, cubic curve's fitting level is the best. The relationship equation between the dielectric constant and moisture content of pine (Figure 6 above 25%) is:

$$\varepsilon = 11.727\text{-}45.725\omega + 72.995\omega^2\text{-}29.178\omega^3. \quad (5)$$

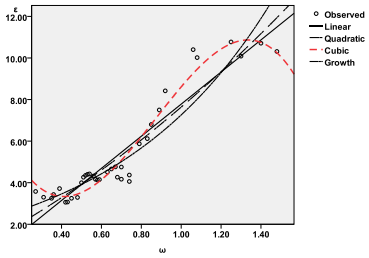


FIGURE 6 Curve fitting pine moisture and dielectric constant

TABLE 5 Pine model and parameter estimation (above 25%)

equation	The model summary					parameter estimation			
	R^2	F	df_1	df_2	Sig.	b_0	b_1	b_2	b_3
linear	.842	207.991	1	39	.000	.200	8.248		
secondary	.843	102.196	2	38	.000	–.306	9.755	–.914	
cubic	.950	235.228	3	37	.000	11.727	–45.725	72.995	–29.178
growth	.857	232.948	1	39	.000	.770	1.291		

ω is the independent variables, ε is the dependent variable

The various parameters in Tables 6 and 1 are the same. Equation's inspection level α is 0.001, the value of *Sig*. is 0.000. Four equations of Eucalyptus are significant. The highest fitting (R^2) is cubic curve, R^2 is 0.951. Therefore, cubic curve's fitting level is the best. The relationship equation between the dielectric constant and moisture content of Eucalyptus is:

$$\varepsilon = -0.365 + 9.426\omega + 10.976\omega^2 - 15.777\omega^3. \quad (6)$$

TABLE 6 Eucalyptus model and parameter estimation

equation	The model summary					parameter estimation			
	R2	*F*	*df1*	*df2*	*Sig.*	*b0*	*b1*	*b2*	*b3*
linear	.945	1277.662	1	75	.000	–.343	10.791		
secondary	.949	692.827	2	74	.000	–.634	13.658	–4.682	
cubic	.951	469.925	3	73	.000	–.365	9.426	10.976	–15.777
growth	.738	211.522	1	75	.000	–1.108	5.957		

ω is the independent variables, ε is the dependent variable.

Pine overall fitting curve is shown in Figure 7. By the fitting degree of R^2, the piecewise fitting degree of pine is larger than the whole. So we choose the sub-fitting equation.

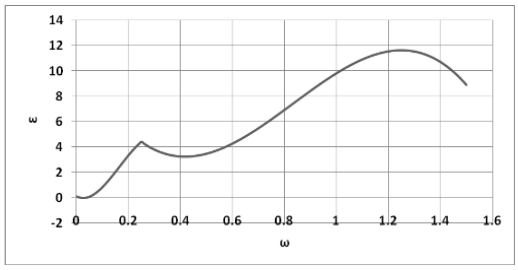


FIGURE 7 Curve fitting pine moisture and dielectric constant

From cubic curve in the Figure 8 we can see, the dielectric constant increases as the moisture content increases from 0 to 65%.

The main work of this paper is to use the TRU trees radar system to achieve accurate measurement of timber internal defects. Because in the measurement process of intermediary the radar system's electric constant value is fixed, the system exists accuracy defects. The results show that if different dielectric constants can be chosen, the location of detected wood internal defects is different. And as the moisture content of lumber changes, its dielectric constant changes.

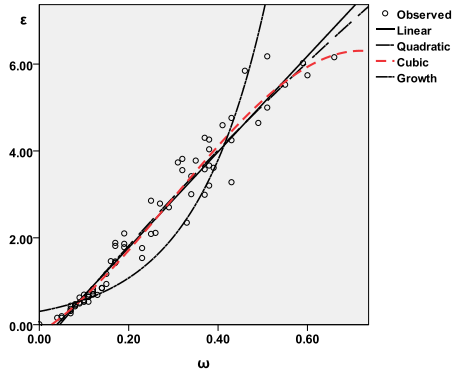


FIGURE 8 Curve fitting eucalyptus moisture and dielectric constant

According to the dielectric constant values of wood at different moisture content obtained in this paper, taking pine as an example, the results of analysis are applied. A piece of pine whose moisture content is 100% was chosen. A hole is measured in 10.5 cm from the bark. The defect position was gotten with TRU trees radar system, as shown in Figure 8. When the moisture content is over 25%, the pine's dielectric constant is obtained by this article. The relationship equation:

$$\varepsilon = 11.727 - 45.725\omega + 72.995\omega^2 - 29.178\omega^3. \quad (7)$$

If the moisture content is 100%, the dielectric constant is 9.819. The permittivity of TRU trees radar detection system was corrected to 9.819, measured defect location again, as shown in Figure 9. Compare Figure 8 to Figure 9, defect location in Figure 8 is 9.5 cm, error is about 9.5%. But defect location in Figure 10 is 10.7 cm, error is about 1.9%. Therefore, the measurement position using the model was closer to the actual position of a defect. It was found that application of the model has improved the accuracy of the TRU trees radar system.

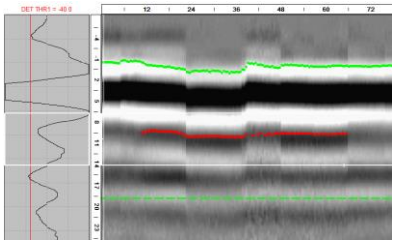


FIGURE 9 defect locations of pine

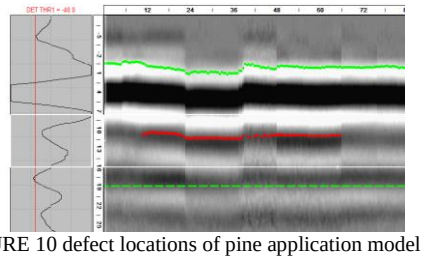


FIGURE 10 defect locations of pine application model

4 Conclusions

Relational models of four typical wood moisture content and dielectric constant were obtained, by the dielectric constant measuring instrument and the weighing method. The result shows that, for poplar, willow, eucalyptus, distribution curves of the relation between moisture content and dielectric constant were cubic, but for Pine, distribution curve was piecewise fitted with two different cubic. The dielectric constant increases exponentially with increasing the moisture content from 0 to Fiber saturation point (FSP), while it does not have much effect within the range of FSP to 100%, the concussion of measurement point was caused by measurement errors. When the moisture content is over 100%, the dielectric constant increases a bit linear upward trend.

Therefore, based on the relational model of moisture content and dielectric constant obtained, the accuracy of measurement instrument has been improved, meanwhile, experimental results showed that the model can improve the accuracy of the instrument effectively.

Acknowledgements

This work was supported by "the Fundamental Research Funds for the Central Universities (BLX2013008)".

References

[1] Jol H M 2008 Ground penetrating radar theory and applications *Elsevier*
[2] Saarenketo T, Scullion T 2000 Road evaluation with ground penetrating radar *Journal of applied geophysics* **43**(2) 119-38
[3] Neal A 2004 Ground-penetrating radar and its use in sedimentology: principles, problems and progress *Earth-science reviews* **66**(3-4) 261-330
[4] Bai B, Jian Z 2001 Advances and applications of ground penetrating radar measuring technology *Chinese Journal of Rock Mechanics and Engineering* **20**(4) 527-31
[5] Butnor J R, Pruyn M L, Shaw D C, Harmon M E, Mucciardi A N, Ryan M G 2001 Detecting defects in conifers with ground penetrating radar: Applications and Challenges Forest Pathology **39**(5) 309-22
[6] James W L 1975 Dielectric Properties of Wood and Hardboard: Variation with Temperature, Frequency, Moisture Content, and Grain Orientation (No. FSRP-FPL-245) *Forest Products Laboratory Madison WIS*
[7] Torgovnikov G I 1993 Dielectric properties of wood and wood-based materials
[8] Kabir M F, Daud W M, Khalid K, Sidek H A A 1998 Dielectric and ultrasonic properties of rubber wood: Effect of moisture content, grain direction and frequency *Holz als Roh-und Werkstoff* **56**(4) 223-7
[9] Hamiyet S, Ay N 2004 Dielectric properties of hardwood species at microwave frequencies *Journal of Wood Science* **50**(4) 375-80
[10] Şahin K H 2009 Thermal and dielectric properties of pine wood in the transverse direction *Bioresources* **4**(4) 1663-9
[11] Pallant J 2001 SPSS survival manual: A step-by-step guide to data analysis using SPSS for Windows (Version 10) *Allen & Unwin*

Authors

Jinxia Lv, born in January, 1989, Beijing, P.R. China

Current position, grades: Master graduate student at the School of Technology, Beijing Forestry University, China.
University studies: BSc in Automation at Beijing Forestry University in China.
Scientific interests: automatic control, nondestructive testing.
Publications: more than 2 papers.
Experience: teaching experience of 1 years, 3 scientific research projects.

Lin Gao, born in February, 1958, Beijing, P.R. China

Current position, grades: associate professor at the School of Technology, Beijing Forestry University, China.
University studies: Doctor degree in Engineering at Beijing Forestry University in China.
Scientific interest: automatic control, nondestructive testing.
Publications: more than 25 papers.
Experience: teaching experience of 30 years, 4 scientific research projects.

Jain Wen, born in June, 1981, Beijing, P.R. China

Current position, grades: lecturer at the School of Technology, Beijing Forestry University, China.
University studies: Phd in Mechatronic Engineering at Beijing Jiaotong University in China.
Scientific interest: nondestructive detection, electromechanical control system.
Publications: more than 10 papers.
Experience: teaching experience of 3 years, 3 scientific research projects.

Physical parameter identification and wave force inversion research of bridge pier structure

Zijian Wang[1]*, Liming Wu[2]

[1]*Chongqing University of Science & Technology, Chongqing, China, 401331*

[2]*Chongqing Technology and Business Institute, Chongqing, China, 400052*

Received 7 September 2014, www.cmnt.lv

Abstract

According to total compensation composite inversion algorithm, physical parameter identification and inversion of wave time interval are done under the condition of input information of isolated pier in wave force part being unknown. Based on the characteristics of large-diameter piers in shallow water which are affected by wave force and combining with "statistical average method" in probability theory, structural physical parameter identification and wave inversion calculation are done under the condition of unknown wave force. Analysis of numerical values shows that results of structural physical parameter identification cater for accuracy requirements which is feasible for inversion method of parameter identification of large-diameter piers in shallow water with unknown input information thus providing a new method for structural physical parameter identification research of river-spanning and sea-crossing piers affected by wave force and providing references for engineering application.

Keywords: bridge pier structure, parameter identification, unknown input, load inversion, time domain correlation

1 Introduction

Performance of river-spanning and sea-crossing bridges in service gradually degenerates under the effect of load and environment. Although degeneration process is slow, structural damage is sudden and fragile which keeps high randomness and unpredictability giving considerable difficulties to safety forecasting. Therefore bridges' health monitoring and preservation including detecting the current condition of bridges in service and judging their damage condition are essential. Correctness of bridge structural health monitoring results mainly depends on whether recognition algorithm could accurately and effectively identify the real form of structure from the actual recorded signal.

At present most recognition algorithms are established on the basis of recorded information being complete which is not suitable for damage detection of river-spanning and sea-crossing bridge structures. The reason is that environmental excitation is generally taken as vibrating load in physical testing. However, interaction positions of these vibrating loads are uncertain which brings difficulties for measurement and is difficult to guarantee the completeness of recorded information. Moreover limited by testing cost, not enough number of sensor could be used to record the response messages at all positions with independent freedom degree in the structure, which is also difficult to guarantee the completeness of response. Therefore, how to identify the structural dynamic properties under the condition of testing information being incomplete or unknown is a difficult problem in the research of bridge dynamic detection technology theory at present. According to this, this paper takes irregular wave excitation simulated by linear amplitude superposition method as example and utilizes total compensation composite inversion algorithm to do identification on isolated piers under the following two input conditions:

1) When wind power is low, only underwater part of isolated bridge pier is affected by wave. While force on the part above the water is approximate to 0. Identification is done on it under the condition of input information being unknown.

2) Large-diameter piers in shallow. water are totally submerged. They are done with unknown input information system identification affected by wave force. Identification results show that total compensation composite inversion algorithm, in which wave excitation is manually simulated could better invert wave effect, do large-diameter unknown input system identification and wave force inversion.

2 Numerical simulation of random wave force of isolated piers

2.1 NUMERICAL SIMULATION OF IRREGULAR WAVE OF LINEAR SUPERPOSITION METHOD

Suppose that the known input wave force $f(t)$ is Gaussian random process whose average value i=0 and wave spectrum density function $S(\omega)$ is known. Then numerical simulation could be done on manual wave through three methods of constant amplitude superposition method, various amplitude superposition method and autoregressive method. This paper utilizes various amplitude superposition method, in which Fourier rapidly transforms helping to greatly save

***Corresponding author** e-mail: 121687678@qq.com

computation time. Comparing with constant amplitude superposition method, it presents certain improvement and keeps unconditionally stable. Specific derivation and simulation process is as follows [1]:

One random wave process could be described as:

$$\eta(t)=\sum_{n=1}^{\infty}a_n\cos(k_n x-\omega_n t-\varepsilon_n). \quad (1)$$

In terms of a fixed point, it is advisable that x=0, then:

$$\eta(t)=\sum_{n=1}^{\infty}a_n\cos(\omega_n t+\varepsilon_n). \quad (2)$$

In order to determine the of each a_n,ω_n component wave, wave spectrum $S_{\eta\eta}(\omega)$ may be used to make it in which ε_n is the evenly random distribution number in the range of $0\sim 2\pi$. Take wave spectrum $s_{\eta\eta}(\omega)$ as target spectrum whose energy covers the vast majority in the range $\omega_1\sim\omega_{m+1}$ and other parts are rejected. It is defined that $\Delta\omega_i=\omega_{i+1}-\omega_i$ and $\varpi_i=\frac{1}{2}(\omega_i+\omega_{i+1})$. Superpose m cosine waves of wave energies in m sections whose wavefront equation is:

$$\eta(t)=\sum_{t=1}^{m}\sqrt{2S_{\eta\eta}\varpi_i\Delta\omega_i}\cos(\omega_i t+\varepsilon_i). \quad (3)$$

According to [2], fluctuate horizontal velocity and accelerated velocity of water particle to acquire the following equations:

$$\mu(t)=\sum_{t=1}^{m}\sqrt{2S_{\eta\eta}(\omega_i)\Delta\omega_i}\,\omega_i\frac{\cosh(k_i z)}{\sinh(k_i d)}\cos(\omega_i t+\varepsilon_i), \quad (4)$$

$$\mu(t)=-\sum_{t=1}^{m}\sqrt{2S_{\eta\eta}(\omega_i)\Delta\omega_i}\,\omega_i^2\frac{\cosh(k_i z)}{\sinh(k_i d)}\cos(\omega_i t+\varepsilon_i). \quad (5)$$

Wave force Equation (6) is affected on piles of per unit length deduced according to Morison.

$$f(t)=f_D(t)+f_I(t)=\frac{\pi}{4}\rho C_m D^2\dot{u}(t)+\frac{1}{2}\rho C_m D^2 u(t)|u(t)|. \quad (6)$$

Suppose that the total height of this pier is *l*, height of the below-surface part is *d*, weight is concentrated on the single-degree-of - freedom system of z_0 height centre. Then time-varying Equations (7)-(9) of the simplified total inertia force of single degree of freedom, total drag force item and total wave force are shown as Equation (9):

$$P_I(t)=\frac{\pi}{4}\rho C_m D^2\dot{\mu}(t)d, \quad (7)$$

$$P_D(t)=\frac{1}{2}\rho C_D D\mu(t)|\mu(t)|d, \quad (8)$$

$$F_l(t)=\frac{1}{2}\rho C_D D f(z_0)\sum_{i=1}^{m}2S_{\eta\eta}(\omega_i)\Delta\omega_i\omega_i^2\frac{\cosh(k_i z)}{\sinh(k_i d)}\cos(\omega_i t+\varepsilon_i)\left|\frac{\cosh(k_i z)}{\sinh(k_i d)}\cos(\omega_i t+\varepsilon_i)\right|-\frac{\pi}{4}\rho C_m D^2 d f(z_0)\sum_{t=1}^{m}\sqrt{2S_{\eta\eta}(\omega_i)\Delta\omega_i}\,\omega_i^2\frac{\cosh(k_i z)}{\sinh(k_i d)}\sin(\omega_i t+\varepsilon_i), \quad (9)$$

$$F_j(t)=\frac{\pi}{4}\rho C_m D^2\frac{\cosh(k_i z)}{\sinh(k_i d)}\sum_{i=1}^{m}\sqrt{2S_{\eta\eta}(\omega_i)\Delta\omega_i}\,\omega_i^2\sin(\omega_i t+\varepsilon_i)dz,\ j=1,2,\ldots. \quad (10)$$

Total wave force $F_l(t)$ in Equation (9) is the wave force vibration equation, which is acquired after wave is transformed to the top of pier. Substitute this wave force vibration equation into dynamic equation, in which response spectrums of displacement, velocity and acceleration of this pier structure could be acquired through calculating this differential equation.

2.2 SIMPLIFIED CALCULATION OF WAVE FORCE OF LARGE-DIAMETER PILE IN SHALLOW WATER

2.2.1 Water Depth Keeping Great Influence on Wave Motion

Wave in deep water (*d*≥*0.5L*), *d* means water depth, *L* is wave length, the same below.) is called deep-water wave. Wave in shallow water (*0.05L*≤*d*<*0.5L*) is called shallow-water wave. Wave in extra-shallow water (*d*<*0.5L*) is called extra-shallow wave. According to [3], suppose that wave period *T* or wave number *K* in shallow water do not change with water depth namely $K=\frac{2\pi}{L}$.

2.2.2 Large-Diameter Components

Aiming at large-diameter components, the influences of diameter of pier on wave force usually utilize diffraction theory proposed by MacCamy and Fuchs from USACE in 1954. In terms of relatively big-size structures (*D*≥0.2*L*, *D* is the diameter of the structure), viscosity effect could be neglected. Wave force is main the inertia force [3].Thereby formula of simplified wave force of large-diameter piers in shallow water taking *dz* height as unit is as Equation (10). Here *dz* is the height of pier which is divided into single degree of freedom.

j is the dividing segments taking dz as height and $j = d / dz$. z_j is the height from centre of j to the bottom of pier. Other parameters are the same as the above.

It is deduced from the simplified Equation (10) that wave force ratio affected on any adjacent unit is determined only by $\cosh(kz_j)$:

$$\alpha_j = \frac{F_{j+1}(t)}{F_j(t)} = \frac{\cosh\left(kz_{j+1}\right)}{\cosh\left(kz_j\right)}. \tag{11}$$

3 Algorithm theory and identification steps

3.1 PART OF UNKNOWN EXTERNAL EXCITATION MATRIX

Electrical Dynamic equation of multiple-degree-of-freedom system is:

$$M\ddot{u} + C\dot{u} + Ku = F(t), \tag{12}$$

where M, C, K are respectively weight, damping and stiffness matrix. $F(t)$ is the external excitation vector. u, $\dot{u}$ and $\ddot{u}$ are respectively structural displacement, speed and acceleration response vector.

Suppose that M is the known diagonal moment. Transpose the Equation (12) to acquire that:

$$C\dot{u} + Ku = F(t) - M\ddot{u} = P. \tag{13}$$

Lead P to become system input vector.

1) Part of unknown input external excitation matrix could be shown as:

$$F(t) = \left[F_k(t) \;\; F_u(t)\right]^{'}. \tag{14}$$

where $F_k(t)$ is external excitation of known time interval information. $F_u(t)$ is external excitation of unknown time interval information. According to Reference [4], Equation (13) could be transformed as:

$$H\theta = P, \tag{15}$$

where H is system response matrix namely the response of system under excitation of external system $F(t)$. θ is the system physical parameter vector waiting for being identified. According to Equations (13) and (14), system input vector of P is acquired:

$$P = \left[P_k \;\; P_u\right]^{'}. \tag{16}$$

Total compensation inversion algorithm identifies system parameter θ according to Equation (15) and inverts unknown input vector P_u under the condition of system response matrix H and part of input P_k being known. Inversion algorithm is done on the part of unknown input according to [5] whose main steps are:

Artificially define initial value of any structural parameter vector to be $\hat{\theta}_0$;

Substitute the given initial value $\hat{\theta}_0$ into Equation (15) to calculate out that:

$$\tilde{P}_0 = H\hat{\theta}_0, \tag{17}$$

where $\tilde{P}_0$ could be expressed as:

$$\tilde{P}_0 = \left[\tilde{P}_k \;\; \tilde{P}_u\right]^{'}. \tag{18}$$

Replace the estimated $\tilde{P}_k$ in Equation (21) with known excitation time interval P_k to acquire the revised system input vector:

$$\hat{P}_0 = \left[P_k \;\; \tilde{P}_u\right]^{'}. \tag{19}$$

According to Equation (18) perform the calculations on the revised $\hat{P}_0$ vector based on least squares criterion to acquire the new estimated value of structural parameter:

$$\hat{\theta}_1 = \left(H^{'}H\right)^{-1} H^{'}\hat{P}_0. \tag{20}$$

Determine whether $\hat{\theta}_1$ caters for all convergence conditions. All parameters cater for the following formula:

$$\left|\frac{\hat{\theta}^i_n - \hat{\theta}^i_{n-1}}{\hat{\theta}_{i-1}}\right| \le \varepsilon, \tag{21}$$

where $\hat{\theta}^i_n$ is the i system parameter estimated value of the n iteration. ε is the artificially given precision value. If all conditions cater for Equation (21), calculation finishes, in which the current system parameter identification result is the final calculation result. If not, new $\hat{\theta}_1$ is taken as initial value, in which Equations (16)-(21) are repeated.

3.2 UNKNOWN EXTERNAL EXCITATION MATRIX

$$F(t) = \left[F_u(t) \;\; \alpha F_u(t)\right]^{'}, \tag{22}$$

where $F_u(t)$ is wave force of the unknown input unit. $\alpha F_u(t)$ is the wave force of adjacent unknown input unit. α is the ratio of wave forces between two adjacent units. They are calculated through Equation (11).

Similarly it could be transformed as follows according to Equation (15):

$$H\theta = P. \tag{23}$$

According to Equation (18), system input vector is acquired:

$$\tilde{P}_0 = \left[\tilde{P}_{ui} \;\; \tilde{P}_{u(i+1)}\right]^{'}, \tag{24}$$

where $\tilde{P}_{ui}$ and $\tilde{P}_{u(i+1)}$ are respectively the unknown input vectors of i and i+1 degrees of freedom or units.

Main steps of inversion algorithm under similar part of unknown input are:

Initial value of artificially given any structural parameter vector is $\hat{\theta}_0$.

Substitute the given initial value $\hat{\theta}_0$ into Equation (16) to calculate out that:

$$\tilde{P}_0 = H\hat{\theta}_0, \tag{25}$$

where $\tilde{P}_0$ can be expressed as:

$$\tilde{P}_0 = \left[\tilde{P}_{ui} \quad \tilde{P}_{u(i+1)}\right]^{'}. \tag{26}$$

After $\tilde{P}_0$ is calculated, $F(t)$ in Equation (22) becomes known, through which estimated values of $\tilde{F}_u(t)$ and $\alpha\tilde{F}_u(t)$ could be acquired according to Equation (22). Generally values of $\tilde{F}_u(t)$ in estimated $\tilde{F}_u(t)$ and $\alpha\tilde{F}_u(t)$ are unequal which does not cater for the equal result deduced in Equation (11). At this moment statistical average method in probability is done on $\tilde{F}_u(t)$, in which the average $\bar{F}_u(t)$ is utilized to acquire the new estimated input force $\hat{P}$ process [6] of each unit.

$$\bar{F}_u(t_i) = \frac{1}{N}\sum_{l}^{N}\tilde{F}_{u.l}(t_i), i = 1,2,...,M; l = 1,2,...,N, \tag{27}$$

where N is degree of freedom or unit number. M is sampling points.

Utilize the new $\bar{F}_u(t)$ to acquire the revised vector $\hat{P}_0$. Then new parameter estimated value could be acquired under least squares criterion according to Equation (20).

Determine whether the new calculated $\hat{\theta}_1$ and the original $\hat{\theta}_0$ cater for the convergence condition of the given precision ε. Suppose that $\hat{\theta}_1^i$ is the i identification value of structural parameter. If all parameters cater for the following equation:

$$\left|\frac{\hat{\theta}_1^i - \hat{\theta}_0^i}{\hat{\theta}_0^i}\right| \le \varepsilon. \tag{28}$$

Then parameter estimated value of this step $\hat{\theta}_1$ is taken as the final calculation result. Otherwise it would be taken as new parameter initial value to repeat steps 2~5 up to convergence.

Utilize the finally identified parameter $\hat{\theta}_1$ to invert the practical wave force excitation according to Equation (16).

Calculation steps under the above-mentioned two input conditions could be expressed more directly by Figure 1.

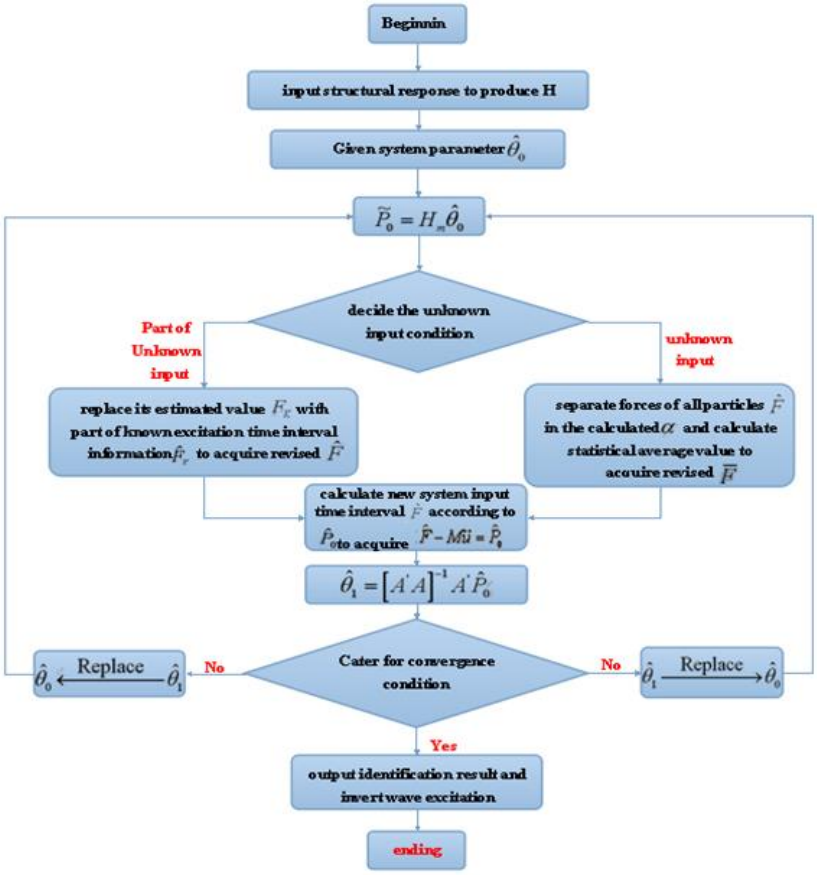


FIGURE 1 Flow diagram of inversion calculation

It is seen from a series of algorithm steps above that physical parameter identifications of pier structure under condition of part of wave input being unknown and of large-diameter piers in shallow water under the condition of wave input being unknown according to total compensation composite algorithm are done whose concept is clear. Thought is of clarity and identification steps are easy for programming realization.

4 Example analysis

4.1 EXAMPLE ENGINEERING PROFILE

Establish the simplified model of isolated pierstructure Figure 2.

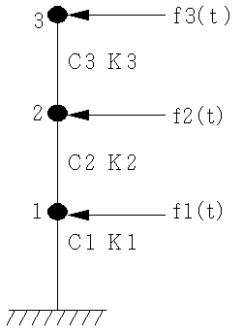


FIGURE 2 Simplified model of isolated pier structure

Suppose that concentrated masses M1 = 18840 kg, M2 = 18000 kg, M3 = 15840 kg, stiffnesses between units K1 = 1.623×10^7 N/m, K2 = 1.082×10^7 N/m, K3 = 0.952×10^7N/m and proportional dampings among units are respectively C1 = 138693 Ns/m, C2 = 108462 Ns/m and C3 = 75786 NS/m. Diameter of the bridge pier is Dm whose heighth is 24m. Here the designed wave height $H = H_{1\%} \approx 1.51H_{1/3} = 9.0m$ and designed wave length $L = \bar{L}$(average wavelength) $= 135.0$m. According to characteristics of wave method, it is acquired that $P_D \le 0.5P_I$. Neglect drag force and only take effects of inertial wave force into consideration. Wave spectrum uses P-M Spectrum. Suppose that displacement, speed and acceleration response of each node are all known. It is required to identify all dampings and stiffness parameters and invert the input wave force. Utilizing $Newark-\beta$ Method to calculate dynamic response, in which sampling point M is 1000 and sampling time interval $dt = 1.0s$. Suppose that all mass parameters in calculation process are known and real parameters of structure are shown in Table 1. Physical parameter identification and effect load inversion are done under the following conditions.

TABLE 1 Parameter list of each node of bridge pier

Node	Weight(kg)	Damping(kg/ms)	Stiffness(N/m)
1	0.1298e+07	0.7788e+06	0.2384e+06
2	0.1400e+07	0.8400e+06	0.2540e+06
3	0.1584e+07	0.9504e+06	0.3010e+06

4.2 PHYSICAL IDENTIFICATION AND WAVE FORCE INVERSION UNDER PART OF INPUT UNKNOWN AND UNKNOWN INPUT

When diameter of pier d is 9.0m and water depth is 20m, wave time intervals of node1 and node2 affected by irregular wave force which is simulated by variable amplitude superposition method are shown in Figure 3 in which correlation coefficient $\alpha_1 = 1.14$ of $f1(t)$ and $f2(t)$. It is known through numerical calculation that after response, the assumed input wave forces $f1(t)$ and $f2(t)$ are unknown. At this moment known force of pile segment above water surface $\alpha_1 = 1.14$. Thereby a problem of parameter identification under part of input being unknown and input load inversion comes into being.

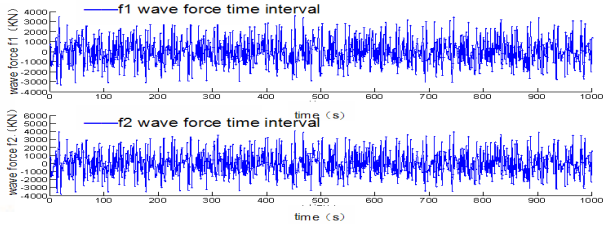


FIGURE 3 Time interval curve of wave force exerted on nodes

Stiffness and damping matrix of bridge pier structure in this case are shown as (29):

$$K = \begin{bmatrix} k_1+k_2 & -k_2 & 0 \\ -k_2 & k_2+k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix}, \quad C = \begin{bmatrix} c_1+c_2 & -c_2 & 0 \\ -c_2 & c_2+c_3 & -c_3 \\ 0 & -c_3 & c_3 \end{bmatrix}. \tag{29}$$

Concrete matrix expressions of H, θ and P deduced by Equation (16) are shown as Equation (30):

$$H(t_i) = \begin{bmatrix} x_1(t_i) & x_1(t_i)-x_2(t_i) & 0 & \dot{x}_1(t_i) & \dot{x}_1(t_i)-\dot{x}_2(t_i) & 0 \\ 0 & x_2(t_i)-x_1(t_i) & x_2(t_i)-x_3(t_i) & 0 & \dot{x}_2(t_i)-\dot{x}_1(t_i) & \dot{x}_2(t_i)-\dot{x}_3(t_i) \\ 0 & 0 & x_3(t_i)-x_2(t_i) & 0 & 0 & \dot{x}_3(t_i)-\dot{x}_2(t_i) \end{bmatrix}, \tag{30}$$

$$P(t) = [P(t_1), P(t_2), P(t_3), ..., P(t_M)]', \tag{31}$$

$$P(t_i) = \begin{bmatrix} f_1(t_i) - m_1\ddot{x}_1(t_i) \\ f_2(t_i) - m_2\ddot{x}_2(t_i) \\ 0 \end{bmatrix}, \tag{32}$$

$$\theta = [k_1, k_2, k_3, c_1, c_2, c_3]', \tag{33}$$

According to the above-mentioned conditions, responses of displacement u, speed $\dot{u}$ and acceleration $\ddot{u}$ of structure affected by $f2(t)$ and $f3(t)$ are acquired utilizing Nemark Method. Meanwhile Node 3 not affected by wave load is taken as known part of input information of input load $f3(t) = 0$. Without regard to noise, parameter results identified by the method of this paper are shown in Table 2. Taking the influence of noise on the actually acquired data into consideration, influence of white noise is added into the response acquired by structure to simulate the actually measured results, in which added noise level is determined by percentage between response peak value and noise peak value [7] in order to test the sensitive degree of this algorithm on the influence of noise. Here parameter identification and load inversion are considered under two conditions of 1% and 6%. Parameter identification results under the condition of noise are shown in Table 3 and Table 4.

It is seen from Table 3 and 4, and Figures 4-7 that under condition of part of unknown input wave force being time domain related, method of this paper could not only identify physical parameter of pier but also accurately invert input load of each node. In terms of 6% of noise level, identified parameters are utilized to do inversion on the input wave forces on Nodes 1 and 2. Comparing with actually input wave force, the maximum error is 6.6149% whose results are shown in Figures 4-7. Obviously, its inversion results are ideal.

TABLE 2 Parameter identification list without noise (iteration 169 times, ε=0.000002)

Node	Damping identification value (kg/ms)	Damping error (%)	Stiffness identification value (N/m)	Stiffness error (%)
1	0.77879e+06	0.012	0.23839e+06	0.004
2	0.83998e+06	0.002	0.25399e+06	0.004
3	0.95040e+06	0.000	0.30100e+06	0.000

TABLE 3 Parameter identification list with 1% noise (iteration 172 times, ε=0.000002)

Node	Damping identification value (kg/ms)	Damping error (%)	Stiffness identification value (N/m)	Stiffness error (%)
1	0.77820e+06	0.077	0.23876e+06	0.151
2	0.8527e+06	1.512	0.25063e+06	1.327
3	0.95407e+06	0.386	0.29753e+06	1.153

TABLE 4 Parameter identification list with 6% noise (iteration 171 times, ε=0.000002)

Node	Damping identification value (kg/ms)	Damping error (%)	Stiffness identification value (N/m)	Stiffness error (%)
1	0.80296e+06	3.1022	0.25417e+06	6.6149
2	0.87081e+06	3.6678	0.24319e+06	4.2559
3	0.98396e+06	3.5311	0.32047e+06	6.4684

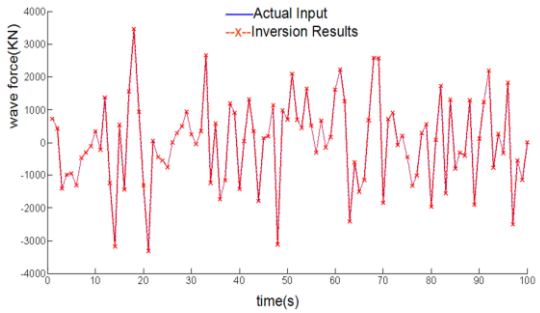


FIGURE 4 Comparison between wave load input and inversion time interval on node 1

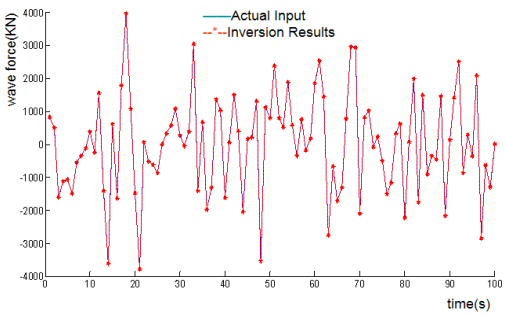


FIGURE 5 Comparison between wave load input and inversion time interval on node 2

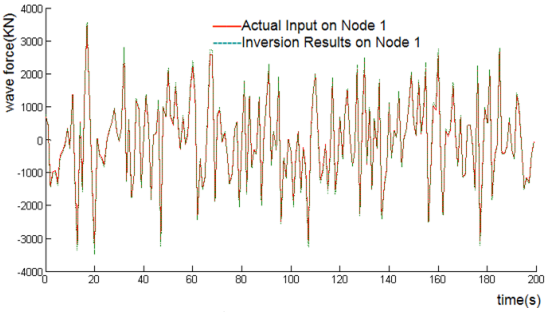


FIGURE 6 Comparison of actual input inversion results on node 1 under 6% noise

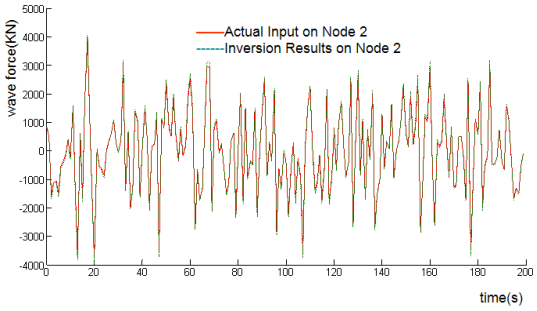


FIGURE 7 Comparison of actual input inversion results on node 2 under 6% noise

When diameter of pier D is 9.0m and water depth is 24m, wave time intervals of Nodes 1, 2 and Node 3, affected by irregular wave force, which is simulated by variable amplitude superposition method are shown in Figure 4, among

which correlation coefficient of $f1(t)$ and $f2(t)$ is $\alpha_1 = 1.14$ and the one of $f1(t)$ and $f3(t)$ is $\alpha_2 = 1.44$. Through numerical calculation it is acquired that input wave forces $f1(t)$, $f2(t)$ and $f3(t)$ are unknown after response. Thereby a typical problem of parameter identifycation and load inversion under unknown load input comes into being. Input wave time intervals on pier nodes are shown in Figure 8.

Similar with the condition above, parameter result identified by the method in this paper are shown in Table 5 without taking noise into consideration. Similarly parameter identification and load inversion under 1% and 6% percentage of noise are considered. Parameter identification results with noise are shown in Table 6 and Table 7.

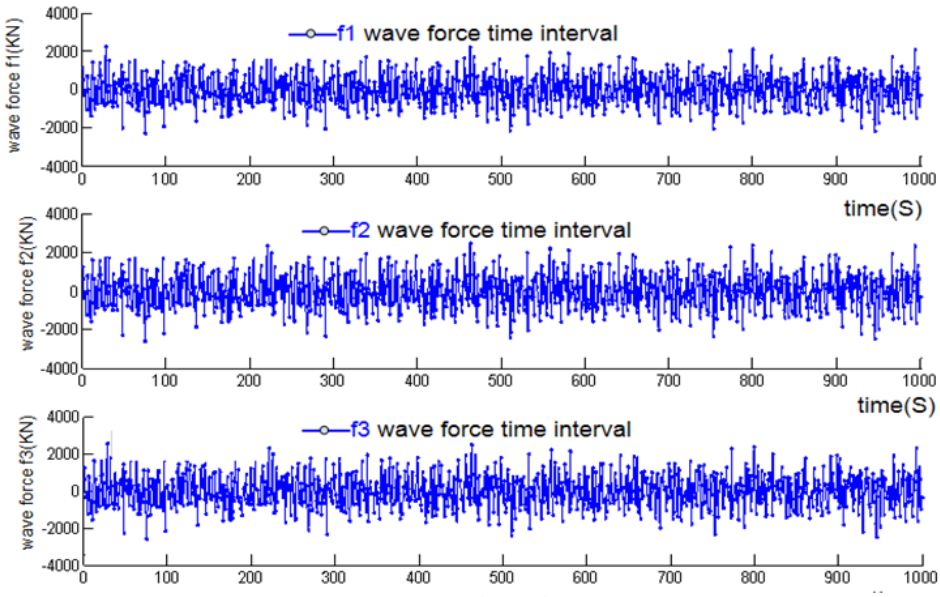


FIGURE 8 Time interval curve of wave force exerted on nodes

TABLE 5 Parameter identification list without noise (iteration 816 times, ε=0.000002)

Node	Damping identification value (kg/ms)	Damping error (%)	Stiffness identification value (N/m)	Stiffness error (%)
1	0.77876e+06	0.0039	0.23837e+06	0.0084
2	0.83995e+06	0.0036	0.25397e+06	0.0079
3	0.95034e+06	0.0063	0.30096e+06	0.0010

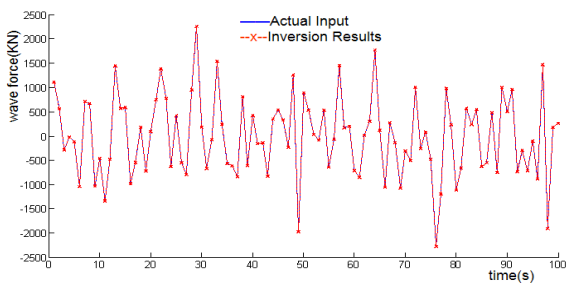


FIGURE 9 Comparison between wave load input and inversion time interval on node 1

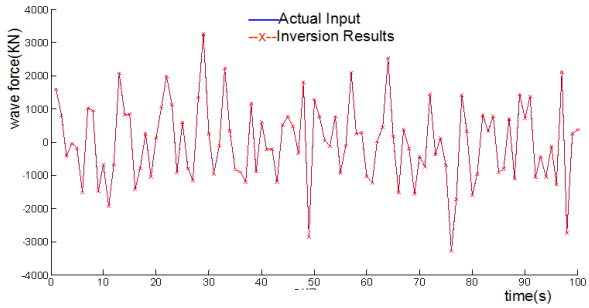


FIGURE 10 Comparison between wave load input and inversion time interval on node 2

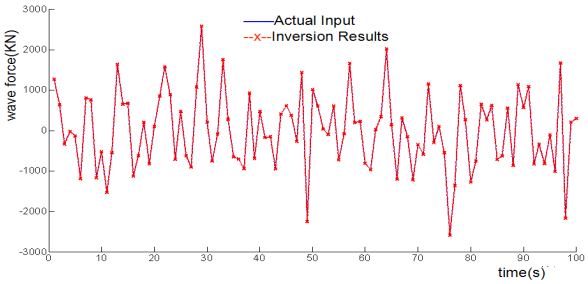


FIGURE 11 Comparison between wave load input and inversion time interval on node 3

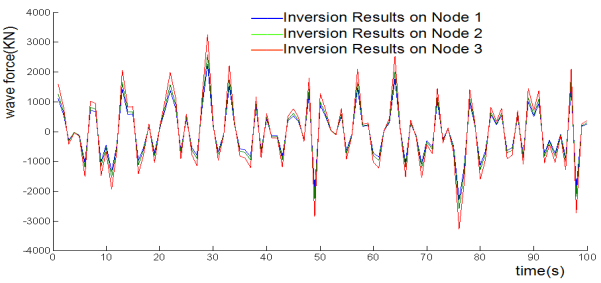


FIGURE 12 Comparison between wave load input and inversion time interval among three nodes

TABLE 6 Parameter identification list with 1% noise (iteration 823 times, ε=0.000002)

Node	Damping identification value (kg/ms)	Damping error (%)	Stiffness identification value (N/m)	Stiffness error (%)
1	0.77645e+06	0.3018	0.23867e+06	0.1133
2	0.84012e+06	0.0143	0.25128e+06	1.0709
3	0.98147e+06	3.2691	0.30019e+06	0.2691

TABLE 7 Parameter identification list with 6% noise (iteration 775 times, ε = 0.000002)

Node	Damping identification value (kg/ms)	Damping error (%)	Stiffness identification value (N/m)	Stiffness error (%)
1	0.7980e+06	2.4653	0.2370e+06	0.5872
2	0.7820e+06	6.9048	0.2685e+06	5.7087
3	1.0292e+06	8.2912	0.3083e+06	2.4252

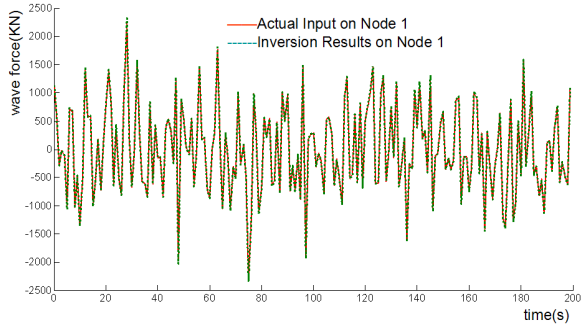


FIGURE 13 Comparison of actual input inversion results on Node 1 under 6% Noise

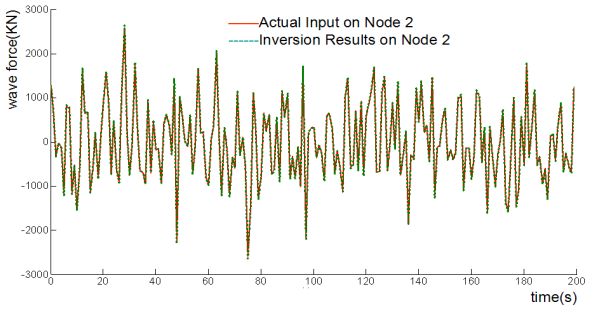


FIGURE 14 Comparison of actual input inversion results on Node 2 under 6% Noise

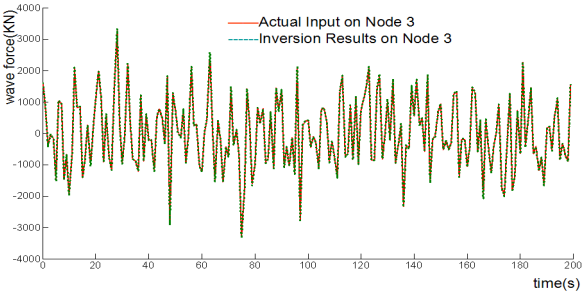


FIGURE 15 Comparison of actual input inversion results on node 3 under 6% noise

It is seen from Tables 2-7 and Figures 13-15 that under the condition of unknown input wave force being time domain related, method of this paper could not only identify physical parameter of pier but also accurately invert input load of each node. In terms of 6% of noise level, identified parameters are utilized to do inversion on the input wave forces on Node 1 and Node 2. Comparing with actually input wave force, the maximum error is 5.7087%. Obviously its inversion results are ideal.

5 Conclusions

Under condition of input wave load being known and unknown input wave load being time domain totally related, this paper does researches on physical parameter identification and load inversion of cantilever bridge pier structure. Suitable node units on piers are chosen to establish simplified computation model, in which characteristic parameters of damping and stiffness chosen from each node unit are taken as targets waiting for being identified. Numerical analysis is done utilizing numerical analysis software MALTAB through specific examples whose analysis conclusions are as follows:

Utilize total compensation method to do physical parameter identification and unknown wave force inversion on bridge piers under condition of part of unknown wave input. Its identification and inversion results present good stability.

In terms of circular piers without considering wave drag force, statistical average method of probability and improved total compensation method are combined together to do physical parameter identification and load inversion under unknown wave force input when wave force time intervals affected on each node are almost totally time domain related and correlation coefficient is α_i. For cantilever bridge pier structure, this algorithm keeps good applicability and accuracy on parameter identifications of damping and stiffness of each node under the influences of different noise signals when shear is out of shape.

Different initial values of physical parameters waiting for being identified keep little influence on identification result. Whether chosen initial values are near to the true values is just related to convergence time and convergence steps.

Without influences of noise, identification precision of physical parameter on each node of bridge pier structure is higher than that with noise existing whose reason is that there exists big difference between order of magnitude of stiffness parameter and the one of damping parameter. This leads identification result of stiffness parameter to be obviously lower than that of damping parameter. Small scope of noise influence keeps both being within the range of precision requirements.

This paper only takes researches of physical parameter identification and load inversion of cantilever bridge piers affected by wave load into consideration. It does not consider influences of bridge superstructure, cross-section shape

of pier, water velocity, wind load, etc. Necessary experiments and actual monitoring are still needed to test or correct the inversion calculation method in order to understand applicability and universality of the proposed method in this paper more clearly.

References

[1] Yu Y X 1981 Numerical simulation of the waves *Chinese Journal of Dalian Institute of science and Technology* **20**(3)84-90 *(in Chinese)*
[2] Wang S Q, Liang B C 2013 Wave Mechanics for Ocean Engineering, *China Ocean University Press Qingdao (in Chinese)*
[3] Li Y C, Teng B 2002 Wave Action on Maritime Structures *China Ocean Press* Beijing *(in Chinese)*
[4] Li J 1996 Analysis and Modeling of Stochastic Structural System *China Science Press* Beijing *(in Chinese)*
[5] Li J, Chen J 1998 Study of Structural System Identification with Incomplete Input Information *Earthquake Engineering and Engineering Vibration* **4**(18) 40-7 *(in Chinese)*
[6] Zhao T, Li J 2004 Research on parameter identification of structure with unknown top-floor load and wind load *Chinese Journal of Computational Mechanics* **2**(21) 202-8 *(in Chinese)*
[7] Xie X Z, Yi W J 2005 Updated Total Compensation Method for Solving Composite Inversion Problem of Structural *Dynamic System* **22**(1) 28-32 *(in Chinese)*

Authors

Zijian Wang, born in November, 1979, Chongqing City, P.R. China

Current position, grades: lecturer at the School of Zi-Jian Wang, Chongqing University of Science and Technology, China.
University studies: BE in Civil Engineering at Shenyang University in China. ME and DE in Civil Engineering at Chongqing Jiaotong University in China.
Scientific interest: civil engineering, numerical simulation.
Publications: more than 15 papers.
Experience: teaching experience of 2 years, 2 scientific research projects.

Liming Wu, born in December, 1981, Chongqing City, P.R. China

Current position, grades: lecturer at the School of Liming Wu, Chongqing Technology and Business Institute, China.
University studies: BE in Civil Engineering at Shenyang University in China.
Scientific interests: civil engineering, geotechnical engineering.
Publications: more than 12 papers.
Experience: teaching experience of 5 years, 3 scientific research projects.

Load spectrum for creep-fatigue life prediction of viewport used in human occupied vehicle

Yiting Kang[1, 2]*, Yali Feng[3], Jinggao Lin[4], Wenming Zhang[1]

[1]*School of Mechanical Engineering, University of Science and Technology Beijing, 30 Xueyuan Road, Beijing 100083, China*

[2]*CONCAVE Research Centre, Department of Mechanical & Industrial Engineering, Concordia University, Montreal, Canada*

[3]*School of Civil and Environmental Engineering, University of Science and Technology Beijing, 30 Xueyuan Road, Beijing 100083, China*

[4]*China Ocean Mineral Resources Research & Development Association, 1 Fuxingmenwai Street, Beijing 100860, China*

Received 6 September 2014, www.cmnt.lv

Abstract

Acrylic plastic, polymethylmethacrylate (PMMA), are widely employed as the material for the viewports of human occupied vehicle (HOV) which usually dives into deep sea. The service life of viewport is critical to the reliability and safety of HOV. In order to predict life of viewport in design stage, mathematical statistics method is applied to establish the load spectrum for viewport. It is found that ALVIN (America) HOV's dive-depth data is in a skewed distribution, and a piecewise function combined Gumbel and Weibull distributions is proposed for data fitting. HOV undertakes long-term and cyclic load in service, which will cause damage on viewport, so a creep-fatigue load spectrum is established and applied for JIAOLONG (China) HOV's viewport, which integrates both the dive-frequency of each depth range and the duration that maximum stress acts. The proposed method for determination of creep-fatigue load spectrum could thus be considered to be employed for failure analysis and life prediction of modern HOV's viewport.

Keywords: viewport, human occupied vehicle, distribution function, creep and fatigue failure, load spectrum

1 Introduction

Viewport is the main tool to observe and investigate underwater world for scientists and engineers occupied in a HOV. A crucial evaluation standard of a HOV is the number and vision of its viewport. The manned spherical shell of ALVIN HOV (America) has a main viewport and four auxiliary viewports, which can ensure a broad vision in deep sea [1].

Acrylic plastic, PMMA, is the choice material of underwater window rated for depths to 11000m [2]. It is widely employed as the viewport's material due to its optical property, corrosion-resistance, and great strength-weight ratio [3]. The viewport configurations mainly include flat disc, conical frustum, spherical sector, hemisphere, and hyper-hemisphere [4, 5]. HOV need dive to different depths during its service life to finish different scientific investigations, which generates alternating stress and creep strain on viewport due to cyclic and long-term external load [6]. So creep and fatigue are the main failure forms of HOV's viewport [7]. In order to predict its life, a load spectrum for both creep and fatigue analyses should be built at first.

PMMA is a typical viscoelastic material whose fatigue properties are strongly influenced by temperature and cyclic frequency. The creep of PMMA may occur when exposed to the long-term and high load. The creep behaviour of PMMA was theoretically and experimentally investigated by lots of researchers for the characterization, predictive models, and temperature effect [8-10]. Also its fatigue property is studied for the crack propagation, frequency effect, and life prediction [11-13]. Creep damage accumulates in every cycle with a long time, so the fatigue life is affected by creep. The models for life predicting considering both fatigue and creep effects have been proposed [14, 15]. These works are instructive to future research on life prediction of HOV's viewport. But they cannot be used directly due to the simple structure of PMMA specimen and load spectrum. A comprehensive analysis of creep and fatigue behaviours of PMMA viewport is still lacking. The first step for life prediction is to build a real load spectrum for the viewport.

The dive depth of a HOV for different scientific investigations is a random process about time, so the external load on viewport is also a group of random loads with time span. Mathematical statistics method is applied to analyse dive data to determine the load spectrum of viewport because of the random character of external load.

Professor Sasaki, a Japanese scholar, has made a data statistics on dive frequency of Yomiuri submersible whose maximum dive depth is 300m. Chen [16] has also analysed Yomiuri's data and found an obvious bimodal distribution of dive frequency and depth. Li [17] proposed that Gumbel distribution function can be used to determine fatigue load spectrum of spherical shell used in HOV. Although some studies have been made on load spectrums of submarine and submersible, there are still several problems to be solved. First, a more suitable distribution function should be propo-

***Corresponding author** e-mail: kangyiting@sina.com

sed for HOV, which works in deep sea. Second, a load spectrum concerning both time and frequency should be determined for HOV's viewport because its failure behaviour is not only fatigue but also creep.

In this paper, a load spectrum for life prediction of HOV's viewport is determined. Firstly, ALVIN's dive-depth data is collected for analysis and statistics at first. Moreover, Gumbel and Weibull distribution functions are applied for data statistics, and two fitting curves are compared to propose a more suitable function. Finally, the load spectrum for JIAOLONG HOV's viewport is determined based on the new function where both time and frequency are considered.

2 Collection of dive data

The maximum dive-depth of 7 HOVs are more than 4500m, namely MIR I/II and RUS (6000m, Russia), SHINKAI 6500(6500m, Japan), NAUTILE (6000m, France), ALVIN (4500m, America), and JIAOLONG (7000m, China).

The dive-depth data of ALVIN is selected to analyse (Table 1) because it is relatively more comprehensive compared to the others mentioned above. Only total numbers and maximum dive-depth of the others can be found, which is not enough to build a reliable load spectrum.

TABLE 1 Dive data of ALVIN HOV

Depth range (m)	Frequency	Proportion (%)	Accumulative proportion (%)
0-300	161	3.8	3.8
300-600	99	2.4	6.2
600-900	227	5.4	11.6
900-1200	141	3.4	15
1200-1500	202	4.8	19.8
1500-1800	169	4.0	23.8
1800-2100	377	9.0	32.8
2400-2700	1311	31.2	78.2
2700-3000	209	5.0	83.2
3000-3300	153	3.6	86.8
3300-3600	137	3.3	90.1
3600-3900	256	6.1	96.2
3900-4200	122	2.9	99.1
4200-4500	38	0.9	100

3 Data fitting of dive-depth

ALVIN's dive data is in an asymmetric and skewed distribution (Figure 1), so Weibull and Gumbel distribution functions are selected for data fitting. The fitting degrees of the two curves are compared to find out a proper function to establish load spectrum of ALVIN.

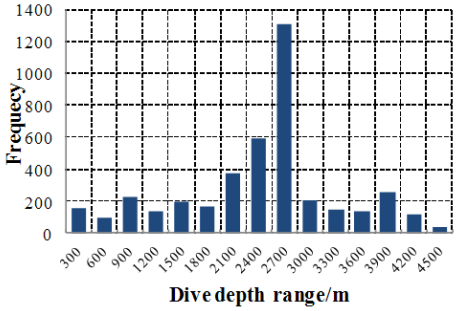


FIGURE 1 Histogram of ALVIN's dive data

3.1 WEIBULL DISTRIBUTION

The frequency distribution with different shapes can be described well by Weibull function with three parameters whose probability density and distribution functions respectively are:

$$f(x)=\frac{\beta}{\eta-\gamma}\left(\frac{x-\gamma}{\eta-\gamma}\right)^{\beta-1}\cdot\exp\left[-\left(\frac{x-\gamma}{\eta-\gamma}\right)^{\beta}\right], \quad (1)$$

$$F(x)=1-\exp\left[-\left(\frac{x-\gamma}{\eta-\gamma}\right)^{\beta}\right], \quad (2)$$

where γ is location parameter, β is shape parameter ($\beta > 0$), and η is scale parameter ($\eta > 0$).

It is obvious that location parameter γ of ALVIN's data distribution is 0 according to Figure 1, so only the shape and scale parameters, β and η, are required to be solved. Equations (1) and (2) can be simplified to

$$f(x)=\frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\eta}\right)^{\beta}\right], \quad (3)$$

$$F(x)=1-\exp\left[-(\frac{x}{\eta})^{\beta}\right]. \quad (4)$$

The shape and scale parameters could be estimated by graphic method. Firstly, Equation (4) should be transformed by taking the logarithm twice of both the left and right-part:

$$\ln\ln\left(\frac{1}{1-F(x)}\right)=\beta(\ln x-\ln\eta). \quad (5)$$

Assuming $Y=\ln\ln(1/(1-F(x)))$, $X=\ln x$ and $B=\beta(\ln\eta)$, Equation (5) can be expressed as:

$$Y=\beta X-B. \quad (6)$$

Obviously, Equation (6) is a line equation, so shape and scale parameters of Weibull distribution, β and η could be obtained from its gradient and intercept.

The confidence intervals of β and B under the confidence level of 1-α are respectively:

$$Q_{\beta}=[\beta-t_{1-\frac{\alpha}{2}}(n-2)\cdot\sigma_{\beta},\beta+t_{1-\frac{\alpha}{2}}(n-2)\cdot\sigma_{\beta}], \quad (7)$$

$$Q_{B}=[B-t_{1-\frac{\alpha}{2}}(n-2)\cdot\sigma_{B},B+t_{1-\frac{\alpha}{2}}(n-2)\cdot\sigma_{B}]. \quad (8)$$

3.2 GUMBEL DISTRIBUTION

Gumbel function is a theoretical model whose extreme value is in asymptotic distribution, and its probability density function and distribution function respectively are:

$$f(x)=\frac{1}{\varphi}\cdot\exp\left[\frac{x-\lambda}{\varphi}-\exp(\varphi)\right], \quad (9)$$

$$F(x)=1-\exp\left[-\exp\left(\frac{x-\lambda}{\varphi}\right)\right], \quad (10)$$

where φ is shape parameter (φ>0), and λ is scale parameter.

The shape and scale parameters of Gumbel distribution can also be obtained by transforming Equation (10) to:

$$\ln\ln\left(\frac{1}{1-F(x)}\right)=\frac{1}{\varphi}(x-\lambda). \quad (11)$$

Assuming $Z=\ln\ln(1/(1-F(x)))$, $C=1/\varphi$ and $D=\lambda/\varphi$, Equation (11) can be also transformed to a line equation:

$$Z=Cx+D. \quad (12)$$

So the shape and scale parameters of Gumbel distribution, φ and λ, could be solved from the gradients and intercept of Equation (12).

The confidence intervals of C and D under the confidence level of $1-\alpha$ are respectively:

$$Q_C=[C-t_{1-\frac{\alpha}{2}}(n-2)\cdot\sigma_C, C+t_{1-\frac{\alpha}{2}}(n-2)\cdot\sigma_C], \quad (13)$$

$$Q_D=[D-t_{1-\frac{\alpha}{2}}(n-2)\cdot\sigma_D, D+t_{1-\frac{\alpha}{2}}(n-2)\cdot\sigma_D]. \quad (14)$$

3.3 COMPARISION OF CURVES FITTED USING WEIBULL AND GUMBEL DISTRIBUTION

Curve of ALVIN's dive data fitted employing Weibull distribution is shown in Figure 2, which indicates that the fitting degree is much higher when the dive-depth is over 2700m. β and η are solved (β=3.1059, and η=2599.2), so the distribution function can be obtained

$$f(x)=\frac{3.1059}{2599.2}\left(\frac{x}{2599.2}\right)^{2.1059}\cdot\exp\left[-\left(\frac{x}{2599.2}\right)^{3.1059}\right], \quad (15)$$

$$F(x)=1-\exp\left[-\left(\frac{x}{2599.2}\right)^{3.1059}\right]. \quad (16)$$

The curve fitted by Gumbel distribution is also shown in Figure 2, which suggests it agrees well with the origin data when the dive depth is less than 2700m. φ and λ are solved (φ = 725.42, and λ = 2634.6), so the distribution function can be obtained:

$$f(x)=\frac{1}{725.42}\exp\left[\frac{x-2634.6}{725.42}-\exp\left(\frac{x-2634.6}{725.42}\right)\right], \quad (17)$$

$$F(x)=1-\exp\left[-\exp\left(\frac{x-2634.6}{725.42}\right)\right]. \quad (18)$$

Q_β and Q_B in Weilbull distribution under the confidence level of 80% are [1.2985, 1.7366] and [8.5725, 12.9942]. Q_C and Q_D in Gumbel distribution under the confidence level of 95% are [0.0009, 0.0011] and [-2.9553, -2.5156].

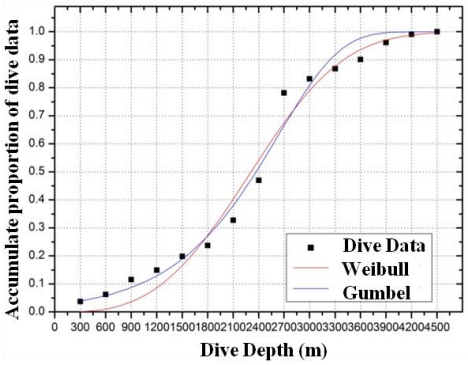


FIGURE 2 Comparison of two curves fitted by Weibull and Gumbel

When dive depth (x) is less than 2700m, fitting degree of the curve "Gumbel" is much better, while that of "Weibull" is superior when x is more than 2700m. So distribution function can be expressed with the combination of Gumbel and Weibull:

$$F(x)=\begin{cases}1-\exp\left[-\exp\left(\frac{x-2634.6}{725.42}\right)\right], x<2700\\ 1-\exp\left[-\left(\frac{x}{2599.2}\right)^{3.1059}\right], x\geq 2700\end{cases}. \quad (19)$$

Dive frequency (n) of ALVIN in the depth-range [d_1, d_2] can be solved by:

$$n=\begin{cases}N\left\{\exp\left[-\exp(\varsigma_1)\right]-\exp\left[-\exp(\varsigma_2)\right]\right\}, x<x_p\\ N\left\{\exp\left[-(\xi_1)^{3.1059}\right]-\exp\left[-(\xi_2)^{3.1059}\right]\right\}, x\geq x_p\end{cases}, \quad (20)$$

where N is the total dive frequency, $\varsigma_i=\frac{d_i-2634.6}{725.42}$, $\xi_1=\frac{d_i}{2599.2}$, $i=1,2$ and $x_p=2700$.

Result of n solved by Equation (20) is shown in Table 2, which clearly demonstrates Weibull and Gumbel are respectively applicable for data fitting in different depth-ranges. And it is necessary to develop a new function with the combination of Weibull and Gumbel.

TABLE 2 Dive data of ALVIN solved by Weibull and Gumbel functions

No.	Depth-range (m)	Frequency	Proportion (%)		
			Real	Weibull	Gumbel
1	0-300	161	3.8	0.1	3.9
2	300-600	99	2.4	0.9	1.9
3	600-900	227	5.4	2.6	2.9
4	900-1200	141	3.4	5.0	4.2
5	1200-1500	202	4.8	7.9	6.0
6	1500-1800	169	4.0	10.8	8.2
7	1800-2100	377	9.0	12.9	10.9
8	2100-2400	595	14.2	13.9	13.5
9	2400-2700	1311	31.2	13.4	15.0
10	2700-3000	209	5.0	11.5	14.4
11	3000-3300	153	3.6	8.7	10.9
12	3300-3600	137	3.3	5.9	5.9
13	3600-3900	256	6.1	3.5	1.9
14	3900-4200	122	2.9	1.8	0.3
15	4200-4500	38	0.9	0.8	0.0

The sum of squares of the difference between fitting results and real data are:

$$E_j = \sum_{i=1}^{15} (P_{ji} - P_i)^2 , \quad j = W, G, P , \tag{21}$$

where P_{ji} and P_i are respectively the fitting result and real value in No. i depth-range, and subscript j indicates the different fitting curves: W (Weibull distribution), G (Gumbel distribution), and P (combined both W and G). E_W, E_G and E_P are listed in Table 3. Values of E_W and E_G are almost equal, however that of E_P is much smaller comparing to E_W and E_G. The distribution equations combined both Weibull and Gumbel is more effective for data fitting of dive-depth.

TABLE 3 Sum of squares of difference between fitting result and real data

Parameter	E_W	E_G	E_P
Value	0.049733	0.046639	0.037577

4 Load spectrum for viewport of JIAOLONG HOV

Based on the above analysis of ALVIN's dive data, a piecewise function, which combines Gumbel and Weibull distributions is selected to determine the load spectrum of viewports used in JIAOLONG HOV for creep and fatigue analysis.

Maximum dive-depth of JIAOLONG is 7000m, which is greater than that of ALVIN, so the independent variable in Equation (11) should be transformed to get the new distribution function:

$$F(x) = \begin{cases} 1 - \exp\left[-\exp\left(\dfrac{4.5x/7 - 2634.6}{725.42} \right) \right], x < 4200 \\ 1 - \exp\left[-\left(\dfrac{4.5x/7}{2599.2} \right)^{3.1059} \right], x \geq 4200 \end{cases} , \tag{22}$$

Dive frequency n of JIAOLONG in the depth-range [d_1, d_2] can be calculated by:

$$n = \begin{cases} N\left\{ \exp\left[-\exp(\varsigma_1) \right] - \exp\left[-\exp(\varsigma_2) \right] \right\}, x < x_p \\ N\left\{ \exp\left[-(\xi_1)^{3.1059} \right] - \exp\left[-(\xi_2)^{3.1059} \right] \right\}, x \geq x_p \end{cases} , \tag{23}$$

where $\varsigma_i = \dfrac{4.5d_i/7 - 2634.6}{725.42}$, $i = 1, 2$, $\xi_i = \dfrac{4.5d_1/7}{2599.2}$, $x_p = 4200$, and d_1 and d_2 are between 0 and 7000m.

The data of JIAOLONG's viewport in every depth-range solved by Equation (23) are listed in Table 4.

The fatigue load spectrum of JIAOLONG when it finishes a dive to 7000m is shown in Figure 3.

When JIAOLONG HOV serves, both creep and fatigue damages may occur on its viewports because of the effect of alternating load and long-term load. Creep is a time dependent structural behaviour of acrylic. As a result, the duration of external load should be taken into account in the establishment of its load spectrum. In normal conditions, underwater time of HOV does not exceed 8 hours, and diving and floating speed are about 40m/min. So the duration that maximum stress acts is:

$$t = 8 - \frac{x}{1200} . \tag{24}$$

TABLE 4 Dive data of JIAOLONG HOV (a depth-range: 500m)

No.	**Depth-range (m)**	**Frequency**	**Proportion (%)**
1	0-500	5	3.9
2	500-1000	2	1.9
3	1000-1500	4	2.9
4	1500-2000	5	4.2
5	2000-2500	8	6
6	2500-3000	10	8.2
7	3000-3500	14	10.9
8	3500-4000	17	13.5
9	4000-4500	19	15
10	4500-5000	11	8.7
11	5000-5500	7	5.9
12	5500-6000	4	3.5
13	6000-6500	2	1.8
14	6500-7000	1	0.8

Load spectrum considering both the dive frequency and duration that maximum stress acts for creep and fatigue analysis and life prediction of JIAOLONG HOV's viewports is illustrated in Figure 4 and Table 5.

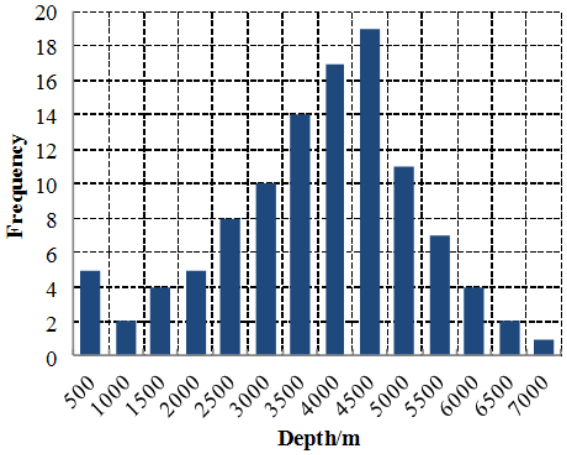


FIGURE 3 Fatigue load spectrum of JIAOLONG HOV

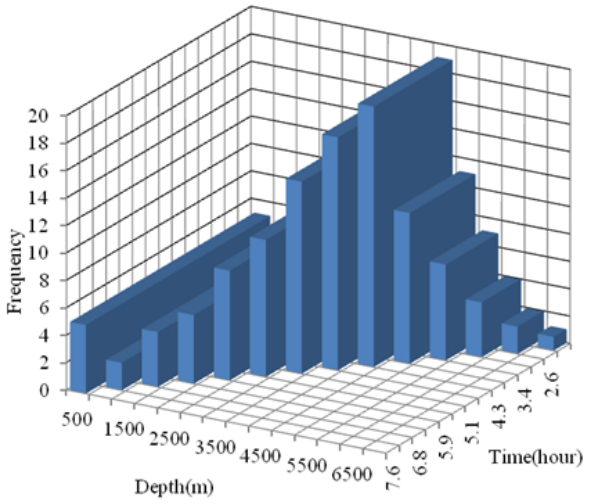


FIGURE 4 Creep and fatigue load spectrum of JIAOLONG HOV

TABLE 5 Data in creep-fatigue load spectrum of JIAOLONG HOV

Depth-range (m)	Frequency	Proportion (%)	Duration (h)
0-500	5	3.9	7.6
500-1000	2	1.9	7.2
1000-1500	4	2.9	6.8
1500-2000	5	4.2	6.3
2000-2500	8	6	5.9
2500-3000	10	8.2	5.5
3000-3500	14	10.9	5.1
3500-4000	17	13.5	4.7
4000-4500	19	15	4.3
4500-5000	11	8.7	3.8
5000-5500	7	5.9	3.4
5500-6000	4	3.5	3.0
6000-6500	2	1.8	2.6
6500-7000	1	0.8	2.2

5 Conclusions

In order to determine a load spectrum for failure analysis of viewport used in JIAOLONG HOV, dive data of ALVIN HOV (America) is collected for analysis and statistics due to its completeness and comparability. It has been found that the dive data has a skewed distribution feature, so both Weibull and Gumbel distributions are selected and compared for ALVIN's data fitting. The results indicate that Weibull and Gumbel function are respectively suitable for fitting data of different depth-ranges, so a piecewise distribution function is proposed to determine the load spectrum with the combination of two distribution functions. Both creep and fatigue are the main failure forms for HOV's viewport made by PMMA, so a load spectrum concerning dive time and frequency, which can be used for, creep and fatigue analysis is determined for JIAOLONG's viewport at last.

Acknowledgements

The authors gratefully acknowledge the support of National High-tech R&D Program (863 Program, 2011AA09A101).

References

[1] Lin J, Feng Y, Zhang W, Kang Y 2013 Review on Viewport Used in Human Occupied Vehicle *Ship Engineering* **35**(3) 1-5 *(in Chinese)*
[2] Stachiw J D 1970 Conical Acrylic Windows under Long Term Hydrostatic Pressure of 20,000 psi *Journal of Engineering for Industry* **92**(1) 237-56
[3] Stachiw J D 2004 Acrylic plastic as structural material for underwater vehicles *International Symposium on Underwater Technology* 289-96
[4] American National Standard American Society of Mechanical Engineers *Safety Standard for Pressure Vessels for Human Occupancy* PVHO-l 1977 (www.asme.org, Codes and Standards)
[5] Stachiw J D 2003 Handbook on Acrylics for Submersibles, Hyperbaric Chambers, and Aquaria *Best Publishing Company: Flagstaff*
[6] Stachiw J D 1976 Spherical Shell Sector Acrylic Plastic Windows with 12,000 ft Operational Depth for Submersible ALVIN *Journal of Engineering for Industry* **98**(2) 523-36
[7] Lin J, Zhang W, Feng Y, and Kang Y 2013 Displacement and Stress Analysis of Viewport Used in Human Occupied Vehicle Based on Rayleigh-Ritz *Journal of Ship Mechanics* **17**(6) 635-44
[8] Arnold J C and White V E 1995 Predictive models for the creep behaviour of PMMA *Material Science and Engineering* **197**(2) 251-60
[9] Gao Z Z, Liu W, Liu Z Q, and Yue, Z F 2010 Experiment and Simulation Study on the Creep Behavior of PMMA at Different Temperatures *Poly-Plastics Technology and Engineering* **49**(14) 1478-82
[10] Chen K, and Hsu R 2007 Evaluation of environmental effects on mechanical properties and characterization of creep behavior of PMMA *Journal of the Chinese Institute of Engineers* **30**(2) 267-74
[11] Carnelli D, Villa T, Gastaldi D, and Pennati G 2011 Predicting fatigue life of a PMMA based knee spacer using a multiaxial fatigue criterion *Journal of Applied Biomaterials and Biomechanics* **9**(3) 185-92
[12] Hoey D, and Taylor D 2009 Comparison of the fatigue behaviour of two different forms of PMMA *Fatigue and Fracture of Engineering Materials and Structures* **32**(3) 261-9
[13] Huang A, Yao W, and Chen F 2014 Analysis of Fatigue Life of PMMA at Different Frequencies Based on a New Damage Mechanics Model *Mathematical Problems in Engineering* Article ID 352676 1-8
[14] Hu Y, Summers J, Hiltner A, Baer E 2003 Correlation of fatigue and creep crack growth in poly(vinyl chloride) *Journal of Materials Science* **38**(4) 633-42
[15] Liu W, and Yang X J 2012 Damage evolution with growing cyclic creep and life prediction of MDYB-3 PMMA *Fatigue and Fracture of Engineering Materials and Structures* **36**(6) 483-91
[16] Chen X 1994 Low-cycle fatigue of submarine and submersible *National Defence Industry Press: Beijing (in Chinese)*
[17] Li X, Liu T, Huang X, and Cui W 2004 Determination of fatigue load spectrum for pressure hull of a deep manned submersible *Journal of Ship Mechanics* **8**(1) 59-70

Authors

Yiting Kang, born in September, 1986, Anhui Province, P.R. China

Current position, grades: postdoctoral Fellow at the CONCAVE Research Centre, Department of Mechanical & Industrial Engineering, Concordia University, Canada.
University studies: Doctoral Degree in Vehicle Engineering from University of Science and Technology Beijing in China.
Scientific interests: vehicle dynamics and kinematics of vehicle suspension.
Publications: more than 10 papers.
Experience: 3 scientific research projects.

Yangli Feng, born in September, 1967, Beijing, P.R. China

Current position, grades: professor at the School of Civil and Environmental Engineering, University of Science and Technology Beijing, China.
University studies: Doctoral Degree Vehicle Engineering from University of Science and Technology Beijing in China.
Scientific interests: equipment for mining resource development, optimization and control of mineral processing, and mineral microorganism technology.
Publications: more than 100 papers.
Experience: teaching experience of 24 years, more than 10 scientific research projects.

Jinggao Lin, born in April, 1986, Liaoning Province, P.R. China

Current position, grades: engineer at the China Ocean Mineral Resources Research & Development Association.
University studies: doctoral degree Vehicle Engineering from University of Science and Technology Beijing in China.
Scientific interest: creep and fatigue analyses of HOV's viewports.
Publications: 5 papers.
Experience: in 3 scientific research projects.

Wenming Zhang, born in August, 1955, Beijing, P.R. China

Current position, grades: professor at the School of Mechanical Engineering, University of Science and Technology Beijing, China.
University studies: Master Degree in Vehicle Engineering from University of Science and Technology Beijing in China.
Scientific interest: design and development of heavy duty mining truck and handling and ride performance analyses of off-highway vehicle.
Publications: more than 100 papers.
Experience: teaching experience of 30 years, more than 10 scientific research projects.

Application of GMM-UBM with an embedded AANN in the acoustic emission signal recognition

Cheng Xinmin[1]*, Li Jing[2], Jiang Yunliang[2], Gao Ge[2]

[1]*School of Information Engineering, Huzhou Teachers' College, Huzhou 313000, Zhejiang, PR China*

[2]*School of Information Science and Engineering, Southeast University, Nanjing 210096, Jiangsu, PR China*

Received 12 June 2014, www.cmnt.lv

Abstract

In this paper, we propose to recognize the Acoustic Emission (AE) signal, by using a Gaussian Mixture Model-Universal Background Model (GMM-UBW) with an embedded AANN. The AANN based GMM-UBW combines the learning ability of neural network and strong distribution capabilities of GMM. And, it trains the model parameters alternatively in order to approach the maximum likelihood. For illustrating the effectiveness of the proposed recognition method for the AE signal, an experiment is conducted. In the experiment, three cases of AE signal are considered, namely with no rub impact, slight rub impact and serious rub impact. The experimental results reveal that the AANN based GMM-UBW outperforms the GMM, with respect to the recognition rate, for any case of AE signal. For the case of slight rub impact, the GMM-UBW and that with embedded AANN both have the worst recognition performance, among the three cases. And, the proposed method has the biggest improvement for this case.

Keywords: acoustic emission; Auto-Associative Neural Network; Gaussian Mixture-Universal Background Model

1 Introduction

The acoustic signal processing has always been a hot topic in the world [1-4]. The acoustic emission is a typical application. The rub-impact between rotor and stator is a common fault in rotating machines [5]. Effective recognition of the rub-impact plays an important role in safe and stable running of the rotating machines [5]. The AE recognition can realize the fault recognition of the AE source by using the AE signal [6-9]. And, it is a non-destructive detection method [6-9]. By employing the tool of pattern recognition [10], the AE signal can be properly classified. The AE recognition can realize the effective fault recognition and the evaluation of the safety performance of the correspondding material.

Gaussian Mixture Model-Universal Background Model (GMM-UBM) plays an important part in the field of pattern recognition (e.g., speech recognition) [11-16]. It can approach the AE data correctly in AE recognition and decrease the total difference between various signals. The artificial neural network (ANN) can simulate the human brains' structure and faculty using parallel processing technology [10]. It can approximate almost any non-linear mapping. And thus, in the AE recognition, it plays an important role where the model construction cannot be reached and in the comprehensive recognition. ANN is an effective tool in analysing the AE signal [10]. Auto-Associative Neural Network (AANN) possesses a special symmetric structure, and is proven to be a successful algorithm in signal detection [15-16].

In order to improve the performance of the AE signal recognition, the AANN is embedded into the GMM-UBM. Firstly, the feature vectors are input into GMM-UBM after being processed by AANN transfer. The output results are used to modify the AANN model through error back propagation. As two models are trained alternatively, the maximum likelihood can be approximated and a superior recognition rate can be achieved.

2 AE recognition based on GMM-UBM with an embedded AANN

2.1 APPLICATION OF AANN IN THE AE RECOGNITION

AANN, proposed by Kramer in 1990s, is a special forward neural network, where the identity mapping is adopted. The topology structure of the AANN is illustrated in the following Figure 1.

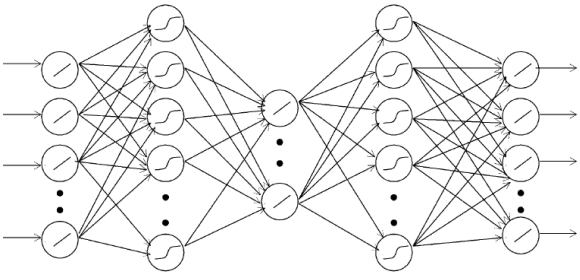


FIGURE 1 Auto-Associative Neural Network

The topology space of the AANN is set through associative learning. Then, the data coding and compression are implemented in the network bottleneck. Afterwards, the characteristic information of the process is obtained and the

* ***Corresponding author*** e-mail: chengxinmin@hutc.zj.cn

noise is filtered. Finally, the output (approximately) identical to the input is obtained. After the model training, the signal with noise being cancelled can be obtained by inputting the original signal with noise into the model. In the compression process, the characteristic dimension is reduced and the signal in higher-dimension space is projected into lower-dimension space. The linear combination of the principle components is utilized to represent the signal's high-order statistical property. And thus, the AANN can be taken as a non-linear principal component analysis (PCA) model.

In the field of pattern recognition, the data in the feature space generally has a complex distribution. However, as for the GMM, the data in the feature space are represented by statistical quantities of first and second orders and the mixture weights, and number of mixture is generally confined. So, the AANN, which is a non-linear model possessing multi-layer perceptron, is considered to be a practical model suitable for processing complex sound signal. It can obtain the error curve of the feature space, which matches the distribution of the training data. The approximation ability of the network to data can be observed, by training the error curve.

In the following, we will deduce the rule for training the model parameters of the AANN model. The basic idea of the training rule of the neural network is gradient descent method. Let $\vec{w}$, D and d be the weights, the training set and an instance in the set, respectively. The error E is defined as follows:

$$E(\vec{w}) = \frac{1}{2}\sum_{d\in D}(t_d - o_d)^2 , \tag{1}$$

where t_d and o_d are the expected and true outputs. Starting from some set of initial coefficients, the gradient descent method can be utilized to modify the coefficients of the model.

$$\Delta w_i = -\eta\frac{\partial E}{\partial w_i} = \eta\sum_{d\in D}(t_d - o_d)x_{id} . \tag{2}$$

For each layer of a neural network, the weights of the input vector $net_j = \sum w_{ji}x_{ji}$ is transformed through the sigmoid function. The training process from i th hidden layer to the j th output layer is illustrated as follows:

$$\begin{aligned}\frac{\partial E_d}{\partial w_{ji}} &= \frac{\partial E_d}{\partial net_j}\cdot\frac{\partial net_j}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j}\cdot x_{ji} \\ &= \frac{\partial E_d}{\partial o_j}\cdot\frac{\partial o_j}{\partial net_j}\cdot x_{ji} = -(t_j - o_j)\cdot o_j(1-o_j)\cdot x_{ji}\end{aligned} . \tag{3}$$

Thus, the weight update quantity is:

$$\Delta w_{ji} = -\eta\frac{\partial E_d}{\partial w_{ji}} = \eta(t_j - o_j)\cdot o_j(1-o_j)\cdot x_{ji} = \eta\delta_j x_{ji} , \tag{4}$$

where $\delta_j = -\dfrac{\partial E_d}{\partial net_j} = (t_j - o_j)\cdot o_j(1-o_j)$ is the error term.

Similarly, for the k th output layer, the training process from the i th input layer to the j th hidden layer is illustrated as follows:

$$\begin{aligned}\frac{\partial E_d}{\partial net_j} &= \sum\frac{\partial E_d}{\partial net_k}\cdot\frac{\partial net_k}{\partial net_j} \\ &= \sum(-\delta_k)\frac{\partial net_k}{\partial o_j}\cdot\frac{\partial o_j}{\partial net_j} \\ &= \sum(-\delta_k)w_{kj}\cdot o_j(1-o_j) = \eta\delta_j x_{ji}\end{aligned} . \tag{5}$$

Thus, the weight update quantity is

$$\Delta w_{ji} = -\eta\frac{\partial E_d}{\partial w_{ji}} = \sum\eta\delta_k w_{kj}\cdot o_j(1-o_j)x_{ji} = \eta\delta_j x_{ji} , \tag{6}$$

where $\delta_j = -\dfrac{\partial E_d}{\partial net_j} = \sum\delta_k w_{kj}\cdot o_j(1-o_j)$ is the error term.

Summing up the above deduction, the weight update process of the AANN is provided as follows:

Step 1. Set the numbers of layers and nodes for each layer, and the initial weight coefficients.

Step 2. For the output o_k of the training instance $(k \quad t_k)$, calculate the error term from hidden layer to output layer $\delta_k = -(t_k - o_k)\cdot o_k(1-o_k)$.

Step 3. Calculate the weight update quantity $\Delta w_{kj} = \eta\delta_k x_{kj}$ from hidden layer to output layer and update the weight value w_{kj}.

Step 4. Calculate the error term from input layer to hidden layer $\delta_j = \sum\delta_k w_{kj}\cdot o_j(1-o_j)$ according to δ_k.

Step 5. Calculate the weight update quantity from input layer to hidden layer $\Delta w_{ji} = \eta\delta_j x_{ji}$ and update w_{ji}.

Step 6. If the convergence conditions are satisfied, then terminate the iteration. Otherwise, go to step 2.

2.2 IMPLEMENT GMM-UBM IN THE AE RECOGNITION

GMM is a linear combination of multiple Gaussian probability density functions. Any continuous distribution can be approximated with desired precision by using GMM, as long as the number of components in the model is large enough. So, GMM has a wide range of applications, such as speech recognition, image denoising, detection of chemical machining process and mechanical fault diagnosis. For a D dimensional feature vector x_t, the M th order probability density function of GMM is given as follows:

$$p(x/\lambda) = \sum_{i=1}^{M} w_i \cdot p_i(x) . \tag{7}$$

Gaussian Mixed density can be expressed by the mean value, covariance matrix, and mixed weight parameterization of the total members' density as follows: $\lambda_i = \{ p_i,\ u_i,\ \sum_i \},\ i = 1,\ 2,\ 3,\ ...,\ M$.

The covariance matrix can be reduced to a diagonal matrix if the individual components are assumed to be independent. Then the D dimensional Gaussian probability density function can be represented as follows:

$$p_i(x) = \frac{1}{(2\pi)^{D/2} \left|\sum_i\right|^{1/2}} \exp\left\{-\frac{1}{2}(x-u_i)' \sum_i^{-1} (x-u_i)\right\}. \quad (8)$$

After obtaining the training vector, the iterative expectation maximization (EM) algorithm can be employed to estimate the maximum likelihood model.

$$p(i \mid x_t, \lambda) = \frac{p_i b_i(x_t)}{\sum_{k=1}^{M} p_k b_k(x_t)}. \quad (9)$$

Iteratively calculate the weight coefficients, mean value and variance according to the following equations.

$$\overline{p_i} = \frac{1}{N} \sum_{t=1}^{N} p(i \mid x_t, \lambda), \quad (10)$$

$$\overline{u_i} = \frac{\sum_{t=1}^{N} p(i \mid x_t, \lambda) x_t}{\sum_{t=1}^{N} p(i \mid x_t, \lambda)}, \quad (11)$$

$$\overline{\sigma_i^2} = \frac{\sum_{t=1}^{N} p(i \mid x_t, \lambda) x_t^2}{\sum_{t=1}^{N} p(i \mid x_t, \lambda)}. \quad (12)$$

In the pattern recognition of GMM-UBW, verification vectors are input into the AE model and background model, in the test, and the ratio of the two models is calculated. If the ratio is higher than the prescribed threshold, the category of this AE source model is accepted. Otherwise, it is rejected. In other words, the background model provides the model for all the training instances, and sums up the model trend of all the input instances. It works as the denominator in the recognition, and has a normalizing effect on the total test voice scores [3]. Compared with GMM, GMM-UBM reduces the effect of rating difference between different voice signals, to a great extent. And thus, it has superior recognition performance. Moreover, because the AE training data is not enough, the number of GMM's mixture components cannot be set big enough. However, this number should be large enough in order to improve the recognition. The GMM-UBW trained through enough speech data can reduce the limitedness of the model's mixture component number.

GMM-UBW was originally applied to the text-independent speaker identification system. And, it also has applications, combined with support vector machine (SVM) theory, in emotional speech recognition and language recognition.

2.3 GMM-UBM STRUCTURE EMBEDDED WITH AANN

According to the error curve of AANN, a probability distribution matching the given data can be obtained. The distribution relates to the network cell and a kind of gain coefficient. The data distribution is not considered for this AANN. The error curve can only be utilized to verify the training ability of the network model. It, however, cannot provide information for signal construction in mathematical model. The GMM cannot transform the data, change its distribution. Further, the model itself has some limitedness and the matching precision cannot reach the desired requirement, due to the assumption of independence.

In view of the GMM-UBW's strong ability of data distribution and the AANN's excellent learning ability, these two kinds of pattern recognition methods are combined together. And, we proposed to recognize the AE signal by using a GMM-UBW with a AANN being embedded.

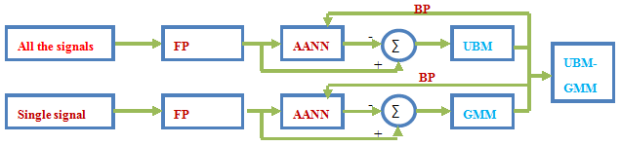


FIGURE 2 GMM-UBM embedded with AANN topping structure

As is illustrated in Figure 2, the training structure is composed of UBM and GMM. And the final output is GMM-UBW model. After extracting the characteristic parameters of the training sound signal, the feature vector is transformed through AANN. The difference between the output of the transformation and the original feature vector is input into the hybrid model. The modified output of the hybrid-model feedbacks the AANN and re-modify the AANN. The above process is repeated until the convergence conditions are satisfied. As for the UBW, the characteristic parameters of all the sound signal are input while the characteristic parameters of GMM are input one by one.

The specific training process is given as follows:

Step 1. Set the values of initial parameters and the convergence conditions of AANN and GMM.

Step 2. Input the characteristic parameters and obtain the error vector.

Step 3. Update the model parameters p_i, u_i and σ_i^2 of GMM while fixing the model parameters of AANN.

Step 4. Input the error vector into the modified GMM and obtain the likelihood probability. Transmit the error to the AANN before backwards and update its model parameter w_i.

Step 5. If the convergence conditions are satisfied, then terminate the iteration. Otherwise, turn to step 2 and continue the above process.

In the above process, the training processes for UBM and GMM are almost the same. The only difference exists in that the parameters of GMM, which is trained after UBM, are optimized based on UBM. This can save a certain amount of training time.

Construct AANN-GMM for each category (the existence of rub-impact or not). After the model training is complete, the recognition of the tested AE signal can begin. Then, input the feature vector into the AANN-GMM and obtain the likelihood probability of the categories. Finally, the recognition results are obtained.

3 Experiments and Discussions

The experimental data in this paper comes from the rotor system equipped with a rub impact bracket. The experimental equipment is illustrated in Figure 3. The rotor rub impact signal is produced by adjusting the nut bolt. The sampling frequency is 1MHz. The feature parameter method is adopted in the experiment. The feature of the AE signal is represented by seven simplified wave feature parameters. These seven parameters include the average vibration amplitude of rub impact signal, the maximum amplitude, the dynamic range of vibration amplitude and the reconstructed signal's energy of the first four contact points of the decomposed signal through wavelet package. The AE signals of no rub impact, slight rub impact and serious rub impact are provided in Figures 4, 5 and 6, respectively.

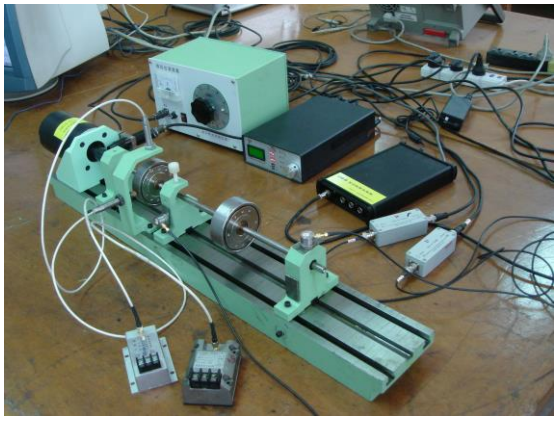


FIGURE 3 Rotor system

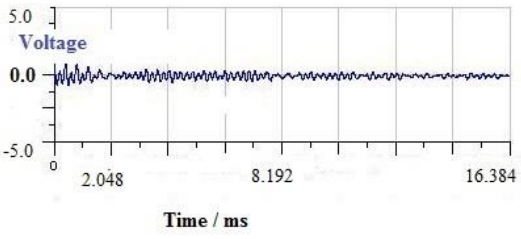


FIGURE 4 AE signal with no rub impact

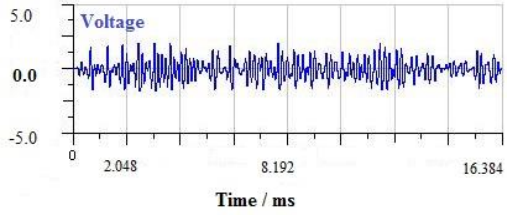


FIGURE 5 AE signal with slight rub impact

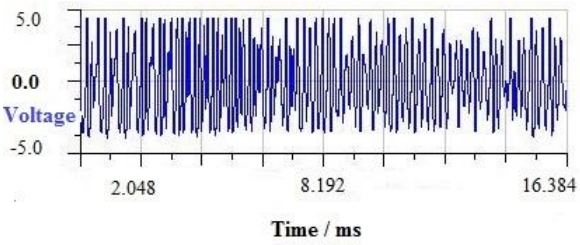


FIGURE 6 AE signal with serious rub impact

The performance of the GMM-UBW and that of the GMM-UBW embedded with AANN are analysed by using the same training data and the same test data. By adjusting the nut bolt, the three cases of AE signal, namely no rub-impact, slight rub-impact and serious rub-impact (all with respect to the rotating axis), are produced. The recognition performance of the above two models are evaluated for the above three cases of AE signal. The recognition rate comparison of the two models is provided in the following Figure 7. And, the recognition improvement of GMM-UBW with embedded AANN, with respect to GMM-UBW, is provided in Figure 8.

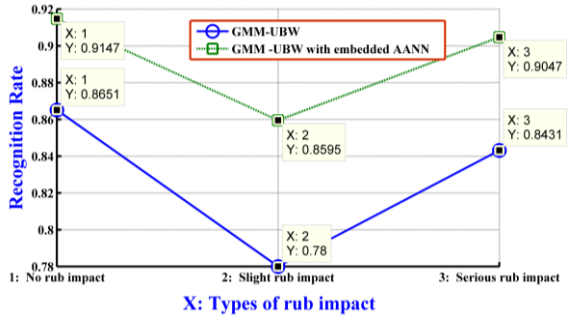


FIGURE 7 Recognition comparison between UBM-GMM and UBM-GMM embedded with AANN

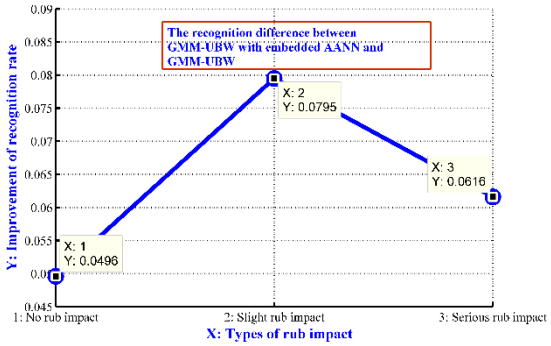


FIGURE 8 Recognition improvement of UBM-GMM embedded with AANN, with respect to UBM-GMM

The experimental results reveal that the GMM-UBW embedded with AANN outperforms the GMM-UBW for the three cases of AE signal separately. Specifically, the recognition rates are improved by 5.75%, 10.10% and 7.76% for the three cases (i.e., no rub impact, slight rub impact and serious rub impact), respectively. The results also reveal that for the case of slight rub impact, the GMM-UBW and that with embedded AANN both have the worst recognition performance, among the three cases. And, the proposed method

has the biggest improvement for this case. This indicates that the two models are not sensitive to the sound caused by the AE source. So, further work is required to extract more suitable feature vector, to improve the model structure and optimization algorithm and to better the fault detection efficiency of the AE method.

4 Conclusions

The GMM-UBW with an embedded AANN is proposed to recognize the AE signal and its recognition performance is studied in this paper. The proposed recognition method exploits the superior learning ability of the AANN to train the parameters of the whole model, by means of a two-stage training method. In the training process, the parameters of the GMM-UBM and the AANN are iteratively improved in order to obtain a superior likelihood probability. The experimental results reveal that the recognition rate of the rub-impact can be improved by using the proposed method.

Acknowledgments

The authors would like to thank the anonymous referees for constructive recommendations. This research was supported by the Zhejiang Provincial Natural Science Foundation of China under Grant No. LY14F010010, the National Natural Science Foundation of China under Grant No.61370173, the Zhejiang Provincial Natural Science Foundation of China under Grant No.R1090244, the Zhejiang Provincial Public Technology Applied Research Projects of China under Grant No. 2014C31084, the Key Innovative Research Team of HuZhou: Internet of Things - Technology & Application under Grant No. 2012KC04.

References

[1] Ruiyu Liang J X, Jian Zhou, Cairong Zou, Li Zhao 2013 An improved method to enhance high-frequency speech intelligibility in noise *Applied Acoustics* **74**(1) 71-8
[2] Liang Rui-Yu, Xi Ji, Zhao Li, Zou Cai-rong, Huang Cheng-wei 2012 Experimental study and improvement of frequency lowering algorithm in Chinese digital hearing aids *Acta Physica Sinica* **61**(13) 134305(1-11)
[3] Wang Q, Zhao L, Qiao J, Zou C 2010 Acoustic feedback cancellation based on weighted adaptive projection subgradient method in hearing aids *Signal Processing* **90**(1) 69-79
[4] Liang R, Zou C, Zhao L, Wang Q, Xi J 2012 Experimental study on enhancement method for high-frequency hearing loss in Chinese digital hearing aids *Acta Acustica* **37**(5) 527-33
[5] Wang S B, Chen X F, Li G Y, et al 2014 Matching Demodulation Transform With Application to Feature Extraction of Rotor Rub-Impact Fault *IEEE Transactions on Instrumentation and Measurement* **63**(5) 1371-83
[6] Ernst R, Dual J 2014 Acoustic Emission Localization in Beams Based on Time Reversed Dispersion *Ultrasonics* **54**(6) 1522-33
[7] El-Alej M E, Corsar M, Mba D 2014 Monitoring the presence of water and water-sand droplets in a horizontal pipe with Acoustic Emission technology *Applied Acoustics* **82** 38-44
[8] Zarouchas D, van Hemelrijck D 2014 Mechanical characterization and damage assessment of thick adhesives for wind turbine blades using acoustic emission and digital image correlation techniques *Journal of Adhesion Science And Technology* **28**(14-15) 1500-16
[9] Ruoyu Li, David He 2012 Rotational Machine Health Monitoring and Fault Detection Using EMD-Based Acoustic Emission Feature Quantification *IEEE Transactions on Instrumentation and Measurement* **61**(4) 990-1001
[10] Ruoxiang Yi, Shifeng Liu, Rongsheng Geng Application of Artificial Neural Network to Acoustic Emission Testing *Nondestructive Testing* **24**(11) 488-96
[11] Qiuwen Wang 2011 *Rapid Speaker Recognition Based on GMM-UBW* Harbin Institute of Technology, Master Thesis
[12] Dan Qu, Binxi Wang, Xin Wei 2003 Automatic Language Identification based on GMM-UBW *Signal Processing* **19**(1) 85-8 *(In Chinese)*
[13] Yongming Huang, Guobao Zhang, Fei Dong 2011 Speech Emotion Recognition based on Two Kinds of GMM-UBW Multidimensional likelihoods and SVM *Application Research of Computers* **28**(1) 98-101
[14] Gang Yu, Changning Li, Jun Sun et al 2010 Machine Fault Diagnosis based on Gaussian Mixture Model and Its Application *The International Journal of Advanced Manufacturing Technology* **48**(1/4) 205-12
[15] Cunbao Chen, Li Zhao 2010 Speaker Verification Based on GMM-UBM with Embedded Auto-associate Neural Network *Journal of Applied Sciences* **28**(1) 38-43
[16] Yegnanarayana B, Kishore S P 2002 AANN: an alternative to GMM for pattern recognition *Neural Networks* **15**(3) 459-69

Authors

Cheng Xinmin

Current position, grades: Master, the Associate Professor of Huzhou Teachers' College
Scientific interest: Intelligent signal processing, Information fusion
Experience: an expert in the field of information fusion

Li Jing

Current position, grades: Master, pursuing the Ph.D. degree in AE signal processing at southeast University
University studies: She received the B.A. degree in electronic information from Hebei University of Engineering, Handan, China, in 2001 and 2005, and the M.Sc. degrees in communication from China University of Mining & Technology, Xuzhou, China, in 2005 and 2008.
Scientific interest: AE signal processing

Jiang Yunliang

Current position, grades: Doctor, the Professor of Huzhou Teachers' College
Scientific interest: GIS, Information fusion
Experience: an expert in the field of GIS and information fusion

Gao Ge

Current position, grades: Master, the Associate Professor of Yancheng Institute of Technology
University studies: She received the B.A. degree information engineering from Southeast University, Nanjing, China, in 2012.
Scientific interest: AE signal processing

Research on method to promote performance of multiple injections diesel engine

Xiaorong Wang[1]*, Chenpeng Liu, Bo Liu[2], Wenxian Tang[1], Jian Zhang[1]

[1]*Institute of mechanical engineering, JiangSu University of Science and Technology, Zhenjiang 212003, China*

[2]*Hudong Heavy Machinery Co., Ltd, Shanghai 200129, China*

Received 12 June 2014, www.cmnt.lv

Abstract

Based on certain type of two-stroke, low-speed diesel engine for marine propulsion, we, in our research project, developed a 3-D model of in-cylinder combustion processes in a diesel engine. Thus, we were able to compare simulated in-cylinder pressure against the experimental pressure under single injection with a maximum error tolerance of 0.36%, the results of which imply that the two are basically consistent. Subsequently, based on the same injection timing, a comparison was made between the performance of a diesel engine using single injection and one using dual injection. The results indicate that the additional injection can lower NOx by58.9% and soot by 30.1%. Moreover, the piston's workload is reduced by 10.1%. Through calibrating injection quality centre, adjusting injection timing of secondary injection reasonably, we gained the method improving performance of combustion and secondary injection diesel engine. By verifying, the method can promote working ability of secondary injection diesel engine effectively. After adjusting injection timing, working ability of secondary injection is as well as it of single injection, but emissions reduce sharply, thereinto, NOx reduced 12.7%, soot reduced 41.1%.

Keywords: Marine low-speed two-stroke diesel engine; injection timing; double injection; diesel performance

1 Introduction

In recent years, the government formulated the worldwide emissions regulations, Tier II regulations had been put into force in 2011, the regulations limits the NOx emission of Marine diesel engine under 14.36 g/kw*h, it fell about 20% than Tier I, and Tier III regulations will be carried out in 2016, when the NOx emissions will be reduced to less than 3.4 g/kw*h, and it drop nearly 80% than Tier I. Diesel engine manufacturers need to look for balance from the engine performance and emissions. Fuel injection technology development makes the stringent requirements possible [1]. Mixing process of gas and air, combined effect of combustion and emission, both impact the performance of diesel engine directly [2-6].

On the basis of the high pressure common rail fuel injection system of diesel engine, AnShijie [7] and others build a common rail diesel multiple injection performance test platform, and carried out the engine test and research about influence of multiple injection on diesel engine performance, the results show that using common rail technology and multiple jet drops off diesel engine exhaust temperature 50 ℃, and noise down 15 db, carbon smoke emissions improves significantly. Zhou Lei [8] and others researched the applicability of the fuel and air mixing process on large eddy simulation model in the prediction of engine early injection. The results show that the injection time moved up is advantageous to the fuel and air homogeneous mixing, mixture distribution is more uniform. WangHu [9] and others carried out the research about repeated injection and EGR coupling control experiment on the influence of the combustion and emission characteristics of diesel engine, compared with single injection, multiple injections coupling EGR control can significantly reduce carbon emissions, and fuel consumption rate debasing. Wang Benliang [10] put forward a proposal for multiple injection coordinated control strategy, aimed at pilot injecttion, main injection and post injection, and designed the control algorithm of multiple injection coordination controller, and has carried on the experimental study, the results show that the design of multiple injection coordinated control strategy can effectively realize the coordinated control on pre-injection, main injection and post injection under different conditions. Muammer Ozkan [11] and others researched the effects of heating and power efficiency in cylinder under different injection strategies, the experimental results show that the reasonable injection timing in advance can ensure the thermal efficiency and power capability and reduce NOx emissions. Mohan, B, et al. [12] discussed all kinds of injection strategy: including different curves of injection pressure, injection curve, injection timing and injection interval Angle, the influence on engine performance and emissions, the results show that pilot injection can reduce combustion noise and NOx emission, the injection interval smaller Angle can promote the soot oxidation, the appropriate injection timing can improve engine performance. So select the reasonable injection timing, and multiple injections can improve the diesel engine performance and reduce emissions.

* ***Corresponding author*** e-mail:wrr31@qq.com

In this paper, we take a Marine low speed two stroke diesel engine as the research object, establish the combustion model in cylinder of diesel engine, do the combustion simulation research under the condition of single jet to the diesel engine, and compared the results with experimental data. Contrast research the diesel engine performance under the condition of single injection and secondary injection, put forward the reasonable selection method of injection timing in reducing emissions under the premise of guaranteeing the power capability, it provides references for balancing diesel engine performance and emissions.

2 Establishment of the model and meshes generation

Taking a 350 mm two-stroke low-speed Marine diesel engine as the research object, for the four whole nozzle injector, spray holes are uniformly distributed. The operational parameters of the model engine are presented in Table 1. Combustion chamber space is composed by the cylinder liner, cylinder head, valve base and piston of diesel engine, cylinder grid as shown in Figure 1. Meshing is completed by special module ES- ICE of STAR - CD, the grid diagram as shown in Figure 1, geometry model and grid of top dead centre as shown in Figure 2, the top dead centre grid number is 32415.

TABLE 1 Basic parameter of a marine diesel engine

Parameter name/unit	parameter value
Cylinder power /kw	653
Diesel engine cylinder diameter /mm	350
Stroke number	2
Compression ratio	17:1
Swirl rate	1.2
Combustion chamber	Exposure to ωtype (with BUMP ring)
Intake type	pressure intercool

FIGURE 1 Diesel engine cylinder grid

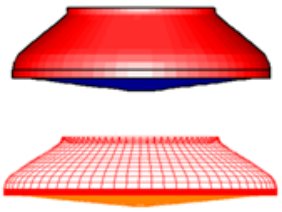


FIGURE 2 Grid of piston at top dead centre

Diesel engine cylinder gas phase flow model is based on classical hydrodynamics compressible navier-stokes equation of viscous fluid, namely according to the quality, composition, momentum and energy conservation law and the ideal gas state equation, a set of partial differential equations to describe the flow process in cylinder. The turbulence model used high Reynolds number k - Epsilon model; spray simulation adopted Huh atomized model, the Reitz - Diwakar droplets broken model and Bai bump wall model; The ignition model is compression ignition Shell model; Combustion model is laminar flow turbulence characteristic time combustion model and the EBU - LATCT model, and Zeldovich NOx generation model.

3 Initial boundary conditions

Initial conditions were given out by one dimensional simulation calculations of the machine; the initial moment pressure in cylinder is 0.485 MPa, the temperature is 363K. Boundary conditions are set according to the experience: cylinder head temperature 823K, the piston temperature 773K, and valve bottom temperature 823K. The boundary of the cylinder wall temperature is divided into three parts, the cylinder expansion direction, before one sixth is 623K, after a third of 548K, the remainder is 498K; the initial vortex ratio in cylinder is 5.5.

4 Model validations

To verify the model and algorithm, the combustion chamber combustion process was simulated under the condition of single jet; the fuel injection time is from the beginning of the 353 oCA, calculated pressure curve in cylinder is shown in Figure 3.

The Figure 3 shows that the compression pressure calculated is 13.04 MPa, explosion pressure is 16.59 MPa. And the original machine test compression pressure is 13.0 MPa; explosion pressure is 16.53 MPa, the compression stress error and the explosion pressure error were 0.31% and 0.36% respectively. The simulation result is in accordance with the test result, and it verifies the effectiveness of the algorithm.

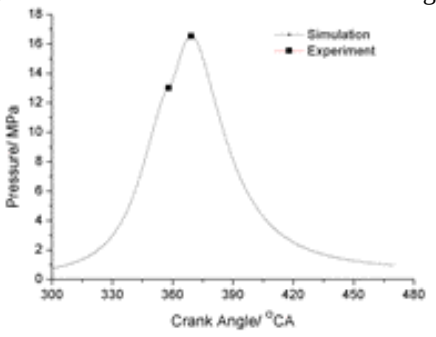


FIGURE 3 Cylinder pressure curve

5 Diesel engine performance contrastive analysis

In order to analyse the influence to diesel engine performance of different fuel injection law at the same injection current, combustion chamber combustion process was simulated under the condition of single injection and secondary injection respectively, simulation process is from exhaust valve closing time to exhaust valve open time. Injection timing of single injection and secondary injection are both 353oCA, fuel injection law is shown in Figure 4.

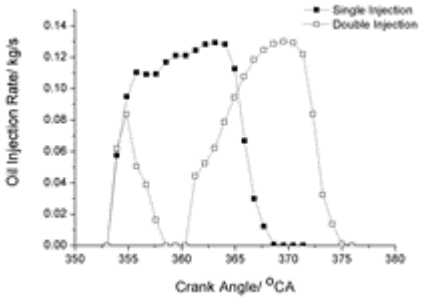
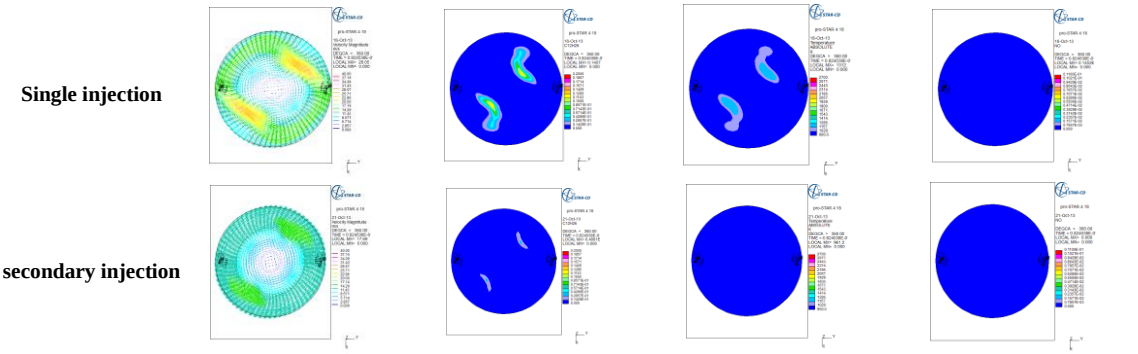


FIGURE 4 Fuel injection law

5.1 COMBUSTION SIMULATION RESULTS IN CYLINDER

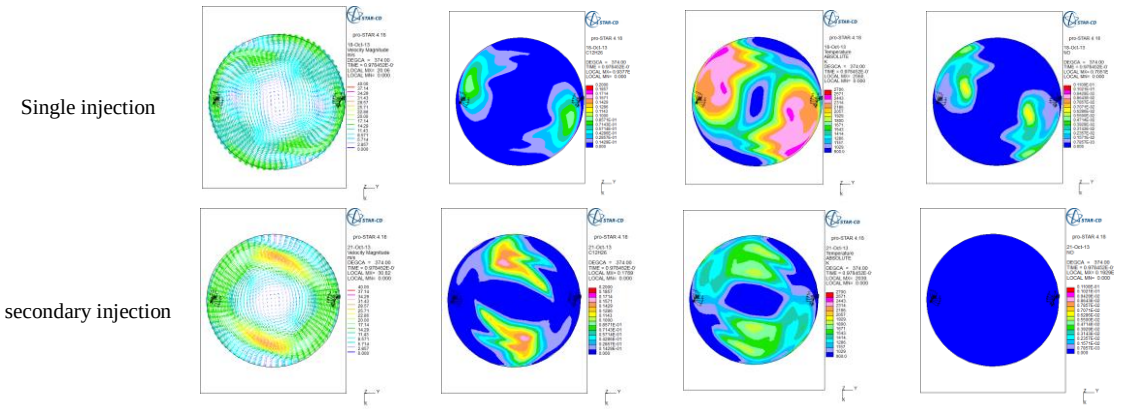
Figure 5 is the distribution of oil hole in the plane (z = 27 mm) of the velocity vector diagram and fuel concentration field, temperature field and the NOx concentration field, under different fuel injection law, the piston runs to the top dead centre namely 360 oCA.



(a)Velocity field **(b) fuel concentration field** **(c) temperature field** **(c) NO_x concentration field**

FIGURE 5 Each variable field distribution in cylinder at 360°CA

Figure 6 shows that under the condition of single jet, affectted by the eddy current in cylinder, the velocity field gradually become circular distribution, while the fuel injection ended in the 368oCA, but gas has not yet been fully burning, so local high, the centre temperature of 2571K, lead to distortion near the high velocity vector field. Under the condition of secondary injection, fuel injection has just concluded at the moment, high-speed injection fuel in cylinder, droplet spray evaporation form two strands of strong turbulence, the fuel injection rate ear along the axis is as high as 40 m/s, the gas concentration is about 15.7%, because gas is not yet fully mixed with air, combustion is not sufficient, so the centre of local high temperature is 1972K only. Due to under the condition of single jet local temperature is too high, creating a large number of NOx, the centre concentration is as high as 0.63%; With less NOx formation under the condition of secondary injection.

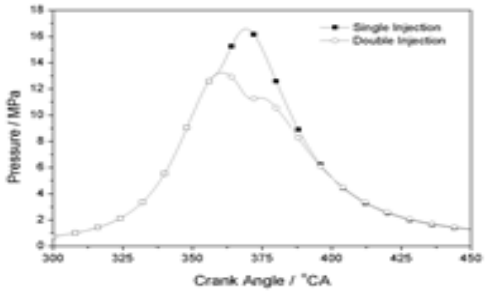


(a)Velocity field (b) fuel concentration field (c) temperature field (c) NO_x concentration field

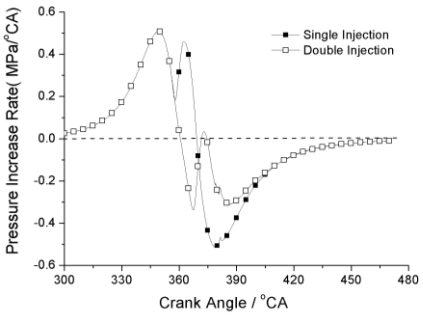
FIGURE 6 Each variable field distribution in cylinder at 374°CA

5.2 DIESEL ENGINE PERFORMANCE ANALYSIS

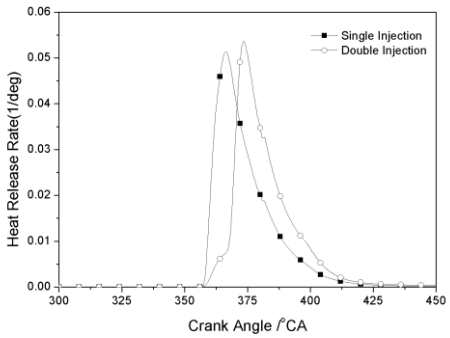
Figure 7 compares the same injection timing and different fuel injection law, combustion and emission performance curve, from top to bottom in turn is average pressure curve in cylinder, the pressure rise rate curve and heat release rate curve, the curve of average temperature in cylinder, total quality curve NOx and soot total quality and contrast figure to do work.



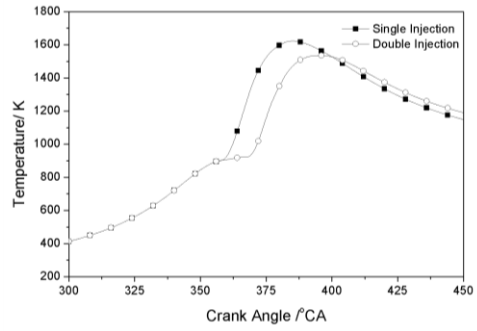
a) Average pressure curve in cylinder



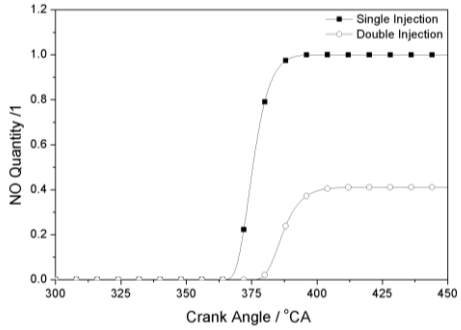
b) Pressure rise rate curve in cylinder



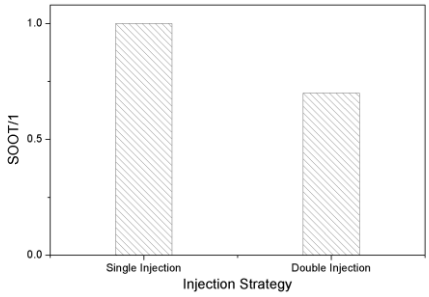
c) Heat release rate curve



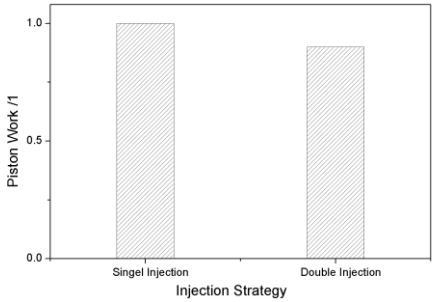
d) Average temperature curve in cylinder



e) NO_x total mass curve



f) Total quality comparison diagram of soot



g) Work contrast figure

FIGURE 7 Diesel engine performance contrast figure

The Figure 7 a) and b) shows that under the condition of single injection and secondary injection ignition delay period is short, about 1oCA, fuel injection quantity is moderate, and it has a good gas mixture; Under the condition of single jet urgent combustion period for 354 oCA - 362 oCA, most high pressure rise rate is 0.45 MPa/oCA, and under the condition of secondary jet urgent combustion period is 354 oCA -358.5 oCA (pre-injection end), due to pre-injection fuel quantity is less, pressure rise rate slowly reducing, maximum pressure is 13.2 MPa only at the moment of 360oCA. Under the condition of single jet, slow burning period is 360 oCA - 369 oCA, maximum pressure is 16.5 MPa, when the secondary injection conditions, it enters the combustion period after the top dead centre, main injection stage appeared the second peak pressure, only 11.3 MPa (374oCA), lower than the first peak pressure. In a word, main injection period of the secondary injection is in the after burning period, the fuel heats in the lower expansion ratio, and the diesel engine efficiency greatly lowered.

From figure 7 c) we know, under the condition of single injection and secondary injection, heat release starting point and end point is basic same, but the secondary injection has high peak heat release rate than single injection, the secondary injection has the superior combustion performance. From figure 7 d) average temperature curve in cylinder, the secondary injection cylinder temperature curve is always lower than that of single injection, and the temperature peak appeared late, it shows that the secondary injection power capability is less than single injection.

Figure 7 e) is NOx formation history curve, we know from the figure, compared with the single injection, under the secondary injection, the formation of NOx starting moment is later, and the total generate amount is only 41.1% of the single jet, is reduced by 58.9%, this is due to under the condition of the secondary injection the average temperature in cylinder is lower, and the peak temperature appear later.

Figure 7 f) is the soot emissions contrast figure, soot under the condition of the secondary injection is only 69.9% of the single injection, reduced by 30.1%, that under the condition of secondary injection gas and oxygen mixed fully, generated less soot. Figure 7 g) shows that the secondary injection ability of doing work is 89.9% of the single injection.

To sum up, due to injection timing of the secondary injection later, main injection in the later period, led to the decrease of the diesel engine power capability, but the secondary injection has the superior combustion performance, and the NOx and soot emissions are lower than that of single injection. If you want to improve the diesel engine work ability under the condition of secondary injection, it needs reasonable adjustment of injection timing.

6 The secondary injection results after adjustment of injection timing analysis

6.1 REASONABLE SELECTION OF INJECTION TIMING

The nature of diesel engine is to change the chemical energy of fuel into the kinetic energy of the crankshaft, the Figure 7 c) heat release rate curve shows that the secondary injection has the superior performance of combustion heat release than single injection, after adjusting the injection timing, the ability of doing work of the secondary injection could reach or exceed the single injection. Due to secondary injection main combustion process is located in the after combustion period, expansion ratio is larger, work performance is poor, so it should be based on the heat release rate curve, adjusting injection timing of the secondary injection reasonably, make the heating centre of gravity (that is, the heat release rate curve integral centroid)coinciding. Because the heat release rate is a process quantity, it is difficult to control, in this paper, by adjusting the cumulative injection quality centre under the condition of secondary injection (i.e., 50% of cumulative injection quantity), with a single jet overlap, such basically guarantee the both exothermic centre of gravity coincided.

Figure 8 is cumulative injection quantity curve, cumulative injection quality centre is the dotted line in the fig-ure, the accumulative fuel injection quality centre of single jet and secondary injection is 6.9oCA and 13.6oCA respectively, to make the cumulative injection quality centre coincided, it needs the injection timing of secondary injection 6.7oCA in advance

Equations should be placed flush-left with the text margin and should be preceded and followed by one line of white, using *Microsoft Equation Editor*. Formulas fonts and sizes must match the text of your document.

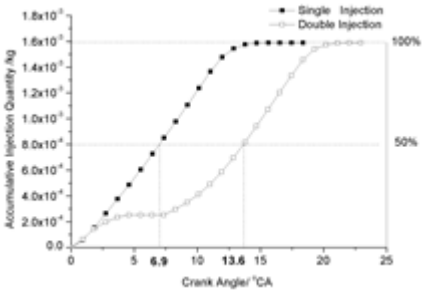
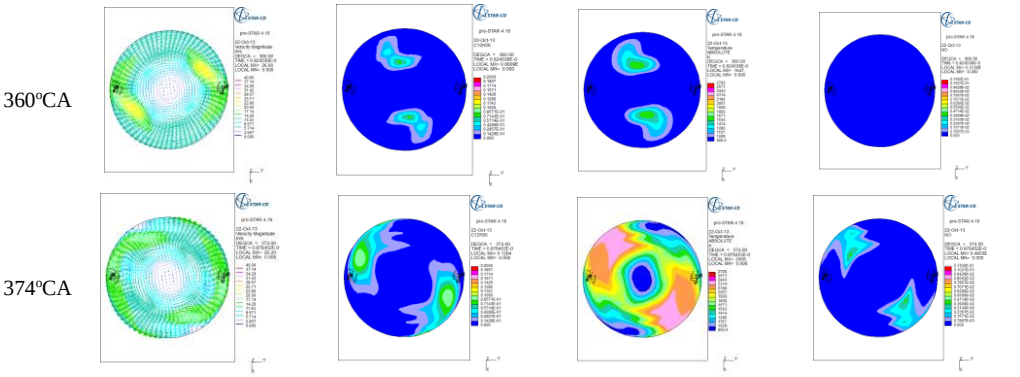


FIGURE 8 Cumulative fuel injection quantity curve

6.2 FLOW FIELD ANALYSES AFTER THE ADJUSTMENT OF INJECTION TIMING IN CYLINDER

Figure 9 for the adjustment of injection timing, under the condition of the secondary injection, 360oCA and 374oCA variables of the field.



(a) Velocity field (b) Fuel concentration field (c) Temperature field (c) NO_x concentration field

FIGURE 9 Distribution of variables in each field after injection timing adjusted

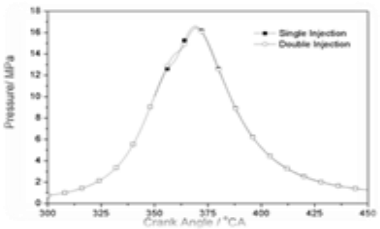
Figure 9, in the 360oCA moment, compared with the single injection, after adjusting the injection timing of secondary injection, centre speed of gas jet is 22.8 m/s, lower than that of single injection; Fuel concentration field distribution is slightly larger than single injection, the concentration of gas jet centre is about 7.9%, it shows that gas and air mixed uniformly under the condition of secondary injection. By the temperature field, the high temperature range is larger than single injection, and centre temperature is 1543K, higher than that of single injection, obviously that the secondary injection combustion performance is good, at this time there is no NOx formation.

At 374oCA, compared with single injection, the difference of velocity field between the secondary injection after adjusting and single jet is small; Range of fuel concentration field distribution is quite, and there is less difference of concentration; the distribution of temperature field is pretty, but the highest temperature is 2314K, only less than single injection, it shows that the secondary injection effectively reduces the local high temperature, temperature tends to be uniform; The NOx distribution range is small, centre concentration is only 4%, and the range is very small, that the secondary injection effectively reduced NOx formation.

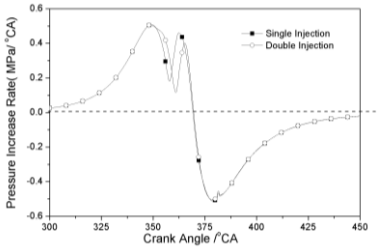
In short, after the adjustment of injection timing, oil and gas mixed more evenly, local temperature rise reduced, and it inhibits the formation of NOx.

6.3 PERFORMANCE ANALYSIS AFTER ADJUSTMENT OF INJECTION TIMING

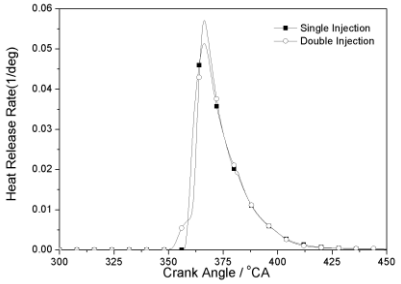
Figure 10 is after the adjustment of injection timing, the diesel engine performance comparison between the secondary injection and single injection.



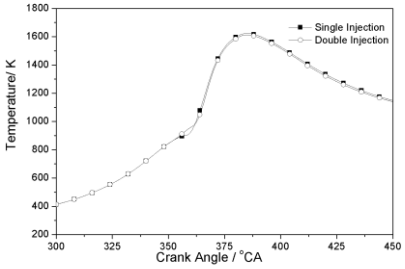
a) Average pressure curve in cylinder



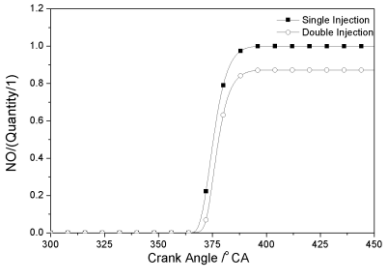
b) Pressure rise rate curve in cylinder



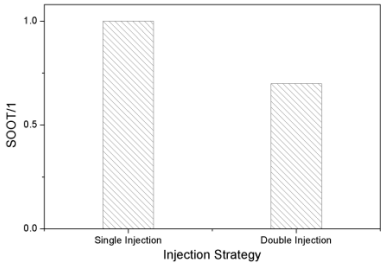
c) Heat release rate curve



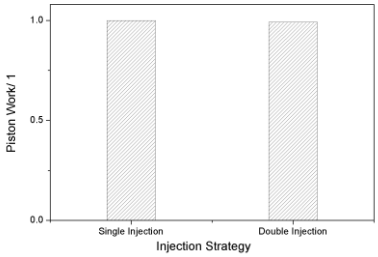
d) Average temperature curve in cylinder



e) NO_x total mass curve



f) Total quality comparison diagram of soot



g) Work contrast figure

FIGURE 10 Diesel engine performance contrast figure

The figure 10 a) and b) shows that after the adjustment of injection timing, the secondary injection ignition delay period is longer, about 8oCA, due to the fuel pre-injection quality is less, gas mixing fully, it is to lay a good foundation for the subsequent combustion; Urgent combustion period is 354oCA - 360oCA, premixed gas combust together almost at this stage, cylinder temperature rise sharply, as the main fuel injecting in cylinder, the process of fuel atomization and evaporation shorted, gas diffused and mixed combustion with oxygen, cylinder pressure rise rapidly, the high pressure rise rate is 0.4 MPa/oCA, lower than that of single injection, it shows that the secondary injection pressure change is moderate; slow-burning period is 360oCA - 365.2oCA, during this stage the peak pressure is 16.4 MPa, and was equal to that of single injection; Compared with the single injection, after combustion period of the two are basically identical.

By figure 10 c) and d), the secondary injection and single jet reach exothermic peak almost at the same time, and peak value of the secondary injection is higher than single injection, it shows that the secondary injection fuel burning more fully; the cylinder temperature curve change rule is consistent, and the secondary injection peak temperature is only 1612K, lower than that of single injection of 1623K, it shows that secondary injection combustion in cylinder is more even.

By figure10 e), NOx generation total amount of the secondary injection is 87.3% of the single jet, was reduced by 12.7%, it shows that the secondary injection high temperature area is small, NOx formation is less; By figure10 f), soot of the secondary injection is only 58.9% of single injection, down 41.1%, that shows gas of the secondary injection mixed fully in slow-burning period, therefore emissions reducing. By figure10 g), the secondary injection ability of doing work is 99.1% of single injection.

In conclusion, after adjusting the injection timing reasonably, the secondary injection has almost the same power capability as single injection, but NOx emissions reduce by 12.7%, soot emissions reduce by 41.1%. That shows by changing cumulative injection fuel quality to adjusting the injection timing of secondary injection, compared with single injection, it can ensure the ability of doing work and greatly reduce emissions.

7 Conclusions

1. Set up a combustion simulation model of marine low speed two stroke diesel engine, simulated and analysed the compression, spray, combustion, emission process of marine low speed diesel engine, and compared with the test data, the maximum error is less than 0.36%.

2. At the same injection timing, compares the diesel engine performance of the single injection and secondary injection, the results show that compared with single jet, the secondary injection caused the reduction of NOx by 58.9%, lower soot by 30.1%, but the ability to make the work fell by 10.1%.

3. The method of adjusting the cumulative injection quality centre to adjust the injection timing of secondary injection reasonably is given, it can enhance the power capability of diesel engine; And combustion simulation was done, the results show that compared with single injection, the secondary injection can ensure the ability of doing work, at the same time, reduce NOx by 12.7% and soot by 41.1%.

References

[1] Musculus M P B, Miles P C, Pickett L M 2013 Conceptual models for partially premixed low-temperature diesel combustion *Prog Energ Combust* **39** 246-83

[2] Zhuang J, Qiao X, Bai J, Hu Z 2014 Effect of injection-strategy on combustion, performance and emission characteristics in a DI-diesel engine fueled with diesel from direct coal liquefaction *Fuel* **121** 141-8

[3] Herfatmanesh M R, Lu P, Attar M A, Zhao H 2013 Experimental investigation into the effects of two-stage injection on fuel injection quantity, combustion and emissions in a high-speed optical common rail diesel engine *Fuel* **109** 137-47

[4] Macian V, Payri R, Ruiz S, Bardi M, Plazas A H 2014 Experimental study of the relationship between injection rate shape and Diesel ignition using a novel piezo-actuated direct-acting injector *Appl Energ* **118** 100-13

[5] Barabás I, Predicting the temperature dependent density of biodiesel–diesel–bioethanol blends *Fuel* **109** 563-74

[6] Finesso R, Spessa E 2014 Ignition delay prediction of multiple injections in diesel engines *Fuel* **119** 170-90

[7] AnShiJie, Chang Hanbao, Xu Hongjun 2011 The experimental study on common rail system repeatedly injected to marine diesel engine performance improvement *Journal of wuhan university of technology (transportation of science and engineering edition)* **35**(5) 1005-9

[8] Zhou Lei, Xie Maozhao, Jia Ming, etc. Large eddy simulation of fuel injection combination in the process of engine early spray *Journal of central south university (natural science edition)* **43**(9) 3678-86

[9] Wang Hu, Yao Mingfa, Zheng Zunqing, etc 2010 Multiple injection and EGR coupling control experimental study on the influence of performance and emissions of the engine *Transactions of csice* **28**(1) 26-32

[10] Wang Benliang, Pei Hailing, Tang Weixin, etc. 2013 High pressure common rail fuel injection system multiple jet coordinated control strategy research *Chinese Internal Combustion Engine Engineering* **04** 18-22

[11] Zkan M, Zkan D B, Zener O, Y Lmaz H 2013 Experimental study on energy and exergy analyses of a diesel engine performed with multiple injection strategies: Effect of pre-injection timing *Appl Therm Eng* **53** 21-30

[12] Mohan B, Yang W, Chou S K 2013 Fuel injection strategies for performance improvement and emissions reduction in compression ignition engines - A review *Renewable and Sustainable Energy Reviews* **28** 664-76

Authors

Wang Xiaorong, born in 1982

Current position, grades: lecturer, Dr, master tutor
University studies: school of mechanical and engineering, Jiangsu university of science and technology,
Scientific interests: computational fluid dynamics and Marine diesel engine combustion and performance study

Chenpeng Liu, born on March 23, 1987, China

Current position, grades: graduate at Jiangsu University of Science and Technology, China.
University studies: B.S degree in machine manufacture and its automation from Henan Mechanical and Electrical Engineering College, 2009, and received the M.S degree in mechanical engineering from Jiangsu University of Science and Technology, China in 2014.
Scientific interests: include CFD analysis, engine performance and emission research.

Bo Liu, born on April 11, 1982, P. R. China

Current position, grades: senior engineer at China Shipbuilding Power Engineering Institute Co., Ltd., Shanghai, China.
University studies: B.S degree in thermal energy and power engineering from Shanghai Jiao Tong University, China in 2004, and received the Ph. D degree in power machinery and engineering from Shanghai Jiao Tong University in 2011.
Scientific interests: performance and combustion development of diesel engine.

Wenxian Tang, born on September 6, 1962, China

Current position, grades: The dean of Mechanical Engineering at Jiangsu University of Science and Technology College, China.
University studies: B.S degree in mechanical manufacturing technology and equipment from Zhenjiang Shipbuilding Institute, 1987, received the M.S degree in machinery manufacturing from Jiangsu University, China in 1994, and received the PhD in Mechanical design and theory from Shanghai University, 2004, Post-doctoral research work in mechanical engineering from Nanjing University of Science and Technology, 2004-2005.
Scientific interests: Digital development of complex products.

Chenpeng Liu, born on March 23, 1987, China

Current position, grades: graduate at Jiangsu University of Science and Technology, China.
University studies: B.S degree in machine manufacture and its automation from Henan Mechanical and Electrical Engineering College, 2009, and received the M.S degree in mechanical engineering from Jiangsu University of Science and Technology, China in 2014.
Scientific interests: include CFD analysis, engine performance and emission research.

An improved hand vein image acquisition method based on the proposed image quality evaluation system

Jun Wang, Guoqing Wang*, Ming Li, Wenhui Yu, Hao Tian

School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, 221116, China

Received 12 June 2014, www.cmnt.lv

Abstract

In the part of hand vein image acquisition of vein recognition, we basically cannot get the images that are suitable for later process due to both the subjective factors and the objective factors. After a deep analysis on the main reasons including the different thickness of hands and mass of things on the hand, we propose the design of hand vein acquisition device based on the adaptive modulation control. The key part of the design is the proposed double-processing image quality assessment system: the first part of which involves the process of quality assessment based on the effective information of multidimensional histogram and the normalized information entropy; the second part involves adding the grey image variance and quantities of cross points with the proper weights, then two process will be made according to the final score: if the score is lower than the set threshold then image will be discarded, and on the other hand the brightness degree of the LED groups of the device will be adaptively adjusted according to the D-value of the score and the threshold so that we can get the high-quality hand vein image for later feature extraction and matching and get a higher recognition rate.

Keywords: hand vein recognition; adaptive control; two-set quality evaluation system; D-value

1 Introduction

With the development of biometric feature recognition and informative technology, the traditional identify technology cannot meet the demand of security. Vein recognition has become a new trend of biometric recognition with some great advantages [1-4]. And the currently main feature image evaluation method focuses on the fingerprint, iris and images of videos [5-7]. After a deep research and study on the existing method for image evaluation, we respectively take the statistical characteristics and structural characterristics after deburring as the feature parameters for evaluation function. The criterion of the first-set quality evaluation system we set is that: Calculating the two-dimensional histogram of hand vein images firstly, then we determine whether to store the image or not according to the spread condition of the pixel around the area of G_1=G_2; and then we calculate the information entropy of the stored image and compare the value with the pre-set threshold to get the second-stored vein image whose value is larger than the threshold (we set it 4.0 according to the recognition rate). In the second-set quality evaluation system, we firstly get the grey variance and number of cross points of the selected hand vein images, and then adopt the pre-set weight to calculate the final parameters by adding them up. The optimized weights we set are respectively 0.5826 and 0.4174 according to the recognition rate. And in the final, we get the final parameter to determine whether to leave the hand vein images for feature extraction, the normalized threshold we set in the final determination part is 0.55 according to the experiment result. And then we design the improved smart-control acquisition device based on the evaluation system, which can get the high-quality hand vein images for later feature extraction and matching and get a higher recognition rate.

2 Theory of quality evaluation system for hand vein images

In the procedure of capturing hand vein images, we usually could not get the high-quality hand vein images for later process, so it is essential for feature extraction and recognition to import quality evaluation function to realize the effective choice of the original vein images.

After a deep research and study on the existing method for image evaluation, we respectively take the statistical characteristics and structural characteristics after deburring as the feature parameters for evaluation function [8]. The explanation of the parameters is as follows:

(1) Statistical characteristics: The method censuses the spread parameters including the variance and entropy of the grey images based on the theory of transforming the original image into grey images, while the variance reflects the contrast degree and the entropy reflects the content of effecttive information.

(2) Structural characteristics: After the pre-process and the deburring process on the hand vein images, we choose the quantities of cross points of the images and the extracted ROI area as the evaluation parameters. On the one hand, we can know whether the images we obtain contain the proper amount of cross points for recognition; On the other hand,

* ***Corresponding author*** e-mail: wangguoqingcumt@163.com

we can get the idea of the amount and effectiveness of the structural information of vein images.

3 Two-set of evaluation systems on hand vein images

3.1 FIRST-SET PART OF THE EVALUATION SYSTEM

In the first part of the evaluation system, we focus on the global spread and the global effectiveness of image information that are closely related with human vision, and we choose the two-dimensional distribution histogram and information entropy to realize the first-set evaluation.

3.1.1 Choice of images based on the two-dimensional distribution histogram

As the global information and features of images, the histogram of grey vein images can show the frequency distribution of grey of pixel and we adopt this parameter as one of the evaluation system. The traditional one dimensional grey histogram is the statistical distribution of isolated pixel. And we find that if the correlation between pixels could be got and analysed, we could get a better choice on the images. So we introduce the multi-dimensional distribution histogram for analysis.

In the theory of image processing, we define the multi-dimensional distribution histogram as follows: count the frequency of grey image *f(x,y)*, and then get $Num\left(G_1,G_2,G_3,\ldots,G_k\right)$, $G_1,G_2,G_3,\ldots,G_k \in \Omega$, and the Num stands for the frequency amount that when it meets the requirement that the pixel is G_1 and the *K-1* around the pixel is $(G_1,G_2,G_3,\ldots,G_k)$, and then we get the definition of *K* dimensional distribution histogram, and then normalizing the histogram value according to the max of frequency $max(Num(G_1,G_2,G_3,\ldots,G_k))$ to get $Norm(G_1,G_2,G_3,\ldots,G_k)$, which is called *K-D* normalized histogram.

$$Norm\left(G_1,G_2,G_3,\ldots,G_k\right) = \frac{Num\left(G_1,G_2,G_3,\ldots,G_k\right)}{\max\left(Num\left(G_1,G_2,G_3,\ldots,G_k\right)\right)} . \quad (1)$$

In the definition, we normalized the *Num* and max one of the frequency values to get the *K-D-log* normalized histogram.

$$Norm_{\log}\left(G_1,G_2,G_3,\ldots,G_k\right) = \frac{\log(Num\left(G_1,G_2,G_3,\ldots,G_k\right))}{\log(\max\left(Num\left(G_1,G_2,G_3,\ldots,G_k\right)\right))} . \quad (2)$$

The max one refers to the peak value of grey frequency value, and for calculating and reckoning compare we use the algorithm of LOG to realize the transformation based on the idea of the fractal dimension of complexity measure. And according to the advantages of statistics and distribution, we can get the key parameter to realize the choice of vein images for later analysis.

On the basis of the above definition, we define the certain pixel as G_1 and the pixels in the right-side region around it as G_2, and the criterion we set is: as for the high-quality vein images, the character of its two-dimensional histogram is that the frequency value of the region $G_1 = G_2$ has a higher peak value and higher distribution density while the character of two-dimensional histogram for low-quality images is a lower peak value and scattered distribution in the region of $G_1 = G_2$.

3.1.2 Choice of images based on the information entropy

The definition of entropy is proposed by Clausius in 1865 which is used for describing the disordered state of thermodynamic system, and then it was introduced into the field of information engineering. After a long period of development and improvement, the experts in the field of image information propose the definition of information entropy to represent the average amount of statistical information of the system [9-10].

In the proposed definition, we import the definition of complexor, and it is made up of $\{x_1, x_2, x_3, ..., x_n\}$, and then we can get the following definition on the condition of supposing x_i=v and the probability $p_i=P(x_i)$:

$$H\left(v\right) = -\sum_{i=1}^{n} p_i \log_z p_i \cdot \quad (3)$$

In the above definition, H(v) stands for the probability distribution function of the defined random variables.

In the field of digital image processing, the spread of different pixel points in the images are different in images, which results in different shape features of the vein images. The amount of statistical information will be much more with the richness of grey or color of the images, which means the quality will be better. According to this the rule, we could know that the much more difference of grey level and the more the statistical information, the better the hand vein image is. Supposing that the statistical probability of grey value m is P_m, and we can get the following calculating formula:

$$H\left(v\right) = \sum_{T=0}^{z33} p_m \log_z p_m \cdot \quad (4)$$

In the above definition, H stands for the information entropy of hand vein image, and the value range of m is [0, 255]. P_m stands for the probability of grey value m, and the criterion is that the higher the value of H, the better and the higher quality of the images.

In conclusion, the criterion of the first-set quality evaluation system we set is that: Calculating the two-dimensional histogram of hand vein images firstly, then we determine whether to store the image or not according to the spread condition of the pixel around the area of $G_1 = G_2$; and then we calculate the information entropy of the stored image and compare the value with the pre-set threshold to get the second-stored vein image whose value is larger than the threshold (we set it 4.0 according to the recognition rate).

3.2 SECOND-SET PART OF THE EVALUATION SYSTEM

After the process of the first-set evaluation system, we obtain the preliminary hand vein images that can be used for subsequent analysis. Then we design algorithm to get the grey variance and numbers of cross points of the selected hand vein images, and then combine the two parameters by setting optimized weights, and then we get the final sample images for feature extraction and matching.

3.2.1 Calculation of grey variance

In this section, we define the selected images as the matrix M * N, and define *I(I,j)* as the grey value of the pixel on the line i and column j. Then we define Mean as the average value of the grey value, which can reflect the degree of brightness; On the other hand, we define Var as the grey variance of the image that reflects the contrast between the foreground and background of the images. The criterion is that: the bigger the value of the Mean and the Var, the better the vein image is.

$$Mean = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i, j), \tag{5}$$

$$Var = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left[I(i, j - Mean) \right]^z . \tag{6}$$

3.2.2 Calculation of cross points

Although we have set the first-set part and the first one of the second-set of the evaluation system to select the high-quality vein images, there are still some selected images with low-quality bringing down the recognition rate, and after a deep analysis on the selected images, we find that there are great differences between the images after the pre-process on (especially the deburring process) the selected images, so we design the algorithm to calculate the number of the cross points. Then we find that the quality of the images differs with the number of cross points and the larger the number is the higher quality the image is.

In the pre-process of deburring, we improve it as the conditional deburring. The character, of which is that the connected points of lines and the turning points and T-shape branch points can be well kept after the deburring. But after a sudden amplification transformation we find that the conditional deburring can't realize the complete single pixel processing, which is shown as follows:

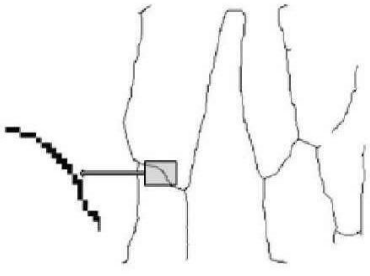


FIGURE 1 Amplification of the vein images after conditional deburring

After a deep analysis on the phenomenon mentioned above, we adopt the template process algorithm after the self-defined improved conditional deburring process on the vein images, and it can remove the single-pixel point to get the real cross points of the images. We divide the existing non-single pixel into two kinds, and as for the first kind shown in Figure 2, we need to cut off the non-single pixel according to the final matching because that these pixels are redundant.

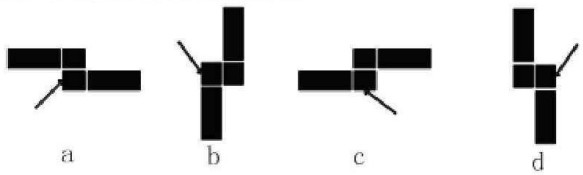


FIGURE 2 The first kind of non-single pixels

The template we adopt is shown in Figure 3, the centre of the template is the reference central point. In the chart, "1" stands for the point in the targeted images while "0" stands for that in the background images, and "x" stands for the points of the targeted images that are used for later analysis on the one hand, on the other hand, it stands for the points of the background images.

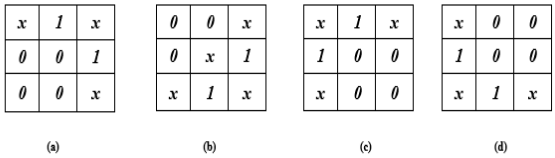


FIGURE 3 The designed template for the first kind of non-single pixels

The criterion we set here is that we firstly choose to cut off some pixels related to the template's central point's so as to decrease the width of the vein, and at the same time we cut off the points the arrow pointing to in Figure 3 (a). Then we adopt the other three pre-defined templates to realize cutting the other points off to leave the pixels that meet the requirements. Then we adopt the iterative operation with the existing operation until the first non-single pixels are cut off completely.

Then we can conclude from the analysis that the second kind of non-single pixels are as follows:

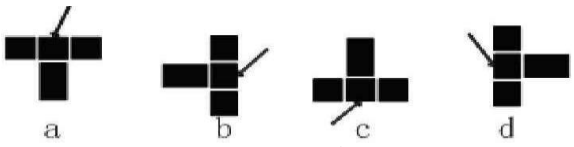


FIGURE 4 The second kind of non-single pixels

In this part, we design another template that are suitable for the second kind of non-single pixels based on the design of the first kind of templates.

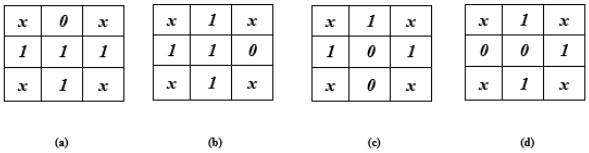


FIGURE 5 The designed template for the second kind of non-single pixels

After the process of cutting the pseudo cross points off, we can get the following vein images as shown in Figure 6 for later analysis. On the other hand, we find that the cross points of the vein image can be easily obtained after the designed templates' process, and we could design a formula to get the number of the cross points so that we can determine whether the images can be used for recognition.

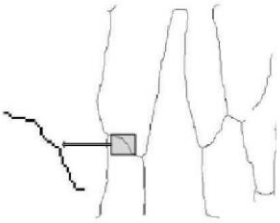


FIGURE 6 The local amplification effect image after the templates' process

In the second-set quality evaluation system, we firstly get the grey variance and number of cross points of the selected hand vein images, and then adopt the pre-set weight to calculate the final parameters by adding them up. The optimized weights we set are respectively 0.5826 and 0.4174 according to the recognition rate. And in the final, we get the final parameter to determine whether to leave the hand vein images for feature extraction, the normalized threshold we set in the final determination part is 0.55 according to the experiment result.

4 Realization of improved smart-control acquisition device based on the evaluation system

Through the process of the proposed evaluation system, we find that the final recognition rate improves a lot after leaving some low-quality vein images. So we import the thought of feedback modulation based on the evaluation system to realize the smart-control of the capturing device and increase the recognition rate in the end.

4.1 DISCUSSION ON THE CONTROL METHOD ON THE CAPTURING DEVICE

According to the imaging principle, we could know that the vein images are obtained on the condition of specified wavelength of light source, so it is required that the voltage should remain in a stable value, and we have realized keeping the homogeneity and stability of the near-infrared light source by adding the control on the power-supplying circuit and equipped the capturing device with light guide plate. And concluding from the circuit theory we can know that the strength of the light source is proportional to the electrical current, so we choose to control the light strength by the way of controlling the electrical current.

We find that the changing trend of the quality evaluation parameter will go up with the increase of the current in a certain range, but it will go down after a certain peak value, so we can realize the control on the capturing of hand vein image to obtain high-quality image by the control on the electrical current. The method we adopt is the PWM control on the circuit.

4.2 REALIZATION OF THE DEVICE

The procedure of capturing hand vein images is as follows: set the parameters of the capturing device by the self-programed software, then starting capturing and calculating the parameters of the evaluation system by the self-programed software, and then take actions mentioned above to obtain the high-quality vein images.

The concrete procedure of the control is as follows:

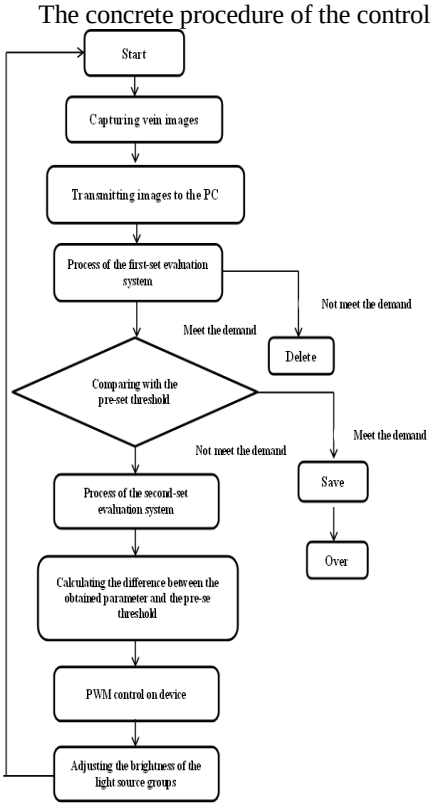
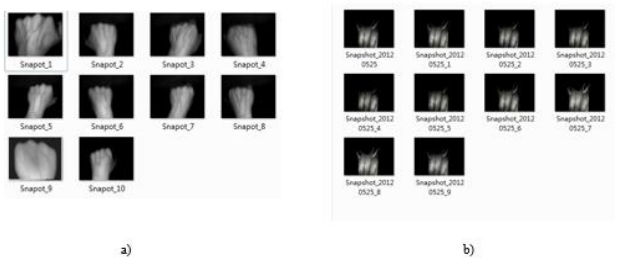


FIGURE 7 The flow chart of the smart-control system

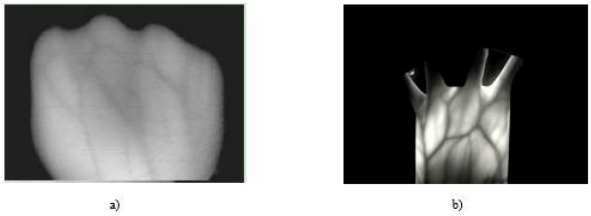
5 Experiment result and discuss

At the beginning of the study on hand vein recognition, we design a capturing device to establish the database of hand vein images for later recognition, but it cannot get a good recognition result, the beginning vein image acquisition method concentrates on transmission imaging mode uses simple control method without feedback signals to realize the control on the LED groups.



a) b)

FIGURE 8 Databases of the hand vein images of one volunteer: a) Acquisition device before optimized; b) Acquisition device after optimized



a) b)

FIGURE 9 The certain sample of the databases: a) Acquisition device before optimized; b) Acquisition device after optimized

Based on the above disadvantages of the current devices and after a deep analysis on the main reasons including the different thickness of hands and mass of things on the hand, we propose the design of hand vein acquisition device based on the adaptive modulation control. The key part of the design is the proposed double-processing image quality assessment system: the first part of which involves the process of quality assessment based on the effective information of multidimensional histogram and the normalized information entropy; the second part involves adding the grey image variance and quantities of cross points with the proper weights, then two process will be made according to the final score: if the score is lower than the set threshold then image will be discarded, and on the other hand the brightness degree of the LED groups of the device will be adaptively adjusted according to the D-value of the score and the threshold so that we can get the high-quality hand vein image for later feature extraction. The following two databases of hand vein images are respectively obtained by the two generation set of devices, and it is obviously that the smart-control device can get a better images.

Acknowledgments

The research described in this paper has been supported by three main projects: The Fundamental Research Funds for the Central Universities of China (Grant No: 2011QNB21); The Innovative engineering developing on graduates of Jiangsu Province in China (Grant No: CXZZ11-0293); The Chinese Coal Industry Association of science and technology guiding plan project (Grant No: MTKJ2011-306); The Innovative Training Project of the China university of mining and technology (Grant No: XZD1029014185).

References

[1] Hashimoto J 2006 Finger vein authentication technology and its future *Symposium on VLSI Circuits, Kawasaki* 5-8
[2] Wang J G, Yau W Y, Suwandy A, et al 2008 Person recognition by fusing palm print and palm vein images based on "Laplacian palm" representation *Pattern Recognition* **41**(5) 1514-27
[3] Crisan S, Tarnovan I G, Crisan T E 2007 A low cost vein detection system using near infrared radiation *IEEE Sensors Applications Symposium, San Diego: IEEE* 51-6
[4] Wang L Y, Leedham C G 2005 A thermal hand vein pattern verification system *Lecture Notes in Computer Science* **3687** 58-65
[5] Lu Wen, Gao Xinbo, Wang Tisheng 2008 A Natural Image Quality Assessment Metric Based on Wavelet-Based Contourlet Transform *Acta Electronica Sinica* **36**(2) 303-8
[6] Zhao Xianyun, Cai Anni 2006 Fingerprint Image Quality Analysis *Journal of Computer-Aided Design & Computer Graphics* **15**(5) 644-50
[7] Pan Lili, Xie Mei 2007 The Algorithm of Iris Image Quality Evaluation *IEEE Trans on Circuits and Systems for Video Technology* **14**(2) 616-9
[8] Benoit A, Le Callet P, et al. 2008 Quality assessment of stereoscopic images *Journal on Image and Video Processing* **2008** 1-13
[9] Pal S K, King R A 1981 Image enhaneement using smoothing with fuzzy sets *IEEE Transmission Systems, Manand Cyberneties* **11**(7) 494-501
[10] Pal S K, King R A 1983 On edge detection of X ray image using fuzzy sets *IEEE Transaction on Pattern Analysis and Machine Intelligenee* **5**(l) 69-77

Authors

Jun Wang, born in January, 1981, Jiangsu Province

Current position, grades: associate Professor in China University of Mining and Technology now.
University studies: M.Sc. in Control Engineering (2013) and is studying for a doctor's degree at School of Information and Electrical Engineering, China University of Mining and Technology.
Scientific interests: robot, biometric and Innovation Education.

Guoqing Wang, born in April, 1990, Hebei Province

Current position, grades: postgraduate of Control Engineering in China University of Mining and Technology.
Univesrity studies: BE in Information Engineering (2014).
Scientific interests: Pattern Recognition and Intelligent System and Biometric Features Recognition

Ming Li, born in March, 1962, Anhui Province

Current position, grades: professor and Ph.D. supervisor in China University of Mining and Technology now.
Univesrity studies: doctor degree from China University of Mining and Technology in 2002.
Scientific interests: intelligent detection and system

Wenhui Yu, born in September, 1990, Anhui Province

Current position, grades: Now he is a student in China University of Mining and Technology.
University studies: BE in Information Engineering (2014).
Scientific interests: Biometric Features Recognition and intelligent detection.

Hao Tian, born in January, 1992, Inner Mongolia Province

University studies: now is a junior student studying at School of Information and Electrical Engineering, China University of Mining and Technology
Scientific interests: robots control technology and Pattern Recognition

Pure electric buses status information compression and transmission methods basis on optimized Huffman coding algorithms

Jingfeng Yang[1, 2, 3], Yong Li[4], Nanfeng Zhang[1, 5], Jiarong He[6], Yueju Xue[1*]

[1]*South China Agricultural University, Guangzhou, 510642, China*

[2]*Guangzhou Public Transport Data Management Centre, Guangzhou, 510620, China*

[3]*Guangzhou Transport Information Control Centre, Guangzhou, 510620, China*

[4]*Guangzhou Institute of Geography, Guangzhou, 510070, China*

[5]*Guangzhou Entry-exit Inspection and Quarantine. Guangzhou, 510623, China*

[6]*South China University of Technology, Guangzhou, 510641, China*

Received 1 October 2014, www.cmnt.lv

Abstract

To strengthen safety supervision, satellite positioning information is needed to be uploaded as well as battery status information of pure electric buses. For the unbalanced coverage of mobile communication base station, data loss and incomplete data are prone to emergence when a large amounts of real-time data transmission through 3G/GPRS network in the poor communication environment of subregion, which is difficult to achieve large-scale farmland information collection and transmission simultaneously. Therefore, optimized Huffman coding compression algorithm for text-oriented information is promoted in this paper. The whole process test results of data exchange and decompression show that, the data compression algorithms can realize data compression effectively. The data sampling period is relatively smaller and its adjacent data is closer, the higher compression rate can be obtained by corresponding compression algorithms, and it can the basically ensure the decompression data without distortion. This will improve the efficiency of information transmission of transportation, as well as to ensure the integrity of information, which is of great significance to realization of transportation energy conservation.

Keywords: vehicle information, data compression, pure electric buses, battery status information

1 Introduction

The key operational vehicle information safety supervision has been continuously strengthened along with the national, provincial governments at all levels. The standard terminal with satellite positioning function were forcibly mounted on passenger bus, tourist charter, dangerous goods vehicles, coaches, trucks, heavy trucks, buses, taxis and so on. Some cities also require passenger bus, tourist chartered buses, taxis and other installation equipment using a video camera in order to enhance information security regulatory standards. Regardless the using of satellite positioning terminal or video camera equipment, it must bear a lot of the cost of wireless data transmission. If using new energy vehicles, such as electric buses, hybrid buses, electric taxis, etc., due to the need to upload more types of data to the regulatory status of the platform, its wireless transmission costs will rise at the same time. How to meet the requirements of the competent authorities, and saves the cost of wireless data transmission has become an important problem to be solved at present industry. The compressed data transmission is one of the important ways to solve the problem, and it can ensure the real-time and data quality of data transmission in the city communication environment, many scholars have conducted extensive research in this areas. Vehicle information is now covered by the text, pictures, video and other types of information. In order to improve data transmission quality, many researchers have proposed a variety of compression and transmission methods, for example, Hashemian [1], Sharma [2], Shi [3], Meng [4] et al proposed for the Huffman coding technology for wireless transmission of GPS, including GPS coordinate, records and other text data compression and transmission problems. Hu [5], Jou [6], Wang [7], Barr [8] et al put forward the LZW compression algorithm, which can be optimized by establishing a method to quickly find a dictionary and it can greatly reduce the time of data compression to achieve optimal compression, and improve transmission efficiency. Wavelet compression [9] method for wireless transmission of compressed gradually becomes a hot spot in recent years. In terms of traffic information, based on GML data compressions, Zhang [10] achieved a traffic emergency data and realized a framework of urban traffic emergency system architecture; Through random matrices with restricted isometric conditions of the

***Corresponding author** e-mail address: xueyj@scau.edu.cn

original high-dimensional data onto a low dimensional space, Li [11] realized the data fast and efficient compression, and via a convex optimization algorithm to complete the data decompression in traffic information processing terminal after data transmission; According to the characteristics of traffic data, Zhao [12] using its research and comparative data compression methods based on principal component analysis (PCA) and independent component analysis (ICA) method, respectively. Scholars' research results could furnish some useful suggestions in the transmission terminal data to a certain extent. However, the problem can not be ignored is, because of the unbalanced nature of the distribution of urban mobile communication base stations, urban communication environment vary greatly, the data transmission capacity is in greater differences in different parts of the city, part of the region exists communication blind area. The method proposed by scholars rarely used in acquisition and transmission terminal data on one hand, also rarely take into consideration differrences the problem in the larger urban communication environment for real-time transmission of data, that is, how to realize text, pictures, video and other information integrated transmission while coverage of communication base station is uneven with large difference between the communication environment in urban areas. In addition, the large flow of data transmission needs to consume more energy; low-power transmission [13] has important significance to prolong the length of the various devices work. Therefore, based on priority principle approach for simple algorithms, short time-consuming, minimizing overhead of data collection terminals and the backend server, the pure electric buses status information compression and transmission methods basis on optimized Huffman coding algorithms is proposed in this paper. Data collection vehicle information compression and transmission method can be achieving the optimized by encryption, compression, transmission, reception, decryption process, which is important to improve the transmission efficiency of vehicle information to ensure the integrity of information and transportation industry to achieve energy saving.

2 Optimized Huffman coding algorithms

According to national standards, Ministry of Transportation and the provincial technical standards requirements, the current standard of vehicle information including position, speed, altitude, license plates, start/flameout, lights, driving status, drivers and other types of text that collected through the data collection terminals. In addition, the new energy automotive vehicle information needed to collect additional battery status such as temperature, and overall cell voltage, current and other battery management information. The data with different format, size, and type will be uploaded via the data exchange and the wireless communication module of the vehicle terminal.

Vehicle status terminal with satellite positioning usually set 15 seconds for the data collection period when collecting for pure electric bus vehicle information. The collecting cycle for battery status information of vehicle battery management system is a relatively short period. Part of collecting cycle of vehicle battery management system is in less than five seconds, whose data generated in terms of quantity and frequency are more than the data collected by vehicle status terminal with satellite positioning.

The whole process from the point of view of data transmission, namely, data collection, transmission, reception, storage throughout the process point of view, when compared to a large number of simultaneous transmission vehicle terminal especially, the main pressure is focused on how the backend server receives a large number of short, transmission, integration of data transmitted by the terminal.

The way to solve the problem of large amounts of data transmission and access mainly includes three. First, data optimization in vehicle terminal, extract the key significance with application data, the remaining data is temporarily stored in the terminal memory device. Second, data are identified in the data receiving platform to extract of critical data, but it must guarantee the reliability of data transmission. The last, data are compressed by vehicle terminal firstly, and decompressed on the platform.

The compression-decompression method can ensure the integrity of data transmission, but also can reduce the data communication flow, which can save communication costs, but its disadvantages are also very obvious. Optimization transmission of terminal is with less disadvantages, the main disadvantage is it requires the performance processing each group of data compression, which would lead to transmission time delay. Real-time upload of vehicle information is demanding on safety supervision. Moreover, with higher hardware configuration of vehicle terminal at present, the capable of handling large amounts of data vehicle terminal hardware compression technology is promoted. Data collecting, data compressing and transmitting could be performed at the same time, which could not be simultaneously affectted.

2.1 DATA COMPRESSION ALGORITHMS

Due to the limitations of the vehicle terminal hardware conditions, the applications for data transmission compression by vehicle terminal were less in the past for a long period of time in the practical. Images, video and other data that takes up a lot of storage space are usually collected in offline mode, namely, images, video and other data in not considering the real-time requirement when used by off-line way.

Although the vehicle terminal hardware technology continues to improve, but data collection, compression and transmission operate at the same time will occupy the terminal necessarily overhead occupy the terminal. Therefore, the selection of data compression algorithms in this paper compared with traditional data compression algorithms. The main principles include that, it must reach the compression effect and the algorithm should not be too complicated in limited bandwidth resources, which would avoid the influence of compression and transmission efficiency, and not occupy too much memory. Taking up too much memory will affect data collection, as well as other applications occupy a larger memory and use a common transmission protocol.

Based on the operating environment consideration of pure electric bus, the selection of compression algorithm for

operating state of the commercial vehicle in this paper must be followed principles of less calculation, fast compression, simple algorithm and easy to implement. The algorithms must to be reached an optimum compression effect, which cannot be too complicated to avoid affecting the compression efficiency of the transmission under the limited bandwidth resources to meet the requirements of the terminal hardware and communications environment. Text data is the main data for the commercial vehicle terminal collecting (including GPS/Beidou positioning information). Optimized Huffman coding technique [1, 15] for effectively lossless compression is proposed for the GPS/Beidou position data and the other text of commercial vehicles operating state to be transmitted in this paper.

Huffman coding principle is: Set consisting of the text is represented as a set of characters with $I=\{I_1, I_2, I_3 \cdots I_n\}$, where I_x represents different characters of text. Assuming the frequency of a character I_x is F_x, the code length is L_x. To make the total length of code source text file of the shortest, you need to determine the coding mode. Making the minimum value of $\sum_{i=0}^{n} F_x L_x$, this encoding is called Huffman coding [3]. Huffman code based on Huffman tree, Huffman tree construction steps are shown as follows [3, 4].

Step 1. Set the given n weights $\{w_1, w_2, w_3, \cdots, w_n\}$ according to n binary trees, where every tree T_x has only one weight w_x is the root of 1, the left and right sub-tree are empty.

Step 2. Select the root node of the binary trees in F with the minimum weights of sub-tree as to construct a new binary tree, and the root of the new lien binary is the value of its left and right sub-tree root node the sum of the weights.

Step 3. Delete the two trees in F, while the new binary was added to F.

Step 4. Repeat Step 2 and Step 3 until F set up with only one tree, this tree is the Huffman tree.

Typically, a large number of repeated characters are existed, such as GPS data. GPS and the other repeated characters text data included by vehicle information can be regarded as redundant information to be removed. On the basis of Huffman compression coding table for rapid processing after compression by contrasting data, then stored in the storage buffer for post-processing. Among them, Huffman compression coding tables are the number of times pre-generated by the background server for text data in the statistical characters, and pre-stored in the Flash of vehicle terminal. The Realization of the process is shown in Figure 1.

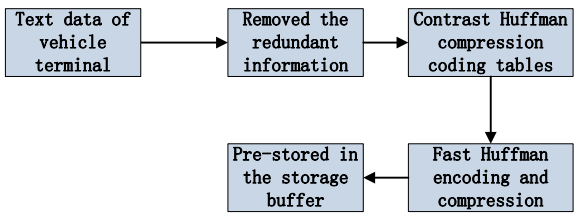


FIGURE 1 the compression flow chart of vehicle terminal text information

Huffman encoding is completely according to the characters of probability to construct coding, there is no error protection function. The algorithm needs to calculate the source character of probability statistics and obtains the probability distribution of the source symbols. This process requires a large cache using the hardware implementation of the algorithm. Generally, it believes that the decoding is complexity, and it is not conducive to the realization of hardware implementation [15, 16]. Huffman coding and arithmetic coding is a typical probability model, many scholars proposed the dictionary model for optimizing the original model, the typical optimizing algorithms such as LZW algorithm [4, 15], etc.

For compression and decompression speed, LZW algorithm is fast, and requires only one scan of the compressed data to be compressed. The compression ratio of the algorithm is very good for the continuous input stream to be compressed with high rate repetition. But the main drawback of this algorithm is the adaptability is rather poor, and the compression rate is less than Huffman algorithm. Moreover, the LZW algorithm usually needs to combine with other algorithms to reach the expected aim for some file with very low complexity. For vehicle operation status data, the data type apparently does not apply to LZW algorithm.

Huffman coding method has certain limitation, such as it need to scan flow of input symbols for two times, be stored or transmitted by Huffman tree before the storage or transmission of Huffman coding results. However, the algorithm has high compression rate, simple and practical, and all the coding has uniqueness while decoding, very suitable for vehicle information data that has high identification requirements. In order to overcome the practical problem of cache, high complexity caused by the Huffman coding, we improved the structure of Huffman tree based on the original. The main steps are as follows:

Step 1. Initialize the established binary Huffman tree, one of the root node has a weight of 0.

Step 2. To encode without the new characters of newly generated two nodes, the connection of the parent node weights add 1, and another node, which is defined as node weights of 0 nodes to define new weight of 0.

Step 3. The new characters have been encrypted through search was coded character location, in the case of the characters of the weight add 1 with coding and compared with the original weight of the same node.

Step 4. Repeat Step 2 and Step 3 until all characters are encoded.

The improved way of Huffman coding by weight set to 0, the original requires two symbols flow is simplified to a scan. The algorithm only needs to scan a probability of 1 characters, and it can only exchange the node between the binary tree needs to node numbering, which reduces the occupation of the cache, and the complexity of the algorithm is decreased. Its deficiencies are mainly embodied in the complete lost of coding error protection function, and the data requirements are relatively high.

2.2 DATA DECOMPRESSION AND PARSING

To distinguish the uploaded data from different vehicle terminal and to ensure uniqueness of uploaded data, uploaded data must be parsed and recognized. Default the parsing code for different vehicle terminal before uploading the data. Read and Match the parsing code by the reception server to achieve data uniqueness verifying. The specific processing is shown as follows.

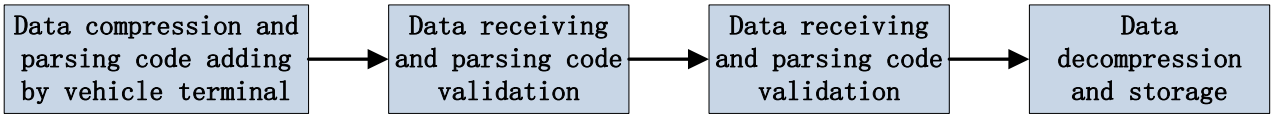


FIGURE 2 The Flow Chart of Data Parsing

Pictured above, it needs to setting parsing code corresponding to vehicle information first. Data compression and parsing code adding are processed by vehicle terminal. Backend server receives the data compression package, and matches the SIM card number, parsing code and other information corresponding to set parsing codes correspondding table. The perfectly matched data will be decompressed and written to the database.

3 Experiments and results

To meet the requirements of real-time data transmission, this paper uses 3G network as a data transmission channel for data transmission. In 3G network, the prerequisite to achieve the theoretical rate is the premise of carrier frequency for data communication, completely and exclusive to a user only. Due to the self-interference system affected by the communication environment, the use of the number of actual formation, the actual data transmission rate and theory of data transmission rate is often of large difference in the vehicle information collection environment.

3.1 DATA TRANSMISSION TEST

Data transmission test mainly completes two test including the network delay test and TCP transmission rate test in the actual environment in city village. Unicom's 3G networks were selected for communication channel for testing in city village environment. Data of vehicle status terminal with satellite positioning and vehicle battery management system were collected and uploaded, then received by server.

Round trip time (RTT) is usually used as a measure of 3G network time delay. Low delay time is of extreme importance in actual application. Lower delay is advantageous for the uplink direction to improve capacity and data throughput, and increase the coverage of high bit rate, obtain higher transmission bit rate. Typically, data transmission testing uses PING method, namely by obtainning response time of sending and receiving data packets to get connected server response time. When sending upload packets in requesting, all the serial number sequence will be added one firstly before sending, the replying information will also be marked the corresponding serial number. Then observe the Ping packet response messages to detect the link for data transmission delay calculation, such as packet loss, grouping repetition and fault sequence etc.

The following table shows the data uploaded test results in a village environment of a city. To be clear, several uploaded tests were started at 10 in the morning, ICMP packets for a total of 120 were send every time, the 3G bandwidth rate of 70-120kbps (city 3G bandwidth rate is about 700-900kbps), TCP transmission rate is close to 3G bandwidth rate.

TABLE 1 Testing results of raw data uploading

Serial Number/Test Items	Maximum Value of Ping Testing	Minimum Value of Ping Testing	Average Value of Ping Testing	Received Volume	Packet Loss Volume	Packet Loss Rate
1 (Packets:120)	7043ms	1190ms	2478ms	115	5	4.17%
2 (Packets:120)	6478ms	948ms	2136ms	117	3	2.50%
3 (Packets:120)	5735ms	1087ms	2332ms	115	6	4.17%
4 (Packets:120)	5845ms	1101ms	2447ms	113	7	5.83%
5 (Packets:120)	6089ms	1017ms	2311ms	116	4	3.33%
Mean	6238.00ms	1068.60 ms	2340.80ms	115.20	5.00	4.00%

TABLE 2 Testing results of compressed data uploading

Serial Number/Test Items	Maximum Value of Ping Testing	Minimum Value of Ping Testing	Average Value of Ping Testing	Received Volume	Packet Loss Volume	Packet Loss Rate
1 (Packets:120)	4546ms	1078ms	2213ms	233	22	8.63%
2 (Packets:120)	4997ms	1210ms	2319ms	230	25	9.80%
3 (Packets:120)	4912ms	1109ms	2190ms	238	17	6.67%
4 (Packets:120)	5221ms	1223ms	2342ms	239	16	6.27%
5 (Packets:120)	4893ms	1005ms	2201ms	234	21	8.24%
Mean	4913.80ms	1125.00ms	2253.00ms	234.80	20.20	7.92%

The result of data packets uploaded test shows that, the maximum delay time is 7043ms, the average delay time of 2340ms, which meets the basic requirements in real time data uploaded. 3G network takes the average packet loss rate 4.17% of data uploading while the network bandwidth is 70-120kbps. Relative to the GPRS data transmission mode, 3G network has been able to meet the requirement of actual application in the setting of upload strategy and selecting transport protocol, but also laid the foundation for the data compression test environment. In addition, it appears from the data uploading

test results, village area of the communication environment is relatively poor, data compression and transmission will still be of importance as well.

Due to environmental influences such as shaking hands error occurred etc., a large amount of data in the case of serious delays caused by transmission network, also had a serious packet loss in the original data uploading test. Direct transmission of original data easily lead to incomplete data, the delay is difficult to reflect the real-time data and so on in 3G networks.

Table 2 is uploading test result by APN for compressed data without identification and screening. The uploading data test method is similar to the original test data uploading method, each transmission is with 255 data packets improved by optimized Huffman coding. The difference between the two methods is to establish the corresponding APN network to upload data using 3G in the same use area.

The results of RTT test in APN network show data compression can effectively reduce the occurrence of packet loss in the area of villages of poor communication environment, while the uploading data is compressed by optimized Huffman coding before transmission.

3.2 DATA COMPRESSION AND DECOMPRESSION TEST

RTT test results showed that the time delay is less, the less the number of packet loss when transmitting a small amount of data. Transmitting a small amount of data of pure electric buses can effectively improve the efficiency of data transmission, but also to ensure the real-time data in the area of poor communication environment.

The data compression testing is shown in Table 3 for data transmission of original, after Huffman Coding, LZW algorithms and optimized Huffman Coding.

The main evaluation on the effect of data compression is mainly based on the compression rate and transmission time. Compression rate is one of the most important indexes to measure the data lossless compression algorithm. Compression rate is a direct measure of data reduction, the running time reflects the complexity of the algorithm, and the combination of the two indexes can be further estimated energy efficiency of communication. Transmission error rate is the rate between data length of accurately praised data had (compressed data containing the decompression process) not received by backend server and the data length before compressing. Packet loss rate is the ratio between the packet numbers of server has not been received and data packets had been transmitted by vehicle terminal.

It should be noted that, according to current technical standards for data collection and transmission requirements, this paper adopts 15s, 10s, 5s three data collection cycle, and the collection time is 300 minutes of pure electric buses collected operating status data as the test data source. This paper also shows the result according to the comparison of the original data, Huffman encoding, LZW algorithm, optimized Huffman coding after data transmission.

TABLE 3 Testing result of data compression

Serial	Data Type	Collect Cycle (s)	Compressed Data Length (kb)	Compression Rate (%)	Transmission Time (s)	Error Rate of Transmission (%)	Packet Loss Rate of Transmission (%)
1	Original	15	335.98	0.00%	49.71	3.71%	6.13%
2	Original	10	479.04	0.00%	75.66	3.99%	6.98%
3	Original	5	943.46	0.00%	169.45	4.65%	8.69%
4	Huffman	15	180.26	46.35%	25.22	2.78%	4.69%
5	Huffman	10	262.68	45.17%	38.41	2.94%	4.99%
6	Huffman	5	511.43	45.79%	91.33	3.05%	5.21%
7	LZW	15	202.32	39.78%	24.33	2.38%	3.89%
8	LZW	10	287.52	39.98%	42.37	2.75%	4.24%
9	LZW	5	569.67	39.62%	104.97	2.98%	4.57%
10	Optimized Huffman	15	142.12	57.70%	19.46	2.24%	3.79%
11	Optimized Huffman	10	204.37	57.34%	27.76	2.39%	3.88%
12	Optimized Huffman	5	408.62	56.69%	68.65	2.49%	4.09%

TABLE 4 Testing result of filtered data compression

Serial	Data Type	Collect Cycle(s)	Compressed Filtered Data Length (kb)	Compression Rate (%)	Transmission Time (s)	Error Rate of Transmission (%)	Packet Loss Rate of Transmission (%)
1	Original	15	115.39	0.00%	17.97	2.56%	1.98%
2	Original	10	180.12	0.00%	28.37	2.62%	2.01%
3	Original	5	367.95	0.00%	57.79	2.89%	2.15%
4	Huffman	15	66.34	42.51%	9.34	2.27%	1.31%
5	Huffman	10	103.43	42.58%	14.03	2.31%	1.44%
6	Huffman	5	207.12	43.71%	29.42	2.35%	1.49%
7	LZW	15	76.53	33.68%	10.64	2.02%	0.89%
8	LZW	10	117.67	34.67%	17.08	2.17%	0.98%
9	LZW	5	236.14	35.82%	31.65	2.29%	1.12%
10	Optimized Huffman	15	53.39	53.73%	7.65	1.83%	0.73%
11	Optimized Huffman	10	82.76	54.05%	9.89	1.97%	0.78%
12	Optimized Huffman	5	169.14	54.03%	21.87	2.08%	0.87%

Experiment result shows that, compressed transmission effect was more notable. The transmission amount of compressed data is reduced greatly, accounting for about between the original amount of data of 1/4-1/3, and the corresponding transmission time is shortened into 1/4-1/3. In addition, it can be seen from the cycle of data collection, data sampling period is relatively smaller and its adjacent data is closer, all kinds of algorithm could get a higher compression rate. In particular, the correlation data has obtained the very good compression rate, which greatly enhances the communication efficiency of system, but also save the cost of communication.

For best results of actual test when using the optimized Huffman coding techniques, the data collection period is 15s. The test results further verify that a lower transmission error rate and transmission loss error can be obtained in the case of the transmission condition of data with less quantity and short transmission time.

The optimized Huffman coding technology has advantages of less calculation amount, high compression rate, and it is suitable for data transmission in the operation area of pure electric bus. For the LZW algorithm, the algorithm has a better compression efficiency for the correlation data in global or local, but because data field correlation of the pure electric bus operation status is small, resulting in limitation of application applied aspects of the object, and thus its compression efficiency is relatively poor in terms of optimized Huffman coding algorithm. To enhance the efficiency of real-time applications of pure electric buses operating status data, and avoid unwanted real-time transmission data occupying a lot of memory and data transmission channel, the paper also carries out data compression and transmission test after the data is filtered and identified, which are shown in Table 4.

The amount of data to be transmitted will be greatly reduced when the filtered and compressed data transmission mode is selected. Real-time data needed to transmitting can be greatly compressed in the transmission mode, which ensure data transmission for real-time application while the error rate of data transmission, data transmission packet loss rate fell to a lower level. Further, it can be seen from the data collection cycle that, the data sampling period is relatively smaller the closer adjacent data is; the data compression rate of various algorithms can be obtained. In particular, an excellent compression rate can be obtained in the case of the data with great correlation, which can greatly enhance the efficiency of communication systems, but also save a lot of communication costs.

In order to further verify the results of data compression algorithm, this paper compares the original data with the parsed decompressed data. Through the data exchanging, vehicle terminal could obtain the speed sensor data. Exchange data through data compression method based on the technology of Huffman coding for data compression, and then the original data file and compressed data files are transmitted to the backend server through 3G network. After transmission, read the received data file for decompression and inverse quantization, and compare the difference between the original data and data contrast reduction before and after data compression.

To be clear, the speed testing data was selected for a passenger bus driving data from 8:00 am to 12:00. An average of 5 seconds data collection cycle was set in the pure electric bus vehicle terminal, and it generated for a total of 2772 (theoretical value should be 2880, because of various reasons in the process of data collection, there are 108 errors, lost or not collected data samples) speed data during the period. Before the data compression, the original speed changing curve is shown in Figure 3. In Figure 3, the horizontal represents the speed of data collection time; the ordinate represents speed value, where are the same as Figure 4. After the data compression and transmission, the parsed speed changing curve is shown in Figure 4.

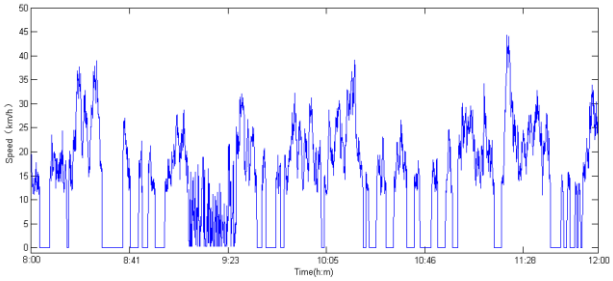


FIGURE 3 The velocity curve before data compression

Figure 4 shows the decompressed and parsed data, which have been compressed, transmitted, parsed, decompressed and warehoused to the backend server.

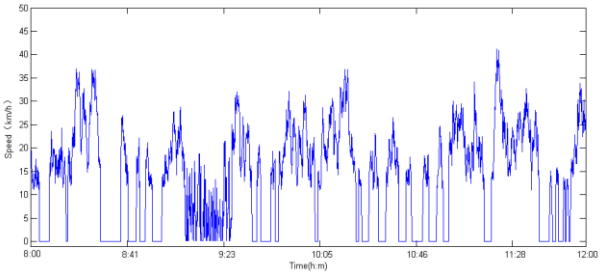


FIGURE 4 The velocity curve after data decompression

From the speed curve of the data before and after compression, the speed error value curve can be seen, the speed data consistent with original data basically before compressed through the process of compression, transmission, parsing, decompression, and data storage. Coding, compression and other data processing technology have met the basic requirements in actual application. To accurately describe the data differences between situation before and after the data compression and transmission, this paper calculated the mean variance to discriminate compression deviation after transmission of the data. After calculation, the mean square difference before and after data compression and transmission is of only 0.2789, which is within a reasonable range. The main reason for the deviation of the data exists in data loss, data failing to parse and the other reason during data transmission.

A total of 2702 data is to accurately have been parsed during 2880 speed data transmitted by vehicle terminal, the proportion of accurate transmission and parsing data is up to 96.25%, which basically meet the requirements of actual application. 108 errors, loss, and no sample data contained

the error data of 25, accounting for 0.87% of the total, loss of data 49, accounting for 1.70% of the total loss rate, no sample data of 34 due to various reasons did not collect, accounting for 1.18% of data should be collected of the total. The availability testing experiment results of compression and transmission algorithm show that, the algorithm has high efficiency and high transmission integrity.

4 Conclusions

Compression algorithm method is researched in this paper to the vehicle information including text information collection types (including position information), image and video to compress the data by using Huffman coding technology. On this basis, all kinds of vehicle information is transmitted based on NMEA-0183 various kinds of vehicle GPS monitoring transmission protocol. Through the process of vehicle information compression, transmission, decompression, parsing and storage whose data from vehicle status terminal with satellite positioning and vehicle battery management system, the compression algorithms are verified by speed data as an example. The algorithm achieved great test results in the experiments, the proportion of accurate data transmission and parse is up to 94.69%, basically meet the requirements of data uploading actual applications in different city region for the urban public transport, passenger transport, taxi and other industries. Meanwhile, the algorithm play an important role for enhancing the efficiency of wireless transmission applications and reducing the cost of data transmission as well as the time, which has the practical significance for the deployment of this base station, the selection of communication mode and the other uses. The data compression algorithms mentioned in this paper provides methods with transmission reliability and minimal distortion. It can provide reliable data transmission method for different mobile base stations coverage of the urban environment, which is important to improve vehicle transmission efficiency of information, ensure the integrity of information and the realization of transport energy conservation.

Acknowledgments

The research is supported by Science Program of General Administration of Quality Supervision, Inspection and Quarantine the People's Republic of China (2014IK183) and Guangdong Research Projects Special Funds of Industry-Academy-Research Cooperation (2012B091100345) and Guangzhou Science and Technology Project (2014Y2-00044).

References

[1] Hashemian R 2004 *IEEE Transactions on Communications* **52**(1) 6-8
[2] Sharma M 2010 Compression Using Huffman Coding International *Journal of Computer Science and Network Security* **10**(5) 133-41
[3] Shi P, Li M, Yu T 2013 Design and Application of an XML Date Compression Algorithm Based on Huffman Coding *Journal of Beijing University of Chemical Technology (Natural Science Edition)* **40**(4) 120-4 *(in Chinese)*
[4] Meng F Y 2010 *Researeh and Application of Huffman Coding in Environmental Protection Real Monitoring System* Thesis of Degree of Master of Ocean University of China 11-13 *(in Chinese)*
[5] Hu P, Zhang J Z 2012 Study and Implementation for Terminal Data Compression Technologies in Remote Fault Diagnosis System *Computer Engineering and Applications* **48**(34) 130-5
[6] Jou J M, Chen P Y 2006 *IEEE Transactions on Communication* **47**(9) 1278-83
[7] Wang Q, Qi C, Luo X 2005 Modified LZW algorithm and its parameters optimization *Journal of Chongqing University of Posts and Telecommunications* **17**(3) 351-5 *(in Chinese)*
[8] Barr K C, Asanovic K 2006 Energy-aware lossless data compression *ACM Transactions on Computer Systems* **24**(3) 250–91
[9] Yang S H, Jia Y, Zhou S W 2013 Virtual nodes-based wavelet compression algorithm for wireless sensor networks Ruanjian Xuebao *Journal of Software* **24**(3) 557-63 *(in Chinese)*
[10] Zhang H Y 2011 *Based on ArcGIS Server Urumuqi Traffic Emergency System Data Compression Technique Research* Thesis of Degree of Master of Xinjiang University *(in Chinese)*
[11] Li Q Q, Zhou Y, Yue Y, et al 2012 Compression method of traffic flow data based on compressed sensing *Journal of Traffic and Transportation Engineering* **12**(3) 113-20 *(in Chinese)*
[12] Zhao Z Q, Zhang Y, Hu J M, Li L 2008 Comparative Study of PCA and ICA Based Traffic Flow Compression *Journal of Highway and Transportation Research and Development* **25**(11) 109-14 *(in Chinese)*
[13] Xu H L, Zhang H, Shen Y, et al 2013 Environment monitoring system for flowers in greenhouse using low-power transmission *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)* **29**(4) 237-44 *(in Chinese)*
[14] Liu X J, Qiu X L, Sun C F, et al 2010 Design and application of knowledge model and PDA-based precision farming system *Transactions of the CSAE* **26**(1) 210-5 *(in Chinese)*
[15] Wang Q, Qi C, Luo X M et al 2009 Modified LZW algorithm and its parameters optimization *Journal of Chongqing University of Posts and Telecommunications* **3**(1) 85-91 *(in Chinese)*
[16] He X M 2011 *Lossless data compression and decompression software and hardware realization* Thesis of Degree of Master of University of Electronic Science and Technology of China 9-11 *(in Chinese)*

Authors

Jingfeng Yang

Current position, grades: post-doctoral student at the South China Agricultural University. Technology expert for transportation information system at Guangzhou Public Transport Data Management Centre and at Guangzhou Transport Information Control Centre.
University studies: MS degree at South China Agricultural University in 2008.
Scientific interests: telematics, GIS application and intelligent traffic, intelligent transportation network and transportation - geographic information system.

Yong Li

Current position, grades: associate professor at Guangzhou Institute of Geography, Guangdong Provincial Academy of Sciences.
Scientific interests: Study on the public traffic system using Geography Information System (GIS).
Experience: modeling and forecasting of the public transit passenger flow, public traffic system of Guangzhou City.

Nanfeng Zhang

Current position, grades: Dr., Senior Engineer, Chief Engineer at Guangzhou Entry Exit Inspection and Quarantine Bureau laboratory.
University studies: PhD degree at South China Agricultural University in 2014.
Scientific interests: mechanical and electrical integration, vehicle detection and fault diagnosis.
Experience: Member of the Society of automotive engineering of China, science and technology expert of Guangdong Provincial Department of Science and Technology, expert of Guangzhou City Auto Repair Association.

Jiarong He

Current position, grades: deputy chief of integrated four divisions at Foshan municipal people's government office, post-graduate student at South China University of Technology.
Scientific interests: software engineering technology.
Experience: participated in big software engineering projects, e.g. at the Department's OA automation construction and the Digital Urban Management projects.

Yueju Xue

Current position, grades: Dr., professor-researcher at the College of Information, South China Agricultural University.
Scientific interests: data mining, big data processing technology.
Scientific interests: data processing technologies, modeling of IOT and intelligent computing.

Allowed capacity calculation of wind farm based on probabilistic constraint

Huiqiu Du[1], Jingbo Wang[2]*, Zhong-hai Lin[1], Pingjian Wang[1], Long Yan[1]

[1]*School of Information and Electronics Engineering, Shandong Institute of Business and Technology, 191 Bin Hai Road, Yantai, China*

[2]*School of Electrical Engineering, Yanshan University, 438 Hebei Western Street, Qinhuangdao, China*

Received 1 October2014, www.cmnt.lv

Abstract

This paper presents a method for penetration capacity calculation of wind farm based on probabilistic constraint. According to the stochastic programming theory, under a confidence interval, a mathematical model that considers the randomness of wind speed is established based on the static safety and power quality constrains. The solving method is proposed according to particle swarm optimization algorithm and stochastic simulation technique too. According to a real power system in Yunnan, the penetration capacity of wind power is calculated under different confidence interval, and the feasibility and effectiveness of this method are verified too. The related factors of affecting the penetration capacity of wind power are analysed, which provide the reference for the planning and design of wind farm.

Keywords: penetration capacity of wind power, chance constraint, stochastic simulation, confidence interval, particle swarm optimization

1 Introduction

Energy crisis and environmental pollution have become the two problems that must be faced in the development of modern society. As a renewable energy, because of low cost and less pollution, wind energy has received great attention around world. In recent years, with the rapid development of power electronic technology, the technology of wind power has made great progress. The national policy of strong support for renewable energy has made a period of rapid development of wind power construction in China. Wind power has becomes one of the most mature, large-scale development and commercialization of new energy technologies [1-14].

The random, intermittence and fluctuation of wind energy result in the wind power dispatching and have a severe impact on grid security, stability and power quality, which is the main factor restricting the installed capacity of wind power and capacity of grid connected. Therefore, to evaluate the maximum wind power capacity of system, that is, allowed capacity of wind power, is the key problem during the planning stage of wind farm.

There are no unified algorithm and formula, because of penetration capacity calculation involving many factors. The system was proposed, based on power structure, load and peak load capacity, which could evaluate the penetration capacity of wind power through power grid [4]. The wind power in Beijing Tianjin Tangshan power grid were assessed comprehensively through comprehensive analysis of the power structure, load characteristics, peak load capacity of thermal power, self capacity, pumped storage power station and tie lines planning, without considering the effect of increasing capacity of wind farm on the system static security. A mathematical model is established for wind power penetration limit in power system based on the static security constrains of power system and a new method for solving the problem is given based on interior point method [5], while there in no considering dynamic constrains related to wind power grid connected operation, which may be the key factor determining the penetration limit of wind power. According to the analysis of environmental uncertainty of wind power grid connected and dependent chance programming theory, the model was established that considered the static operation safety, the traditional power grid output limit constrains, the wind farm power down measures and wind power capacity. The wind power allowed capacity was calculated too [7].

In this paper, we propose a method to calculate the allowed capacity of wind power, based on probability constrains. At the same time, the random of wind power and equation constrains of trend are considered. According to the probability distribution of wind speed, the model is established and solved, based on the Particle Swarm Optimization (PSO) algorithm of stochastic simulation technique.

2 The wind farm and system load model

2.1 WIND FARM

The captured wind power of a fan is a function of wind speed at hub height, and the function can be expressed as

$$p = \frac{1}{2} K \rho S v^3 , \tag{1}$$

***Corresponding author** e-mail: duhuiqiu@126.com

p is the captured wind power; K is the power coefficient and the nonlinear function of the tip speed ratio and pitch angle; S is swept area of wind wheel; v is wind speed at hub height. The relationship between active power generated by the generator and wind speed can be expressed as follow

$$p=\begin{cases}0, v<v_{in}, \text{or } v>v_{out}\\ \dfrac{v^3-v_{in}{}^3}{v_n{}^3-v_{in}{}^3}p_n, v_{in}<=v<=v_n\\ p_n, v_n<v<v_{out}\end{cases}, \tag{2}$$

v_{in}, v_{out}, v_n denote cutting in speed, cutting out speed and the rated wind speed respectively; p_n is the rated power. Because of random of wind speed, the power generated by wind power has the character of random. The probability of wind speed between 0~25m/s is very high and the annual averaged distribution of wind speed in YunNan can be expressed by Weibull distribution function [10]

$$f(v)=\frac{k}{c}\left(\frac{v}{c}\right)^{k-1}\exp\left(-\left(\frac{v}{c}\right)^{k}\right). \tag{3}$$

The probability distribution of wind speed is shown in Figure 1.

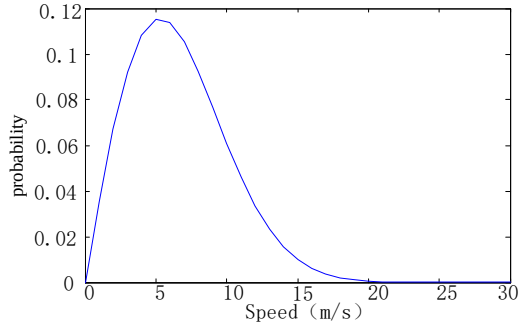


FIGURE 1 The probability distribution curve of wind speed

2.2 SYSTEM LOAD MODEL

Load changes directly affect the wind farm access capacity, experimental results show that, the system load distribution of most area can describe according to the normal distribution function in the long run.

$$\psi(p_{Li})=\frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{(p_{Li}-\mu_i)^2}{2\sigma^2}], \tag{4}$$

where p_{Li} is the system load of the i-th node; μ_i is the average load of the node; σ is variance.

3 Probabilistic constrain model of power system of wind farm

Under the condition of constrains containing random variables, if the decision making under disadvantage may not satisfy the constraints, then the decision making should make the probability of establishment for constraint condition established not less than a certain confidence level [8].

The constraint condition mainly means that power flow do not overload, the node voltage and the output of active power and wattles power of conventional generator are limited. If the loss of net is not considered, the summation of power increment of all nodes in the power grid connected wind farm is zero, in order to meet the power balance. Due to the randomness of active power output from wind power and the volatility of system load, the state variable calculated through system flow is random. The inequality constraints with random variable are expressed as the form of probability, which is established at a specified confidence level. The confidence level has a general value between 0.9 and 1. The value is smaller, the risk is bigger. The value is closer to 1.0, the result becomes more conservative. Taking the accepted largest installed capacity of wind power as goal function, the goal function can be expressed as follow:

$$f=\max(\sum_{i=1}^{n}p_{fi}). \tag{5}$$

The inequality constraints can be expressed as follow

$$\begin{aligned}&\Pr ob\{p_{x\min}\le p_{xj}\le p_{x\max}\}\ge\partial\\ &\Pr ob\{U_{\min}\le U_i\le U_{\max}\}\ge\partial\\ &\Pr ob\{Q_{g\min}\le Q_{gi}\le Q_{g\max}\}\ge\partial\\ &p_{g\min}\le p_{gi}\le p_{g\max}\end{aligned}. \tag{6}$$

The equality constraints can be expressed as follow

$$\begin{aligned}&P_{Gi}+P_{fi}-P_{\mathrm{Li}}=U_i\sum_{j\in i}U_j(G_{ij}\cos\delta_{ij}+B_{ij}\sin\delta_{ij})\\ &Q_{Gi}+Q_{fi}-Q_{\mathrm{Li}}=U_i\sum_{j\in i}U_j(G_{ij}\sin\delta_{ij}-B_{ij}\cos\delta_{ij})\end{aligned}. \tag{7}$$

P*rob* is the probability, p_{xj} is the flow of x line, p_{fi} is the wind power of grid connected at i node; p_{Li} is the system load at i node; p_{xj} is power flow at j line; ∂ is confident probability of inequality constraints.

4 Particle Swarm Optimization algorithm base on stochastic simulation

4.1 THE TECHNIQUE OF STOCHASTIC SIMULATION

The stochastic simulation, i.e., Monte Carlo simulation, provides an effective way to verify the probability constraints. The stochastic variables are sampled from known probability, which provides the basis for decision-making or tests the decision making for system. For chance constraints with stochastic variable

$$\begin{cases}\Pr ob\{p_{x\min}\le p_{xj}\le p_{x\max}\}\ge\partial\\ \Pr ob\{U_{\min}\le U_i\le U_{\max}\}\ge\partial\\ \Pr ob\{Q_{g\min}\le Q_{gi}\le Q_{g\max}\}\ge\partial\end{cases}. \tag{8}$$

The method of random simulation can be expressed as follow: n independent random individual, for example, $v_1, v_2, ...v_n$ are sampled from probability distribution $f(v)$. The n' is the number of satisfying constraints in n trials. According to the law of large numbers n'/n can be used to estimate the value of Equation (8). If $n'/n >= \partial$, the particle satisfies the probability constraints; if $n'/n < \partial$, the particle violates the probability constraints, and the value of fitness is penalized.

4.2 PARTICLE SWARM OPTIMIZATION ALGORITHM

Particle Swarm Optimization algorithm (PSO) stems from the research on group prey behavior of birds was proposed by Kennedy and Eberhart in 1995. Starting from stochastic solution, the optimal solution is searched by means of the iteration and the global optimum can be founded by following the searched optimal value. The vector $X_i = (x_{i1}, x_{i2}, ...x_{in})^T$ denotes the position of particle in space; the vector $p_i = (p_{i1}, p_{i2}, ...p_{in})^T$ denotes the optimal solution of vector of i particle in iteration; the vector $v_i = (v_{i1}, v_{i2}, ...v_{in})^T$ denotes the speed of i particle in N dimensional solution space.

In order to avoid falling into local optimum, the evolution route of particle depends not only on current position also on the previous evolution route. At the same time, the method of linear weight is employed, which ensures fast convergence speed and gets higher convergence precision. The flow- chart of program is shown in Figure 2. The evolution equation of particle is expressed as follow

$$\begin{aligned} v_i(t+1) = {} & w * v_i(t) + c_1 * r_1[Z_{best}(t) - x_i(t)] + \\ & c_2 * r_2[G_{best}(t) - x_i(t)] + \\ & c_1 * r[Z_{best}(t-1) - x_i(t-1)] + \\ & c_2 * r_2[G_{best}(t-1) - x_i(t-1)] \end{aligned}, \quad (9)$$

t is the number of iteration; c_1 and c_2 are learning factor and represent the weight of two optimal solutions respectively; r_1 and r_2 are independent random number between zero and one.

The inertia weight is

$$w = (N-n)\frac{w_{\max}(w_{\max} - w_{\min})}{N}, \quad (10)$$

n is current cycle number; N is total cycle number.

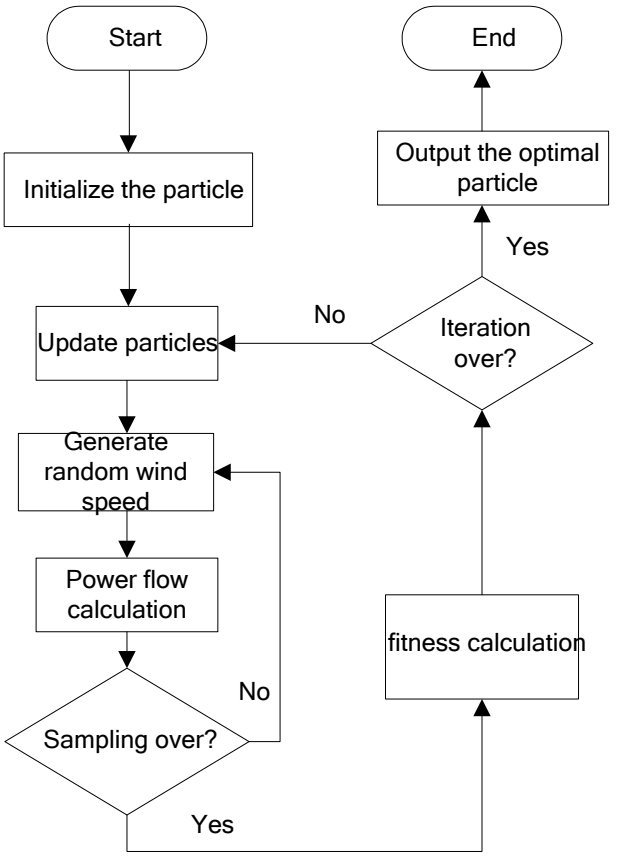


FIGURE 2 The flowchart of PSO based on stochastic simulation

5 Analysis of examples

In this paper a certain practical power grid of Yunnan is used to be analysed and calculated. The system simple topology is shown in Figure 3.

No. 1 generator is defined as the balanced unit. No. 25 and 29 base load generators do not participate in the active power regulation. The rate of wind power is 1.5 MW. The system benchmark capacity is 1000 MVA. No.13, 16, 18, 27 etc. are selected as wind farms and connected grid. The limit and upper limit of active power of conventional unit parameters for optimization are shown in Table 1.

TABLE 1 Conventional unit parameters

Unit node	Lower limit of active power (pu)	Upper limit of active power (pu)
1(Man wan)	0.56	0.94
2(Da chaoshan)	0.243	0.85
3(Zong yanghai)	0.42	0.56

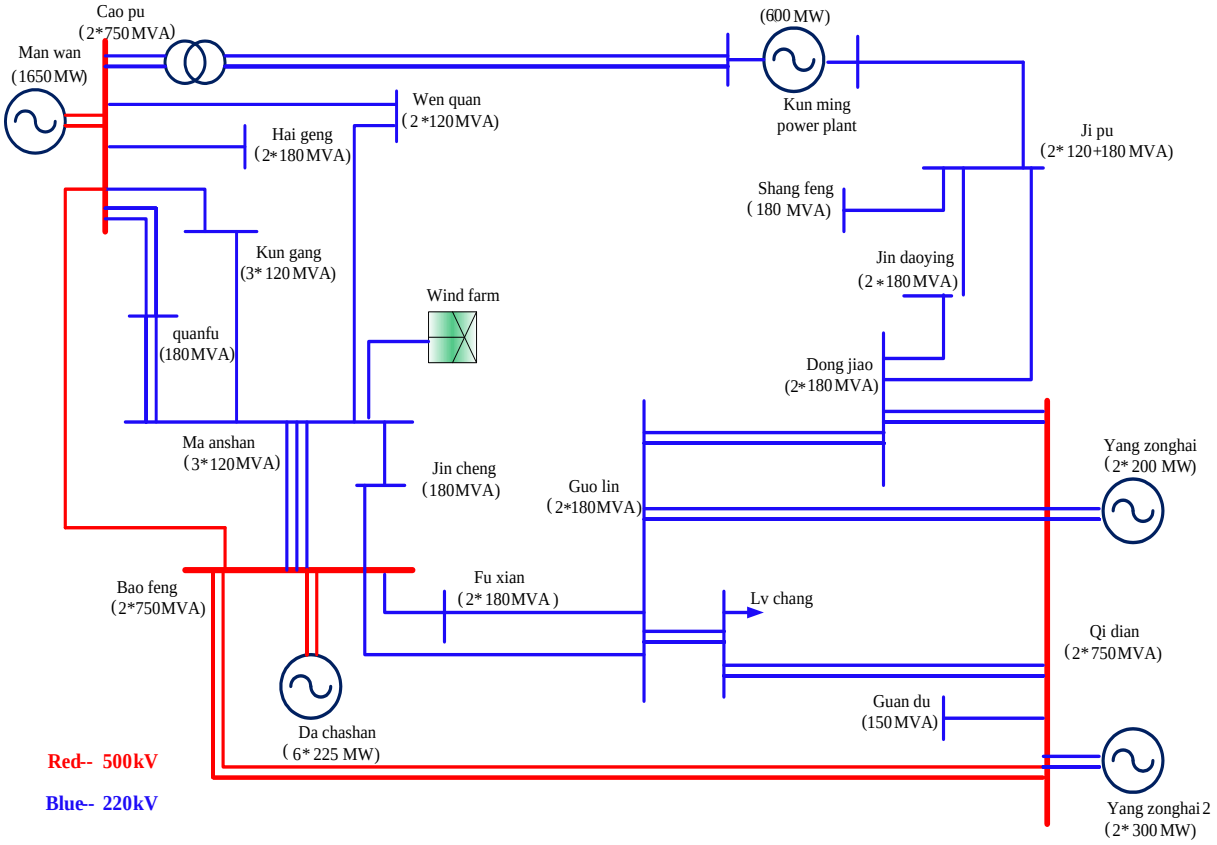


FIGURE 3 The system simple topology of a certain practical power grid of Yunnan

This paper analyses the three factors that affect the wind-power allowed capacity:

(*1*) *Different confidence level and connected grid.* Average wind speed μ =7.5m/s, the shape factor k=2, load variance=1. Wind-power allowed capacities are calculated as shown in Table 2.

TABLE 2 Penetration powers with different confidence levels and connected grids

Connected grid	**Penetration powers (p_u)**		
	α=0.99	*α=0.99*	*α=1.0*
13(Hai geng)	0.5145	0.4095	0.3795
16(Ma anshan)	0.5790	0.4620	0.4260
18(Fu xian)	0.8100	0.6540	0.6065
27(Pu ji)	0.8355	0.6585	0.6075

As seen from Table 3, wind-power allowed capacities are different when wind farms are connected into the system. Different wind-power allowed capacities can be used to determine the best access point for maximizing the acceptance of wind power. Meanwhile, wind-power allowed capacities increase relatively with the confidence probability reducing. The essence of the phenomenon is to allow some low probabilities, which violate constraint conditions, thus conservative calculation results are avoided.

(*2*) *Load levels.* To study, the total load of the system is increased by 10% compared with Table 2. Penetration powers are calculated as shown in Table 3. Due to the demand for electrical energy increasing, wind-power allowed capacities of most nodes increase with load increasing after the system load increases. Figure 4 shows the evolution curve of the optimal solution in the 16th node when α=1.0.

TABLE 3 The effect of load on wind power access capacity

Connected grid	**Wind-power allowed capacity (p_u)**		
	α=0.95	*α=0.99*	*α=1.0*
13(Hai geng)	0.6015	0.4890	0.4500
16(Ma anshan)	0.7875	0.6270	0.5790
18(Guo lin)	0.9150	0.7695	0.7065
27(Pu ji)	1.035	0.8310	0.7675

Compared with Table 2 and Table 3, wind-power allowed capacities of most nodes increase with load increasing after the system load increases. Wind-power allowed capacities of some nodes is not significantly increased because of other restrictions constraints

(*3*) *Wind speed*: The confidence probability is 0.99 in order to investigate effects of different average wind speeds. Other parameters are same as Table 2. Penetration powers are calculated as shown in Table 4 under different average wind speeds.

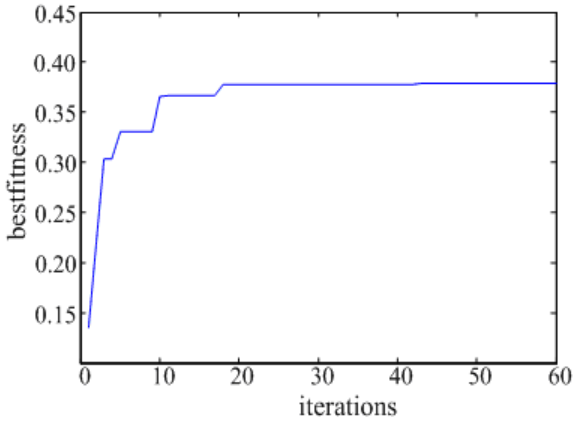


FIGURE 4 Particle-swarm optimal evolutions

TABLE 4 Effect of average wind speed of wind power access capacity

Connected grid	Wind-power allowed capacity (p_u)		
	μ=6.5m/s	μ=7.5m/s	μ=8.5m/s
13(Hai geng)	0.4680	0.4095	0.3915
16(Ma anshan)	0.5130	0.4620	0.4410
18(Guo lin)	0.7425	0.6540	0.6255
27(Pu ji)	0.7440	0.6585	0.6300

Compared with Table 2 and Table 4, the fan output increases with the increase of the average wind speed. The level of system disturbance also increases and the wind-power capacity decreases.

7 Conclusions

In this paper, an analysis method is proposed based on probability constraints. The optimization analysis model of evaluating allowed capacity is established considered wind-power randomness under uncertain environment. The program of MATLAB is developed based on stochastic simulation techniques and particle-swarm optimization algorithms. The analysis of examples demonstrates the feasibility and effectiveness of the algorithm and model. The analytical method of allowed wind-power capacity is also suitable for other renewable energy. The limitation of this method is time consuming because of the use of Monte Cano simulation techniques and only the consideration of the static system constraint. The wind power system dynamic constraint is also a key factor in the decision of the wind power access capability. How to consider these dynamic constraints into the optimization model is the focus of future research.

Acknowledgments

This work was financially supported by the NSFC (61307041) and Natural Science Foundation project of Shandong province of China (ZR2012FL11, ZR2013AL014).

References

[1] Tian Q-q, Zhu L 2011 The countermeasure to accelerate the development of solar energy industry of Yunnan Province *Goufang Zhinan* (3) 125 (*in Chinese*)
[2] Liu Z-w 2007 Research on impact of large grid-connected wind farms to the system and maximum setting capacity Wu Han *Huazhong University of Science and Technology* (*in Chinese*)
[3] Liao P, Li X-y 2008 A Survey on Calculation Method of Wind Power Penetration Limit *Power System Technology* 32(10) 50-3
[4] Sun R-f, Zhang T, Liang J 2011 Evaluation and Applicaion of Wind Power Integration Capacity In Power Grid *Automation of Electric Power Systems* 35(4) 70-6
[5] Lei Y-z, Wang W-s, Yin Y-h, Dai H-z 2001 An Optimization Method for Determining Wind Power Penetration Limit in Power System Under Static security Constrains *Proceeding of the CSEE* 21(6) 25-8
[6] Guo X-r, Wang D-y, Yang G-q 2011 Maximum injection power calculation based on enhanced particle swarm optimization *Electric Power* 44(1) 86-9
[7] Qiao J-g, Xu F, Lu Z-x, Min Y 2008 Optimization Analysis Model of Grid-connected Wind Capacity Based on Chance Programming *Automation of Electric Power Systems* 32(10) 84-7
[8] Charnes A, Cooper W W 1995 Chance-constrained programming *Management Science* 6(1) 73-9
[9] Liao J-l 2008 Calculation and Analysis on The Wind Power Penetration Limit *Storage Transportation & Preservation Of Commodities* 30(5) 104-5
[10] Li Z-y, Wang M, Chen E-y, Gu B-w 1998 Study of Webull-distributing Parameters And Wind Energy Characteristic Value of Wind Speed Around The Parts Of YunNan In Which The Wind Energy Can Be Developed *Acta Energy Solaris Snica* 19(3) 34-9 (*in Chinese*)
[11] Zhang J-h, Yang Q, Li W-g 2010 Maximum Injection Power Calculation Of Wind Farms Connected To Power Systems Based On An Improved Particle Swarm Optimization Algorithms *Acta Energlae Solaris Sinica* 31(5) 630-5 (*in Chinese*)
[12] Wu J, Li G-j, Sun Y-z 2007 Maximum Injection Power Calculation Of Wind Farms Connected To Power Systems Based On Stochastic Programming *Power System Technology* 31(14) 15-9
[13] Wang R, Gu W, Sun R, Li Q 2011 Analysis On Wind Power Penetration Limit Based On Probabilistically Optimal Power Flow *Power System Technology* 35(12) 214-20

Author

Huiqiu Du, China

Current position: associate professor of School of Information and Electronic Engineering, Shandong Institute of Business and Technology, China.
University studies: Master in engineering precision instruments and machinery (1985) from Harbin Normal University.
Scientific interest: robot technology and control.

Wang Jingbo, China

Current position: major in electrical engineering at Yanshan University, Qin Huangdao, China
University studies: master in electrical engineering at Yanshan University, Qin Huangdao, China
Scientific interest: power system economic operation and risk management

Lin Zhonghai, China

Current position: associate professor at Shandong Institute of Business and Technology.
University studies: doctor's degree in electronic science and technology from Xi'an Jiao Tong University, Xi'an, China in 2011.
Scientific interest: semiconductor materials and devices based on nanotechnology.

Wang Pingjian, China

Current position: instructor at Shandong Institute of Business and Technology.
University studies: the doctor's degree in Condensed matter physics from institute of Solid State Physics, Chinese Academy of Science, Hefei, China in 2008.
Scientific interest: theory of condensed state physics.

Yan Long

Current position: instructor at Shandong Institute of Business and Technology.
University studies: doctor's degree in Mechanical and Electronic Engineering from Shandong University, Jinan, China in 2008.
Scientific interest: machine vision and image processing.

Parametric design and simulation analysis of turbine blade

Lida Zhu[1], Hongyu Chen[1], Zhongzi Tian[2], Hai Wang[3]

[1]*School of Mechanical Engineering & Automation, North-eastern University, Shenyang, Liaoning, 110819, P.R. China*

[2]*Beijing Hangke Engine Control System Technology Co., Ltd. Beijing, 102600, P.R. China*

[3]*Shenyang Jin Xin Hongmei Materials Co. Ltd, Shenyang, Liaoning, 110032, P.R. China*

Received 1 October 2014, www.cmnt.lv

Abstract

To improve design quality and efficiency of turbine blade, the geometric model of blade is parameterized by using the quintic polynomial method in this paper. The geometric shape and performance of blade mainly depend on blade profile and cascade with parametric model. First of all, the blade profile lines are designed and generated in different parameters based on the theory of quintic polynomial. The effect of relative parameters on the blade profile is analysed conveniently and rapidly. Then, the geometric shape of blade is generated in the arranged blade profiles according to blade cascade position and consistency. The 3D model of blade is created by the parametric blade profile and cascade. It is shown that parametric model of blade will provide the reference and foundation for application research.

Keywords: turbine blade, quintic polynomial, parametric design, 3D model

1 Introduction

Blade is a main part in aeroengine, which affects directly work performance of engine equipment. Moreover, the geometric shape and performance of blade mainly depend on blade geometric parametric model, so it is important for design process. Since the blade profile is a complex curved surface, the design quality and efficiency have to be improved by some effective methods [1-3].

There are some researchers in parameters design of blade. Song [4] has presented the parametric design of turbine blades based on feature modelling. Bing et al. [5] have developed and achieved three-dimensional design of the mixed-flow pump impeller. The numerical simulation was employed to analyse the effects of the blade parameters. Oh et al. [6] have compared the various parameters in the effects on the cavitation performance of the mixed-flow pump impeller. Yue et al. [7] have presented the multidisciplinary design optimization for aeroengine turbine blade.

This paper is divided into four main sections; the first, entitled the main geometric parameters of blade profile and cascade. The second section: the quintic polynomial designning the blade type line. The next part: the blade type line parameter design. Finally, the 3D model from parametric design.

2 The main geometric parameters of blade profile and cascade

To analyse the blade profile process, it is a necessary to build the parametric model. Blade curve modeling is usually made up of a certain number of sections, which is along the radial in order, while the plain cascade is made up of many same shape blade profiles which is apart some distance. In order to establish the model better, it must consider the cascade in the mass, for the location in the cascade and consistency have an important effect on blade design [7-10]. In this study, the main geometric parameters of cascade are described into two parts: one is the main geometric parameters of blade and another is the blade location's parameters in the cascade, as shown the Figure1.

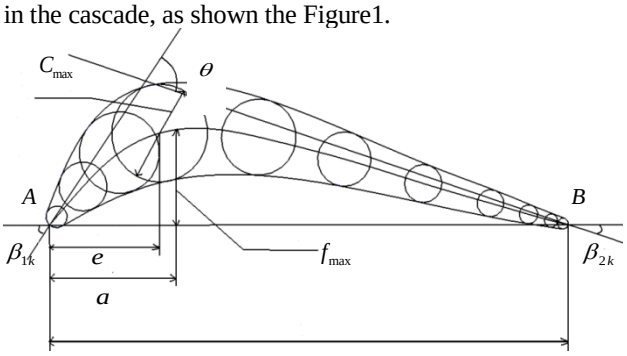


FIGURE 1 Main geometric parameters of the blade

1) Camber line: the line which crosses the centre of inscribed circle usually is called mean camber line. Chord length b: the mean camber line interacts the lending of blade at point A and the edge at point B, the line which links A to B is called chord length, usually is expressed by chord length b.

2) Maximum deflection f_{max} and the relative position: the maximum vertical distance between the mean camber line and chord length is called maximum deflection. Blade front angle β_{1k} and trail angle β_{2k} : the angle between mean camber line tangent in the front point A and chord length is called blade front angle, the angle between mean camber line tangent in the edge point B and chord length is called

trail angle. Maximum thickness $C_{\max}$: the maximum thickness represents blade.

3) Thickness, the maximum diameter of blade inscribe circle is the maximum thickness, the distance between the front point A and maximum inscribe circle centre along the chord length.

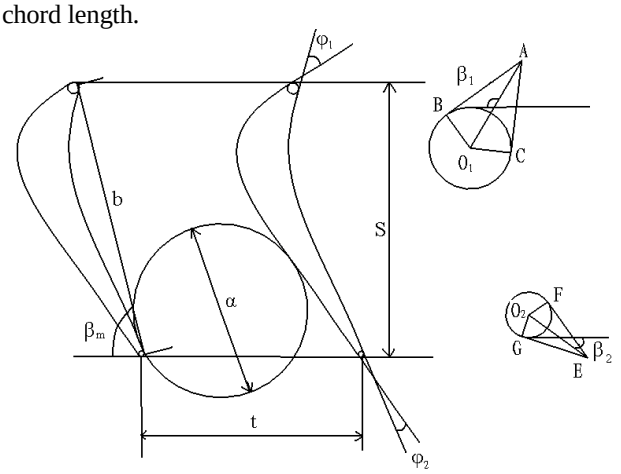


FIGURE 2 Main geometric parameters of blade cascade

4) Blade bending angle θ: the sum of the front angle and the trail angle equals bending angle, the angle between mean camber line tangent in the front point A and in the edge point B represents the bending degree, as the Figure 2 shows: $\theta = \beta_{1k} + \beta_{2k}$. Inlet angle β_1 : the angle between flow direction and frontal line in the leading edge, the frontal line is the line that connects all the blade front point A in the cascade. Exit angle β_2 : the angle between flow direction and frontal line in the trail edge. Blade type installation angle β_m : the angle between the frontal line and chord length represents the blade type's installation in the cascade. Grid pitch t: the distance that two adjacent blade type corresponding points are along the frontal line represents the blade arranges density. The blade section grid pitch shows as: $t = 2\pi R_i / N$. R_i each section radio value. Blade consistency τ : the ratio b/t is called consistency τ , as $\tau = b / t$.

The above parameters are usually used in the blade design, different parameters will design the different blade, but these parameters are not isolated, they can interfere with each other's, sometimes it is impossible that make any parameter to the optimal value, because some of these are incompatible. Such as the selection of chord length and consistency, when selecting the optimal chord length, the blade consistency may be too big or too small, and vice versa. Therefore, parameters selection must be considered synthetically and demonstrated in the test [7, 11-14].

3 The quintic polynomial designing the blade profile

3.1 QUINTIC POLYNOMIAL BLADE MODELLING THEORY

In general, turbine blade type line adopts curve combination, there are some discontinuous curvature derivative points in the curve combination, generating the peak of the velocity and pressure near the blade and degrading the turbine performance. In the analysis of airflow characteristics around the blade: turbine performance mainly depends on the blade surface boundary lawyer flow, while surface curvature radius has an effect on boundary lawyer flow. In order to make the blade surface's velocity and pressure distribute smoothly, it can use the quintic polynomial method to generate the blade type line.

The turbine principle and design parameters determine the flow parameters, formula and experience figure can determine each section geometric parameters, then select the optimal type line. Based on the type line generates the 3D blade. Different parameters generate different blade. The quintic polynomial method can be described that supposes blade type line is a quintic polynomial curve, using special point coordinate and their first derivative and second derivative fits blade section line.

Blade back and blade basin type line are $y_p = f(\mathrm{x})$ and $y_s = g(\mathrm{x})$, which are expressed as follows:

$$\begin{cases} y_p = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 \\ y_s = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4 + b_5 x^5 \end{cases}, \tag{1}$$

y_p and y_s first derivative are expressed as follows:

$$\begin{cases} y_p{}' = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + 5a_5 x^4 \\ y_s{}' = b_1 + 2b_2 x + 3b_3 x^2 + 4b_4 x^3 + 5b_5 x^4 \end{cases}, \tag{2}$$

y_p and y_s second derivative are expressed as follows:

$$\begin{cases} y_p{}'' = 2a_2 + 6a_3 x + 12a_4 x^2 + 20a_5 x^3 \\ y_s{}'' = 2b_2 + 6b_3 x + 12b_4 x^2 + 20b_5 x^3 \end{cases}. \tag{3}$$

Blade basin starting point coordinate $(x_{p1}, \mathrm{y}_{p1})$, first derivative $\mathrm{y}_{p1}{}'$, second derivative $y_{p1}{}''$. Blade basin final point coordinate $(x_{p2}, \mathrm{y}_{p2})$, first derivative $\mathrm{y}_{p2}{}'$, second derivative $\mathrm{y}_{p2}{}''$. Let blade back starting point coordinate $(x_{s1}, \mathrm{y}_{s1})$, first derivative $\mathrm{y}_{s1}{}'$, second derivative $\mathrm{y}_{s1}{}''$. Blade back final point coordinate $(x_{s2}, \mathrm{y}_{s2})$, first derivative $\mathrm{y}_{s2}{}'$, second derivative $\mathrm{y}_{s2}{}''$. Then substituted equation is expressed as follows:

$$\begin{bmatrix} 1 & x_{p1} & x_{p1}^2 & x_{p1}^3 & x_{p1}^4 & x_{p1}^5 \\ 1 & x_{p2} & x_{p2}^2 & x_{p2}^3 & x_{p2}^4 & x_{p2}^5 \\ 0 & 1 & 2x_{p1} & 3x_{p1}^2 & 4x_{p1}^3 & 5x_{p1}^4 \\ 0 & 1 & 2x_{p2} & 3x_{p2}^2 & 4x_{p2}^3 & 5x_{p2}^4 \\ 0 & 0 & 2 & 6x_{p1} & 12x_{p1}^2 & 20x_{p1}^3 \\ 0 & 0 & 2 & 6x_{p1} & 12x_{p1}^2 & 20x_{p2}^3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} y_{p1} \\ y_{p2} \\ y_{p1}^{'} \\ y_{p2}^{'} \\ y_{p1}^{''} \\ y_{p2}^{''} \end{bmatrix}, \tag{4}$$

$$\begin{bmatrix} 1 & x_{s1} & x_{s1}^2 & x_{s1}^3 & x_{s1}^4 & x_{s1}^5 \\ 1 & x_{s2} & x_{s2}^2 & x_{s2}^3 & x_{s2}^4 & x_{s2}^5 \\ 0 & 1 & 2x_{s1} & 3x_{s1}^2 & 4x_{s1}^3 & 5x_{s1}^4 \\ 0 & 1 & 2x_{s2} & 3x_{s2}^2 & 4x_{s2}^3 & 5x_{s2}^4 \\ 0 & 0 & 2 & 6x_{s1} & 12x_{s1}^2 & 20x_{s1}^3 \\ 0 & 0 & 2 & 6x_{s1} & 12x_{s1}^2 & 20x_{s2}^3 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} y_{s1} \\ y_{s2} \\ y_{s1}^{'} \\ y_{s2}^{'} \\ y_{s1}^{''} \\ y_{s2}^{''} \end{bmatrix}. \quad (5)$$

According to the solution set of Equation (4) and Equation (5), the equation of solution will be set to $y_p = f(x)$ and $y_s = g(x)$, and then it will get the blade back and blade basin polynomial expression.

3.2 THE DETERMINING METHOD OF SPECIAL POINT COORDINATE AND DERIVATIVE

When the cascade geometric parameters are known, establish coordinate as the Figure 3 shows, the blade front edge and trail edge circle centre are O_1 and O_2. The inlet edge is in coincidence with y axis and is tangent with circle O_1 at point *d*, exit edge is tangent with circle O_2 .at point *k*, chord length b is tangent with circle O_1, O_2 at point *f*, *h*, y_s and y_p represent a certain point coordinate along the y axis in the blade back and blade basin.

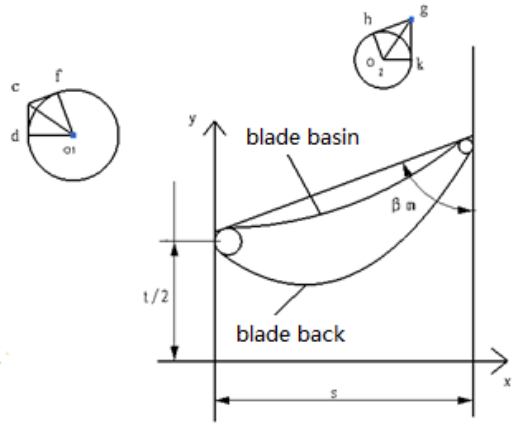


FIGURE 3 Coordinate relations schemes

1) The blade front edge and trail edge key point coordinate:

The front circle centre O_1 coordinates as:

$$\begin{cases} x_{o1} = r_1 \\ y_{01} = \dfrac{1}{2}t \end{cases}. \quad (6)$$

Tangency point *f* coordinates as:

$$\begin{cases} x_f = x_{o1} - r_1 \cos\beta_m \\ y_f = y_{o1} - r_1 \sin\beta_m \end{cases}. \quad (7)$$

The trail edge circle centre O_1 coordinates as:

$$\begin{cases} x_{o2} = s - r_2 \\ y_{o2} = y_f + (s - x_f)\cot\beta_m - r_2 \cot|\beta_m / 2| \end{cases}. \quad (8)$$

Tangency point h coordinates as:

$$\begin{cases} x_h = x_{02} - r_2 \cos\beta_m \\ y_h = y_{o2} - r_2 \sin\beta_m \end{cases}. \quad (9)$$

2) The blade back and blade basin first point coordinate and derivative

Blade back first point *C* coordinates as:

$$\begin{cases} x_{p1} = x_{01} - r_1 \cos|\beta_1 - \varphi_1 / 2| \\ y_{p1} = y_{o1} + r_1 \sin|\beta_1 - \varphi_1 / 2| \end{cases}. \quad (10)$$

Blade basin first point *B* coordinates as:

$$\begin{cases} x_{s1} = x_{01} + r_1 \cos|\beta_1 + \varphi_1 / 2| \\ y_{s1} = y_{o1} - r_1 \sin|\beta_1 + \varphi_1 / 2| \end{cases}. \quad (11)$$

Blade back first point *C* first derivative as:

$$y_{p1}^{'} = \cot|\beta_1 - \varphi_1 / 2|. \quad (12)$$

Blade basin first point *B* first derivative as:

$$y_{s1}^{'} = \cot|\beta_1 + \varphi_1 / 2|. \quad (13)$$

The blade back first point *C* and blade basin first point B second derivative value is 0.

3) The blade back and blade basin final point coordinate and first derivative and second derivative.

Blade back final point *F* coordinates as:

$$\begin{cases} x_{p2} = x_{02} - r_2 \cos|\beta_2 + \varphi_2 / 2| \\ y_{p2} = y_{o2} + r_2 \sin|\beta_2 + \varphi_2 / 2| \end{cases}. \quad (14)$$

Blade basin final point *G* coordinates as:

$$\begin{cases} x_{s2} = x_{02} + r_2 \cos|\beta_2 - \varphi_2 / 2| \\ y_{s2} = y_{o2} - r_2 \sin|\beta_2 - \varphi_2 / 2| \end{cases}. \quad (15)$$

Blade back final point *F* first derivative as:

$$y_{p2}^{'} = \cot|\beta_2 + \varphi_2 / 2|. \quad (16)$$

Blade basin final point *G* first derivative as:

$$y_{s2}^{'} = \cot|\beta_2 + \varphi_2 / 2|. \quad (17)$$

According to the above method, substituting the basic parameters of blade section to the above equals, it will get the quintic polynomial curve, and then determine the type line expression.

4 The blade type line parameter design

The program is developed by using mathematical software in the different parameters to generate the blade section type line, it can save time and clearly find that the effect of each

parameter on blade type line and relationship between the parameters and blade type line.

Using the MATLAB Graphical User Interface module, author develops the blade type parameter interface, which is based on MATLAB research. The reason using the GUI module is that computer operation user interface display in GUI graphics mode. Compared with the early computer command line interface, graphics mode interface makes user more easily acceptable in vision and is convenient to the non-professional user. It allows people to avoid remember a lot of basic commands. People can simply click the window menu to accomplish the design conveniently. Different parameters generate different design in the interface, automatic generating the blade type line. Based on the figure parameters establishes the blade section line: the interface can analyse the interaction relationship between the parameters and the blade type line.

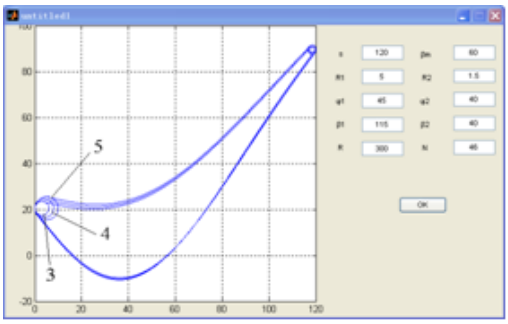


FIGURE 4 Effect of leading edge radius on type line of blade

The installation angle generates the blade type line in different parameters (50°, 60°, and 70°) in Figure 5. The installation angle is smaller, the type line is steeper, and otherwise the type line is smooth.

4.2 THE INFLUENCE OF INTAKE SIDE ANGLE OF AND AIR ANGLE ON TYPE LINE OF BLADE CROSS-SECTION

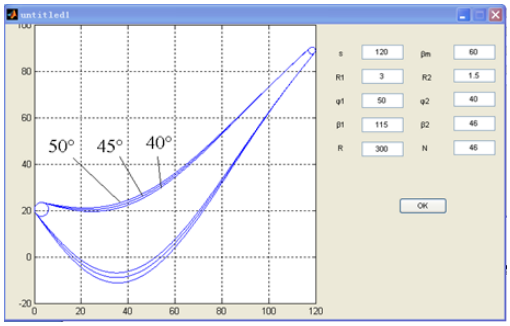


FIGURE 6 Angle of the intake side effects on type line of blade

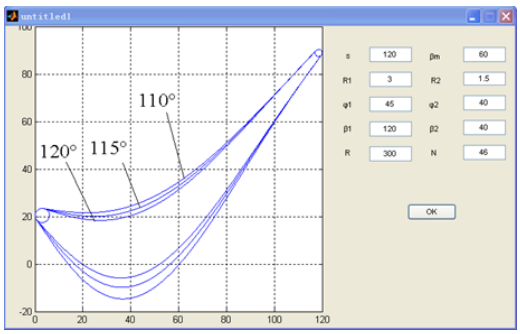


FIGURE 7 Air intake angle effects on type line of blade

4.1 THE INFLUENCE OF EDGE RADIUS AND INSTALLATION ANGLE ON THE TYPE LINE OF BLADE CROSS-SECTION

The Figure 4 shows the radius of different parameters (R1=3, 4, 5) generate the type line curve. The lending edge radius becomes larger; the blade thickness will be bigger, the blade basin and blade back curvature will also increase.

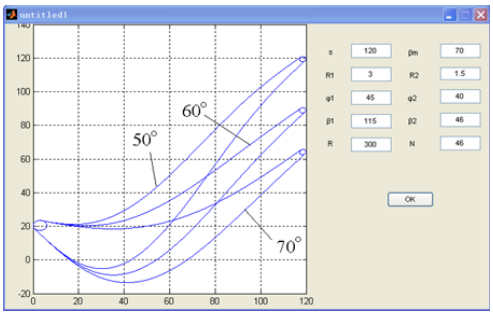


FIGURE 5 Effect of Installation angle on type line of blade

As the Figure 6 shows, with angle of the intake side increasing, blade section thickness will increase, yet the blade bending curvature decreases. In the Figure 7, when the air intake angle is 110°, the blade cross-section type line curvature is minimum, while 120° is maximum. The 3D model is created in different parameters based blade profile and cascade, as shown in Figure 8:

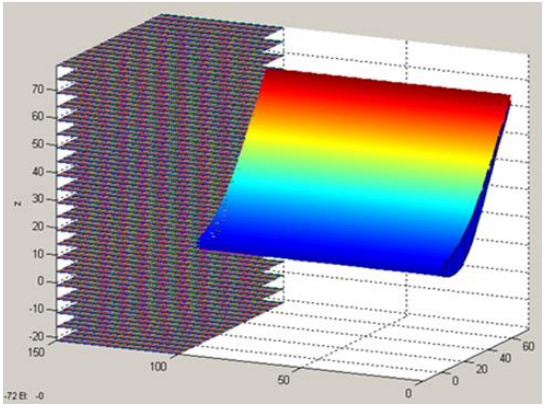


FIGURE 8 3D model of blade in different parameters

5 The 3D model and data exchange

3D model is established by using Unigraphics software based on the blade profile, which results from drawing many curves, ranging them in certain order, and then generating the blade solid model. Utilizing UG law curve function mass, it can draw the quintic polynomial, as shown in Figure 9. Finally, we draw the blade section type line by UG.

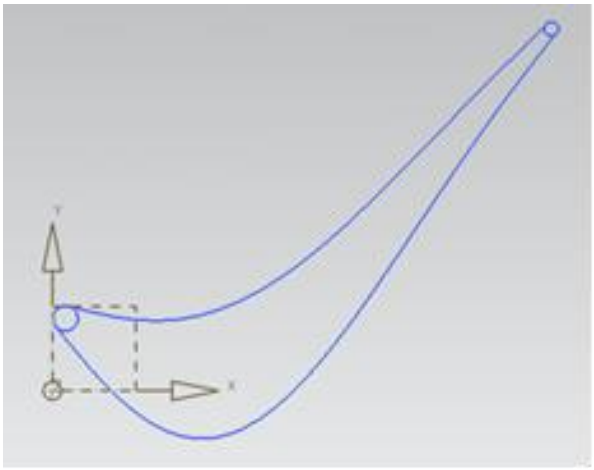


FIGURE 9 Type line of single blade cross-section

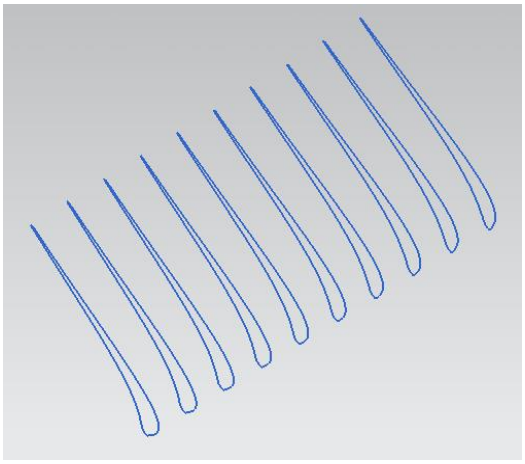


FIGURE 10 Type line of multi-blade cross-section

Many section type line curves are obtained by using the same method, as shown the Figure 10. Generated the type line curve in UG, it will be generated the blade solid model, as shown the Figure 11:

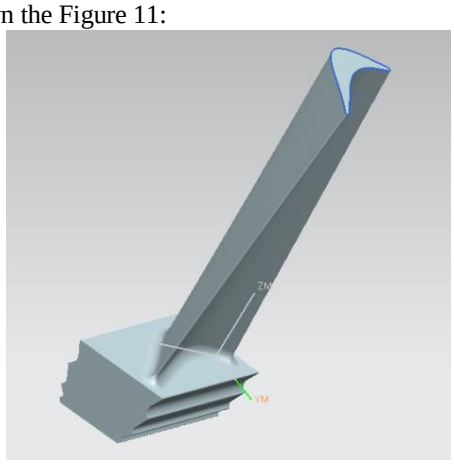


FIGURE 11 Cross-section and three-dimensional model of blade

6 Conclusions

1) In the first part of this paper, the blade design parameters and quintic polynomial method are introduced in detail. The performance of turbine depends on the boundary layer flow on blade surface; In the meanwhile the boundary layer flow is affected by curvature radius of blade surface. In order to get a smooth distribution for surface pressure and speed, quintic polynomial method is a better chose.

2) As the study presented, the blade profile generation interface, whose characters are convenient, rapid, and accurate, combined with quintic polynomial method is used to generate blade profile. As the figures got from simulations with different parameters shows, the type line has correspondent profiles for blade, from which we can understand the influence relation between different parameters clearly.

3) Finally, according to the theory of quintic polynomial method for blade profile line, the blade modeling method is studied, and based on which the blade profile line is generated by MATLAB and the 3D model by UG. Thus, quintic polynomial method can be proved to be a feasible method for generating blade model.

Acknowledgments

This work was supported by (National Natural Science Foundation of China) NSFC (51105072) and (51475087), supported by Fundamental Research Funds for the Central Universities (N130403010).

References

[1] Ismail F, Ziaei R 2002 International Journal of Machine Tools & Manufacture **42** 115-22
[2] Lazoglu I, Boz Y, Erdim H 2011 CIRP Annals-Manufacturing Technology(60) 117-20
[3] Erdem O L, Erhan B 2009 International Journal of Machine Tools & Manufacture 49 1053-62
[4] Song Y, Xi P 2004 Journal of Beijing University of Aeronautics and Astronautics **30**(4) 321-4 (in Chinese)
[5] Bing H, Cao S 2013 Science China Technological Sciences.**56**(9) 2194–206
[6] Oh H W 2010 Proc Instn Mech Engrs Part A: J Power Energy **224**(A6) 881–7
[7] Yue Z, Li L 2007 Beijing: Science press **9** 22-4 (in Chinese)
[8] Budak E 2000 Annals of the CIRP **49**(1) 31-6
[9] Budak E 2012 MM Science Journal 358-65
[10] Ozturk E, Budak E 2010 Journal of Manufacturing Science and Engineering 132 1-13
[11] Song Q, Ai X, Tang W P 2011 The International Journal of Advanced Manufacturing Technology **55**(9-12) 883-9
[12] Budak E, Çomak A, Öztürk E 2013 CIRP Annals-Manufacturing Technology **62**/1 403-6
[13] Biermann D, Kersting P, Surmann T 2010 CIRP Annals-Manufacturing Technology (59) 125-8
[14] Liang X G, Yao Z Q 2011 Computer-Aided Design **43**(8) 971-8

Authors

Lida Zhu, China

Current position, grades: an associate professor at Northeastern University, China.
University studies: PhD degree in Mechanical Manufacturing and Automation from Northeastern University, China, in 2010.
Scientific interest: high-speed turn-milling technology, chatter stability of thin-wall.

Hongyu Chen, China

Current position, grades: a postgraduate at School of Mechanical Engineering & Automation in Northeastern University, China.
University studies: bachelor degree in Mechanical Engineering & Automation from Northeastern University.
Scientific interest: high-speed turning-milling, grinding.

Tian Zhongzi, China

Current position, grades: Beijing Hangke engine control system technology Co., Ltd., China.
University studies: Master degree in Mechanical Engineering & Automation from Northeastern University in 2011
Scientific interest: High-speed turning-milling, Analysis of turbine blade

Wang Hai, China

Current position, grades: a general manager in Shenyang Jin Xin Hongmei Materials Co. Ltd., China.
University studies: bachelor degree in Mechanical Manufacturing and Automation from Northeastern University.
Scientific interest: modeling and simulation analysis.

A speech emotion enhancement method for hearing aid

Shulan Xia*, Jilin Wang

College of Electrical Engineering, Yancheng Institute of Technology, Yancheng 224051, Jiangsu, PR China

Received 1 October 2014, www.cmnt.lv

Abstract

In this paper, emotional perception of the hearing-impaired patients for hearing aid is investigated, and a speech enhancement algorithm is proposed, which is text-independent and requires less and non-parallel training data. In addition, the conversion of prosodic and spectral parameters is also studied. The Eigenvoice Gaussian mixture model (EV-GMM) is used to transform the F0s and spectral parameters, which is built using multiple pre-stored sources emotional and target neutral speech sentences. In the training and testing stages, the duration modification is utilized to improve the performance of EV-GMM training and converted output quality and an adaptive median filter is proposed to smooth the trajectory of the converted speech. Perceptual and objective experiments are presented, simulation results corroborate the effectiveness of the proposed algorithms.

Keywords: hearing aid, emotional speech enhancement, EV-GMM, duration modification

1 Introduction

Emotional communication is very vital for social activities. However, hearing-impaired patients lack this communication for their defective hearings. Although in recent years, the hearing aid research in China is deeper and deeper [1-4], but the little study focused on emotional problems for hearing-impaired patients has yet still been done. The emotion enhancement for hearing aid is an efficient technique to compensate this problem, which refers to transforming the emotional character of the source emotion to enhance the emotional feelings of the hearing-impaired patients.

One strategy for emotion enhancement is based on voice conversion methods. For example, the GMM-based voice transformation algorithm is directly applied to carry out the emotion transformation in [5], it is found the prosody features mainly dominate the emotion state and the transformation of spectral parameters is not sufficient for conveying the required target emotion. In [6], the relevance of the speech components including F0, residual signal and spectral envelope is investigated.

Another alternative method is based on unit selection. In [7], the GMM classification and regression tree (CART) model has been adopted to transform the prosody from neutral speech to target one. In [8], the F0 contours are generated using hidden Markov model (HMM) and the syllable contours are selected from the database using the cost function. In [9], the emotion transformation based on prosodic unit selection is explored and discussed. It is obvious when the corpus is large, these approaches seem to work better than prosodic and spectral voice conversion methods, however, the emotion database is too large, and hard to design and label, etc.

Based on these works, the emotional speech enhancement system, which is completely text-independent and need less non-parallel training data, is proposed. The feature modifications of pitch, duration and spectral envelope are investigated, and an efficient EV-GMM framework is proposed to improve the performance of emotion enhancement. Meanwhile, the duration modification is incorporated in the training and testing phase to improve the enhancement performance, and the efficient adaptive median filter is also adopted to smooth the enhanced speech and reduce the discontinuity problem.

The paper is organized as follows. Section 2 gives a brief introduction of the emotion corpus. Section 3 describes the proposed prosody and spectrum enhancement methods. Then different experimental results are discussed in section 4. Finally, the conclusions are made in section 5.

2 DATASET

A mandarin emotional speech corpus including five types of emotional states (happiness, sadness, anger, fear and surprise) and one neutral state was established for the experiments. Two broadcast professionals including one male named LIN and one female named HUA were hired for the recording. 100 sentences with no apparent emotional tendency were provided for the material. Each sentence in the six simulating speaking styles was uttered resulting total 600 sentences for each speaker. The speech sentences were recorded in a quiet lab environment with 11.025KHz sampling rates and 16 bit precision, each of them has around 3~4ms valid speech.

The corpus for each speaker is divided into two parts in this paper: source set and target set. The first includes four types of pre-stored emotional speaking styles and one testing emotional one, and the latter contains the neutral speech.

3 Prosody and spectrum transformation

It is well known the prosody transformation plays an important role in emotion transformation, so it can be regarded

***Corresponding author** e-mail: xslnj@126.com

as a particular voice transformation focusing on prosodic level. In this section, the transformations of prosodic features including F0 and duration are investigated, and a spectral transformation is also performed to enforce the performance.

3.1 F0 TRANSFORMATION

3.1.1 Baseline F0 transformation

One typical F0 transformation is based on GMM which is previously applied to spectrum transformation [10, 11]. Denoting the F0s of source emotional and target neutral speech by f_x and f_y respectively, f_x and f_y are modelled by a joint GMM, the probability is as follows

$$p(f_x, f_y \mid \lambda) = \sum_{i=1}^{M} \alpha_i N(f_x, f_y, \mu_i, \Sigma_i) . \quad (1)$$

And the converted function can be written as

$$F(f_x) = E(f_y \mid f_x) = \sum_{i=1}^{M} p_i(f_x)[\mu_i^y + \frac{\Sigma_i^{yx}}{\Sigma_i^{xx}}(f_x - \mu_i^x)], \quad (2)$$

$$p_i(f_x) = \frac{\alpha_i N(f_x, \mu_i^x, \Sigma_i^{xx})}{\sum_{k=1}^{M} \alpha_k N(f_x, \mu_k^x, \Sigma_k^{xx})} , \quad (3)$$

where $\boldsymbol{\mu_i} = \begin{bmatrix} \mu_i^x \\ \mu_i^y \end{bmatrix}$ and $\boldsymbol{\Sigma_i} = \begin{bmatrix} \Sigma_i^{xx} & \Sigma_i^{xy} \\ \Sigma_i^{yx} & \Sigma_i^{yy} \end{bmatrix}$ are the mean and covariance matrices of the *i*-th component, and the size of GMM components is *M*. In this method, the parallel utterances of source and target are needed for the model training.

3.1.2 Proposed F0 transformation

The proposed F0 transformation is based on EV-GMM algorithms, which is motivated by the voice transformation based on speaker adaptation [12] and many-to-one voice transformation [13], the EV-GMM has similar form as prior GMM, except for the mean value of source speaking style, which takes the form as

$$\mu_i^x = b_i(0) + B_i \omega , \quad (4)$$

where $b_i(0)$ is the bias value for the *i*-th component, $B_i=[b_i(1),b_i(2),\ldots,b_i(J)]$ is a matrix consisting with *J* basis vectors, and $\omega = [\omega_i(0), \omega_i(2), ..., \omega_i(J)]^T$ is a *J*-dimensional weight vector.

The EV-GMM allows the adaptation of arbitrary input new emotional speech by adjusting the values of the weight vector ω, which is estimated by the maximum likelihood eigendecomposition (MLED) [14] method as follows,

$$\hat{\omega} = \int p(f_x^u, f_y \mid \lambda^{ev}) df_y , \quad (5)$$

where λ^{ev} means the EV-GMM model, and f_x^u is a new source emotional F0 sequence for training. Using the GMM trained with pre-stored source emotional and target neutral speech pairs as the initial model and expectation maximization (EM) algorithm, an adapted EV-GMM for the new source emotional speech can be achieved. The transformation of F0s is directly performed using the adapted F0 EV-GMM model.

3.2 DURATION MODIFICATION

As is well known, different emotions have different speaking rates, which are determined by the number of frames. The simple duration modification based on an average linear ratio is adopted and it takes the form as

$$\hat{D}_n = D_e \frac{\bar{D}_n}{\bar{D}_e} , \quad (6)$$

where D_e and $\hat{D}_n$ are the durations of source emotional and converted speech, $\bar{D}_e$ and $\bar{D}_n$ are the average durations of source emotional and target neutral speech respectively.

Duration as an important emotional feature should be taken into account to separate different emotions. Unfortunately, it is overlooked by traditional voice transformation which mainly bases on spectral and F0 modifications. A new strategy is developed to address this issue in emotional transformation framework, the durations of source emotionnal utterances are modified to map those of target neutral ones in the training process of EV-GMM, the durations of source emotional speech can be seen as a pre-processing module, and It is depicted in Figure 1.

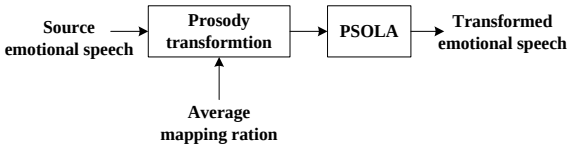


FIGURE 1 A flowchart of duration modification

3.3 SPECTRUM TRANSFORMATION

The spectral transformation works as styles of F0s and the transformation is also straightforward performed using the spectral adapted EV-GMM, which is based on the GMM transformation method as proposed by Kain [6].

3.4 POST-PROCESSING BY RAMF

The emotion transformation is performed on a frame-by-frame basis. One main shortcoming is the discontinuities in the converted speech. A ranked-order based adaptive median filter (RAMF) [15] technique previously applied to image processing has been presented here to remove the unrepresentative points and smooth the converted speech. Assuming *W* is a rectangle filtering window, S_{cur} are the values of current points, $S_{\min}$, $S_{\max}$, S_{med}, are the minimum, maximum and median values of points in the filtering windows. The modified RMAF can be seen a two-level structure: level *A* and level *B*.

Level A: $\begin{aligned} A_1 &= S_{med} - S_{\min} \\ A_2 &= S_{med} - S_{\max} \end{aligned}$. (7)

If $A_1 > 0$ and $A_2 < 0$, then turn to level B, otherwise, increase the size of W to repeat level A until the size of $W > S_{\max}$, then S_{cur} is used as the output.

Level B: $\begin{aligned} B_1 &= S_{cur} - S_{\min} \\ B_2 &= S_{cur} - S_{\max} \end{aligned}$. (8)

If $B_1 > 0$ and $B_2 < 0$, then S_{cur} is used as the output, or S_{med} is adopted as an output.

4 Evaluation experiments

Several objective and subjective experiments were designed to evaluate the performance of the proposed emotion transformation algorithm. On one hand, an objective experiment is conducted to measure the distance of F0 contours between source emotional and target neutral speech. On the other hand, the subjective tests including ABX, mean opinion score (MOS) were conducted by listening test with ten experienced listeners. Five kinds of transformation methods were compared. They are the traditional GMM based transformation method using parallel data, proposed EV-GMM transformation method considering pitch only (PO), pitch and duration (PD), pitch, spectrum and duration (PSD), and pitch, spectrum and duration adding RAMF (PSDR).

In order to train the EV-GMM model, four types of emotions including happiness, sadness, anger and fear were used as pre-stored source speech, and a neutral one was used as target speech, all of them was phonetically balanced utterances and aligned by dynamic time warping (DTW) technique. The first 50 (sentence: 1-50) parallel utterances for each emotional and neutral pairs totally 250 were used for training EV-GMM. The testing emotional transformation is conducted between surprise and neutral, and a traditional GMM based transformation method using surprised and neutral parallel training dataset were used for comparison. Sentences 91-100 of surprise and neutral were used for evaluation, and sentences 51-90 were for the training of tradetional GMM method.

It is noted that F0s were in log-scaled domain, and the 16 order line spectral frequencies (LSFs) were extracted. The sizes of EV-GMMs for F0 and spectrum were optimized as 256 and 512, respectively. the numbers of GMMs for F0 and spectrum were set as 16 and 64, respectively.

4.1 OBJECTIVE EVALUATION

The objective evaluation experiment was performed to assess the performance of prosody transformation, the error measurement in this section is a mean error normalized by the initial F0 distance between source and target speech, which can be written as

$$\xi = \frac{\frac{1}{N}\sum_{i=1}^{N} | y_i - F(x_i) |}{\frac{1}{N}\sum_{i=1}^{N} | y_i - x_i |}, \quad (9)$$

where x_i and y_i are the F0 values for source and target speech respectively, N is the number of frames, and $F(\ )$ refers to the F0 converted function.

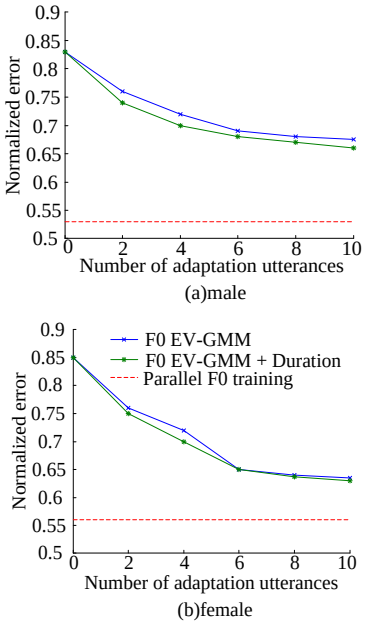


FIGURE 2 Normalized error

Figure 2 depicts the performance of presented prosody transformation method with different number of adaptation data. It can be found that the mean error decreases with more adaptation data, since it correspondences a more accurate F0 modeling. Incorporating the duration modification can greatly enhance the performance of F0 transformation. Moreover, it can be seen when the training data is above four for male and six for female, the errors remain approximately consonant, which means the number of sentences is enough to model the F0 distributions.

4.2 SUBJECTIVE EVALUATIONS

In order to evaluate the similarity between converted and target neutral speech, an ABX test was designed to judge whether X is close to A or B, where X means the converted emotional speech, A and B either the source surprised or target neutral speech.

Table 1 shows the percentages of the converted speech that were closer to the target using the above mentioned methods. It is obvious proposed EV-GMM (with the PCR of 79%) based can achieve satisfactory results compared to traditional GMM using parallel training data (with the PCR of 85%), and the prosody including pitch and duration mainly contribute to the emotion transformation, the spectrum transformation and RAMF module can enhance the transformation performance.

TABLE 1 Percentage of correct responses using ABX test

Methods	Percentage of correct response (%)	
	Male	*Female*
Parallel training	85.1	86.3
PO	69.1	65.3
PD	75.5	72.4
PSD	77.2	76.2
PSDR	79.4	77.5

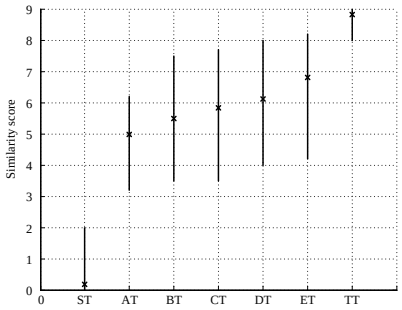


FIGURE 3 Similarity test

A MOS test is also conducted to assess the quality of the converted emotional speech, each converted speech based on GMM and proposed EV-GMM was shown to the listeners, who were asked to rate the quality using 10-point scores with 0 for "totally different" and 9 for "identical". The pairs of speech include source speech (S), target speech (T), converted speech using PO (A), converted speech using PD (B), converted speech using PSD (C), converted speech using PSDR (D) and converted speech using parallel training data (E) respectively. Different sentences were used to make the pairs, so listener can judge the similarity between different speech pairs.

Figure 3 summarized the result of MOS, the mean score in each case was marked "X", and the vertical solid lines indict the variances of scores. The "ET" using parallel training data outperforms other pairs, and "AT" which adopts F0 only for EV-GMM transformation performs worst, the mean score of "DT" is nearer to that of "ET", which indicts the proposed method can achieve an acceptable performance compared with ideal transformation methods using parallel data.

5 Conclusions

A novel emotional speech enhancement method for hearing aid is proposed in the paper, which is based on EV-GMM and relaxes the constraints of parallel training data. The idea of the algorithm is to transforming the emotional character of the source emotion to enhance the emotional feelings of the hearing-impaired patients. Objective test shows the mean error is small and comparable to that using the baseline GMM based transformation using the parallel corpus. The subjective performances also demonstrate the efficiency of presented method that is indicated by listening tests.

Further work will be focused on subjective evaluations for hearing-impaired persons.

Acknowledgments

The work was supported by the National Natural Science Foundation of China under Grant No. 61301219, No. 61375028, No. 61273266 and No. 61301295, the Natural Science Foundation of Jiangsu Province under Grant No. BK20130241.

References

[1] Ruiyu Liang J X, Jian Zhou, Cairong Zou, Li Zhao 2013 An improved method to enhance high-frequency speech intelligibility in noise *Applied Acoustics* **74**(1) 71-8 (*in Chinese*)
[2] Liang Rui-Yu, Xi Ji, Zhao Li, Zou Cai-rong, Huang Cheng-wei 2012 Experimental study and improvement of frequency lowering algorithm in Chinese digital hearing aids A*cta physica sinica* **61**(13) 134305(1-11) (*in Chinese*)
[3] Wang Q, Zhao L, Qiao J, Zou C 2010 Acoustic feedback cancellation based on weighted adaptive projection subgradient method in hearing aids *Signal Processing* **90**(1) 69-79
[4] Liang R, Zou C, Zhao L, Wang Q, Xi J 2012 Experimental study on enhancement method for high-frequency hearing loss in Chinese digital hearing aids *Acta Acustica* **37**(5) 527-33 (*in Chinese*)
[5] Kawanami H, Iwami Y, Toda T, Saruwatari H, Shikano K 2003 GMM-based Voice Conversion Applied to Emotional Speech Synthesis *in Proc Eurospeech, Geneva Switzerland* 2401-4
[6] Barra R, Montero J M, Macias-Guarasa J, Ferreiros J, Pardo J M 2007 On the limitations of voice conversion techniques in emotion identification tasks *in Proc Interspeech* 2233-6
[7] Tao J H, Kang Y G, Li A J 2006 *IEEE Trans on Audio, Speech, and Language Processing* **14**(4) 1145-54
[8] Wu C H, Hsia C C, Liu T H, Wang J F 2006 *IEEE Trans on Audio Speech and Language Processing* **14**(4) 1109-16
[9] Erro D, Navas E, Herndez I, Saratxaga I 2010 *IEEE Trans on Audio Speech and Language Processing* **18**(5) 974-83
[10] Kain A, Macon M W 1998 Spectral voice conversion for text-to-speech synthesis *in Proc ICASSP Seattle USA* 285-8
[11] Inanoglu Z 2003 Transforming Pitch in a Voice Conversion. Framework *Master's thesis St Edmund's College University of Cambridge* 28-32
[12] Mouchtaris A, Spiegel J V D, Mueller P 2004 Non-Parallel Training for Voice Conversion by Maximum Likelihood Constrained Adaptation *in Proc ICASSP Montreal Canada* **1** 1-4
[13] Toda T, Ohtani Y, Shikano K 2006 Eigenvoice Conversion Based on Gaussian Mixture Model *in Proc Interspeech Pittsburgh USA* 2446-9
[14] Kuhn R, Junqua J, Nguyen P, Niedzielski N 2000 *IEEE Trans Speech and Audio Processing* **8**(6) 695-707
[15] Hwang H, Haddad R A 1995 *IEEE Transactions on Image Processing* **4**(4) 499-502

Authors

Xia Shulan, China

Current position, grades: master, the associate professor of Yancheng Institute of Technology, China.
Scientific interest: signal processing, speech signal processing.
Experience: an expert in the field of signal processing.

Wang Jilin, China

Current position, grades: master, the associate professor of Yancheng Institute of Technology, China.
Scientific interest: signal processing, speech signal processing.
Experience: an expert in the field of signal processing.

Optimization and integration method for railway freight stations based on a hybrid neural network model

Yan Sun, Maoxiang Lang*, Danzhu Wang

School of Traffic and Transportation, Beijing Jiaotong University, Haidian District, 100044 Beijing, P.R. China

Received 12 June 2014, www.cmnt.lv

Abstract

Given to the current problems existing in the operation of railway freight stations and the entire railway freight transport network, in order to integrate the railway freight stations and optimize the traditional railway freight transport mode, we first propose a strategy on the optimization and integration for railway freight stations, then design a hybrid neural network model to recognize the operating performance of each railway freight station by classifying them into four ranks based on the proposed strategy. The characteristic of the proposed model is its combination of the respective advantages of unsupervised learning algorithm based neural network and supervised learning algorithm based neural network. Finally, an empirical study from Hohhot Railway Administration is given to verify the feasibility of the proposed model. The simulation results of the empirical study indicate that (1) the accurate recognition of training samples has significant influence on the classification result; (2) the proposed model can recognize the operating performance of the railway freight stations under relatively high accuracy.

Keywords: Railway Freight Station, Optimization and Integration, Hybrid Neural Network Model, SOM Neural Network, Probabilistic Neural Network

1 Introduction

Railway freight stations are the nodes where goods get into or out of the railway network, rail wagons get classified and freight trains get sorted [1]. They play an important role in the railway freight transport organization. And there are enormous railway freight stations dispersed on the wide railway network in China, which provides the freight market with great convenience in a long period. However, with the development of social economy, optimization and adjusttment of industrial structure, and perfection of logistics industry in recent years, the traditional transport organization based on the dispersed layout of railway freight stations cannot adapt the tendency of the freight transport intensive development. Therefore, problems existing in the operation of railway freight stations tend to be obvious, such as insufficient shipment, inefficient operation, and delay of goods transport, waste of transport resources and so no. Railway freight transport hence falls behind compared with other transport modes, for example, highway freight transport. In 2006, China Ministry of Railways of China proposed "Two Integration, One Construction" revolution to overcome these problems. Since then, studies on optimization and integration for railway freight stations have been placed great emphasis on.

Optimization and integration for railway freight stations can be considered as a typical classification problem based on the operating performance of each railway freight station. According to their operating performance, railway freight stations can be classified into different ranks. Railway freight stations in different ranks will play different roles in the railway freight transport network in order to make up a flatter railway network system. Railway freight transport organization can be then further optimized based on this flatter railway network.

As for the optimization and integration methods, besides qualitative methods, some quantitative methods have been applied in the optimization and integration for railway freight transport resources, including Analytical Hierarchy Process (AHP) analysis [2, 3], fuzzy comprehensive evaluation method [4-6] and Pareto (also named ABC) analysis [7]. However, these methods rely on individual experiences of evaluators to some extent and have less comprehensive consideration of various influencing factors. The limitations of these methods in the calculation are apparent, especially when the scale of data and samples is tremendous. Therefore, study on optimization and integration for railway freight stations still has a large research potential.

Recent studies on the classification problem in other research fields have attached great importance to artificial intelligence, due to its well capacity of processing large size of data, high calculation efficiency and accuracy. As one of the most mature and widely used artificial intelligence methods, artificial neural networks have been proved well feasible in many relevant studies. Some studies are presented as follows.

Azami et al. [8] focused on the BP neural network for the recognition of the quality of GPS satellites. In this study, taking geometric dilution of precision as the evaluation index, the GPS satellites was classified into six classes by six modified BP neural networks, and the resilient BP neural network proved the highest accuracy and the least calculation time in the simulation. Park and Cho [9] studied on

* ***Corresponding author*** e-mail: langmaoxiang@yeah.net

the welding quality evaluation and applied LVQ neural network and BP neural network to classify electrode force patterns into five standard patterns. The experiment result in this study indicated the success rate for the testing samples of BP neural network was higher than LVQ neural network, while its calculation time is longer than LVQ neural network. Tan and Du [10] designed a feature extraction model and combined it with RBF neural network to perform the remote sensing image classification. The classification accuracy of the proposed model was verified by comparing with BP neural network and minimum distance classification method. Tambouratzis et al. [11] adopted hierarchical feature map to optimize the basic SOM neural network. In this study, texts classification problem based on their register and author style was used to verify the feasibility of the modified SOM neural network model. Ryoo et al. [12] described a fuzzy neural network and temperature response curve fitting based method to classify the unknown materials. In this experiment for three materials, the superiority of the modified method was verified that the agreement between measured curve and approximated curve was suitable. Using gene expression data, Sun et al. [13] applied discrete wavelet transform-based feature extraction and probabilistic neural network jointly to classify tissues into six classes. In the empirical study, the proposed DWT-based method possessed higher accuracy than the no DWT methods. Wang et al. [14] presented a novel tumor classification approach by combining probabilistic neural network with a neighborhood rough set model. In this study, tumor was classified into five classes by different gene datasets. 4-fold accuracy of colon dataset, leukemia dataset, and SRBCT dataset was up to 96.77%, 100% and 100%, respectively. Debska and Guzowska-Świder [15] proposed an artificial neural network based classification model to recognize the quality of beer. In this study, thirteen elements were selected as the characteristic of the quality of beer. RBF neural network model and MPL neural network model were applied to undertake the classification. The simulation result indicated that artificial neural network techniques allowed then discrimination between qualities of beer samples with up to 100% of correct classifications. Huang and Pan [16] used a probabilistic neural network to carry out the classification of operating performance of the enterprises. In this study, the input data of the neural network were selected by data mining technique. After inputting fifteen variables from five aspects, probabilistic neural network could give higher classification accuracy compared with BP neural network.

On the basis of the previous studies above, in Section 2, we propose a strategy on optimization and integration for railway freight stations in Section 2. In Section 3, we analyse the characteristic of supervised learning algorithm based neural network and unsupervised learning algorithm based neural network. In Section 4, by combining the respective advantages of the two kinds of neural networks, we definite a hybrid neural network model for the recognition of operating performance of the railway freight stations and present its modelling process. In Section 5, an empirical study from Hohhot Railway Administration is used to verify the feasibility of the proposed model and explain the importance of the accurate recognition of training samples. Finally, the conclusions of this study are drawn in Section 6.

2 Strategy on optimization and integration for railway freight stations

2.1 OPERATING PERFORMANCE OF RAILWAY FREIGHT STATIONS

Due to the difference of the environment where a railway freight station locates and the transport resources it owns, railway freight stations will perform different operating performances in the railway freight transport organization.

In this study, we use four ranks to distinguish the operating performance of the railway freight stations by referring some relevant studies on enterprise management [16, 17]. Operating performance rank designed in this study is as shown in Table 1.

TABLE 1 Operating Performance Rank of Railway Freight Stations

Rank	Operating Performance	Representation
I	Excellent	(1,0,0,0)
II	Good	(0, 1, 0, 0)
III	Medium	(0, 0, 1, 0)
IV	Poor	(0, 0, 0, 1)

In order to describe the characteristic of a railway freight station and further apply it to recognize which rank the operating performance of this railway freight station belongs to, we mainly take four aspects into consideration, including social environment factors, freight operation factors, transport location condition factors and freight transport facility factors, and establish an evaluation index system, which is as shown in Figure 1. If the i^{th} (i=1, 2, …, 16) index is represented by x_i, the characteristic of a railway freight station can be described by an index vector $X = (x_1, ..., x_i, ..., x_{16})$.

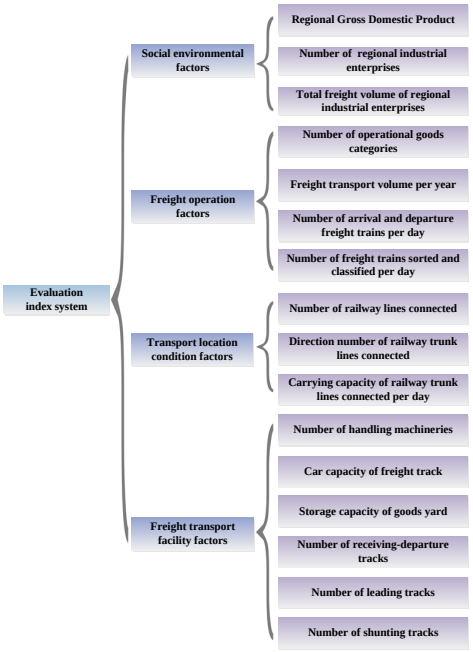


FIGURE 1 Evaluation Index System of the Operating Performance of Railway Freight Stations

2.2 OPTIMIZATION FOR THE RAILWAY FREIGHT TRANSPORT ORGANIZATION BASED ON THE OPERATING PERFORMANCE OF RAILWAY FREIGHT STATIONS.

Based on the ranks in Table 1, railway freight stations are classified into 4 ranks. Due to the poor operating performance of the railway freight stations in Rank IV, they should be closed and the transport resources occupied by them should be recycled by railway freight stations in Rank I, II and III. As for railway freight stations from Rank I to III, the roles they play in the railway network are given as follows.

Railway freight stations in Rank I are the 1st layer nodes in the railway network. They are the sites where through freight trains from different directions get sorted or classified. They locate in the crossing of railway trunk lines and possess adequate facilities corresponding with the sorting and classification of the freight trains. Railway freight stations in Rank II are the 2nd layer nodes in the railway network. They organize the highway-railway intermodal transport and undertake loading and unloading, storage, sorting, flitting and distribution of goods. Therefore, railway freight stations in Rank II should have well railway brunch line transport conditions and possess adequate facilities including handling machines, freight tracks and goods yard, as well as better freight operation environment. Railway freight stations in Rank III are 3rd layer nodes. They are the most widespread nodes in the railway freight network. These stations are the end of the internal freight transport chain. Their main operations include providing pickup and home-delivery services and conducting the freight marketing business. The railway freight network can be divided into several freight organization zones composed of "one 1st layer railway freight station, a few 2nd layer railway freight stations and several 3rd layer railway freight stations" as shown in Figure 2.

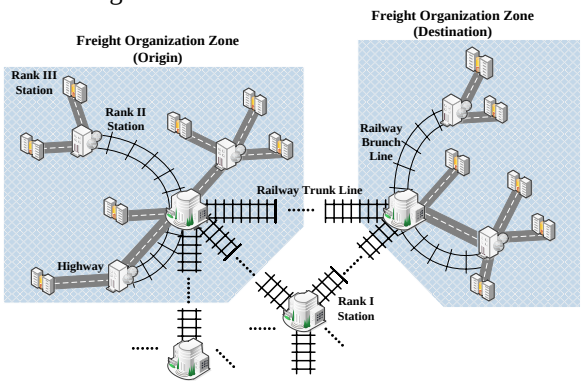


FIGURE 2 Optimized Operation Mode of Railway Freight Transport Network

According to the analysis above, the operation mode of railway freight transport network can be optimized as shown in Figure 3. Figure 3 indicates the basic transport organization mode in the railway freight transport network. In the transport organization, railway transport plays a dominated role and is the main force of the transport organization. Highway transport is the supplement and its participation can enhance the flexibility of the transport organization and extend the freight transport service chain as well. Different composite modes can be selected based on the goods volume and the centralization degree of the goods flow destinations.

In the optimized operation mode, the dispersed railway freight transport resources can be integrated, the advantage of highway and railway in short and long distance transport can be taken, and the operation of railway freight transport can be simplified.

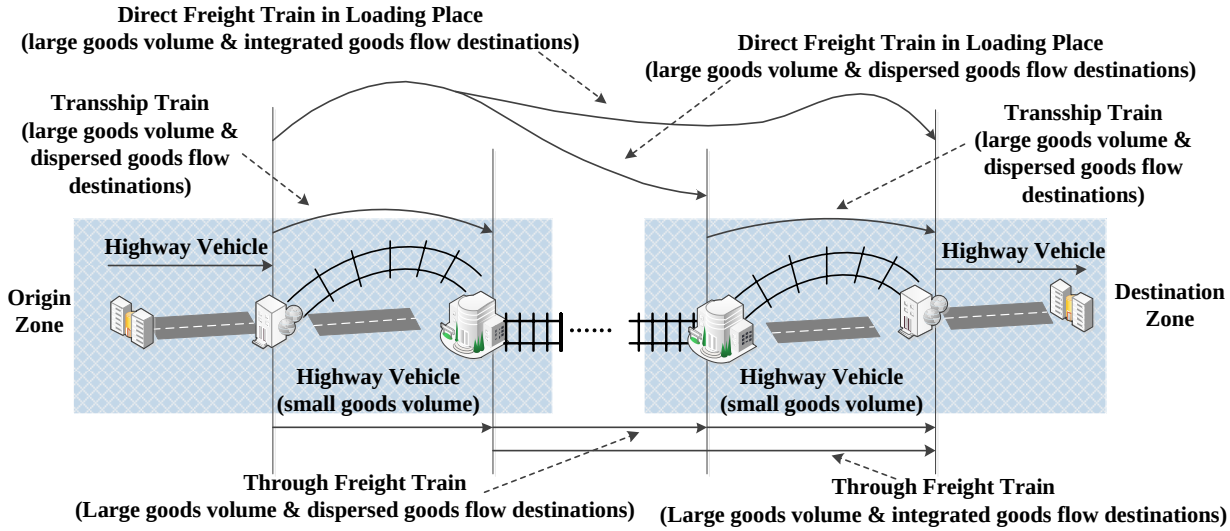


FIGURE 3 Transport Mode in the Optimized Railway Freight Transport Network

3 Supervised and unsupervised learning algorithm based neural networks

According to the learning algorithm, artificial neural networks can be mainly classified into 2 categories, including supervised learning algorithm based neural networks and unsupervised learning algorithm based ones. Their respective characteristic is stated as follows.

3.1 SUPERVISED LEARNING ALGORITHM BASED NEURAL NETWORKS

The common characteristic of the supervised learning based neural networks is the requirement of their learning process for the training samples whose outputs should be determined in advance. Once the training samples and their outputs are given, the classification result will be a certainty. Therefore, the classification accuracy depends on the recognition of the training samples. In the previous study, the training samples were usually selected by individual experiences, which may result in the conflicts or mistakes due to different experiences and will reduce the classification accuracy, especially when the classification ranks are various and the data size of the experimental sample set is tremensdous.

Supervised learning based neural networks include BP neural network, linear neural network, RBF neural network and its deformation modes (generalized regression neural network and probabilistic neural network) and so on.

Compared with other neural networks, probabilistic neural network has many advantages in the classification, such as simplified operation, faster convergent rate, higher stability and better fault tolerance [18, 19].

3.2 AN UNSUPERVISED LEARNING ALGORITHM BASED NEURAL NETWORK

SOM (Self-Organizing Map) neural network adopts an unsupervised and competitive learning algorithm to map the multi-dimensional data onto a 2-dimensional map [20, 21], which can classify the samples into several clusters without determining the training samples and their expected output in advance and can avoid subjective affect from individual experiences consequently. However, the output value of the neural network depends on its parameter setting, including the learning rate and the number of competitive neurons and maximum iteration times, which results in the uncertainty of the classification result.

4 Modelling of the hybrid neural network model for the recognition of the operating performance of railway freight stations

4.1 DEFINITION OF THE HYBRID NEURAL NETWORK MODEL

Generally, the selection of training samples and recognition of their outputs are determined by the researchers, for example, studies presented in [8-16]. Due to the different experience of different researchers, the selection and recognition of the training samples may be various, which leads to conflicts and even mistakes existing in the classification result. Sometimes, the wrong recognition of some training samples leads to the relatively low classification accuracy. It easily happens especially when there is tremendous information to be processed.

Therefore, based on the analysis in Section 3, we design a hybrid neural network model to recognize the operating performance of the railway freight stations. The hybrid neural network consists of two parts: SOM neural network for the preliminary recognition and probabilistic neural network for the accurate recognition. The respective advantages of the supervised learning algorithm based neural network and the unsupervised learning algorithm based neural network can be combined in this model.

According to the preliminary recognition result by SOM neural network, we can select accurate recognized samples as training and testing samples, and use them to train and test the probabilistic neural network. Finally, a well-constructed probabilistic neural network can be gained to recognize the validation samples. The block diagram of the hybrid neural network is as shown in Figure 4.

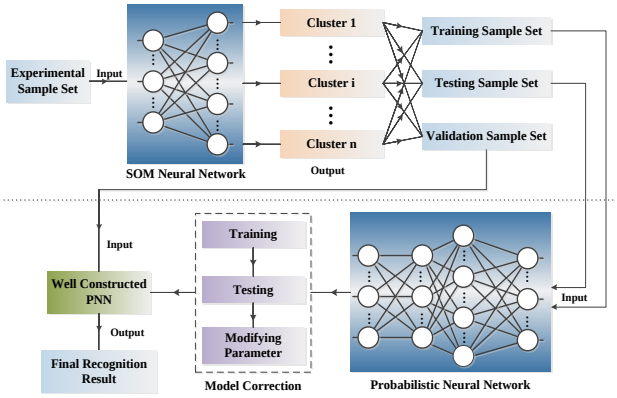


FIGURE 4 Block Diagram of the Hybrid Neural Network Model

4.2 MODELLING PROCESS OF SOM NEURAL NETWORK

SOM neural network is a 2-layer artificial neural network composed of an input layer and a competitive layer. The learning process of SOM neural network consists of three sub processes: competition process, cooperation process and adaption process [21]. Its modelling process is presented as follows.

Step 1. Initialize the SOM neural network.

In this step, the number of competitive neurons (n) and the number of maximum iteration times (T) are determined. Weights connecting the input neurons to the competitive neurons are set in range [0, 1] randomly. In the t^{th} training process, w_{ij}^t is the weight connecting the i^{th} input neuron to the j^{th} competitive neuron, so the weight vector of the j^{th} competitive neuron is $W_j^t = (w_{1j}^t, ..., w_{ij}^t, ..., w_{mj}^t)$, where m is the number of input neurons and equals the number of the evaluation indexes. In this study, m=16.

Step 2. Input the index vector X of a railway freight station by the input neurons.

Step 3. Calculate the Euclidean distance between X and W_j^t by Eq.1.

$$dist_j^t = \left\| X - W_j^t \right\| = \sqrt{\sum_{i=1}^{m} (x_i - w_{ij}^t)} \quad j = 1,2,..,n \tag{1}$$

The Euclidean distance reflects the degree that a competitive neuron matches the index vector of a railway freight station. It will be configured as a parameter in the next step.

Step 4. Gain the winning competitive neuron and its neighbour neuron set.

In this step, if the Euclidean distance between the index vector and the weight vector of the j^{th} competitive neuron satisfies Eq.2, the j^{th} competition neuron is defined as the winning neuron and can be represents by j^*. Its neighbour neuron set $N_{j^*}^t$ can be then determined.

$$dist_j^t = \min\{dist_k^t\}_{k=1}^n, \tag{2}$$

Step 5. Modify weights by Eq.3.

$$w_{ik}^{t+1} = \begin{cases} w_{ik}^t + \alpha_t \cdot (x_i - w_{ik}^t) & k \in \{j^*, N_{j^*}\} \\ w_{ik}^t & else \end{cases} \quad k = 1,2,...,n \quad i = 1,2,...,m \tag{3}$$

where α_t represents the learning rate of the neural network and $\alpha_t \in [0,1]$. Its value decreases with the iteration process.

Step 6. Judge the termination condition is attained or not.

If $t<T$, set $t=t+1$, and then repeat Step 2 to Step 6, otherwise stop the algorithm and give the output value of the input index vector X.

Step 7. Calculate the output value of the competitive layer.

In this step, the output value of the k^{th} competitive neuron and the entire competition layer can be calculated by Eq.4 and Eq.5, respectively.

$$y_k^t = \begin{cases} 1 & k \in \{j^*, N_{j^*}\} \\ 0 & else \end{cases}, \quad k = 1,2,...,n \tag{4}$$

$$y_t = \sum_{k=1}^{n} y_k^t . \tag{5}$$

When inputting the index vectors of different railway freight stations into the neural network, if the output values of the competition layer are equal, these railway freight stations belong to the same rank, otherwise they belong to different ranks.

4.3 MODELLING PROCESS OF PROBABILISTIC NEURAL NETWORK

Probabilistic neural network is a 4-layer feed-forward neural network composed of an input layer, a pattern layer, a summation layer and an output layer. It is a type of classifier based on Bayesian decision and Parzen estimation [22]. Its modelling process of is presented as follows.

Step 1. Input the index vector X of a railway freight station by the input neurons whose number equals the number of evaluation indexes.

Step 2. Calculate the output value of the pattern neurons.

In this step, the ith pattern neuron corresponds with the i^{th} training sample X_i. Using radial basis function as its transfer function, the output value of the ith pattern neuron can be calculated by Eq.6.

$$y_{P_i}(X) = \exp(-\left\| X - X_i \right\|^2 / 2\delta^2), \quad i = 1,2,...,n \tag{6}$$

where n represents the number of training samples.

Step 3. Calculate the output value of summation neurons.

Weights connecting the i^{th} pattern neuron to the j^{th} summation neuron are set as Eq.7.

$$w_{ij} = \begin{cases} 1 & training\ sample\ i \in Rank\ j \\ 0 & else \end{cases}, \quad j = 1,2,...,k \quad i = 1,2,...,n \tag{7}$$

where k represents the number of operating performance ranks.

Using $y_{P_i}(X)$ $(i = 1,2,...,n)$ as its input, the summation neuron will do weighted summation and output the value by Eq.8.

$$y_{S_j}(X) = \sum_{i=1}^{n} [w_{ij} \cdot y_{P_i}(X)], \quad j = 1,2,...,k \tag{8}$$

Step 4. Gain the classification results by output neurons.

The probability of railway freight station Z ∈ Rank j can be calculated by Eq.9.

$$P(R_j \mid Z) = \frac{P(Z \mid R_j) \cdot P(R_j)}{P(Z)}, \tag{9}$$

where $P(Z \mid R_j)$ represents the conditional probability of X, P(Z) is a constant, which has no effect on the classification results. $P(R_j)$ represents the prior probability of Rank j, and can be calculated from training sample set by Eq.10.

$$P(R_j) = \frac{n_j}{n}, \tag{10}$$

where n_j represents the number of training samples that belong to Rank j.

Parzen estimation is applied to estimate the unknown $P(Z \mid R_j)$ by the training sample set according to Eq.11.

$$P(Z \mid R_j) = \frac{\sum_{i=1}^{n} \exp(-\left\|X - X_i^j\right\|^2 / 2\delta^2)}{\sqrt{(2\pi)^2 \cdot \delta^d \cdot n_j}} = \frac{y_{S_j}(X)}{\sqrt{(2\pi)^2 \cdot \delta^d \cdot n_j}}, \quad (11)$$

where d represents the number of evaluation indexes, δ represents the smoothing parameter, and X_i^j represents the ith training sample that belongs to Rank j.

When all prior probabilities of the railway freight station belonging to each rank are all equal, there is $P(Z \mid R_j) \propto y_{S_j}(X)$. Therefore, according to "winner takes all" algorithm [23], the output value of the output neurons can be determined by Eq.12.

$$y_j(X) = \begin{cases} 1 & y_{S_j}(X) > y_{S_q}(X) \\ 0 & else \end{cases}, \quad j, q = 1, 2, ..., m \quad j \neq q \quad (12)$$

Besides, the output values of the probabilistic neural network satisfy Eq.13.

$$\sum_{j=1}^{k} y_j(X) = 1. \quad (13)$$

5 Empirical study: evidence from hohhot railway administration

5.1 DATA COLLECTION AND PRE-PROCESSING

There are 108 railway freight stations managed by Hohhot Railway Administration. Part of their initial data of the evaluation index system are as shown in Table 2.

Before simulation, the initial data should be normalized into range [0, 1] by Min-Max technique (Eq.14) in order to improve the both classification accuracy and calculation efficiency of the hybrid neural network model [24].

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \qquad i = 1, 2, ..., 16, \quad (14)$$

where $x_{\max} = \max\{x_1, ..., x_i, ..., x_{16}\}$ and $x_{\min} = \min\{x_1, ..., x_i, ..., x_{16}\}$. x_i and x_i' represent the data before and after normalization, respectively.

TABLE 1 Initial Data of the Railway Freight Station Samples

Station Name	**1**	**2**	**3**	**……**	**106**	**107**	**108**
	Baotouxi	*Hohhot*	*Fengzhen*	*……*	*Wuhaixi*	*Ceke*	*Salaqi*
x_1	7417253	2321449	921899	……	1201539	1789072	1511753
x_2	2	13	3	……	9	5	6
x_3	235.6	2050.5	664.7	……	641.1	168.6	1756.9
x_4	4	4	4	……	6	10	5
x_5	299.9	1984.1	700.7	……	699.4	239.3	1767.4
x_6	183	28.6	0	……	101	82	26
x_7	109	18	0	……	18	9	4
x_8	3	2	1	……	2	2	1
x_9	2	2	2	……	2	2	2
x_{10}	122	95	102	……	306	146	102
x_{11}	0	22	0	……	1	1	0
x_{12}	16	127	61	……	57	124	0
x_{13}	0	245	1157	……	4647	0	733
x_{14}	31	26	15	……	7	13	4
x_{15}	4	3	0	……	3	1	1
x_{16}	29	3	2	……	11	6	0

5.2 MODEL SIMULATION

In this study, we use Matlab R2012b to perform the model simulation by a Lenovo laptop with Intel Core i5 3235M 2.60GHz CPU and 4GB RAM. In the SOM neural network simulation, the topology of the SOM neural network is set as shown in Figure 5 and the maximum iteration times is set to 500.

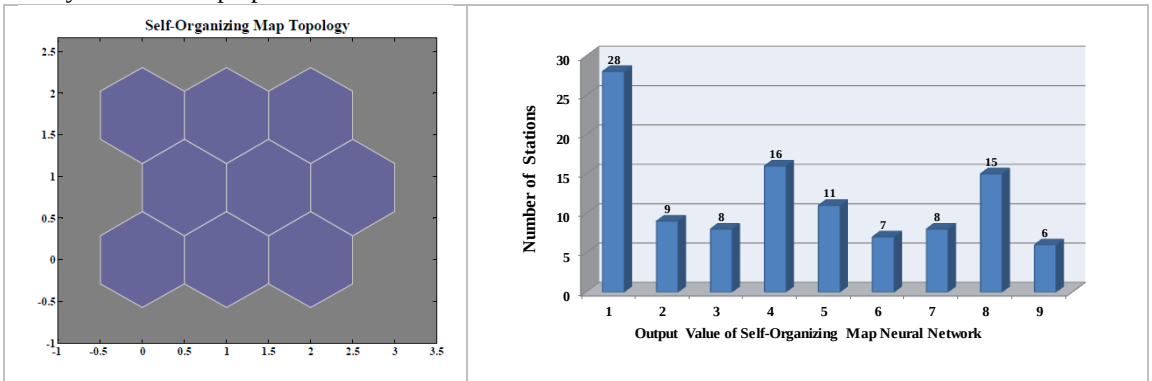


FIGURE 5 Topology of SOM Neural Network

FIGURE 6 Preliminary Recognition Result of the SOM Neural Network

When completing the SOM neural network simulation, the preliminary recognition result is as shown in Figure 6.

Based on the preliminary recognition result, the training samples and testing samples we select are as shown in Table 3. Their expected output corresponds with the binary representation in Table 1.

TABLE 3 Selection of Training Samples and Testing Samples

Rank	Training Samples	Testing Samples
I	Baotouxi, Hohhot, Baotoubei, Linhe	(1) Wuhaixi, (2) Erlian
II	Gongjiban, Baotou, Baotoudong, Fengzhen, Salaqi, Wanshuiquan	(3) Baita, (4) Jiningnan, (5) Saihantala, (6) Wuhaibei
III	Hantaichuan, Hantaichuanbei, Xiangshawan, Erdaoshahe, Gongwusu	(7) Jilantai, (8) Kabuqi, (9) Laoshidan, (10) Lasengmiao, (11) Lasengzhongmiao
IV	Kundulunzhao, Bikeqi, Gongmiaozi, Chasuqi, Toudaoqiao	(12) Tuguiwula, (13) Mingan, (14) Ejina, (15) Xinghe, (16) Daluhao

After several testing, the smoothing parameter we set finally is 1.5. The testing result is as shown in Figure 7. As we can see from Fig. 7, the ratio of correct and wrong number of testing samples is 15:1, which indicates the correct recognition rate is relatively high.

5.3 COMPARISON AND ANALYSIS

In order to verify the importance of the selection of training samples on the recognition result, assume BaotouIVi is mistaken for Rank II, Baotoudong is mistaken for Rank I, Hantaichuan is mistaken for Rank IV, and Toudaoqiao is mistaken for Rank III, then the testing result is as shown in Fig. 8.

The ratio of correct and wrong number of testing samples is 9:7. The comparison between Figure 7 and Figure 8 indicates that the wrong selection of a few training samples can result in the severe reduction of the correct recognition rate of the testing samples, which will further lead to the mistaken in the recognition of the operating performance of the entire railway freight station samples.

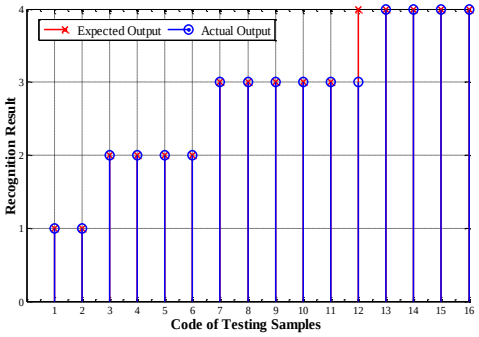


FIGURE 7 Testing Result of the Probabilistic Neural Network

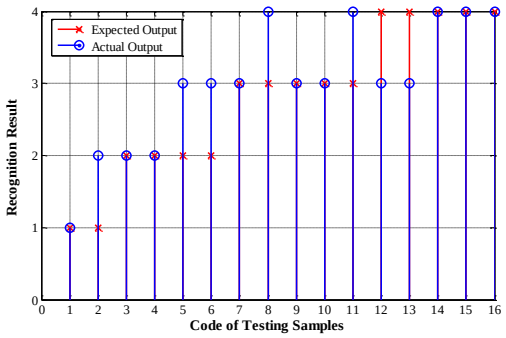


FIGURE 8 Testing Result with Some Wrong Training Samples

In order to verify the performance of the proposed model, we also utilize a 3-layer BP neural network to recognize the operating performance of the testing samples.

The parameters of BP neural network are set as follows. The number of hidden neurons is 40, the number of maximum iteration is 500, and the goal of training accuracy is 10^{-8}. The training performance of BP neural network and its testing result are as shown in Figure 9 and Figure 10, respectively. The ratio of correct and wrong number of testing samples is 13:3, which is lower than the probabilistic neural network.

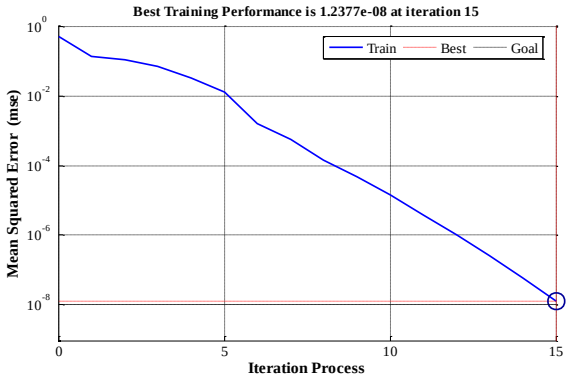


FIGURE 9 Training Performance of BP Neural Network with Given Parameters

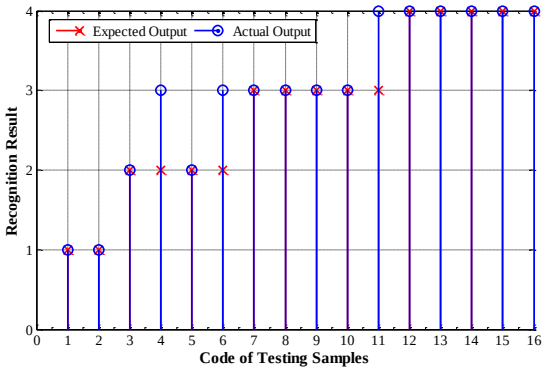


FIGURE 10 Testing Result of BP Neural Network

5.4 FINAL RECOGNITION RESULT

The final simulation result of the hybrid neural network is as shown in Table 4. There are 7 railway freight stations in Rank I, 21 railway freight stations in Rank II, 53 railway freight stations in Rank III, and 27 railway freight stations in Rank IV.

TABLE 4 Operating Performance of the Railway Freight Stations

Rank	Railway Freight Stations	Total Number
I	Baotouxi, Hohhot, Baotoubei, Linhe, Wuhaixi, Erlian, Jining	7
II	Gongjiban, Baotoudong, Baotou, Fengzhen, Salaqi, Wanshuiquan, Baita, Jiningnan, Saihantala, Wuhaibei, Dalahai, Baiyunebo, Ceke, Wuhai, Bayangaole, Jiningnan, Zhenglanqi, Duolun, Zhelimu, Xilinhaote, Zhenglanqi	21
III	Hantaichuan, Hantaichuanbei, Xiangshawan, Erdaoshahe, Gongwusu, Jilantai, Kabuqi, Laoshidan, Lasengmiao, Lasengzhongmiao, Dongxing, Wulateqianqi, Dongshenxi, Guchengwan, Xinxiancheng, Taigemu, Miaoliang, Chagannuoer, Wurigentala, Benjing, Hayehutong, Taosihao, Gujiabao, Minzu, Suji, Zhurihe, Qisumu, Jiangui, Wulashan, Baotounan, Dongshengxi, Eerduosi, Xinjie, Hohhotnan, Suhongtu, Hangjinhouqi, Tukemumiao, Baiyinkulun, Huade, Jinjiadian, Jingpeng, Kailu, Lindong, Linxi, Sanggendalai, Mengtagenlai, Sharinai, Pingandi, Yuzhoudi, Zhelimu, Zhengxiangbaiqi, Baiyunebonan, Xilinhaotexi	53
IV	Kundulunzhao, Bikeqi, Gongmiaozi, Chasuqi, Toudaoqiao, Tuguiwula, Mingan, Ejina, Xinghe, Daluhao, Taigemu, Baiyinchagan, Qixiaying, Baoziwan, Taoboqi, Zhuozishan, Benhong, Wuyuan, Chabuga, Baiyunebonan, Ehensuwu, Bayinhua, Liangsha, Baotumu, Huhewenduer,Nongkensituan, Dalatenan	27

6 Conclusions

In this study, we propose a strategy on optimization and integration for railway freight stations and design a hybrid neural network model to recognize the operating performance of the railway freight stations. Following conclusions can be drawn from the simulation results.

(1) The model can recognize the operating performance of each railway freight station under relatively high accuracy by classifying them into four ranks according to the proposed strategy.

(2) The accurate selection of the training samples has a significant influence on the recognition of the operating performance of the railway freight stations, which can be seen from Figure 7 and Figure 8. The unsupervised learning algorithm base neural network - SOM neural network can help us select the training samples and testing samples accurately, if necessary.

(3) Compared with the widely used BP neural network, probabilistic neural network performs higher recognition accuracy of the operating performance of railway freight stations.

However, in this study, there are still some imperfections presented as follows. Therefore, further research is required.

(1) The transport mode in the optimized railway freight transport network (See in Figure 3) limits to the qualitative analysis. How to organize the freight transport in the optimized railway freight transport network needs to be further analysed quantitatively.

(2) The optimization and integration method for the railway freight stations we propose in this study is currently in the design stage. Its feasibility needs to be further verified in the practical transport production.

Acknowledgments

This study was supported by the National Natural Science Foundation Project (No.71390332-3), the Science and Technology Development Plan of Ministry of Railways (No.2012X012-A) and the Scientific Research Plan of Hohhot Railway Administration (No.2012.2A003) of P.R. China.

References

[1] Caprara A, Malaguti E, Toth P 2011 A freight service design problem for a railway corridor *Transportation Science* **45**(2) 147-62
[2] Zhao Y, Pu Y, Zhong C Z 2005 Optimizing railway logistics structure by fuzzy mathematical theory and analytic hierarchy process *China Railway Science* **26**(3) 119-23 (*in Chinese*)
[3] Wang L L, Zhang X C, Hai T 2007 Weighted Cluster Analysis on Macroscopic Distribution Areas of Railway Dangerous Goods Stations *Journal of the China Railway Society* **29**(6) 101-5 (*in Chinese*)
[4] Lang M X, Zhong M, Liu Z Y 2002 Study on the Closing of Private Railway Sidings by Using Fuzzy Comprehensive Evaluation Method *Chinese Journal of Management Science* **10**(z1) 107-10 (*in Chinese*)
[5] Wang D Z, Li X F, Lang M X 2011 Application of Fuzzy Comprehensive Evaluation on Optimization and Integration of Railway Special Lines *Railway Transport and Economy* **33**(1) 22-6 (*in Chinese*)
[6] Li X F, Lang M X 2010 A Fuzzy Comprehensive Evaluation Model for the Optimization and Integration of the Industrial Railway Sidings *Proceedings 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering* 236-9
[7] Wei R 2007 *Study on the adjusting method for the layout of railway freight stations* Master Thesis, Beijing Jiaotong University, Beijing, China (*in Chinese*)
[8] Azami H, Mosavi M R, Sanei S 2013 Classification of GPS satellites using improved back propagation training algorithms *Wireless Personal Communications* **7**(2) 789-803
[9] Park Y J, Cho H 2004 Quality evaluation by classification of electrode force patterns in the resistance spot welding process using neural networks *Proceedings of the Institution of Mechanical Engineers Part B* **28**(B11) 1513-24
[10] Tan K, Du P J 2008 Hyper spectral remote sensing image classification based on radical basis function neural network *Spectroscopy and Spectral Analysis* **28**(9) 2009-13
[11] Tambouratzis G, Hairetakis N, Markantonatou S, Carayannis G 2003 Applying the SOM model to text classification according to register and stylistic content *International Journal of Neural Systems* **13**(1) 1-11

[12] Ryoo Y J, Lim Y C, Kim K H 2001 Classification of materials using temperature response curve fitting and fuzzy neural network *Sensors and Actuators A: Physical* **94**(1-2) 11-8
[13] Sun G M, Dong X Y, Xu G D 2006 Tumor tissue identification based on gene expression data using DWT feature extraction and PNN classifier *Neurocomputing* **69**(4-6) 387-402
[14] Wang S L, Li X L, Zhang S W, Gui J, Huang D S 2006 Tumor classification by combining PNN classifier ensemble with neighborhood rough set based gene reduction *Computers in Biology and Medicine* **40**(2) 179-89
[15] Debska B, Guzowska-Świder B 2011 Application of artificial neural network in food classification *Analytica Chimica Acta* **705**(1-2) 283-91
[16] Huang J C, Pan W T 2010 Forecasting classification of operating performance of enterprises by probabilistic neural network *Journal of Information & Optimization Sciences* **31**(2) 333-45
[17] Zhang L Y, Zhang X 2013 Multi-objective team formation optimization for new product development *Computers and Industrial Engineering* **64**(3) 804-11
[18] Ma Y L, Ma X Z, Feng Z J, Zhang L 2007 Fault diagnosis of high-pressure heater system based on radial basis probabilistic neural network *Journal of North China Electric Power University* **34**(5) 81-4 (*in Chinese*)
[19] Guo X Y 2007 Using probabilistic neural network to diagnose the fault of self-propelled gun engine *Measurement & Control Technology* **26**(8) 804-11 (*in Chinese*)
[20] Behbahani S, Nasrabadi A M 2009 Application of SOM neural network in clustering *Journal of Biomedical Science & Engineering* **2**(8) 637-43
[21] Ghaseminezhad M H, Karami A 2011 A novel self-organizing map (SOM) neural network for discrete groups of data clustering *Applied Soft Computing Journal* **11**(4) 3771-8
[22] Specht D F 1990 Probabilistic neural networks *Neural Networks* **3**(1) 333-45
[23] Zhao L T, Chai T Y, Diao X K, Yuan D C 2012 Multi-class classification with one-against-one using probabilistic extreme learning machine *Lecture Notes in Computer Science* **7368** 10-9
[24] Yagiz S, Sezer E A, Gokceoglu C 2012 Artificial neural networks and nonlinear regression techniques to assess the influence of slake durability cycles on the prediction of uniaxial compressive strength and modulus of elasticity for carbonate rocks *International Journal for Numerical and Analytical Methods in Geomechanics* **36**(14) 1636-50

Authors

Yan Sun, born in 1990, Shuguang, Shandong Province, P.R. China

Current position, grades: Ph.D Candidate of School of Traffic and Transportation, Beijing Jiaotong University.
University studies: majored in Traffic and Transportation, and received his Bachelor Degree in Engineering in 2013 at School of Traffic and Transportation, Beijing Jiaotong University.
Scientific interest: intermodal transport planning and management.
Publications: two articles as the first author, including one journal article in Journal of Industrial Engineering and Management (2014 Vol.7 No.2) and one conference article in VMEIT 2014.

Maoxiang Lang, born in 1969, Gaotang, Shandong Province, P.R. China

Current position, grades: professor and Ph.D advisor at School of Traffic and Transportation, Beijing Jiaotong University.
University studies: He received his Bachelor Degree in Engineering in 1991 at Shanghai Tiedao University, and then received his Master and Ph.D Degree in 1994 and 2002, respectively, at Beijing Jiaotong University.
Scientific interest: transportation and logistics management, transportation marketing management and modern railway freight technology and management.
Publications: He has published more than 60 articles.

Danzhu Wang, born in 1985, Qinhuangdao, Heibei Province, P.R. China

Current position, grades: Ph.D Candidate of School of Traffic and Transportation, Beijing Jiaotong University.
University studies: Bachelor and Master Degree in Engineering in 2007 and 2010, respectively, at School of Traffic and Transportation, Beijing Jiaotong University
Scientific interest: design of railway freight/logistics service productions.
Publications: 5 articles.

PH sensor design based on potentiometric analysis method

Song Yu[1]*, Liu Hua[2]

[1]*Jilin Technology College of Electronic Information, Jilin, 132021, China*

[2]*Calibration Centre of State Grid Heilongjiang Electric Power Company Limited, Helongjiang, 150030, China*

Received 1 March 2014, www.cmnt.lv

Abstract

In order to accurately, continuously detect solution *pH* on line, regarding C8051F020 as the control centre, the output voltage of glass electrode is amplified through high input impedance circuit utilizing potentiometric method to obtain related voltage signal of *pH* value. Standard solution and temperature sensor ADT7301 are utilized for system calibration and moving average filter of the detected *pH* values, thereby achieving real-time online detection of *pH* value. Detection results are accurate and reliable; low power consumption and miniaturation also facilitate rapid collection and analysis of data under different environments.

Keywords: pH detection, C8051F020, glass electrode

1 Introduction

The *pH* value is a basic parameter for the acid concentrations of the solution, and an important standard determining the quality of a variety of water system such as industrial water, environmental water and domestic water. Common methods of detecting *pH* parameters are chemical analysis, paper strip analysis and potential analysis. Chemical analysis and paper strip analysis cannot achieve online real-time monitoring, so in this research, potentiometric method is utilized to design a portable sensor for rapidly detecting the *pH* value of water samples, which is suitable for operating personnel to duly analyse and estimate water quality on the spot. The sensor detects *pH* value by composite glass electrode, amplifies weak voltage signal through signal processing circuit and converts the signal into standard voltage signal. Temperature compensation is needed in *pH* detection, ADT7301 digital temperature sensor is utilized for temperature detection in this research, which can effectively improve the accuracy of detection.

2 Principle of *pH* detection

Detection of *pH* is to measure the concentration of hydrogen ions, and the definition of *pH* is the negative logarithm of hydrogen ion concentration, namely, $pH = -\log[H^+]$. The *pH* value 0 indicates strong acidic solution, while *pH* 14 indicates strong alkaline solution. At room temperature, the solution with $pH = 7$ is neutral solution. In this research, pH is measured through potentiometric method. Glass electrode and silver-chloride electrode are regarded as the indicating electrode and reference electrode, respectively, and the two electrodes are encapsulated together to form a composite glass electrode. After the sensor is inserted into the test solution, composite glass electrode and test solution can form a primary battery, and the positive pole and negative pole of primary cells are two output lead wires of composite glass electrode. In measuring the *pH* of the solution, the electrodynamic force between glass electrode and reference electrode will change with the change of hydrogen ion concentration in sample solution. If the electric potential difference between two electrodes is zero, then *pH* value of test solution is 7, and the potential of *pH* composite electrode is 0V. According to Nernst equation, the relationship between output electrodynamic force of primary battery, absolute temperature of test solution and pH value of test solution is as below [1].

$$E = E_0 + KT(pH_x - pH_s), \tag{1}$$

wherein, E is the output electrodynamic force of primary battery; constant E_0 is the potential difference determined by electrode material, internal reference solution, internal reference electrode and liquid junction potential; constant K is the Nernst factor; T is the absolute temperature of sample solution; pH_x is the pH value of the test solution; constant pH_s is the pH value of buffer solution in the composite glass electrode.

Through Equation (1), the output electromotive force of original battery is generated by pH value and temperature of test solution together, so after simultaneously measuring output electrodynamic force and solution temperature of primary battery, pH value of the test solution can be calculated according to Equation (1).

3 System hardware design

The overall system structure is shown in Figure 1. C8051F020 microcontroller is the control core of the system.

***Corresponding author** e-mail: yusongjlmail@126.com

Through signal processing circuit with high input impedance, weak electrical signals of *pH* detection electrode are sent into the built-in A/D module of microcontroller, conducting temperature compensation through thirteen-bit digital temperature sensor ADT7301. External advanced ferroelectric memory chip FM31256 is regarded as non-volatile memory to automatically save data when the system is powered down. Human-machine interface is composed of matrix keyboard and LCD dot matrix, which can communicate with the host computer through RS232 bus.

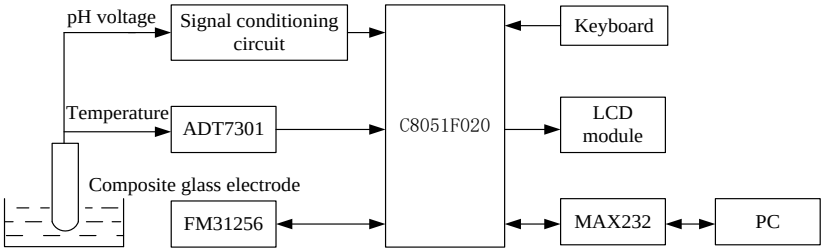


FIGURE 1 Block diagram of *pH* detection system

3.1 STRUCTURE AND PERFORMANCE CHARACTERISTICS OF C8051F020

C8051F020 is a mixed-signal system microcontroller produced by Cygnal Company [2], the instruction set of which is fully compatible with MCS-51. Compared with previous 51-series microcontrollers, C8051F020 has a lot of new features, but its reliability and speed should been improved. Generally, the change in the *pH* value of solution is a slow transition process, which does not need A/D converter with high sampling rate. C8051F020 microcontroller has twelve-bit successive approximation A/D converter and sampling holder, connecting with the outside world through an eight-port analog multiplexer. The A/D converter has a nonlinear accuracy of 0.25LSB and a sampling rate up to 100kbit/s at the highest resolution. Therefore, there is no need to configure an appropriative A/D converter if C8051F020 microcontroller has been chosen as system MCU. Meanwhile, C8051F020 microcontroller has built-in in-system programming Flash (64K bytes), RAM (4352 bytes) and E^2PROM (512 bytes), so we do not need to design additional extenders and data memories, and the on-chip E^2PROM is directly utilize to store system parameters. C8051F020 has six operating modes and automatically enters the sleep state when the system is not working, thus effectively lowering the power consumption of the system.

3.2 DETECTION MODULE OF *PH*

In detection module, *pH* composite electrode and amplification circuit constitute the detection unit. If the *pH* value of test solution is distributed in the range of 0-14, *pH* composite electrode will output bipolar analog signals. However, due to the high internal resistance of *pH* composite electrode $\left(10^8\Omega \sim 10^{10}\Omega\right)$, the output signals are relatively weak, usually hundreds of millivolts. Therefore, output signal needs amplification and translation to meet the input range of A/D analog-digital conversion circuit; composite *pH* electrode has high resistance, so the key of detection circuit is to achieve high input impedance and consider impedance matching, which requires input impedance of detection circuit to maintain in the range of $10^{12}\Omega \sim 10^{13}\Omega$, thus effectively reducing detection errors. Detection circuit of *pH* value is shown in Figure 2.

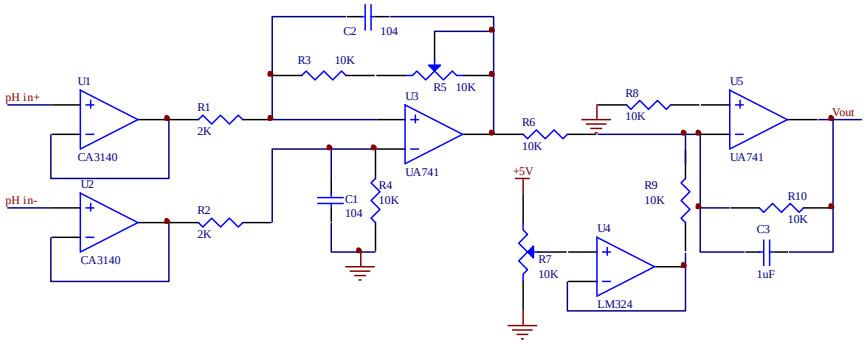


FIGURE 2 *pH* detection circuit

In order to reduce detection noise and improve system stability, an operational amplifier CA3140 is utilized in this research, which is a BIMOS operational amplifier developed by the Radio Corporation of America. On the integrated chip, piezoelectric PMOS transistor technology, the advantages of high-voltage bipolar transistor, MOS/FET input and the technology of bipolar output are utilized by combining the excellent performance of COS/MOS op amp for the first time. COS/MOS op amp is characterized by high input impedance, low bias current, low noise and high gain, and mainly utilized to complete impedance matching, reduce detection noise and improve system stability. In order

to improve input impedance of the whole transmitter, transforming part of input impedance is made up of CA3140 with input impedance up to $10^{14}\Omega$. Each CA3140 is connected to a voltage follower, constituting the first stage circuit to improve input impedance. Besides, due to the symmetrical structure, CMRR can be effectively improved if two devices have the same parameters, and the common-mode signal and offset error signals can also offset.

Generic op amp UA741 is chosen as the second-stage differential amplifier, and amplification can be deduced by Equation (2).

$$Vout = (\frac{R_1 + R_3 + R_5}{R_2 + R_4})V_+ - \frac{R_3 + R_5}{R_1}V_-, \quad (2)$$

where $Vout$ is output of U_3 in Figure 2; V_+ and V_- are the output voltages of two voltage followers. Amplification can be set by changing the resistance value of variable resistor. Through the addition operation between the amplified signal and output of LM324 follower circuit, the output signal of second-stage differential amplifier is added to a fixed voltage, thus realizing level conversion and finally achieving the output of 0.5V-4.8V signal through the reverse amplifier circuit composed of UA741.

3.3 TEMPERATURE DETECTION

In this design, 13-bit digital temperature sensor ADT7301 is utilized to detect temperature. The sensor, with no other external circuitry, does not need to convert analog signal to digital signal, and it can directly output digital quantity. ADT7301 makes the system structure more simple and reliable. ADT7301 and microcontroller C8051F020 utilize SPI communication mode; the serial interface consists of four wires-C/S, SCLK, DIN and DOUT. Temperature detection circuit is shown in Figure 3.

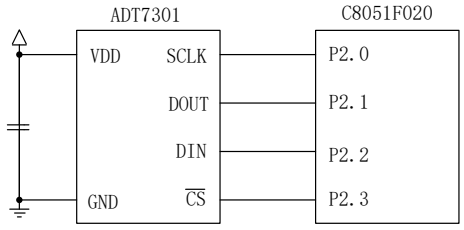


FIGURE 3 The temperature detection circuit

A bandgap temperature sensor for temperature detection and a thirteen-bit A/D converter [3] are integrated in ADT7301 chip, which has a low supply current, high temperature conversion precision, low power and flexible, convenient serial interface. Crystal oscillator is also integrated in AD7301, so clocks can be directly connected to serial port at work, with no need of A/D conversion. The chip has two modes-normal operating mode and power-saving mode. In normal operating mode, the internal clock oscillator will drive automatic conversion timing, so that the chip electrifies the analog circuitry once per second for a temperature conversion, which generally takes 800us. After the conversion, chip analog circuit will automatically power off, and then automatically power on after one second. Therefore, the latest temperature conversion value can always be obtained in temperature value register. Temperature value register is a fourteen-bit read-only register utilized to store temperature conversion result of ADC, and the result is composed of thirteen-bit binary complement code and one-bit sign bit, the top digit. This sensor has a detection accuracy of ± 0.5℃, temperature resolution of 0.03125℃ and detection range of -40℃~+150℃.

By setting control register ADT7301, the chip can be set to power-saving mode. In power-saving mode, the on-chip oscillator is turned off without temperature conversion in ADT7301 until the normal operating mode. By writing zero to the control register, the chip will restore to normal operating mode. Power-saving mode of C8051F02 microcontroller can effectively reduce the power consumption of the system.

4 Electrode calibration and temperature correction

4.1 ELECTRODE CALIBRATION

Before use, *pH* sensor should be set to zero. The specific way is to insert *pH* composite electrode into the solution with temperature of 25℃ and *pH* value of 7, and adjust the zeroing end of *pH* detection circuit until output voltage signal of amplifying circuit is 0V. Then the glass electrode needs calibration. Due to reasons including the manufacturing process of glass electrode, the actual values of parameter E_0 and K in Equation (1) are different from their theoretical values, changing with the aging of electrodes. Therefore, in order to accurately detect *pH* value of the solution, standard buffer solution with the known *pH* value should be utilized for the correction of above-mentioned parameters. The calibration method is to select standard buffer solution $pH = 4.01$ and $pH = 9.18$ [4]. Specific means are as follows: the *pH* values of two standard buffer solutions are supposed as pH_1 and pH_2, respectively, and the output electrodynamic forces are E_1 and E_2; two solutions are calibrated under the same temperature, thereby obtaining relationship between output electrodynamic force E and pH value through Equation (1), shown in Equation (3).

$$\begin{aligned} E_1 &= E_0 + KT(pH_1 - pH_s) \\ E_2 &= E_0 + KT(pH_2 - pH_s) \end{aligned}, \quad (3)$$

$$K = \frac{E_2 - E_1}{T(pH_2 - pH_1)}. \quad (4)$$

For the test solution,

$$E_x = E_0 + KT(pH_x - pH_s). \quad (5)$$

Therefore, the *pH* value of test solution can be obtained.

$$pH_x = pH_1 + \frac{E_x - E_1}{KT}. \quad (6)$$

The *pH* value of two standard buffer solutions pH_1, pH_2, the corresponding electrodynamic force E_1, E_2 and the calculated parameter K are stored in E^2PROM. And *pH* of the test solution can be obtained through Equation (6). During the process of *pH* detection, the temperature of standard solution and test solution should be kept close and constant to prevent sudden changes in temperature.

4.2 TEMPERATURE CORRECTION

Based on Equation (1) and Equation (6), *pH* electrode output voltage changes with the change in the temperature of test solution, while *pH* value of the solution has no correlation with the temperature. Therefore, the temperature should be accurately detected to ensure the accuracy of instruments. In this research, a software method is utilized to correct thirteen-bit output of ADT7301. Theoretically, the digital quantity of ADT7301 has a corresponding relationship with temperature, shown in Equation (7).

$$T = N \cdot \frac{190}{2^{13}} + (-40). \quad (7)$$

At temperature *t*, the actual digital quantity of ADC will change from the ideal value N to N' due to errors. If N' is substituted into Equation (7) for calculation without correction, there will be errors in the detection results. In this research, the actual digital quantity N' of ADT7301 is corrected into ideal value N. Then N is substituted into Equation (7), thus eliminating detection errors. Equation (8) is the correction equation.

$$N = \frac{N_{150} - N_{-40}}{N'_{150} - N'_{-40}}(N' - N'_{-40}) + N_{-40}, \quad (8)$$

wherein, N_{150} and N_{-40} are the ideal digital quantity of ADT7301 at 150℃ and -40℃, the values of which are 2^{13}-1 and 0, respectively; N'_{150} and N'_{-40} are the actual digital quantity of ADT7301 at 150℃ and -40℃, the values of which are measured and stored in E^2PROM. In actual detection, the digital quantity needs correction through Equation (8) before being substituted into the Equation (7), thus calculating the final temperature value [5].

5 System workflow and software filtering

According to requirements and characteristics of portable *pH* detection, system software can complete self-diagnostic of hardware and system initialization, handle keyboard commands and interface tasks, and conduct data acquisition and processing, temperature compensation and alarm. Figure 4 is the flow chart of main program. To reduce errors in detection data and ensure the stability and reproducibility of detection, using the method [6], the combination of moving arithmetic average filter and anti-pulse interference average value is utilized in the detection subroutine.

Assuming that $u(n)$ is the *pH* detection result at the n-th time, and L is the length of the moving average filter window, the filter output $x(n)$ at the n th time can be calculated through Equation (9).

$$x(n) = \frac{1}{L-2}\left\{\sum_{i=1}^{L} u(n+i) - \max_{1\le i\le L} u(n+i) - \min_{1\le i\le L} u(n+i)\right\}, \quad (9)$$

wherein, $\max_{1\le i\le L} u(n+i)$ is the maximum value; $\min_{1\le i\le L} u(n+i)$ is the minimum value.

According to Equation (9), we need to sample L times in detecting a *pH* value, obtain L *pH* values with little differences and form the basic sequence, $u(n+1), u(n+2), \ldots, u(n+L)$. Then the first value of the sequence is removed after every sampling; the remaining values sequentially move forward by one bit; the sampled new value will be inserted at the last bit of the sequence, thereby obtaining a new sequence. After deleting the maximum and minimum value of new sequence, the average of remaining $L-2$ values is calculated as $x(n)$, the final result of detection. The value of L is determined by A/D conversion rate and the required stability time of detection. Based on actual tests and references [6], the value is $L = 9$ in this research.

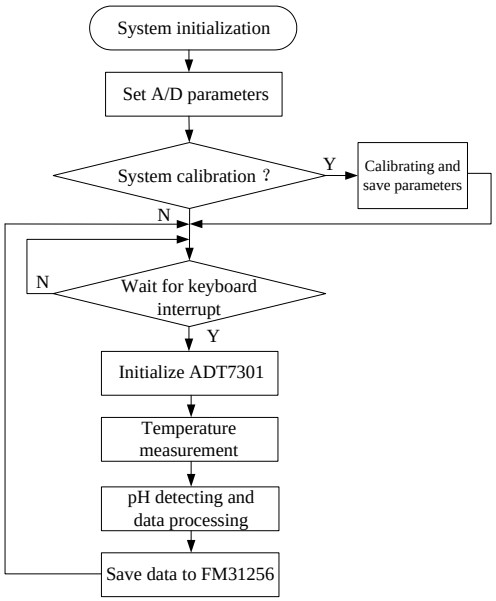


FIGURE 4 Flowchart of *pH* detection sensor

6 Experiment results and analysis

The standard buffer solution with the *pH* of 4.01 (Hach Company) was selected in this research. The experiment was carried out at about 25℃. The standard solution was put in a container with ice water or hot water to change the temperature of buffer standard and verify the accuracy of temperature compensation. In order to verify the reproducibility

of sensor probe and the entire system, each standard solution was measured six times; the probe was put in the solution for 5 minutes each time; *pH* values were determined after the readings of *pH* sensor is stable. Table 1 is a part of the experimental results. According to Table 1, the difference between the detected six *pH* values is less than 0.008, showing good reproducibility of the sensor; the difference between detected value at different temperatures and theoretical value is less than 0.006, indicating that temperature has little influence on the detection results.

TABLE *pH* value of experimental data detection sensors

Standard solution		Times						Theoretical value
pH	*Temperature /℃*	*1*	*2*	*3*	*4*	*5*	*6*	
4.01	5	4.002	4.006	3.999	4.002	3.998	4.003	4.00
	25	4.014	4.011	4.009	4.013	4.015	4.014	4.01
	35	4.035	4.033	4.029	4.031	4.032	4.034	4.03

7 Conclusions

A method for online *pH* detection is proposed in this research with complete design of hardware and software programs. Impedance matching method is utilized to eliminate the influence of high resistance in glass electrode, and electrodes are calibrated through standard buffer solution, thus eliminating zero drift and electrode deviations. Besides, the output of temperature sensor is calibrated by software to effectively improve the accuracy of temperature detection. The accuracy and repeatability of sensors are effectively improved by establishing *pH* value sequence and utilizing moving average filter to process the *pH* values, realizing long-term, on-line and high-precision detection at the range of -40℃~+150℃ in the *pH* detection system.

References

[1] Wyatt D L 2000 *pH* measurement: a guide to electrode selection, focusing on the reference system *Proceedings of Annual ISA Analysis Division Symposium* **33** 55-6
[2] Silicon Laboratories Inc C8051F020/1/2/3 mixed signal ISP flash micro controller. C8051F020 datasheet. 2005
[3] Analog Device 13-Bit ±0.5°C Accurate Micro Power Digital Temperature Sensor in 6-Lead SOT-23 ADT7301 datasheet 2004
[4] Yehya H, Ghallab 2004 A novel *pH* sensor current mode read-out circuit using operational floating current conveyor *Proceedings of the 2004 International Conference* 262-26
[5] Spitzer P, Wemer B 2002 Improved reliability of *pH* measurements. Analytical and Bioanalytical Chemistry **374**(5) 787-95
[6] Pandey S, Kal S 2000 Simple model for an Electrolyte-Oxide-Semiconductor *pH*-sensor *Proceedings of SPIE - The International Society for Optical Engineering* **75**(1) 719-22

Authors

Yu Song, born in May, 1966, Jilin, China

Current position, grades: a senior engineer in Jilin Technology College of Electronic Information, China.
Scientific interest: power automation, electrical automation.
Publications: 20 papers.

Hua Liu, born in June, 1968, Helongjiang, China

Current position, grades: a senior engineer in Calibration Centre of State Grid Heilongjiang Electric Power Company Limited, China.
Scientific interest: advanced metering infrastructure and electrical automation.
Publications: More than 15 papers.

Digital controller for electric vehicle synchronous motor rotor

Weihua Chen*

State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources

North China Electric Power University, Beijing 102206, China

Received 1 March 2014, www.cmnt.lv

Abstract

At present, the number of electric vehicle is increasing, and, in case of big vehicle speed change, the vehicle motor cannot be used efficiently. In order to resolve this problem, it is necessary to improve the topological structure and control strategy, and design a new converter. In this paper, we apply a two-channel synchronous Buck converter. The main circuit is to achieve the maximum power output by the stage regulation, while the improved converter topology realizes the overall system function. The experimental results show that the new converter has superior performance in big vehicle speed change, and has the high quality energy output. This research has an important practical significance to improve the utilization of renewable energy.

Keywords: power system, electric vehicle, digital controller, inverter, converter

1 Introduction

With new kind of vehicle developing, electric vehicle is getting more and more attention of researchers [1]. Now the utilization of drive motor for electric vehicle is usually in the way of the asynchronous motor mode. Asynchronous motor is with low price, but the poor performance. So its application is limited [2-4]. While synchronous motor has a lot of advantages, such as good mechanical characteristic and adjustable air gap. It may operate efficiently in the big speed range with greatly improved power efficiency and reduced operation and maintenance costs. Synchronous motor has become an important development direction in drive motor for electric vehicle, characterized by being more economical, convenient, and practical.

The research on the utilization of synchronous motor for ele-ctric vehicle mainly concentrates on synchronous motor physical structure, while paying little attention to how to control synchronous motor. The resulting consequence is the fact that synchronous motor does not operate perfectly for electric vehicle. Traffic jam is very serious in many large cities all over the world, vehicles have to start and strop frequently, so the traditional synchronous motor control cannot gain synchronous motor effective performance. This research aims to improve the performance of electric vehicle under big speed range by designing a new excitation converter control strategy for synchronous motor [5-7].

Some research has been done on the topology structure and control strategy of the excitation converter. The excitation converter used the two-channel synchronous Buck circuit topology. On the control strategy, phased control strategy and adaptive PI control method based online parameter adjustment were used. DSP chip TMS320F2812 was used as the control core to realize the switch controlling, data processing, communication and other functions. Through the improvement of excitation converter topology and control strategy, we can improve the performance of ecitation converter and the performance of synchronous motor under big speed range, and promote the development and utilization of electric vehicle.

2 Excitation converter topology

The excitation converter adopts the two-channel synchronous Buck parallel connection, as shown in Figure 1. The purpose of the two-channel synchronous Buck parallel connection is to increase the output current ripple frequency and reduce its amplitude. It can improve the dynamic response speed by reducing the output filter inductor. Under the same power output condition, the two-channel synchronous Buck parallel connection preferment [8-13].

Double closed loop is composed of voltage loop and current loop independently, as shown in Figure 1. It can realize the constant pressure output or constant current output. Firstly, Voltage loop sample the output voltage of *Vo*, and then compare it with the set voltage value *Vr*. The difference between the two voltages is put into the adaptive PI regulator, shown in the dashed box. Current loop is with the same voltage loop. Adaptive PI regulator can adjust proportion coefficient KP and integral coefficient KI. This can be used to realize cycle current or voltage control and overcurrent protection. Voltage loop work during quick response stage in the converter, while current loop work during the steady constant current stage in the converter.

***Corresponding author** e-mail: cwhwc@163.com

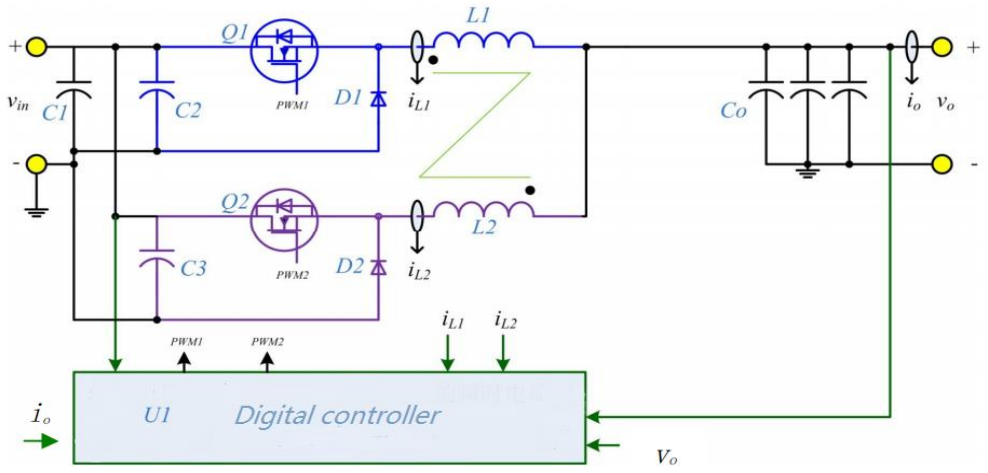


FIGURE 1 Two channel synchronous Buck converter main circuit topology diagram

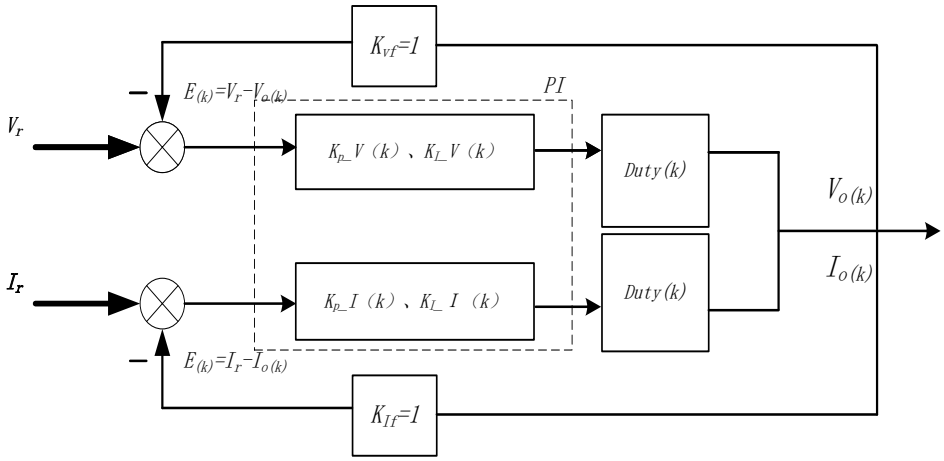


FIGURE 2 Double closed loop control structure diagram

Here introduced several parameters: *Vr* is the symbol of converter expected output voltage value (equal to the desired output current of *Ir* with the equivalent resistance of Re product, Re identification, see section third). *Vin* is the converter input voltage *Vo* converter; the actual output voltage. *D* is the power tube gate driving signal duty ratio of PWM wave; *ΔV* is expected output voltage and the actual output voltage difference, namely Delta *V*=*Vr*-*Vo*.

In electric vehicle, torque speed requirements change quickly. The rotor excitation current need to response quickly, namely the excitation converter can make rapid response to the different output needs. At the same time the overshoot and steady state error of the output are smaller. But the PI controller with fixed parameters is difficult to meet the requirements. This paper presents a simple, efficient adaptive PI controller to solve the problem [14, 15].

The mathematical equations for the adaptive PI controller are

$$\begin{aligned}\Delta u(k) &= u(k)\text{-}u(k\text{-}1)\\ &= K_p(e(k)\text{-}e(k\text{-}1)) + K_I e(k)\end{aligned}, \tag{1}$$

where the two parameter K_P, K_I tuning is adaptive PI control key points.

In order to make the converter voltage output can quickly track the desired voltage value, and the start-up phase voltage overshoot in a proper range, let the coefficient K_p is expected to duty ratio divided by the voltage difference

$$K_p = \frac{d}{\Delta V} \tag{2}$$

and let

$$d = \frac{V_r}{V_{in}}. \tag{3}$$

So coefficient K_p is

$$K_p = \frac{k_1 \times (k_2 \times V_r - V_{start})}{V_{in} \times (k_2 \times V_r - V_o)}, \tag{4}$$

where k_1 is the correction coefficient, k_2 is the actual voltage output and the expected voltage output ratio from fast response stage switching to a steady flow stage. By many tests, k_2 is from 0.85 to 0.9. Because the voltage close loop control

begin at the end of soft starting, so we need to subtract voltage V_{start} at the soft start end time.

In addition, let

$$K_I = \frac{K_P}{k_3}, \tag{5}$$

where k_3 is the coefficient of integral coefficient.

3 Converter control strategy

3.1 CONVERTER CONTROL STRATEGY

The work state of converter is divided into soft start stage, fast response stage and steady flow stage. So, the converter control strategy is divided also.

1) Soft start stage. The converter begins by soft start. The voltage output rises from zero to 8 percent of desired voltage value. In this stage, the duty ratio of PWM increases gradually, while the converter voltage output and current output and increases slowly. The purpose of the soft start is to preheat the power devices and circuits. Improving output steadily to prevent overshoot accident. On the other hand, it can identify load parameter. If we do not set the soft start of the converter in the start-up stage, the power switching devices work at the condition with the duty cycle of 100%. So there will be a lot of surge current into the output capacitor. Sometimes, the output voltage overshoot are large. And the surge current may also damage switch tube and other devices. Using the strategy of soft start can eliminate the surge current and avoid the overshoot of output voltage.

2) Fast response stage. In the fast response stage, in order to make the converter voltage output and current output can reach near the given value rapidly; voltage loop with adaptive PI controller is used to reduce the values of voltage output and the current output rise time effectively.

3) Steady flow stage. When the value of voltage output reaches ninety percent of the expected voltage output value, it is the time to be steady flow stage. At this stage, a current loop with adaptive PI controller is used.

3.2 PARAMETER IDENTIFICATION

In the soft startup stage, the control system gets the converter values of voltage output and current output through sensors. The excitation coil resistance is calculated on-line sometimes to obtain the equivalent resistance value with the method of least squares. The biggest advantage of online parameter identification is that effects factors such as temperature, rotation and others on the excitation coil resistance would be contained in the identification parameters naturally once the parameters identification finished. And it is easy to implement and calculation. But each the identifycation parameter may be different. The purpose of parameter identification is to provide the basis parameters for adaptive PI control.

4 Communication and protocol

The integrated control system of electric vehicle covers the parameters of motor stator, rotor, battery's working condition, vehicle speed, and the operation of the driver. The CAN bus technology is used to connect the device for obtaining the above parameters. It is the key to realize the electric vehicle set control. As a multi machine serial communication protocol, the CAN protocol is mainly used for network communication with multi node. The data frames for the CAN protocol is shown in Figure 3. As a practical application, designers need to develop communication protocol at the top level satisfied with their own project. Designers need to fill the data field in the data frame and to set the standard communication rate.

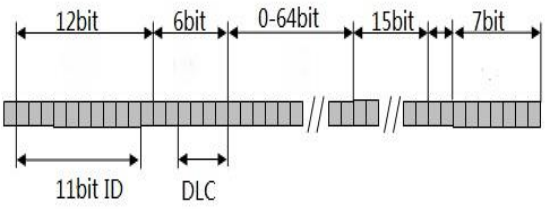


FIGURE 3 Schematic diagram of CAN standard frame format

Considering the communication between excitation converter system and other electric vehicles modules, the standard data frame is used in the paper. The data length is 8 bytes. A kind of simple and effective communication protocol top layer is made. Five communication frames are designed. They are frame 1, frame 2, frame 3, converter output frame, converter device temperature frame respectively. Frame 1, frame 2, and frame 3 are sent from the stator side controller to converter. Converter output frame and converter device temperature frame are sent from converter to the stator side controller. As shown in Table 1-5.

TABLE 1 CAN communication protocol - control frame 1

Name	Begin-end	Unit	Reset value	Smallest value	Biggest value
Inverter_Fault _flag	3-4	-	0	0	3
Converter_Reset _Request	5-6	-	0	0	3
Converter_Enable _Disable	7-8	-	0	0	3
Field_Voltage _Limit	17-32	V	250	0	250
Field_Current _Limit	33-48	A	20	0	20

TABLE 2 CAN communication protocol - control frame 2

Name	Begin-end	Unit	Reset value	Smallest value	Biggest value
Output_Current _Command	1-16	-	0	0	200
Command _Check_Sum	17-25	-	0	0	255

TABLE 3 CAN communication protocol - control frame 3

Name	Begin-end	Unit	Reset value	Smallest value	Biggest value
Converter_Fault _Signal	1-4	-	0	0	15
Converter _Output_Status	5-6	-	0	0	3
Converter_StatusValid	7-8	-	0	0	3
Converter_Field _Voltage_Limit	17-32	V	0	0	250
Converter_Field _Current_Limit	33-48	A	0	0	20

TABLE 4 CAN communication protocol - control frame 4

Name	Begin-end	Unit	Reset value	Smallest value	Biggest value
Field_Coil_Resistance	1-16	Ω	0	0	1000
Real_Converter _Output_Voltage	17-32	V	0	0	250
Real_Converter _Output_Current	33-48	A	0	0	20
Real_Converter _Output_Power	49-64	W	0	0	2500

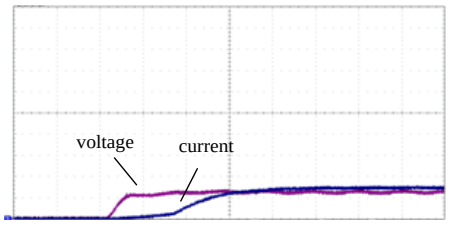
TABLE 5 CAN communication protocol - control frame 5

Name	Begin-end	Unit	Reset value	Smallest value	Biggest value
Field_Coil_Temperature	1-16	℃	0	-50	250
Switching_Component_Temperature	17-32	℃	0	-50	250
Inductor_Temperature	33-48	℃	0	-50	250

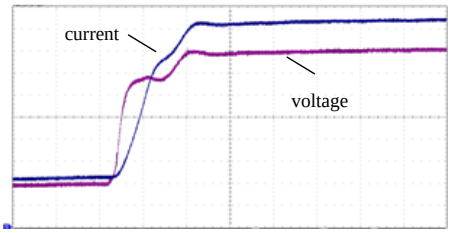
In this design physical implementation of CAN communication relies on the CAN module meets the CAN2.0B protocol standard built-in TMS320F2812 chip. The CAN module communication rate is up to 1Mbps. And it is of low power mode and CAN bus wake-up function. Considering the practical engineering, the configuration of CAN bit rate is 500bps. In the experiment, the PC with USB2CAN function is used for the simulation of stator side controller and converter doing CAN communication and control. It is obviously that the CAN communication is of good reliability.

5 Case studying

A prototype is built to make the related tests. And the test results are recorded. Figure 4 and Figure 5 is the converter start-up test wave form and dynamic response wave form respectively. Figure 6 is the steady state ripple wave form converter. In the experiment, resistors and inductors are used to simulate converter load. The rated load is a resistor of 35Ω and a series inductance 650mH.

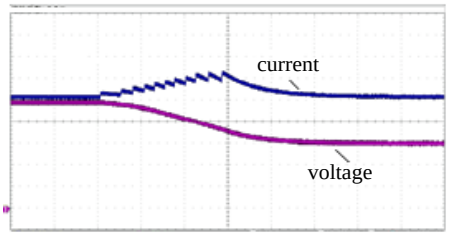


a) Soft start stage

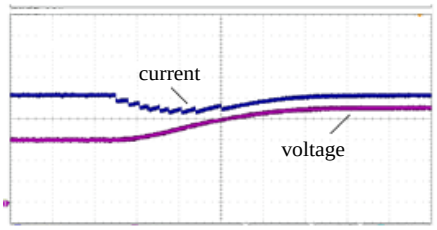


b) Regulating stage

FIGURE 4 Test waveforms at rated load starting



a) Load decreasing



b) Load increasing

FIGURE 5 Test waveforms at rated load starting

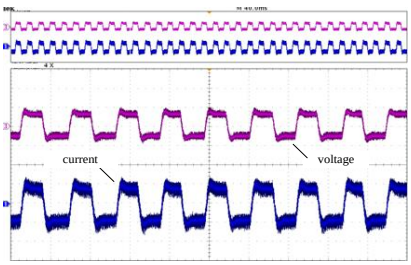


FIGURE 6 Voltage and current output ripple peak value

As we can see from Figure 4, the output values increase rapidly during the converter start-up stage. The response time is less than 100ms. The voltage overshoot is less than 15V. The current overshoot is less than 1A. It can be seen from Figure 5 that converter dynamic responses rapidly when load changes. The response time is less than 200ms. It can be seen from Figure 6 that the converter steady state voltage ripple is less than 1V and that the steady state current ripple is less than 50mA.

In addition, the converter power and efficiency is test, as the results shown in Table 6. In the two long tests, converter has a stable output. The values of output power are more than 800 watts. The values of efficiency are more than 95%. The above values satisfy the requirements of power and efficiency.

5 Conclusions

This paper describes the main design ideas of the synchronous motor rotor excitation converter for electric vehicles. The two channels synchronous Buck interleaves. The topology of converter is optimized. The control stage and adaptive PI converter are used. It can quickly achieve the desired output value during the different output demand. Making CAN communication protocol is not only to meet the communication needs of converter and stator side controller, but also provides reference and space for electric vehicles to build the more complex internal communication network in the future. The prototype is with good output and regulation performance. It meets the basic requirements of application in vehicle rotor excitation converter.

References

[1] Ferdowsi M 2009 Vehicles fleet as a distributed energy storage system for the power grid *Proceedings of IEEE Power & Engineering Society General Meeting Calgary Canada* 1-2

[2] Kato S 2010 Design optimization of interior permanent magnet synchronous motors for HEV & EV *SAE International Journal of Engines* **3**(1) 956-63

[3] Nasiri H 2011 Dynamic modelling and simulation of transmitter based series-parallel HEV applied to Toyota Prius 2004 *10th International Conference on Environment and Electrical Engineering* 1-4

[4] Song Y H, Yang X, Lu Z X 2010 Integration of plug-in hybrid and electric vehicles: experience from China *Proceedings of Power & Engineering Society General Meeting Minneapolis MU USA* 1-5

[5] Hashermnia N, Asaei B 2008 Comparative study of using different electric motors in the electric vehicles, international conference on electrical machines *Vilamoura Rortugal Portuguesa* 1844-8

[6] Jeon C, Choi M K, Yoo J 2012 The effect of social influence on consumers' hybridelectric vehicles adoption in Korea and China *International Conference on Advanced Communication Technology* 336-40

[7] Song Y 2010 Integration of plug-in hybrid and electric vehicles: Experience from China *IEEE Power and Energy Society General Meeting* 1-6

[8] Wen X 2011 Electric vehicle key technology research in China *International Aegean Conference on Electrical Machines and Power Electronics* 308-14

[9] Zhang, J 2011 Application of PMSM vector control system in PHEV *Journal of Jiangsu University* **32**(2) 157-62

[10] Consoli A, Scarcella G, Scelba G 2010 *IEEE Transactions on Industrial Applications* **46**(1) 121-9

[11] Bolognani S, Petrella R, Prearo A 2011 *IEEE Transactions on Industrial Applications* **47**(1) 105-14

[12] Kim S, Yoon Y D, Sul S K 2010 Parameter independent maximum torque per ampere (mtpa) control of ipm machine based on signal injection *2010 25th Annual IEEE Applied Power Electronics Conference and Exposition, Palm Springs* 103-8

[13] Xu W, Zhang Y, Zeng C, Xiao X 2012 Rotor field oriented control of linear induction machine based on fuzzy self-adapting PI controller *IEEE International Symposium on Industrial Electronics* 731-6

[14] Han SH 2010 *IEEE Transctions on Energy Conversion* **25**(2) 441-9

[15] Jeon C, Choi M K, Yoo J 2012 The effect of social influence on consumers' hybridelectric vehicles adoption in Korea and China *International Conference on Advanced Communication Technology* 336-4

Author

Weihua Chen, born on May 13, 1978, China

Current position, grades: a researcher in North China Electric Power University.
University studies: M.S. and Ph.D. degree from North China Electrical Power University and Zhejiang University, China in 2003 and 2007.
Scientific interest: automation, renewable energy, and power system stability and control.
Publications: 2 papers.

Effect factor research on monitoring wear of piston ring based on magnetoresistive sensor

Zhangming Peng[1, 2, 3], Huihua Cai[3], Qingchen Sun[4]*, GuojinChen[1, 2], Shaohui Su[1, 2]

[1]*School of Mechanical Engineering, Hangzhou Dianzi University, Hangzhou, 310018*

[2]*Zhejiang Provincial Key Laboratory of Ship and Port Machinery Equipment Technology, Hangzhou, 310018*

[3]*Yangfan Group CO.LTD, Yangfan Ship Design&Reseach Institute, Zhoushan, 316100*

[4]*East China of Lloyd's Register*

Received 1 August 2014, www.cmnt.lv

Abstract

Piston ring is one of the important parts in diesel engine, and its excessive wear reduces the sealing performance of the combustion chamber, the method on monitoring wear of piston rings with the magneto-resistive sensor is researched from the simulation viewpoint, the research shows that the amplitudes reduce with the increase of the piston ring wear. The theoretical basis for monitoring piston ring wear with a MR sensor is provided.

Keywords: piston ring wear, magneto-resistive sensor, diesel engine

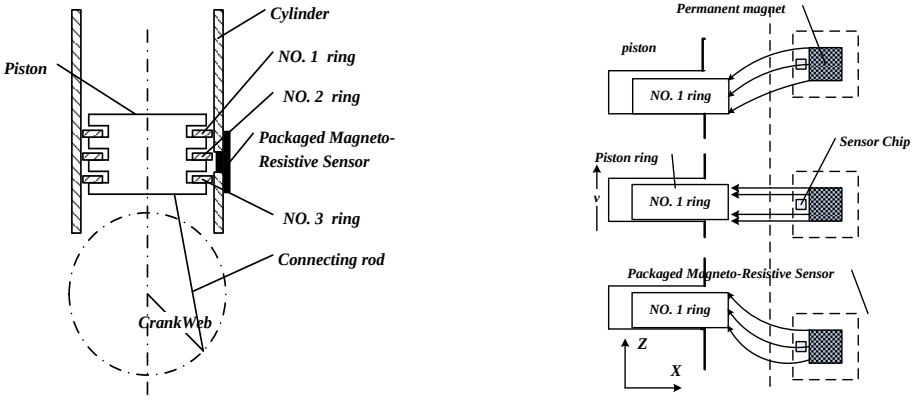
1 Introduction

Piston ring is one of the important parts in diesel engine, and its excessive wear reduces the sealing performance of the combustion chamber, which reduces the performance and the reliability of the diesel engine. In order to obtain the real-time operation status of the piston ring, the on-line monitoring methods should be researched. The wear of the piston ring can be measured by a magnetic sensor, and some monitoring system can achieve the monitoring function at present [1, 2].

Literature [3] studied the online monitoring method for the general piston rings based on the magneto-resistive sensor. On the research foundation of the literature [3], the effect of piston ring motion was researched in the paper, and it provides the theoretical basis for the practical application in the diesel engine.

2 Detection principle of magneto-resistive sensor

When a diesel engine works, the piston and the piston ring reciprocate within the cylinder and a magneto-resistive (MR) sensor is installed on cylinder liner to measure magnetic field intensity, as shown in Figure 1a. Figure 1b shows the distributions of the magnetic flux density when the piston ring passes the MR sensor. The change of the magnetic field intensity in Z direction can be found according to dealing with monitoring data. If the piston rings wear, the amplitudes of the magnetic field intensity on monitoring point would change respectively. And then the output amplitudes of the MR sensor change also, so that the wear of the piston rings can be detected.



1a 1b

FIGURE 1 Sketch map of magnetic lines of force

***Corresponding author** e-mail: qingchen.sun@lr.org

According to detection principle, the piston ring wear can be monitored with the amplitude change of magnetic field intensity, which may be bound up with its radial motion, including speed and acceleration, and has nothing to do with the axial movement, so the simulation must be finished to find out the relation between the radial motion and the wear of the piston ring.

3 Dynamics simulation on piston ring

3.1 DYNAMIC ANALYSIS ON PISTON RING

It needs to computer the balance equation of piston ring to find its radial motion. Figure 2 shows the force analysis of piston ring [4, 5].

The force of ring in axial direction includes:

$F_{a,con}$, the force between the ring groove and ring, including the oil film pressure $F_{a,oil,con}$ and the contact pressure between the ring groove and ring $F_{a,gro,con}$.

$F_g = mg$, the gravity of ring.

$F_{a,gas}$ the axial cylinder pressure.

$F_{a,fric}$ the friction force between the cylinder wall and piston ring.

The axial equation is given as:

$$m\frac{d^2y}{dt^2} = F_{a,con} + F_{a,gas} + F_{a,fric} + mg \ . \tag{1}$$

The force in radial direction includes:

$F_{r,con}$, the force between the cylinder wall and ring, including the oil film pressure $F_{r,oil,con}$ and the contact pressure between the cylinder wall and ring $F_{r,cyl,con}$

$F_{tension} = k_r(h' + h_0)$ the tension of piston ring, where k_r is the radial elastic coefficient of the piston ring.

$F_{r,gas}$ the radial cylinder pressure.

$F_{r,fric}$ the friction force between the ring and groove,

$F_{rl,gas}$ the gas force acting on the ring surface near cylinder wall due to the less lubricating oil.

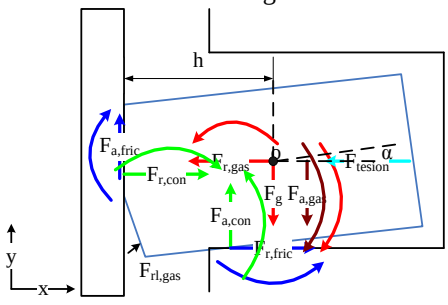


FIGURE 2 Force diagram of piston ring

The radial Equation is given as:

$$m\frac{d^2h}{dt^2} = F_{r,con} + k_r(h' + h_0) + F_{r,gas} + F_{r,fric} + F_{rl,gas} \tag{2}$$
$$= F_{r,oil,con} + F_{r,cyl,con} + k_r(h' + h_0) + F_{r,gas} + F_{r,fric} + F_{rl,gas}$$

The Torque equation is given as:

$$I\frac{d^2\alpha}{dt^2} = M_{a,con} + M_{a,gas} + M_{a,fric} + M_{r,con} + M_{r,fric} + M_t \ , \tag{3}$$

where $M_t = k_t\alpha$, $k_t = \dfrac{Eb^3 \ln\left(\dfrac{D}{d}\right)}{3(D+d)}$.

I is the moment of inertia, α the angle of inclination, E the modulus of elasticity, b the axial height, D the outer diameter, d the inner diameter.

The axial motion equations of the piston and piston ring can be described as:

$$y = y_p + y_{rp} \ , \tag{4}$$

$$\frac{dy}{dt} = \frac{dy_p}{dt} + \frac{dy_{rp}}{dt} \ , \tag{5}$$

$$\frac{d^2y}{dt^2} = \frac{d^2y_p}{dt^2} + \frac{d^2y_{rp}}{dt^2} \ , \tag{6}$$

where y is the axial displacement of piston ring centre of gravity, y_p the axial displacement of piston, y_{rp} the axial displacement between piston ring and piston.

According to the relationship between y_p and crank angle, piston ring motion trajectory can be found.

3.2 THE RADIAL MOTION SIMULATION ON PISTON RING

The wear limit of the piston ring is 4 mm for the marine diesel RTA52U, so the wear rage of the piston ring was set from 0 mm to 4 mm with a 1 mm step, thus the piston ring motions were simulated in various wear conditions according to the previous equations.

The radial velocity and displacement of piston rings were found during a working cycle of diesel engine. Figure 3 shows the radial velocity and displacement curves of the piston rings without wear.

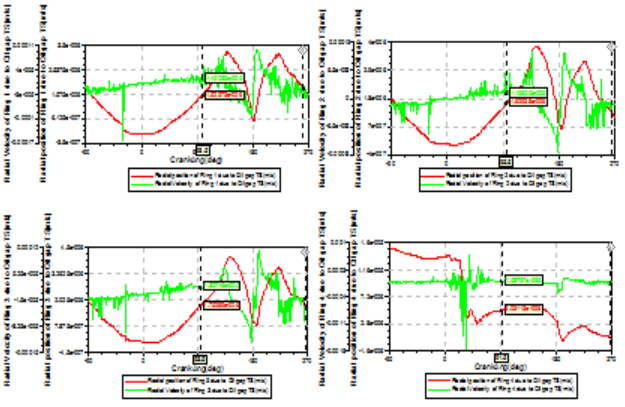


FIGURE 3 The radial velocity and displacement of piston rings

According to simulation results, it can be found that the change tendencies of the radial velocity and acceleration are almost the same for the piston rings except NO.4 ring (oil ring), and the minimal radial displacement occurs at the maximum combustion pressure.

When the piston rings passed the monitoring point in various wear conditions, the radial motions of piston rings were found, as shown in Table 1.

TABLE 1 Motion parameters for ring passing sensors (100 r/min)

Piston ring	Wear value 0mm		Wear value 1.0mm	
	Radial speed(m/s)	*Radial displacement (m/s2)*	*Radial speed(m/s)*	*Radial displacement (m/s2)*
NO.1	1.5e-5	1.5e-6	1.5e-5	1.8e-6
NO.2	1.2e-5	1.68e-6	1.2e-5	2.2e-6
NO.3	1.8e-5	1.8e-6	1.8e-5	2.1e-6
NO.4	7.6e-6	5.0e-6	7.6e-6	5.0e-6
Piston ring	**Wear value 3.0mm**		**Wear value 4.0mm**	
	Radial speed(m/s)	*Radial displacement (m/s2)*	*Radial speed(m/s)*	*Radial displacement (m/s2)*
NO.1	1.5e-5	1.6e-6	1.5e-5	1.6e-6
NO.2	1.3e-5	1.9e-6	1.3e-5	1.9e-6
NO.3	1.8e-5	2.1e-6	1.8e-5	2.1e-6
NO.4	6.9e-6	2.4e-5	6.9e-6	2.4e-5

4 Simulation on magnetic field intensity

A finite element model consisting of a piston, piston rings, a cylinder liner, and a magneto-resistive sensor was established on the basis of actual size, the magnetic field intensity of the monitoring point could be simulated. The piston ring motions of finite element model were set according to the previous results. Thus, the magnetic field intensity was simulated under different wear conditions, as shown in Figure 4. The relation between the amplitudes of the magnetic field intensity and the piston ring wear is shown in Figure 5. It shows that there is corresponding relationship between the amplitudes and the ring wear, and the amplitudes decrease with the increase of the ring wear.

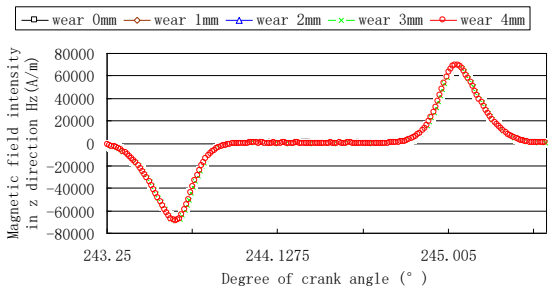


FIGURE 4 Calculation waveform

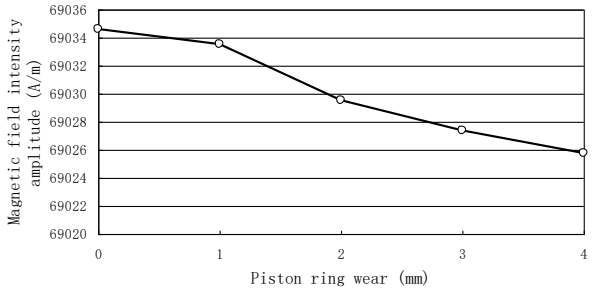


FIGURE 5 The relation between amplitudes and ring wear

5 Conclusions

The magnetic field intensity on the monitoring point is simulated for RTA52U marine diesel engine during a working cycle, the method on monitoring wear of piston rings with the magneto-resistive sensor is researched from the simulation point of view, it can be found the ring radial movements do not affect the amplitudes of the magnetic field intensity, and there is corresponding relationship between the amplitudes and the ring wear. Furthermore, the amplitudes reduce with the increase of the piston ring wear. It provides the theoretical basis for monitoring piston ring wear with a MR sensor.

References

[1] Zhao J 2006 Study on fault diagnosis of fuel system of diesel engine based on fractal theory *(in Chinese)*
[2] Zhong Q-M 2008 Experimental research on monitoring and diagnosing for diesel engine *Wuhan University of Technology: Wuhan (in Chinese)*
[3] Yang J-G, Peng Z-M, Yu Y 2010 *Transactions of CSICE* **28**(1) 85-9
[4] Krisada Wannatong, Somchai Chanchaona, Surachai Sanitjai 2008 Simulation modelling practice and theory **16** 127-46
[5] Tain T, Noordzij L B, Wong V, Heywood J B 1998 *Journal of Engineering for Gas Turbines and Power* **120** 843-54

Authors

Peng Zhangming, born on September 17, 1977, Hubei Province, China

Current position, grades: lecturer of Hangzhou Dianzi University, doctor's degree.
University studies: Wuhan university of Technology.
Scientific interest: monitoring and control of diesel engine.
Publications: 2 papers.

Cai Huihua, born on May 26, 1974, Anhui Province, China

Current position, grades: general manager.
University studies: Wuhan university of Technology.
Scientific interest: ship engineering.
Publications: 10 papers.

Sun Qingchen, born on July 29, 1980, Jiangsu Province, China

Current position, grades: marketing manager for East China.
University studies: Wuhan university of Technology.
Scientific interest: marine engineer.
Publications: 3 papers.

Guojin Chen, born on February 5, 1961, Zhejiang Province, China

Current position, grades: PhD, professor of school of Mechanical Engineering in Hangzhou Dianzi University.
University studies: XiDian University.
Scientific interest: mechatronics theory and technology, control theory and technology.

Shaohui Su, born on September 7, 1978, Henan Province, China

Current position, grades: associate professor.
University studies: Zhejiang University
Scientific interest: product data management, innovation design.
Publication: 10 papers.

Steel sheet location tracking and automatic sorting self-adaption control model

Xuechao Liao[1, 2*]

[1]*College of Computer Science, Wuhan University of Science and Technology, Wuhan 430081, P. R. China*

[2]*Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430081, P. R. China*

Received 1 Sept 2014, www.cmnt.lv

Abstract

The sorting control system is the critical procedure in the final product line of cold rolling mill. Taking the tinning steel sheet sorting control system of shearing production line of WISCO's cold rolling plant as an example, the process of steel sorting control system is introduced. The running principle and control concept of steel sheet position tracking and sorting are particularly presented. The practical application proves that the system can position the tinning steel sheet which is moving in high speed, which is of high controllability and reliability.

Keywords: sorting control, position tracking, magnet system, self-adapt regulating, dropping posture control

1 Introduction

The primary task of cold rolling plant shearing production line [1] is to automatically detect the sheet defects, and shear the sheet to specified length, then automatically separate the finished products from the inferiors, finally pack manually and deliver. The steel sheet sorting system [2] is the key step in the whole production process. The Figure 1 and Table 1 are the process flow of sheet sorting system ().

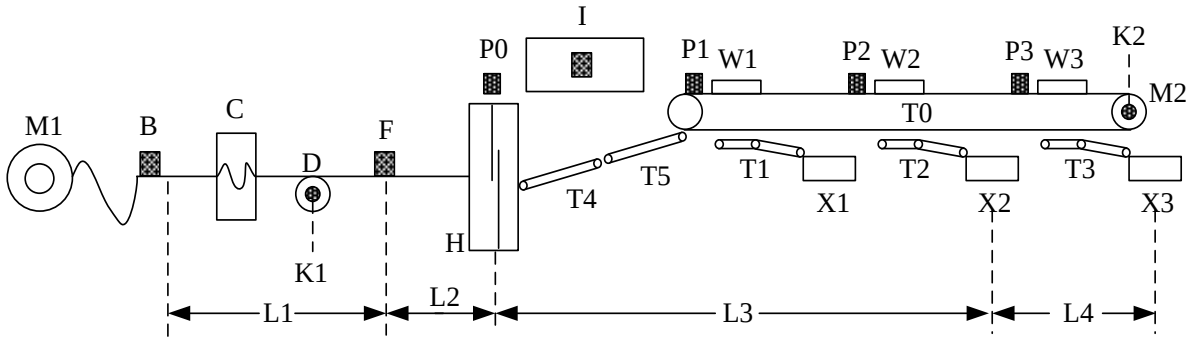


FIGURE 1 Sheet sorting system process flow

TABLE 1 Symbols description of sheet sorting system process flow

Symbol	Description	Symbol	Description
M1	Decoiler	T0	Main transmitting belt
K1	Speed encoder	T1,T2,T3	1#,2#,3# Stacking table belt
B	Thickness gauge	T4,T5	4#, 5# Entry belt
C	Straightener	W1,W2,W3	1#,2#,3# Stacking table magnet system
D	Pinch roll(S roll)	P0	Flying shear phototube
F	Pinhole detector	P1,P2,P3	1#,2#,3# Stacking table phototube
H	Flying shear	X1	1# Inferior stacking table
M2	Recoiler	X2,X3	2#,3# Quality stacking table
K2	Speed encoder	I	Examining table

***Corresponding author** e-mail: Liaoxuechao2008@sina.com.cn

2 Process flow and control principles of the system

2.1 SYSTEM PROCESS FLOW

The system process flow [3] is: steel coil is uncoiled by decoiler, then going through the thickness gauge {B}, straightener, pinch roll and pinhole detector {F}, and be feeded into flying shear. The steel coil is cut to specified length steel sheets by flying shear, and then transmitted to belt transportting system to automatically select [4]. The steel sheets are separated into quality products or the inferiors, and then stacked to specific stacking table, transmitted to packaging roll, finally packing and transport manually [5].

When a new steel coil is uncoiled and sheared [6], the previous sheets are identified as inferior sheet, the operator press the inferiors falling button and drop the previous sheets into inferior stacking table [7].

The thickness gauge {B} detects the sheets' thickness on-line [8]. If the sheet thickness exceeds the positive and negative deviation range, the certain segment sheets are identified as inferior products. The pinhole detector {F} includes left and right pinhole detecting units [9]. If any detecting unit inspects the pinhole defect, the certain sheet is identified as inferior products. On the examining table [I], the quality control inspectors inspect each sheet. If the sheet has surface defect, the certain sheet is identified as inferior products. All the defect sheets will be stacked into inferiors stacking table.

The flying shear cut the steel coil into specified length sheets [10], the sheets are conveyed into main transmitting belt {T0} by the 4#, 5# Entry belt {T4, T5}. Several permanent magnet steel bodies are installed in the middle part of main transmitting belt so as to attract the sheets just beneath the main transmitting belt. There are three magnetic field control systems {W1, W2 and W3} in the main belt above the stacking table belt [11]. When the system working in certain time frame, the system's magnetic field can counteract the permanent magnet steel bodies' magnetic field. Then the sheets under the permanent magnet steel bodies will drop onto the stacking table belt, and transmit to corresponding stacking tables.

The sheet box can be automatic adjusted to appropriate length and width, so as to contain the sheets. When the sheet quantity of quality stacking table reach the specified quantity, the sheets can be automatic packaged and transported.

2.2 SYSTEM PROCESS FLOW

The key points of the sorting control system are:

1) The system should accurately track the position of sheets, which is transmitting at a high speed (300 m/min), so as to find the inferior sheets detected by the pinhole detector and thickness gauge and make it drop into the inferior stacking table.

2) The system should be adaptive to the speed of three belts {T0, T4, T5} according to the sheet's length. The sheets' distance can be extended to a proper value, so as to be easy for the phototube to track the sheets.

3) Because the system's running speed is very high, the magnetic control system should control the sheets' dropping location and dropping location accurately according to the sheets' length and running speed, so as to avoid the pile-up accident.

3 Self-adaption sorting control model

3.1 DEFECT SHEETS POSITION TRACKING

The key point of the whole sorting control system is sheets' position tracking, especially the inferiors sheets' position tracking.

The three stacking table and flying shear {P0, P1, P2 and P3} are equipped with four phototubes, through which the sheets number can be counted and tracked.

Supposing:

- N_0: flying shear phototubes' counting number;
- N_1, N_2, N_3: P1, P2, P3 phototubes' counting number;
- L_1: the distance between thickness gauge {B} and pinhole detector {F};
- L_2: the distance between pinhole detector {F} and flying shear{H};
- L_0: the length of sheet.

We define a array R to save the defect sheet counting number.

If at the T_1 time, system detected a thickness defect sheet [12], and then thickness defect sheet counting number at the current time can be calculated according to formula (1) and storage to array R:

$$X_1 = \text{RND\_}(\frac{L_1 + L_2}{L_0}) + N_0 + 1. \tag{1}$$

If at the T_2 time, system detected a pinhole defect sheet [13], and then pinhole defect sheet counting number at the current time can be calculated according to Equation (2) and storage to array R:

$$X_2 = \text{RND\_}(\frac{L_3}{L_0}) + N_0 + 1, \tag{2}$$

where the symbol RND. represents rounded down operation.

The distance L_3 is actual sheet length between pinhole detector and the head of the next sheet.

Namely: $L_3 = L_2 + L_4 - L_5$, where L_4 is the distance between flying shear and the head of the next sheet.

The system real-time collects the counting pulse signal of pinch roll {D} speed encoder {K1}. Supposing that the counting pulse value of K1 per rotation is n_1, pinch roll's diameter is Q_1. Then corresponding distance s1 of each K1 pulse can be calculate as:

$$s_1 = \frac{Q_1 \times \pi}{n_1}, \tag{3}$$

Figure 2 is the schematic plot of calculating distance L_3. Supposing, at the time of P0 rising edge, counting value of K1 is M_1. At the time of F rising edge, counting value of K1

is M_2. The distance L_4 can be calculated as: $L_4 = (M_2 - M_1) \times s_1$.

The distance L_5 is distance correction factor, $L_5 = V_l \times 0.1666$, where V_l is pinch roll's line speed.

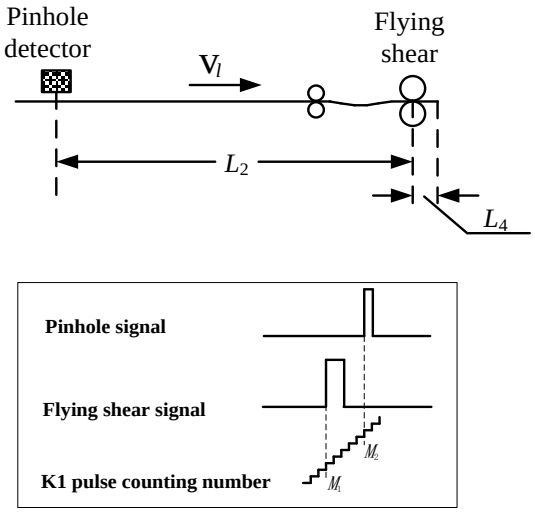


FIGURE 2 Schematic plot of calculating distance L3

Assuming the pinhole locating between the two sheets, the system make the following processing, supposing distance:

$$\Delta L = L_3 - RND\_[L_3 / L_0], \tag{4}$$

if $\Delta L < 0.1$, then the current sheet and previous sheet are all defected sheets, the counting number X_2 and X_2-1 will be stored into the array R;

if $\Delta L > 0.9$, then the current sheet and next sheet are all the defect sheets, the counting number X_2 and X_2+1 will storage into the array R;

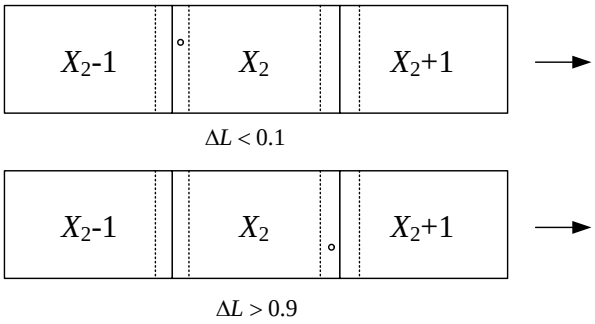


FIGURE 3 The situation of the pinhole locating between two sheet

Every time array R receiving a new element, system will sort the array R to maintain the defected sheet counting number value from small to large.

When P1 detects the counting number $N1$ to be equal to the detect sheet counting value in array R, mean: $N_1 = R[i]$, then system will drop this defect sheet into inferior stacking table.

3.2 TRANSMITTING BELT SPEED CONTROL

The system should adaptive control the speed of three belts [T0, T4 and T5] according to the sheet's length [14].

The distance between tow sheets should be extended to a proper value, so as to be easy to phototube tracking the sheets [14]. Consequently, the sorting control system can run in high speed. Thus, the speed of S roll and six belts should be real-time controlled.

Supposing:

-V_0: the max speed of main belt T0; V3
-V_1: the max speed of S roll; V0
-V_4: the max speed of belt T4; V1
-V_5: the max speed of belt T5; V2
-V_s: the max speed of stacking belt T1, T2, T3.

The speed of V_1 can be ascertained according to the sheet length. Then the other speed can be calculated as follows:

$$V_0 = V_1(1 + k_0), \tag{5}$$

$$V_4 = V_1(1 + k_4), \tag{6}$$

$$V_5 = V_1(1 + k_5), \tag{7}$$

where k_0, k_4, k_5 is the speed scale factor, which can be calculated through algorithm of Figure 2. p_0, p_4, p_5 is the setting initial value of k_0, k_4, k_5, which is the experience value according to main belt length and running speed. In this sorting system, p_0=0.25, p_4=0.1, p_5=0.2.

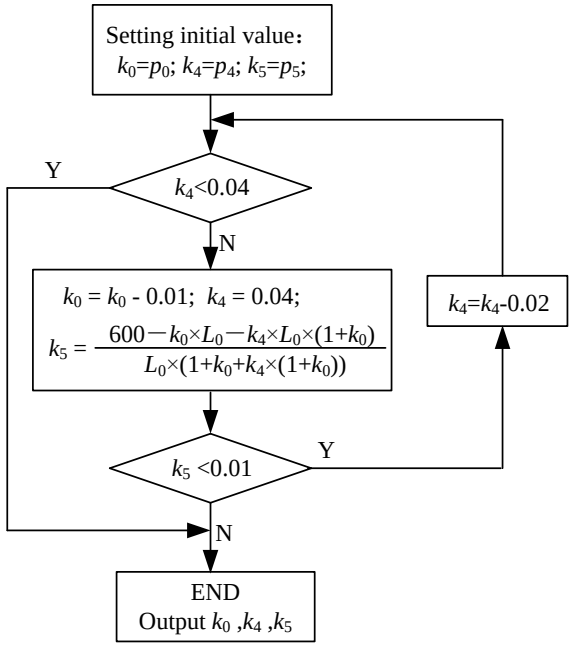


FIGURE 4 speed scale factor calculating algorithm flow

Through the above algorithm, the belt speed is getting higher and higher (the specific relationship is $V_0 > V_2 > V_0 > V_1$). So the sheets after cutting by flying shear can be extended to a certain distance (approximately 600mm), so as to be easily tracked by phototubes.

When the sheets drop into the stacking table belt, the belt conveys it to the stacking box in a lower speed V_s, V_s=V_0/3.5.

3.3 SHEETS DROPPING CONTROL

Because of the sheets running at the very high speed (the max speed can reach 300m/min) in the line, every dropping position and posture of the sheet must be accurately controlled, to avoid the sheets scratching with each other and piling up in the stacking box [14].

Take the 1# stacking table as an example, the principle of how to control the dropping location and posture is described as follows:

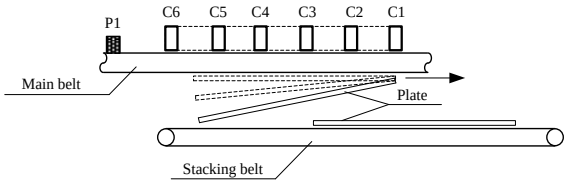


FIGURE 5 dropping control schematic diagram of 1# stacking table magnetic system

Supposing the distance between P1 and electromagnet C6 is S, the sheet length is L_0, the line speed of T0 is V_0.

The system real-time collects the pulse counting signal of the main belt encoder. Supposing that the counting pulse value of K2 per rotation is n_2, main belt roll's diameter is Q_2. Then corresponding distance *S2* of each K2 pulse can be calculated as:

$$S_2 = \frac{Q_2 \times \pi}{n_2}. \tag{8}$$

When phototube P1 detects the sheet, the system starts counting. Supposing its' counting number is *m*. When *m* satisfies the Equation (8), system will consider that the sheet reaches the dropping location, 1# Stacking table magnet system working, and the sheet drops to the 1# stacking belt.

$$m \times S_2 \geq L_0 + S_2 - V_0 \times t, \tag{9}$$

where *t* is a experience delay time.

At first, C6 losses of field, and then C5, C4, C3, C2, C1 successively loss of field delayed by a certain T period. So the sheet will drop smoothly and steadily to the stacking table belt (the sheet tail dropping firstly, the sheet nose dropping later.)

When *m* satisfies the Equation (9), system will consider that the sheet already dropped to the stacking belt, 1# Stacking table magnet system will stop working.

$$m \times S_2 \geq L_0 + S_2 + S_1, \tag{10}$$

where S_1 is the experience delay distance.

Because the main belt speed is 3.5 times higher than the stacking belt, the current dropping sheet tail will fold to the stacking belt, and it's head will fold to previous dropping sheet's head, so as to avoid the sheets scratching with each other in high speed. The sheets fold with each other and transmit to the stacking box through stacking belt.

4 Some methods of improving sorting accuracy

In order to improve the sorting control accuracy, the following hardware and software methods can be adopted:

1) Because of the time interval between pinhole signal and phototube signal is very short and the signals are easy to be interfered, the system firstly adopts limiting shake filtering method to process these signals, then call corresponding interrupt service program to respond to these signal, so as to improve the sorting system response speed of defected sheets.

2) During the running process, system will real-time regulate the relevant parameters (mainly including delay time parameter *t* and delay distance S_1) according to the running speed and sheets separation distance.

3) In order to ensure all the defected sheets will drop into the inferior stacking table and improve the system accuracy, system adopt the following methods to dispose defect signals:

- Pinhole defect signal: when pinhole defect appears between two sheets, these two sheets will be dropped into the 1# inferior stacking table.

- Thickness gauge signal: Supposing the thickness defect signal appear at T_1 time and disappear at T_2 time, the sheets passing thickness gauge during these time (T_1 to T_2) will all drop into the 1# inferior stacking table, as well as the next three sheets.

5 Conclusions

Sorting control system is an important part of cold rolling shear production line, the system section speed and accuracy is especially worth to be further in-depth studied. The essay, based on the sorting control principle, detailed analyses the realizing method of self-adaption control model in details, mainly including sheets high speed running, sheets real-time accurately tracking, sheets dropping posture accurately control and sheets defect detecting etc. The sorting control model can be applied to the similar industrial production lines.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61174107 and 61100055), Natural Science Foundation of Hubei Province (Grant No. 2011CDB233).

References

[1] Baba Y 2004 On automation of sheet sorting - by vacuum belt conveyance *Kami Pa Gikyoshi/Japan Tappi Journal* **58**(1) 41-5
[2] Nakamura H 2007 The efficiency progress with sheet sorting *Kami Pa Gikyoshi/Japan Tappi Journal* **61**(11) 10-13
[3] Grachev V G, Solodovnik F S, Kuzmina L I, Novosad P G, Yakovlev E A 1999 Electromagnetic devices for transporting, sorting and laying hot rolled sheets at units of metallurgical production *Tyazheloe Mashinostroenie* **5** 32-3
[4] Thaxton C S, Calantoni J 2006 Vertical sorting and preferential transport in sheet flow with bimodal size distributions of sediment *Proceedings of the Coastal Engineering Conference* 3056-65

[5] Nakamura H 2007 The approach for no manual sorting system *Kami Pa Gikyoshi/Japan Tappi Journal* **61**(1) 84-7
[6] Pennington N J 2003 Laser cell combines cutting and sorting *Modern Metals* **59**(8) 19-21
[7] Baba A 2003 Automation of the sorting work of planographic sheets - Planography automatic sorting machine with suction belt feeding *Kami Pa Gikyoshi/Japan Tappi Journal* **57**(2) 17-21
[8] Khotomlyanskii A L, Goloborodko V G, Shebanits E N, Svetlakova V A 1974 Statistical Analysis of thickness variations in coils of light-gauge sheet steel *Steel in the USSR* **4**(1) 46-8
[9] Zeze M, Tanaka A, Tsujino R 2001 Formation mechanism of sliver-type surface defect with oxide scale on sheet and coil *Tetsu-To-Hagane/Journal of the Iron and Steel Institute of Japan* **87**(2) 85-92
[10] Brine B J 1977 1,880 mm oscillating trapezoidal shear line *National Conference Publication-Institution of Engineers Australia* 229-32
[11] Chang B, Bai S, Du D, Zhang H, Zhou Y 2010 Studies on the micro-laser spot welding of an NdFeB permanent magnet with a low carbon steel *Journal of Materials Processing Technology* **210**(6-7) 885-91
[12] Jitsukawa Masaharu, Mitsunari Motonobu, Tanabe Shunichi, Okaimi Yuji, Sekiguchi Katsumasa, Hosoya Yoshihiro, Furuta Akihiko 1990 Equipment and operation of NK-EFL and the quality of NKE-CORE *NKK Technical Review* **60** 16-23
[13] Goto Hiroshi, Akao Noboru, Hara Nobuyoshi, Sugimoto Katsuhisa 2007 Pinhole defect density of Cr Nx thin films formed by ion-beam-enhanced deposition on stainless steel substrates *Journal of the Electrochemical Society* **154**(4) 189-94
[14] Zhu X-Y, Zhang D-H, Peng LG, Wang KY, Song J 2013 Speed optimization technology of process control in pickling line and tandem cold mill *25th Chinese Control and Decision Conference CCDC 2013* 837-40

Author

Liao Xuechao, born in 1979, Hubei Province, China

Current position, grades: the lecturer of the College of Computer Science and Technology at Wuhan University of Science and Technology.
University studies: Master degree from Wuhan University of Science and Technology in 2006.
Scientific interest: industry process control, electric motor and motor fault diagnosis.

Probabilistic neural Network with statistical feature for fault diagnosis of permanent magnet motor

Lei Dong[1, 4], Weimin Li[1], Weiguo Zhao[2]*, Yunfei Chen[3]

[1]*School of Mechanical Engineering, Hebei University of Technology, Tianjin 300130, China*

[2]*School of Water Conservancy and Electric Power, Hebei University of Engineering, Handan 056038, China*

[3]*School of Control Science and Engineering, Hebei University of Technology, Tianjin 300130, China*

[4]*Tianjin Navigation Instruments Research Institute, Tianjin 300131, China*

Received 1 June 2014, www.cmnt.lv

Abstract

Permanent magnet motors are very important components in commercially available equipments and industrial applications due to high reliability and robust performance, and it is important to take an appropriate and effective approach to diagnose fault for them. The implementation of probabilistic neural network (PNN) with the statistical features for permanent magnet motor is developed in this paper, and the statistical features are determined according to the stator current characteristics of motor to effectively reduce dimensionality of sample space. The experimental results demonstrate that, compared with RBF network, the proposed method is more effective in identifying various types of faults.

Keywords: Permanent Magnet Motor, Probabilistic Neural Network, Statistical Feature

1 Introduction

Nowadays motors are the most extensively applied to all types of industries owing to their high reliability, simple construction and robust design [1]. Especially for permanent magnet motors, they have been used in many modern industrial fields owing to their high efficiency, low weight place and high torque [2]. But they are always exposed to a variety of complex environments and conditions which are accompanied with the natural aging process of any machine, moreover, they are very sensitive to the strict constraints because of the environment of embarked systems, thus causing the motor various failures. In this way, Fault detection of electrical machines has received extensive attentions in recent years. There are many methods to detect mechanical and electrical problems in motors, which include vibration, stator current, magnetic flux density, etc. Direct observation and measurement method is used in detecting traditional motor fault diagnosis, but it is unable to meet the requirements of modern motor manufacturers [3]; Parameter estimation method needs to establish a precise dynamic model of the motor and identify the motor electromagnetic parameters through the model, thus resulting in unclear fault characteristics [4, 5].

Recently BP neural network is developed for fault diagnosis of motors and many electric power systems [6-8]. However, BP neural network has the shortcomings of slow convergence speed and a tendency to the local optimum which have a serious impact on its generalization ability to fault diagnosis [9, 10].

PNN, as a subgroup of Radial Basis Neural Networks, is good suitable for dealing with classification problems, it is based on Parzen's method of density estimation and Bayes' decision strategy. The most important advantage of PNN is the simple structure, training manner and only one free parameter. In PNN, the smoothing factor has to be adjusted by the user and it can be adjusted at run time without the requirement of network retraining. So a significant contribution in this work is that the composition of statistical features is used to train PNN as a novel classifier for the fault diagnosis of permanent magnet motor, and to evaluate performance of the classifier the BP network is compared in term of the classification accuracy and train time.

2 Probabilistic neural network

The PNN was first proposed by Specht [11], it can be considered as a normalized radial basis function network in which there is a hidden unit centred at every training case, the probability density is the scaled sum of the kernel function for all training samples and the Gaussian function is used frequently. For PNN, his most important advantage is that it has a simple structure and only one free pamaerter named as the smoothing factor; one can adjust it without having to consider network retraining at run time. The probability density function is expressed as follows:

$$\hat{f}_j(\vec{x}) = \frac{1}{(2\pi)^{p/2}\sigma^p m_j}\sum_{i=1}^{m_j} exp\left(-\frac{(\vec{x}-\vec{x}_{ij})^T\cdot(\vec{x}-\vec{x}_{ij})}{2\sigma^2}\right), \quad (1)$$

* ***Corresponding author*** e-mail: zwg770123@163.com

where, $\hat{f}_j(\vec{x})$ is the probability of vector $\vec{x}$ occurring in set m_j, $\hat{f}_j$ is the estimated density for the *j*-th class, $\vec{x}$ is test case, $\vec{x}_{ij}$ is *i*-th training sample of the *j*-th class, *p* is dimensionality of $\vec{x}_{ij}$, σ: is the smoothing factor, m_j is number of training cases in the *j*-th class.

PNN is a four-layered feed-forward network topology that implements Bayes' decision criterion, the simple architecture is depicted in Figure 1.

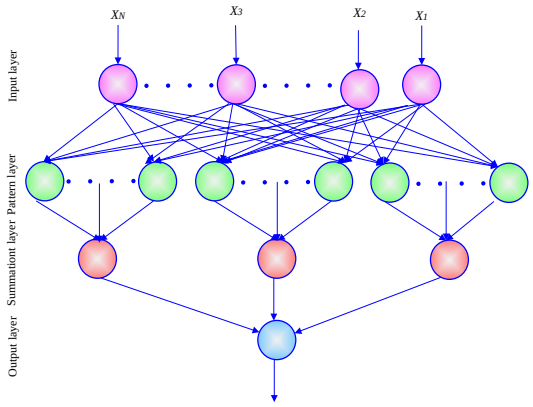


FIGURE 1 Architecture of the PNN

The PNN consists of input layer, pattern layer, summation layer and output layer [12]. The input layer simply distributes the input to the neurons in the pattern layer and does not perform any computation. For the pattern layer, it contains one pattern neuron for each training case, with an exponential activation function, a pattern neuron computes the squared Euclidean distance between a new input vector and the ith training vector of the jth class, the output of the neuron is calculated by the following multiscalar Gaussian function:

$$\hat{f}_j(\vec{x}) = \frac{1}{(2\pi)^{p/2}\sigma^p}\sum_{i=1}^{m_j} exp\left(-\frac{(\vec{x}-\vec{x}_{ij})^T\cdot(\vec{x}-\vec{x}_{ij})}{2\sigma^2}\right). \quad (2)$$

For each class, the summation layer contains a summation neuron, the summation neuron for the first sums the output of the pattern neurons that contain the training cases of the first class, the number of the neurons of the pattern layer is identical to the number of the training samples. The summation neuron for the second class sums the output of the pattern neurons that contain the training cases of the second class, and so on, the activation for each class in the summation neuron is identical to the estimated density function value of this class, the summation layer neurons calculate the maximum likelihood of the pattern vector by using equation (1). Finally, the output neurons of the output layer can obtain the result of the summation neurons. The neurons of the output layers are threshold discriminators that implement Bayes' decision criterion and the output neuron can generate the respective estimated probabilities for the test case, the neuron in the decision layer determines the class belongingness by:

$$l = \arg\min_{1\le k\le c}\{\rho_k\}, \quad (3)$$

where *c* is the number of classes of the training set and *l* is the estimated class.

3 Experimental results

For the permanent magnet motor, the current frequency is very high and the amplitude varies considerably with frequency, and the obtained sample is high-dimensional, it makes many classifiers become very complex and time-consuming in training process. To overcome the above difficulties, one must guarantee that a classifier should have a better classification accuracy with fewer feature dimensionalities.

TABLE 1 Statistical features definition

Feature number	Statistical features	Formula or description
Feature 1	Maximum value	Maximum amplitude value in a given current signal
Feature 2	Minimum value	Minimum amplitude value in a given current signal
Feature 3	Absolute mean value	$\sum_{i=1}^{N}\lvert x_i\rvert / N$
Feature 4	Standard deviation	$\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2/(N-1)}$
Feature 5	Median value	Middle amplitude value separating the greater and lesser halves in a given current signal
Feature 6	Range	Difference in maximum and minimum amplitude values in a given current signal

So in order to reduce the dimensionality of the extracted samples and obtain a low-dimensional feature vectors, we need to select an appropriate sample space and implement feature extraction. Feature extraction is regarded as a process of computation of some measures for the signal, a sample set of statistical features namely Maximum, Minimum, Absolute mean value, Standard deviation, Median value and Range, are selected for the study in Table 1. These statistical features are extracted from the current signals correspondding to various faults; Table 2 describes 6 kinds of faults; Figure 2 shows 6 kinds of different current fault signals, and Table 3 shows the statistical features correspondding to Figure 2; Figure 3 shows the scatter diagrams between any two features, it is clear that the relationship characteristics between any faults are expressed, and there not exist any two features which can distinguish all the faults. A total of 240 data sets are used in the training, each of fault consists of 40 training samples, and the trained network is tested with data sets consisting of 40 test samples, in which each of fault contain 10 test samples.

TABLE 2 Fault definition

Fault numbers	Fault description
Fault 1	The rusty spot in the bearing inner and outer rings and ball
Fault 2	Something wrong with the bearing retainer
Fault 3	A greater pitting in the bearing inner ring
Fault 4	Something wrong with the two bearing concentricity
Fault 5	A phase winding short circuit
Fault 6	No fault

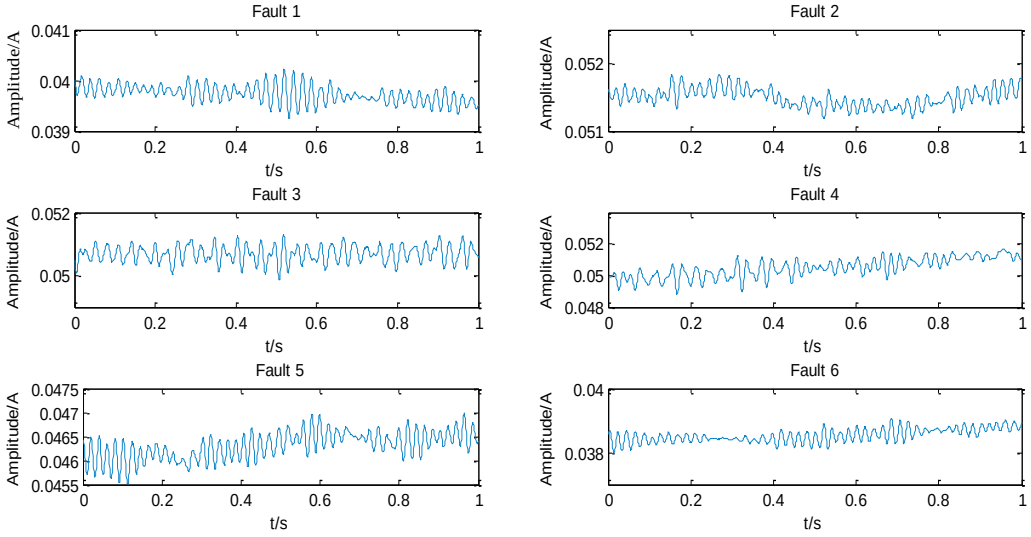


FIGURE 2 6 kinds of different current fault signals

TABLE 3 Statistical features corresponding to Figure 2

Fault number	**Statistical features**						**Output**					
	Feature 1	*Feature 2*	*Feature 3*	*Feature 4*	*Feature 5*	*Feature 6*						
1	0.03968	0.03923	0.03948	0.00010	0.00045	0.03948	1	0	0	0	0	0
2	0.05153	0.05096	0.05117	0.00012	0.00057	0.05117	0	1	0	0	0	0
3	0.05111	0.05013	0.05065	0.00024	0.00099	0.05065	0	0	1	0	0	0
4	0.05341	0.05115	0.05232	0.00049	0.00227	0.05232	0	0	0	1	0	0
5	0.04694	0.04574	0.04633	0.00026	0.00120	0.04633	0	0	0	0	1	0
6	0.03871	0.03825	0.03846	0.00011	0.00046	0.03845	0	0	0	0	0	1

The PNN and statistical features are used in fault classification of the permanent magnet motor, the network architecture and the training process are also taken into account, in addition, the performance of network generally depends on the sizes of the training set and test sets. In the developed classifiers for classification, the smoothing factor plays a viral role in the PNN, and an appropriate smoothing factor is often data dependent, meanwhile, the proper choice for the smoothing factor also improves the accuracy of the PNN. If the smoothing factor is too small, individual training patterns will be regarded only in isolation, and we only can obtain a nearest neighbor classifier, but if the smoothing factor is too high, details of the density can be blent together [13]. In general, there is no uniform approach for solving determine the smoothing factor, but we can determine it by minimizing the error, we use different smoothing factors to evaluate the performance of the PNN.

20 different values that range from 0.1 to 2 are adopted in order to better evaluate the performance to train a more accurate model. The estimation results of the PNN for the training samples using the above given values are depicted in Figure 4. It can be seen that the smoothing factor largely has effects on the training accuracy for PNN, and that the value $\sigma = 0.9$ generates the smallest training error.

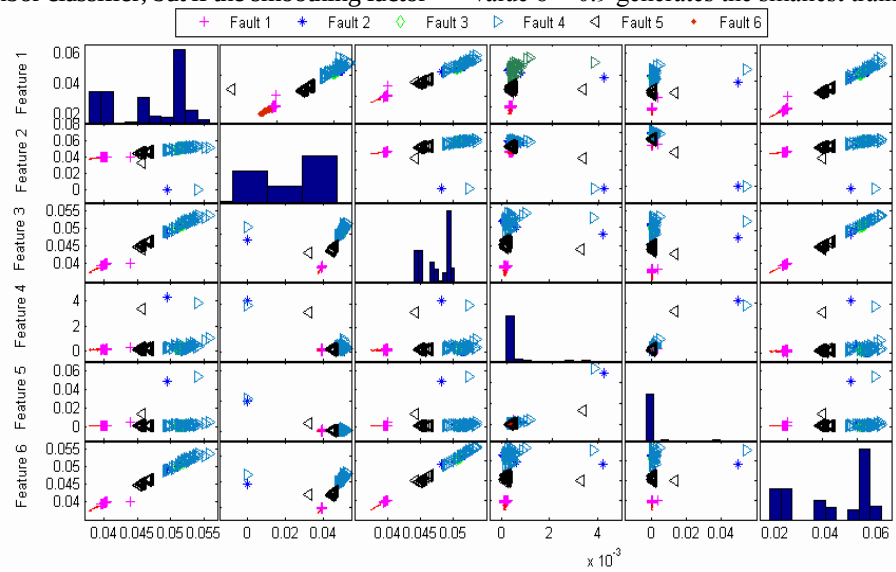


FIGURE 3 Scatter diagram between any two features

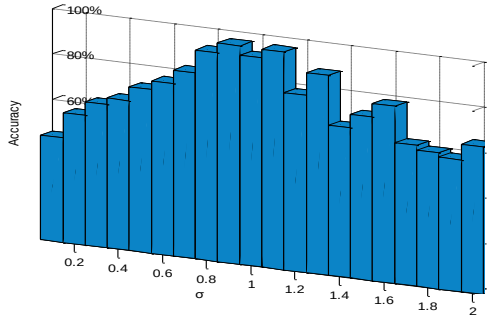


FIGURE 4 Effects of smoothing factor on training error

To compare the performance of the proposed method, the authors also processes RBF neural network for the experiment. The RBF network has three layers, including an input layer, a hidden layer with a nonlinear RBF function for activation and a output layer, the input of the network linearly combines radial basis function with some neuron parameters. For the experiment, the neuron number of the input layer of the RBF network is 6, the neuron number of the hidden layer is self-determining by autonomously increasing other neuron into it, the neuron number of the output layer of the RBF network is 6, and its training parameters are set as follows: the target error is set as 0.001 and the epoch is set as 800, the training process of RBF network is depicted in Figure 5. Then the two trained classifiers are used in testing using 60 testing samples, the comparison results of the between the two classifiers are shown in Table 4, it can be seen that from Table 4, compared with RBF network, the PNN classifier is more timesaving in training and has a higher test accuracy, thus manifesting more robustness, when the representative training samples is enough, a Bayesian optimal classifier must be obtained.

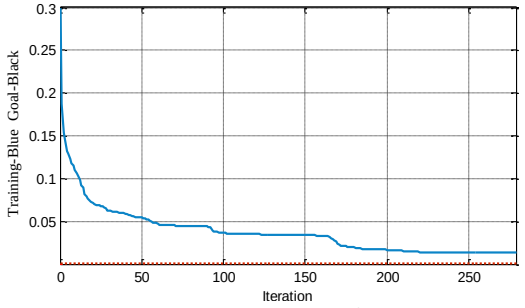


FIGURE 5 Training process of RBF

TABLE 4 Comparison between BP neural network and DDAG-SVM

Method	Training accuracy	Testing accuracy	Training time
PNN	97.62%	95.62%	0.60s
RBF	96.46%	92.08%	17.82s

4 Conclusion

In this study, PNN is used as a robust classifier for diagnosing various faults using the statistical features extracted from the stator current signals of the permanent magnet motor, the extracted statistical features effectively reduce dimensionality of sample space, the comparative experiment shows its the superiority of the performance, and the proposed method in the present work gives a possible application to the monitoring system in industry, which can largely reduce the accidents and ensure the safety of human life.

Acknowledgments

This work is supported by the Science and Technology Research Project of University of Hebei Province No. Y2012016, the Science Research Foundation of Hebei Education Department of China No. 2009422, and the Natural Science Foundation of Hebei Province of China No. E2010001026.

References

[1] Bacha K, Salem Ben S, Chaari A 2012 An improved combination of Hilbert and Park transforms for fault detection and identification in three-phase induction motors *International Journal of Electrical Power & Energy Systems* **43**(1) 1006–16
[2] Huangfu Y G, Liu W G, Ma R Q 2008 Permanent Magnet Synchronous Motor Fault Detection and Isolation Using Second Order Sliding Mode Observer *The 3rd IEEE Conference on Industrial Electronics and Application*, Singapore 639-44
[3] Liu X Q, Zhang H Y, Liu J 2000 Faultdetec-tion and diagnosis of permanent) magnetic DC motor based on parameter estimation and neural network *IEEE Transactions on Industrial Electronics* **47**(10) 1021-30
[4] Wen F, Willett P, Deb S 2000 Signal processing and fault detection with application to CH-46 helicopter data *IEEE Aerospace Conference Proceedings* Big Sky, MT 15-26
[5] Liu M L, Hu X D, Cui S M 2005 Current signal analysis and process used in fault diagnosis of permanent-magnetic DC motor *Journal of Harbin institute of technology* **37**(6) 836-8
[6] El-Sharkawi M A, Niebur D 1996 A Tutorial Course on Artificial Neural Networks with Applications to Power Systems *IEEE Tutorial Course Text, IEEE*, New York
[7] Aneta M, Andrea K, Saso S 2009 The role of artificial neural networks in detection of pulmonary functional abnormalities *Engineering Review* **29**(2) 1-11
[8] Lacroix B, Clegg D 1989 Microprocessor motor protection relay *Proceedings of 4th international conference on developments in power system protection* 235–8
[9] Zhou M R, Nie M Y 2013 Application of the Spectrum Peak Positioning Technology Based on BP Neural Network in Demodulation of Cavity Length of EFPI Fiber Optical Sensor *Journal of Computer and Communications* **1** 67-71
[10] Elif D Ü 2008 Probabilistic neural networks employing Lyapunov exponents for analysis of Doppler ultrasound signals *Computers in Biology and Medicine* **38** 82-9
[11] Specht D F 1990 Probabilistic neural networks *Neural Networks* **3**(1) 109–18
[12] Wang J S, Chiang W C, Hsu Y L, Yang Y T C 2013 ECG arrhythmia classification using a probabilistic neural network with a feature reduction method *Neurocomputing* **116**(20) 38-45
[13] Li P 2011 Structural damage localization using probabilistic neural networks *Mathematical and Computer Modelling* **54** 965-9

Authors

Lei Dong, born on April 6, 1979, Tangshan, Hebei Province, China

Current position, grades: Now she is Ph.D. candidate in School of Mechanical Engineering, Hebei University of Technology
University studies: M.S. degree in Institute of Mechanism, Harbin Institute of Technology, in 2005
Scientific interests: electric machine design and fault diagnosis of motor

Weimin Li, born in April, 1964, China

Current position, grades: professor at Hebei University of Technology, China.
University studies: Ph.D. degree in mechanical manufacture and robotics automation technology from Tianjin University, China in 1998.
Scientific interests: Robots and numerical control technology.

Weiguo Zhao, born on January 23, 1977, Xingtai, Hebei Province, China

University studies: M.S. degree from the School of Computer Science and Software Engineering, Hebei University of Technology in 2005.
Scientific interests: intelligent computing, information processing, and intelligent fault diagnosis.

Yunfei Chen, born on March 29, 1979, Xingtai, Hebei Province, China

University studies: M.S. degree from control theory and control engineering, Hebei University of Technology in 2010.
Scientific interests: electrical automation design and ntellectual electric apparatus.

SVM-Based evaluation model for college laboratory learning

Xiaoling Tan[1]*, Zefu Tan[1], Juan Qu[2], Guangwen Xi[3]

[1]*School of Electronic and Information Engineering, Chongqing Three Gorges University, Chongqing 404100, China*

[2]*School of Mathematics and Statistics, Chongqing Three Gorges University, Chongqing 404100, China*

[3]*Network Centre, Chongqing Three Gorges University, Chongqing 404100, China*

Received 1 June 2014, www.cmnt.lv

Abstract

Evaluation for laboratory learning is based on different factors, while each factor is varied by individuals. Hence it is difficult to express the quantitative nonlinear functional relationship among the evaluation indexes. With limited sample, Support Vector Machine (SVM) could be generalized by compromising between model's complexity and learning ability. That is its advantage on the evaluation of small sample, nonlinear and multi-indexes. It is a good try to apply Support Vector Machine (SVM) to laboratory learning evaluation. With Support Vector Machine (SVM), the relationship between the learning quality and evaluation indexes could be revealed. Experiments show that Support Vector Machine (SVM) model is with high prediction accuracy, faster speed and simple algorithm. It is suitable and more reasonable for laboratory learning evaluation.

Keywords: Support Vector Machine (SVM), evaluation model, laboratory learning quality

1 Introduction

Being different from that of classroom teaching, the valuation of laboratory teaching consists of various factors. An objective, comprehensive, impartial and accurate evaluation is significant to the improving of teaching quality, motivating of learning, and even the optimizing of discipline construction. Evaluation on college laboratory learning combines theory with practice. It works under evaluation index system, while the system is guided by educational theories. Educational theories are the summary of educational practice, which reveal the law of education [1]. On the other side, laboratory learning evaluation is based on different factors, while each factor is varied by individuals. Students have to abide the rules and regulations at the laboratory as well as to be innovative and motivate to experiments. It is also a basic quality for the students to keep the laboratory tidy and safe when they are learning. That shows the laboratory learning is interacted with multi-factors, which should be taken into account at the evaluation [2, 3].

Being lack of approximation capability to nonlinear relationship, the use of multiple linear regression and partial least square method is restricted. As a new technology with the features of nonlinear mapping, learning classifying and real-time optimizing, artificial neural nets (ANN) opened a new way on pattern recognition, nonlinear classification and artificial intelligence. It also has been applied a certain into educational evaluation. However, the features of large calculation, continually local extremum, and low generalization impact the application of ANN [4]. SVM is a new machine learning method based on statistical learning theory. Its topological structure depends on support vector. Differs from that of ANN depended on designer's experience, it solve Ann's problem of high dimension, local minimum and small sample. It covers the advantages of neural network and gray-scale model. After revealing the relationship between the laboratory learning quality and influence factors by learning the existed sample, an accurate and objective evaluation on a specific laboratory learning could be realized by SVM.

Laboratory learning evaluation is essentially a classification by some decision-making mechanism and parameters. To the limited sample, SVM could be generalized by compromising between model's complexity and learning ability. That is its advantage on classification. SVM also overcomes the shortcomings of multi-layer feedforward neural network. It specially aims at the limited samples, not only for an optimal solution of infinite samples, but for the one of the current existed samples. It becomes a quadric form algorithm of optimization finally, which theoretically could achieve a global optimal solution. By the nonlinear transformation, SVM transfers the issue to a high dimensional feature space, and realizes the original nonlinear classification by constructing a linear classification function in the space. That ensures the category is with the ability of generalization, solves the curse of dimensionality and reduces the impact to the speed. To the classification, multi-layer feedforward neural network could realize the nonlinear classification, but the classifier would not be the best. SVM based on statistical theory ensures the optimality of classification theoretically, while the generalization ability of algorithm is further ensured [5]. On the other side, the laboratory learning evaluation is a typical multi-class problem. It is feasible to apply SVM to construct a multi-class model for the evaluation.

* ***Corresponding author*** e-mail: xltan023@hotmail.com

2 SVM

SVM (Support Vector Machine) is a pattern recognition proposed by Vapnik et al in 1995. It is good at solving the problem of small sample, nonlinear and high dimensional machine learning. Its basic idea is: to form an optimal hyper-plane at sample space or feature space to get a max distance between the hyper-plane and the sample sets, so that to gain the maximum generalization ability [6].

2.1 FUNDAMENTAL THEORY OF SVM

SVM is the youngest and most important part of statistical learning theory. It becomes a conventional tool on machine learning as it matures theoretically and practically. So far, it is on the highest performance on application to the text and image classifications of pattern recognition [7].

SVM realizes the following thought: it maps the input vector X to a high dimensional feature space by a nonlinear mapping selected in advance. An optimal classification hyper-plane is formed in the feature space. In order to get an optimal classification hyper-plane in the feature space Z, the display format of Z is not taken into consider, only the calculation of inner product of vectors of SVM and the feature space is needed.

According to functional theories, when kernel function $K(x_i,x_j)$ satisfies Mercer condition, an inner product in a transformation space is corresponded. The basic thought of nonlinear SV is mapping the input vector x_i to a high dimensional Hilbert space (feature space) by a nonlinear mapping selected in advance. An optimal classification hyper-plane is formed in the feature space. The inner product between any two points in the space could be reflected by the kernel function, which is corresponded to the input vectors in the original space.

Different inner product kernel functions of SVM form different classifiers. Here are the major 3 kernel functions.

Firstly, polynomial kernel function. See Equation (1).

$$K(x, x_i) = [(x, x_i) + 1]^p . \tag{1}$$

Secondly, radial basis function (RBF), namely Gaussian kernel function. See Equation (2).

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\delta^2} \right). \tag{2}$$

The significant distinction between RBF classifier and traditional classifier is: the centre of each primary function corresponds to an SV, and its output weight is automatically determined by algorithm. RBF is the most effective kernel function so far.

Thirdly, sigmoid function. See Equation (3).

$$K(x, x_i) = \tanh\left[k(x \bullet x_i) - \delta \right]^q . \tag{3}$$

In this case, SVM acts as a multilayer perception with an implicit strata. The hidden node is automatically determined by algorithm. Moreover, in the algorithm, there is no local minimum, which perplexes neural network [8].

2.2 SVM SORTING ALGORITHM

At present, 3 methods have been proposed to solve SVM multiclass classification problem. They are: One-Against-Others, One-Against-One and Decision Directed Acyclic Graph (DDAG) [9].

2.2.1 One-Against-Others

It was proposed by Vapnik in 1998. The basic thought is: to construct several two-class classifiers and combined them together to accomplish the multi-class classification. Each classifier separates one class from the others, and infers the affiliation of an input *x*. It constructs k SVM two-class sub-classifiers for class k. The i^{th} SVM adopts training samples in the i^{th} class as the positive training samples, while adopting the others as negative ones. The unclassified samples are classified to the class with maximum classification function. Disadvantages: Firstly, every SVM classifier takes all the samples as training samples. When the quantity is large, the training time is long. Secondly, the two-class problems are asymmetric, which means the number of two-class samples are imbalanced.

2.2.2 One-Against-One

It was proposed by Knerr. It constructs all the possible k(k-l)/2 two-class classifiers in the training samples of class k. Each two-class classifier separates class *i* and class j by a simple classification rule. There is many two-class problems in it, so more classifiers are needed than that of One-Against-Others. However, the scale of each problem is much less than that of One-Against-Others. So the training speed is faster. But if k is too large, sub-classifiers will be more and the training speed would slow down.

2.2.3 Decision Directed Acyclic Graph

It was proposed by Platt etc. It is the same with One-Against-One at the training phase. But at the decision phase, DAG started from root node is used. Every internal node is a two-class classifier, and the leaf node is the final class. DDAG is faster at decision than that of One-Against-Others and One-Against-One. However, the result of classification is uncertain. Different classifier as the root node may lead to different classification result.

3 Laboratory learning evaluation model based on SVM

The evaluation includes the aspects of knowledge application, thinking innovation, experiment operation and safety, and experimental attitude. Knowledge application means students' ability of learning, principle theory grasping and practical application. Thinking innovation means innovation consciousness, independently thinking, cooperation,

and exploring to solve problems. Experiment operation and safety means safety awareness, and sanity habit. Experimental attitude means initiative, responsibility, being truthful and practical, and spirit of teamwork [10].

3.1 EVALUATION INDEX SYSTEM OF LABORATORY LEARNING

The students in the laboratory class of Basis of Database at Chongqing Three Gorges University were taken as the sample. Each sample has 7 indexes. Evaluation on knowledge application is based on the experiments and lab reports of database *x1*, form *x2* and programming *x3*. *x4*, the evaluation on thinking innovation is based on comprehensive assessment on the experiment and lab report. *x5*, the one on experiment operation and safety is based on the average value of ordinary checks. *x6*, the one on experimental attitude is based on the comprehensive assessment to the experimental discussion and optional experiments. What's more, a network learning system has been founded to record students' network learning. The average value *x7* based on each online testing is taken as the other index.

According to the above index system, 100 samples in a grade were taken as the dataset of training and testing. The evaluation result was classified into 5 ranks: excellent (identified as 1), good (identified as 2), middle (identified as 3), pass (identified as 4) and failed (identified as 5). Each sample was composed by 7 inputs and 1 target output. To group the data into two, each group randomly consisted of data from all the 5 ranks. The first group with 60 samples was training dataset. The second one with 40 samples was testing dataset. The visualization figure of training data is showed as Figure 1. Its abscissa axis is student's number, the vertical axis is the classification result. Please see Figure 1.

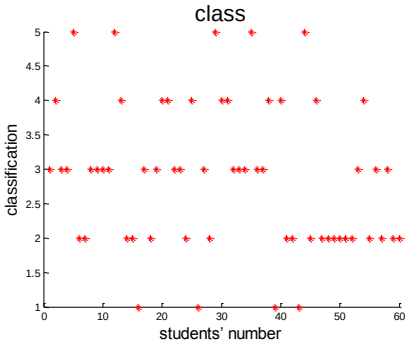


FIGURE 1 Classification of training samples

3.2 NORMALIZATION OF INPUT SAMPLES

It is important to normalize data before adopting SVM, which could eliminate the influence of dimension. Firstly, according to the features of the index data, the big index would weaken the influence of small index to the model if it is not normalized. Secondly, SVM needs to calculate inner product kernel function of sample point. Without normalization, the overlarge value will cause trouble on calculation. In order to reduce the complexity of calculation and shorten the time of training, every data was normalized to [0, 1].

There are many methods of normalization. The frequently-used are MinMax and Exponential Function. MinMax is a linear transformation, by which the original sense of the data could be preferably remained with fewer information loss after normalization. Hence MinMax was adopted to normalize the input samples to [0,1], which is showed as the following. See Equation (4).

$$x_t^{'} = \begin{cases} \dfrac{x_i - \min}{\max - \min}, & if \quad \min < x_i < \max \\ 0, & if \quad x_i < \min \\ 1, & if \quad x_i \geq \max \end{cases} . \tag{4}$$

The original data of evaluation on laboratory learning is showed as Table 1. Please see Table 1.

TABLE 1 Original data of evaluation on laboratory learning

No.	Database x1	Form x2	Programming x3	Thinking innovation x4	Laboratory specification x5	Experimental attitude x6	Network learning x7	classification
1	2	6	6	4	4	5	75	3
2	6	5	5	2	1	5	47	5
3	9	9	7	2	3	5	49	4
4	9	9	7	2	5	5	84	2
5	9	9	7	2	5	5	76	2
6	10	10	7	10	6	5	84	1
7	9	2	6	4	3	5	79	3
8	9	9	6	4	5	5	80	2
9	9	9	7	2	4	5	71	3
10	10	8	7	0	3	5	57	4

The normalized data is showed as Table 2. Please see Table 2.

TABLE 2 Normalized data

No.	Database x1	Form x2	Programming x3	Thinking innovation x4	Laboratory specification x5	Experimental attitude x6	Network learning x7	classification
1	0.6	0.5	0.375	0.2	0	1	0.274	3
2	0.9	0.9	0.625	0.2	0.4	1	0.306	5
3	0.9	0.9	0.625	0.2	0.8	1	0.871	4
4	0.9	0.9	0.625	0.2	0.8	1	0.742	2
5	1	1	0.625	1	1	1	0.871	2
6	0.9	0.2	0.5	0.4	0.4	1	0.790	1
7	0.9	0.9	0.5	0.4	0.8	1	0.806	3
8	0.9	0.9	0.625	0.2	0.6	1	0.661	2
9	1	0.8	0.625	0	0.4	1	0.435	3
10	0	0.8	0.625	0	0.4	0.5	0.661	4

3.3 DESIGN OF MULTI-CLASSIFIED SVM EVALUATION MODEL

The SVM-based laboratory learning evaluation is the classification of learning quality by the pattern classification of SVM.

The evaluation data is formed by the data of 5 ranks, so it is multi-classified. Because the classification patterns are not so many, so One-Against-One was selected. k(k-l)/2 SVM classifiers were constructed to the data of class K. Each classifier only classified two class. This model is simple and has well ability on classification. The well performed RBF kernal was adopted as SVM kernel function.

Two parameters have to be determined in SVM training process: penalty parameter C for mistaken classification, and RBF kernel control parameter. In order to reduce human influence, meanwhile the real-time requirement is not strict in the evaluation system, cross-validation was adopted to search for optimal parameter. It helped the system to be with some adaptivity. The working process of the system is showed on Figure 2. Please see Figure 2.

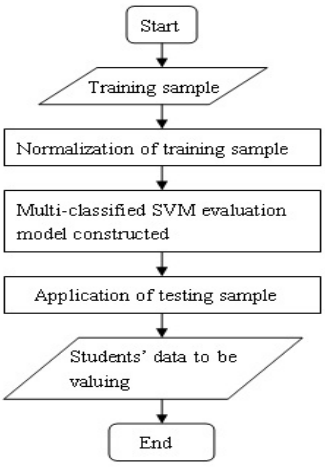


FIGURE 2 Process of evaluation system

4 Experimental result and analysis

The experimental result is showed on Figure 3. In the laboratory learning evaluation of the 40 students, only 6 are not fit the real one. The other 34 are in accordance to the reality. The accordance rate is 85%. It proves that the trained multi-classified SVM evaluation model could simulate teacher's evaluation thought. It could analysis the evaluation data rapidly and reduce the randomness of human evaluation from different teachers, which helps to rank the students more objectively and precisely. Please see Figure 3.

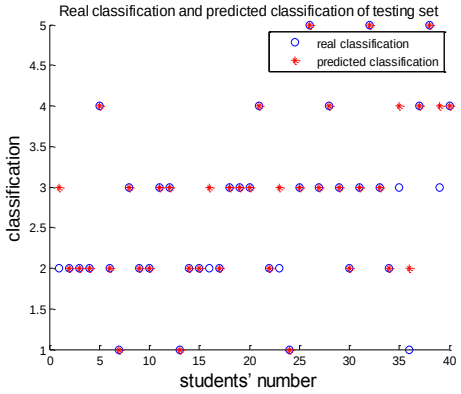


FIGURE 3 Real classification and predicted classification of testing set

5 Conclusions

Because it is difficult to show the quantitative nonlinear functional relationship among the evaluation indexes, the final evaluation rating is usually subjective and one-side. The founding of laboratory learning evaluation system makes the evaluation more logically and reasonably. SVM learning model, with the ability of approximating any nonlinear relationship between input and output, could seek out the relationship between learning quality and the evaluation indexes by learning the given sample. Experiment shows that SVM model is with high prediction accuracy, faster speed and simple algorithm. It is suitable to the laboratory learning evaluation. It is theoretically and practically significant to the development of laboratory teaching.

Acknowledgments

This work was supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No.KJ1401009), Program for Innovation Team Building at Institutions of Higher Education in Chongqing (Grant No.KJTD201320), Research Program for Teaching Reform of Higher Education in Chongqing (Grant No.1203201), Research Program for Teaching Reform of Higher Education in Chongqing Three Gorges University (Grant No.JG120671) and Key Laboratory of Signal and Information Processing at Chongqing Three Gorges University.

References

[1] Huang Y M, Chiu P S, Liu T C, Chen T S 2011 The design and implementation of a meaningful learning-based evaluation method for ubiquitous learning *Computers & Education* **57**(4) 2291-302
[2] Teräs M, Lasonen J 2013 The development of teachers' intercultural competence using a change laboratory method *Vocations and Learning* **6**(1) 107-34
[3] Alaoutinen S, Heikkinen K, Porras J 2012 Experiences of learning styles in an intensive collaborative course *International Journal of Technology and Design Education* **22**(1) 25-49
[4] Satzger B, Kramer O 2013 Goal distance estimation for automated planning using neural networks and support vector machines *Natural Computing* **12**(1) 87-100
[5] Lee K, Cho S, Asfour S 2011 Web-based algorithm for cylindricity evaluation using support vector machine learning *Computers & Industrial Engineering* **60**(2) 228-35
[6] Liu X, Chen Z, Han L, Zhong K 2011 Support Vector Machines for Eco-Environmental Quality Evaluation *Energy Procedia* **13** 6689-95
[7] Wen C J, Zhan Y Z 2011 Support vector data description discriminant analysis *Journal of Jilin University (Engineering and Technology Edition)* **41**(6) 1709-13
[8] Yu L, Yue W, Wang S, Lai K K 2010 Support vector machine based multiagent ensemble learning for credit risk evaluation *Expert Systems with Applications* **37**(2) 1351-60
[9] Brezak D, Majetic D, Udiljak T, Kasac J 2012 Tool wear estimation using an analytic fuzzy classifier and support vector machines *Journal of intelligent manufacturing* **23**(3) 797-809
[10] Lau K H 2014 Computer-based teaching module design: principles derived from learning theories *Medical education* **48**(3) 247-54

Authors

Xiaoling Tan, born on April 22, 1969, China

Current position, grades: associate professor at Chongqing Three Gorges University, China.
University studies: majored in mathematical statistics at Nankai University, China in 1990. **Scientific interests:** data mining, intelligent computing and information security.
Publications: more than 20 research papers.

Zefu Tan, born on January 23, 1969, China

Current position, grades: professor at Chongqing Three Gorges University of China, and a master's supervisor of Chongqing University of Posts and Telecommunications of China.
University studies: Master degree in electromagnetic field and microwave technology from Beijing University of Posts and Telecommunications, China in 2002.
Scientific interests: mobile broadband communication techniques and electronic information technology.
Publications: more than 30 research papers.

Juan Qu, born on August 4, 1984, China

Current position, grades: lecturer at Chongqing Three Gorges University.
University studies: BS degree in 2006, and M.S. degree in Applied Mathematics from Shaanxi Normal University, China in 2009.
Scientific interests: Cryptography, information security, and authentication protocol.
Publications: more than 10 research papers.

Guangwen Xi, born on March 19, 1982, China

Current position, grades: lecturer at Chongqing Three Gorges University.
University studies: BS degree in 2005, and M.S. degree in computer science and technology from Chongqing University, China in 2011.
Scientific interests: computer network security and cloud computing.
Publications: more than 10 research papers.

Usefulness of lethal chromosomes in genetic algorithms solving the constrained optimization problems

Yalong Zhang[1]*, Hisakazu Ogura[2], Xuan Ma[3]

[1]*College of Electrical and Information Engineering, Quzhou University, Quzhou 324000, China*

[2]*Graduate School of Engineering, University of Fukui, Fukui 910-8507, Japan*

[3]*Faculty of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China*

Received 1 June 2014, www.cmnt.lv

Abstract

The infeasible solutions are often generated in population as evolutionary computation solving the combinatorial optimization problems. The number of infeasible solutions impacts the performance of the evolutionary computation searching the optimal solution, in the worst case the algorithm ceases to run. In genetic algorithms, encoding of infeasible solutions is referred to as lethal chromosomes. In this study, we discover a property of lethal chromosomes that: although lethal chromosomes carry out the infeasible solutions in genetic algorithms, their statistical property implies an underlying similarity with the exact solution of the optimization problems. Hereby we propose an operation using statistical property of lethal chromosomes to handle with the lethal chromosomes themselves. Simulation experiments on a large number of test cases demonstrated that it can improves obviously the performance of genetic algorithms to use the statistical property of lethal chromosomes.

Keywords: Genetic Algorithm, Infeasible Solution, Lethal Chromosome, Evolutionary Computation.

1 Introduction

With representing the solution of problem as a chromosome, genetic algorithm (GA) refers the chromosomes that violate the constraint conditions to as lethal chromosomes (LCs). In the population of GA, due to crossover and mutation operations, LCs are often generated with high rates. The greater number of LCs in the population, the worse search performance of the GA.

Early, Iima Hitoshi (1995) investigated the effects of LCs on the performance of the GA, but did not propose a method for handling these problems [1]. Mengchun Xie (1996) proposed an algorithm model to revive the LCs by random crossover and mutation operations [2]. In recent years, researches focusing on the problems associated with infeasible solutions have made some progress. Lyndon While et al. (2009-2013) have been made some achievements on problems associated with infeasible solutions in evolutionary computation [3-7].

In this paper, we propose genetic algorithm on multidimensional knapsack problem (MDKP) that handles the LCs with a reviving operation (RGA). Specifically, we (2009) previously proposed another genetic algorithm using lethal chromosomes on multi-knapsacks problems [8]. But regrettably, we mistakenly referred the multi-knapsacks problems to as multidimensional knapsack problem in [8]. This paper is really to solve the multidimensional knapsack problem, and the reviving operation is totally different.

2 A genetic algorithm using LCs on MDKP

As an NP-hard problem, MDKP is to find a subset of objects that maximizes the total profit while satisfying some resource constraints, which can be formulated as:

$$Maximize \quad \sum_{j=1}^{n} v_j x_j \,, \tag{1}$$

$$s.t. \quad \sum_{j=1}^{n} w_{ij} x_j \le c_i \quad \forall i \in I \,, \tag{2}$$
$$x_j \in \{0,1\} \quad \forall j \in J,$$

where n is the number of objects, m is the number of resources, v_j is the value associated with object j, w_{ij} is the consumption of resource i for object j, c_i is the available quantity of resource i (capacity of knapsacks for the i^{th} resource), and x_j is the decision variable with object j and is set to 1 if j^{th} object is selected (and is otherwise set to 0). $I = \{1, 2, \ldots, m\}$, $J = \{1, 2, \ldots, n\}$.

To solve the MDKP, many researchers consider the GA or other evolutional algorithm. P.C. Chu (1998) presented a GA for MDKP with a repair operator [9]. Günther R. Raidl (1999) proposed a weight-coding in GA for MDKP [10]. Günther R. Raidl (1998) presented an improved hybrid GA for MDKP [11]. Jens Gottlieb (2000) solved the MDKP with a permutation-based GA [12].

* ***Corresponding author*** e-mail: yalong-2008@hotmail.com

2.1 ALGORITHM STRUCTURE

Our RGA has two types of chromosome pools, namely, the *living island* and *lethal island*. The former contains chromosomes, referred to as non-lethal (feasible) chromosomes, which satisfy all constraints. The latter consists of the LCs. In the living island, chromosomes are evolved by genetic operations, and in the lethal island, LCs are revived by reviving operations. This structure is described as following way.

```
1  while initializing population do
2  |  Generate initial population, and then move the
   |  non-lethal chromosomes into the living island, the
   |  lethal chromosomes into the lethal island;
3  end
4  while evolving process do
5  |  for living island do
6  |  |  All of the chromosomes in the living island
   |  |  evolve into the next generation by genetic
   |  |  operations, and the new lethal chromosomes
   |  |  move to lethal island. In this process, the
   |  |  vaccine must be trained;
7  |  end
8  |  for lethal island do
9  |  |  Each chromosome in the lethal island is handled
   |  |  by reviving operations and is then moved to the
   |  |  living island;
10 |  end
11 end
12 Repeat the evolving process until meeting the
   termination conditions of the GA;
```

ALGORITHM 1 Algorithm structure of RGA

We use roulette as selection operation, two-point crossover and mutation rates of 0.05 in genetic operations. Next, we mainly focus on reviving operation, which is performed in *lethal island*.

2.2 REVIVING OPERATION

Idea of the reviving operation is introduced from immune system of biology and medicine. As FIGURE 1, the previous infection led the system to generate the immune memory; the system handles the re-infection with previous memory for avoiding from being infected again.

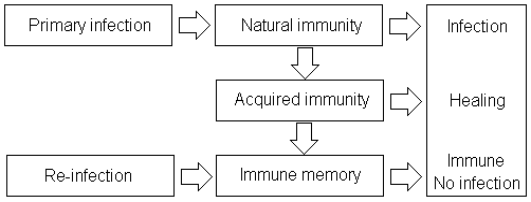


FIGURE 1 Immune system

In this context, we record the statistical information of previous LCs as immune memory, and then handle the succeeding LCs by previous memory. The operation to handle the LCs is called reviving operation in this paper.

As an immune memory, a multi-valued string schema $s_1s_2...s_n$ is constructed and has the initial value that $s_j = 0$, $\forall j \in J$. The $s_1s_2...s_n$ will be used as *immune memory* by reviving operation, so we call it *vaccine*. During the RGA evolving, while every LC referred as $x_1x_2...x_n$ generated, to perform the ALGORITHM 2 flushing the vaccine $s_1s_2...s_n$.

```
   Input: a lethal chromosome x_1 x_2 ... x_n
   Output: flushed vaccine s_1 s_2 ... s_n
1  for (j = 1 to n) do
2  |  if (x_j = 1) then
3  |  |  s_j ← s_j + 1;
4  |  end
5  end
```

ALGORITHM 2 Training the vaccine with lethal chromosomes

By this way the $s_1s_2...s_n$ is a changing string, which is refurbished by every lethal chromosome while evolution of population as long as LCs appearing. As LCs generated are handled by reviving operation in every generation, while the vaccine is refurbished by LCs while the vaccine is used by reviving operation in RGA.

In order to achieve an efficient implementation of the reviving operation, a pre-processing routine is applied to vaccine $s_1s_2...s_n$ that sorts and renumbers genes of LC according to the *increasing* order of s_j's. Assuming $x_1x_2...x_n$ with pre-processing routine is the LC to be revived, proposed reviving operation is described by pseudo-code as following.

```
   Input: a lethal chromosome
   Output: a revived chromosome
1  Let W_i = Σ_{j=1}^{n} w_ij x_j, ∀i ∈ I;
2  for (j = 1 to n) do
3  |  if (x_j = 1) and (W_i > c_i, ∀i ∈ I) then
4  |  |  x_j ← 0;
5  |  |  W_i ← W_i − w_ij, ∀i ∈ I;
6  |  end
7  end
8  for (j = n to 1) do
9  |  if (x_j = 0) and (W_i + w_ij ≤ c_i, ∀i ∈ I) then
10 |  |  x_j ← 1;
11 |  |  W_i ← W_i + w_ij, ∀i ∈ I;
12 |  end
13 end
```

ALGORITHM 3 Reviving operation

3 Computational experiments

A set of standard test data of the MDKP was referred and used by P.C. Chu [9], Günther R. Raidl [10, 11], and Jens Gottlieb [12]. These test data contain 10 instances for each combination of $m \in \{5, 10, 30\}$, $n \in \{100, 250, 500\}$, and $\alpha \in \{0.25, 0.50, 0.75\}$, where $\alpha = c_i / \Sigma_{j=1}^{n} w_{ij}$ being the tightness ratio of instance. Since the exact solution values for most of these problems are unknown, the quality of a solution is measured by the percentage gap of the objective value *fitns* with respecting to the optimal value of the LP-relaxed problem f^{LP}_{max}: $\%\text{-}gap = 100 \times (f^{LP}_{max} - fitns) / f^{LP}_{max}$. The proposed RGA is tested on these 270 MDKP instances and the mean results are shown in TABLE 1.

TABLE 1 Computational results of other algorithms and RGA

Problems	**Average *%-gap***					
All of 270 instances	GA with H1	GA with H2	Swap	Insert	Improved GA	RGA (proposed)
Average	0.589	0.646	0.641	0.629	1.13 - 0.53	0.5472

We also list the available results of other references in TABLE 1 to compare with. The first columns indicate test is on the 270 instances. The next columns in turn report the other results of average *%-gap*. That are *GA with H1* and *GA with H2* proposed by Günther R. Raidl [10], *Swap* and *Insert* are from Jens Gottlieb [12], and the *Improved GA* reported by also Günther R. Raidl [11] for *%-gap* obtained from initial population to $(10^6)^{th}$ generations. The last column reports the results of RGA with the computer condition that CPU is Celeron 1.0 and the algorithms were coded in Visual C++.net (2003).

To analyse the test results, above all we discuss the condition of experiment. The *GA with H1* and the *GA with H2* was tested with population size of 100 and 10^5 solutions had been evaluated. *Swap* and *Insert* had evaluated 10^6 non-duplicate solutions and then get results. *Improved GA* recorded the best solution from initial population to $(10^6)^{th}$ generations. Comparing with them, RGA obtain the results with population size of 50 and running terminates at $(10^4)^{th}$ generation.

Averagely for all instances, RGA obtained the *%-gap* as 0.5472, which is smaller than the results from the algorithm *GA with H1*, *GA with H2*, *swap* and *Insert*, but except the *Improved GA*. *Improved GA* of Günther R. Raidl [11] obtained the best average *%-gap* as 0.534 at $(10^6)^{th}$ generation, it is very great. Because such an experiments to 10^6 generations will cost too CPU seconds to finish for us, we test the RGA only for 10^4 generations and obtained the finial average *%-gap* as 0.5472 with population size 50. Relatively, the population size Günther R. Raidl adopted in *Improved GA* is 100. To compare with *Improved GA* at same generations, we calculated the average *%-gap* obtained also at $(10^4)^{th}$ generation for *Improved GA* according to data in Refs. [11], the result is 0.6211 which greater than 0.5472 despite of large population size 100.

4 Discussion on feature of LCs

To analyse the feature of LCs, we firstly give a definition of *similarity ratio* between two chromosomes. To any two chromosomes X_1 ($x_1^1 x_2^1 \ldots x_n^1$) and X_2 ($x_1^2 x_2^2 \ldots x_n^2$), the *similarity ratio* of them is defined by:

$$similarity\ ratio(X_1, X_2) = 100\frac{1}{n}\sum_{j=1}^{n}(x_j^1 \odot x_j^2), \qquad (3)$$

where, if $x_j^1 = x_j^2$, then $x_j^1 \odot x_j^2 = 1$; otherwise $x_j^1 \odot x_j^2 = 0$. Actually, the *similarity ratio* is used to measure how many the two chromosomes have same value genes.

In order to study feature the vaccine contained, it is hypothesis that at a certain generation $s_1 s_2 \ldots s_n$ is processed to a binary-value string $s'_1 s'_2 \ldots s'_n$ by the way that: for $\forall j \in J$, if ($s_j > t$) then $s'_j \leftarrow 1$, otherwise $s'_j \leftarrow 0$, where t is the threshold be used to classify the s_j's into one or zero, the t is determined as an appropriate value so that it takes the max *similarity ratio* for $s'_1 s'_2 \ldots s'_n$ and *exact chromosome* (*exact solution*) of the problem. Here it is assume that the exact chromosome is known beforehand. We consider the value of *similarity ratio* of $s'_1 s'_2 \ldots s'_n$ and exact chromosome as *feature of vaccine*.

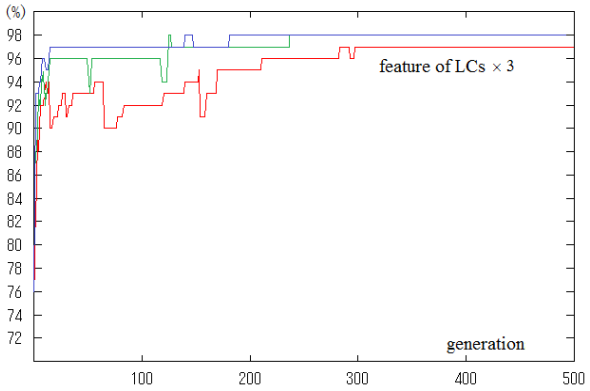


FIGURE 2 Feature of LCs ["feature of LCs" (×3).]

To three MDKP instances picked from OR-library [9] that $m = 5$, $n = 100$, we firstly solve their exact chromosomes with branch and bound method (BBM) to measure the *feature* of vaccine. Their curves of *feature of vaccine* in RGA are shown as FIGURE 2.

We could learn from FIGURE 2 that (i) the *features of vaccine* are increase in the overall; (ii) the *features of vaccine* rise to a high degree at later generation that some to 97 % and some to 98 %. That also means most of genes of exact chromosome could be indicated the value via to classifying the corresponding *bit* of vaccine with a threshold *t*.

5 Conclusions and future research

The paper discovered an important feature of lethal chromosomes, and proposed RGA to use this feature of lethal chromosomes based on reviving operation. Appling RGA to MDKP, a large number of testing results indicate that using lethal chromosomes based on reviving operation could obviously improve search performance of GA. At same time there are also some works should to be done to improve the performances of RGA deeply.

In addition to the method of reviving operations should be researched further in the future, researches should focus on not only lethal chromosomes, but also to excavate the more evolutional resource in infeasible solution regions. On the other hand, if the proposed reviving operations are improved to decrease the time complexity and allow rapider operation, the GA will be applicable to a wider range of field.

Acknowledgments

This work is supported by the Scientific Research Foundation (Project No.Y201430654) funded by the Zhejiang Provincial Education Department (China).

References

[1] Iima Hitoshi, Sannomiya Nobuo 1995 The Influence of Lethal Gene on the Behavior of Genetic Algorithm *The society of instrument and control engineers* **31**(5) 569-76

[2] Mengchun Xie, Tetsushi Yamaguchi, Tomohiro Odaka, Hisakazu Ogura 1996 An Analysis of Evolutionary States in the GA with lethal Genes *The Information and Systems Society, Institute of Electronics, Information and Communication Engineers* **J79-D-II**(5) 870-8

[3] Lyndon While, Philip Hingston 2013 Usefulness of Infeasible Solutions in Evolutionary Search: an Empirical and Mathematical Study *Proc. of 2013 IEEE Congress on Evolutionary Computation*, Cancún, México 1363-70

[4] Deepak Sharma, Prem Soren 2013 Infeasibility Driven Approach for Bi-objective Evolutionary Optimization *Proc. of 2013 IEEE Congress on Evolutionary Computation (CEC)*, Cancun, Mexico 868-75

[5] Tapabrata Ray, Hemant Kumar Singh, Amitay Isaacs, Warren Smith 2009 Infeasibility Driven Evolutionary Algorithm for Constrained Optimization *Constraint-Handling in Evolutionary Optimization Studies in Computational Intelligence* **198** 145-65

[6] Patryk Filipiak, Krzysztof Michalak, Piotr Lipinski 2011 Infeasibility Driven Evolutionary Algorithm with ARIMA-Based Prediction Mechanism *Intelligent Data Engineering and Automated Learning - IDEAL 2011* **6936** 345-52

[7] Maristela Oliveira Santos, Sadao Massago, Bernardo Almada-Lobo 2010 Infeasibility handling in genetic algorithm using nested domains for production planning *Computers &Operations Research* **37**(2010) 1113-22

[8] Yalong Zhang, Xuan Ma, Jousuke Kuroiwa, Tomohiro Odaka, Hisakazu Ogura 2009 A Genetic Algorithm with Utilizing Lethal Chromosomes *Proc. of 2009 IEEE International Conference on Fuzzy Systems*, Jeju Island, Korea 2047-50

[9] Chu P C, Beasley J E 1998 A Genetic Algorithm for the Multidimensional Knapsack Problem *Journal of heuristics* **4**(1) 63-86

[10] Günther R Raidl 1999 Weight-Codings in a Genetic Algorithm for the Multiconstraint Knapsack Problem *Proceedings of the 1999 ACM symposium on applied computing*, Texas, United State 291-6

[11] Günther R Raidl 1998 An Improved Genetic Algorithm for the Multiconstrained 0–1 Knapsack Problem *Proceedings of the 5th IEEE International Conference on Evolutionary Computation*, Anchorage, AK 207-11

[12] Jens Gottlieb 2000 Permutation-Based Evolutionary Algorithms for Multidimensional Knapsack Problems *Symposium on Applied Computing Proceedings of the 2000 ACM symposium on Applied computing*, Como, Italy 408-14

Authors

Yalong Zhang, born on April 6, 1976, China

Current position, grades: works at College of Electrical and Information Engineering, Quzhou University, Quzhou, China; and is involved in research related to intelligent information processing.
University studies: master's course at Xi'an University of Technology, China, in 2007, and received his D.Eng. degree from University of Fukui, Japan, in 2011

Hisakazu Ogura, born on October 1, 1946, Japan

Current position, grades: member of IECE Japan, IPS Japan, and SOFT Japan
University studies: D.Sc. degree from Kyoto University, Japan, in 1977
Scientific interests: in knowledge representation and processing in the fields of artificial intelligence, medical informatics and medical image processing by applying genetic algorithms, artificial neural networks, symbol processing and/or fuzzy theory.
Experience: was a professor of the Department of Human and Artificial Intelligent Systems at the Faculty of Engineering, University of Fukui, Japan.

Xuan Ma, born on February 13, 1962, China

Current position, grades: professor at the Faculty of Automation and Information Engineering, Xi'an University of Technology, China and working on evolutionary computation and intelligent information processing
University studies: D.Eng. degree from University of Fukui, Japan, in 2002

Crop canopy temperature model of ditch-cultivated based on artificial neural network

Min Zhang[1], Qiang Fan[2], Fucang Zhang[3], Xia Li[1], Xuzhang Xue[4], Guodong Wang[1]*

[1]*College of Science, Northwest A&F University, Yangling 712100, Shaanxi, China*

[2]*College of Water Resources and Architectural, Northwest A&F University, Yangling 712100, Shaanxi, China*

[3]*Key laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas of Ministry of Education, Northwest A&F University, Yangling 712100, Shaanxi, China*

[4]*National Engineering Research Centre for Information Technology in Agriculture; Beijing 100097, China*

Received 1 June 2014, www.cmnt.lv

Abstract

Aiming at the mechanism model are influenced by multiple random factors, this paper establishes canopy temperature models based on BP network and RBF network respectively. The models take the temperature, humidity, illumination, soil temperature and ditch depth in the closed greenhouse as input neurons and takes canopy temperature as the output neuron. The results show that both models can well predict ditch-cultivated crop canopy temperature. The mean error between the simulation value and measured value of BP network model is 0.8408℃, and root-mean-square error of 0.5789℃. Actual output and expected output of RBF network model differ little, mean error of 0.2236℃ and root-mean-square error of 0.3496℃. In contrast, RBF network model can more accurately predict crop canopy temperature of ditch-cultivated than BP network model.

Keywords: Ditch-cultivated, Canopy Temperature Model, Artificial Neural Network

1 Introduction

Crop canopy is important space for crops' photosynthesis and transpiration. The change in canopy temperature is influenced by growing environment of crops. It is significant basis to measure water status of crops and provides necessary data support for irrigation control and greenhouse management [1, 2]. So, study on the relationship between crop canopy temperature and each environmental factor and establishment of canopy temperature model are of great significance.

Currently, both Chinese and foreign experts establish multiple simulation models for plant canopy temperature focus on field [3, 4], multi-span greenhouse and Venlo greenhouse [2, 5-8] and study effects of plant canopy temperature under different environments on water status and plant growth conditions. Modeling approach for existing plant canopy temperature is mechanism method, i.e. energy balance method with environmental data as variables and multiple regression equation method. Microenvironment of ditch-cultivated in closed greenhouse is more complex than common greenhouse, and also influenced by multiple random factors. Besides, it is difficult to confirm soil parameters. Thus, modeling with mechanism method greatly influences prediction accuracy of ditch-cultivated crop canopy temperature. ANN has the self-learning ability to simulate human thinking and can realize self-adjustment according to environmental changes. In addition, it owns nonlinear adaptation information processing capability and very strong fault-tolerant capability and can overcome errors caused by parameter selection in mechanism method. ANN expresses dynamic modeling of complex environment through limited parameters and is thus applied widely in information processing, intelligent control and production. In addition, it opens up a new research approach for establishing crop canopy temperature model.

This paper takes potted Chinese cabbage under ditch-cultivated mode in closed greenhouse and adopts BP network and RBF network to establish crop canopy temperature models, respectively. Temperature, humidity, illumination, soil temperature and ditch depth in the closed greenhouse serve as the input of the network and crop canopy temperature in the ditch serve as the output of the network. Then, the relationship between crop canopy temperature under ditch cultivation environment and environmental factors is analysed to provide theoretical support for further optimizing structural design and optimum greenhouse control of ditch cultivation method and analysing plant irrigation and water status.

2 Introduction of artificial neural network

ANN was first put forward by Warren S McCulloch and Water H Pitts in 1943 and owns such advantages as parallel processing, distributed storage, self-learning and associative

* ***Corresponding author*** e-mail:Gdwang211@yahoo.com.cn

memory. It develops on the basis of multiple disciplines including biology, cognitive science, modern neurosciences and computer science and reflects fundamental characterristics of functions of human brain. It is widely applied and developed in artificial intellectual control and optimization, pattern recognition and graphic processing, prediction and intelligent information management, communication and complex system modeling.

2.1 NEURON MODEL

ANN is formed through widely interlinking large quantities of neurons and used to simulate the structure and functions of brain nervous system. Neuron as a basic unit of neural network abstracts and simulates biological neurons. Figure 1 shows a simplified neuron model. It is a nonlinear component with multi-input and single output.

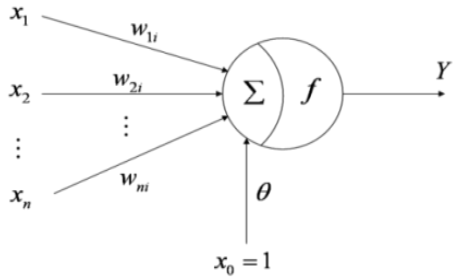


FIGURE 1 The model structure of single neuron

Its input-output relationship can be described as:

$$I_i = \sum_{j=1}^{n} w_{ji} x_j - \theta_i , \tag{1}$$

$$Y = f(I_i) , \tag{2}$$

where, w_{ji} is connection weight from cells j to cells i (for excited state, w_{ji} takes positive value; for inhibitory state, w_{ji} takes negative value); x_j refers to input signal from other cells $(j = 1, 2, \cdots\cdots, n)$; θ_i is the threshold value of neuron unit; n means the number of input signals; Y refers to neuron output; $f(.)$ is a transfer function. The transfer function can be a linear function, but it is usually the nonlinear function like the step function or *S* curve.

ANN explores human image thinking. It simplifies knowledge expression. After network structure is confirmed, ANN can carry out self-learning and self-correction of deviation according to sample data and imply "summarizing rules" gained through self-learning in weight matrix.

2.2 BP NEURAL NETWORK

Back-Propagation model is an error back propagation learning algorithm used for multilayer feed-forward neural network and was proposed by Rumelhart, Hinton and Williams in 1986. BP network contains input layer and output layer as well as one-layer or multi-layer hidden layers. Its structure is shown in Figure 2:

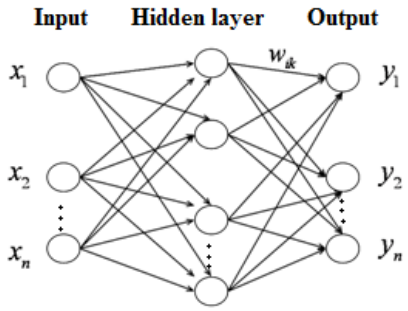


FIGURE 2 The structure of BP networks

When information is inputted in the network, the information is first transmitted to the hidden layer through the input layer, and then transmitted to the next hidden layer after the conversion by the transfer function until information is transmitted to the output layer to output. The information will go through the conversion by the transfer function whenever it passed one layer. The transfer function of BP network should be differentiable anywhere. Usually, *S* function is adopted. For example:

$$f(x) = \frac{1}{(1+e^{-x})} . \tag{3}$$

BP learning algorithm is a learning algorithm with the tutor. It adjusts network connection weight through the deviation between actual output and expected output of the network so that the network can gain expected output for any input. BP learning algorithm trains the network with a group of training samples. Each sample contains input and expected output. During the training, firstly, input information in the training samples is transmitted into the network; the network conducts calculation from the first hidden layer and transfers the calculation result to the next layer until the result is transmitted to the output layer. Every layer of neuron only influences the state of the next layer of neuron. If expected output is not gained in the output layer, calculate the error between the output value of the network and the expected output of the sample; then the network returns the error signal along the original connection passage layer by layer and adjusts connection weight of the nodes in each layer so as to gradually reduce the error until the error meets requirements.

BP network is composed of forward transfer of information and back propagation of the error. Forward transfer is used to calculate network and solve the output for an input. Back propagation is used to transmit the error layer by layer and modify the connection weight so that the network can calculate correctly. Once the network is used to solve practical problems after the training, only forward propagation is needed, without the need of back propagation.

2.3 RBF NEURAL NETWORK

Radial Basis Function is composed of input layer, hidden layer and output layer. Its structure is shown in Figure 3. The nodes in the input layer transmit input signal to the hidden

layer. The nodes in the hidden layer are composed of radial functions like Gaussian function. The basis function in the nodes in the hidden layer influence input signal partially and have local approximation capability. When the input signal approaches the central scope of the basis function, nodes in the hidden layer will generate large output. The nodes in the output layer usually adopt linear function.

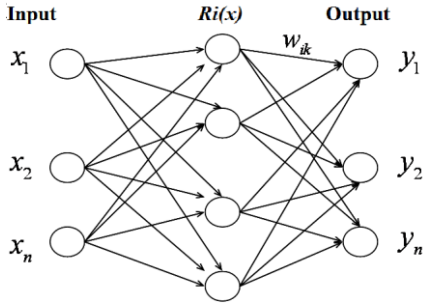


FIGURE 3 The structure of RBF networks

Common basis function in RBF neural network is Gaussian function:

$$R_i(x) = \exp[-\frac{\|x - c_i\|^2}{2\delta_i^2}] \quad i = 1, 2, ..., m, \tag{4}$$

where, c_i is the centre of the i^{th} basis function; δ_i refers to the variable of the i^{th} perception which decides the width of the basis function embracing the central point; $\|x - c_i\|$ is the norm of vector quantity $x - c_i$ and expresses the distance between x and c_i.

RBF neural network mainly adopts stochastic algorithm and self-organizing learning algorithm to select network centre. The two learning algorithms cannot be used for online learning in dynamic input mode. The foundation of valid algorithm is that all possible sample data must be gained in advance. Thus, they are only applicable to offline learning in static mode. Beside, the number of neurons in the hidden layer of RBF network based on above two learning algorithms should be confirmed artificially. Since it is necessary to consider the number of neurons in the hidden layer and a suitable norm, the difficult in solving problems increases.

Zhu Mingxing and Zhang Delong come up with RBF network model based on nearest neighbor-clustering learning algorithm. Nearest neighbor clustering learning algorithm is an online self-adaptation clustering learning algorithm. It is unnecessary to confirm the number of neurons in the hidden layer in advance. RBF network through clustering is optimal. Besides, the algorithm can be used for online learning [9].

According to nearest neighbour clustering learning algorithm, the output of RBF network should be:

$$f(x^k) = \frac{\sum_{i=1}^{M} w_i \exp(-\frac{|x^k - c_i|^2}{r^2})}{\sum_{i=1}^{M} \exp(-\frac{|x^k - c_i|^2}{r^2})}. \tag{5}$$

The radius r decides complexity of dynamic self-adaptation RBF network. If r is smaller, the number of clusters gained is more; the calculation amount is larger; the precision is higher. If r is larger, the number of clusters gained is less; the calculation amount is smaller; the precision is lower. Since r is a one-dimensional parameter, an appropriate r can be usually found through experiments and error information. It is much more convenient than confirming the number of hidden units and a suitable norm. Since each pair of input-output data may generate a new cluster, such dynamic self-adaptation RBF network actually carries out adaptive adjustment of the parameter and structure simultaneously.

3 Brief introduction to basic information

The experiment is conducted in a closed greenhouse in water-saving irrigation experiment station (east longitude 108°04'; northern latitude 34°17'; elevation 506m) of key laboratory of agricultural water-soil engineering in arid region of Northwest A&F University in Yangling, Shaanxi. The closed greenhouse is 3m wide and 12m long. The height of the ridge is 2.5m. It is covered by white polyethylene film. The tested cultivar is Chinese cabbage, which is cultivated with substrate. Substrate moisture in the pot is adequate during seeding. In the growing process, supplement water 2 times of water consumption caused by transpiration at the interval of 2 days.

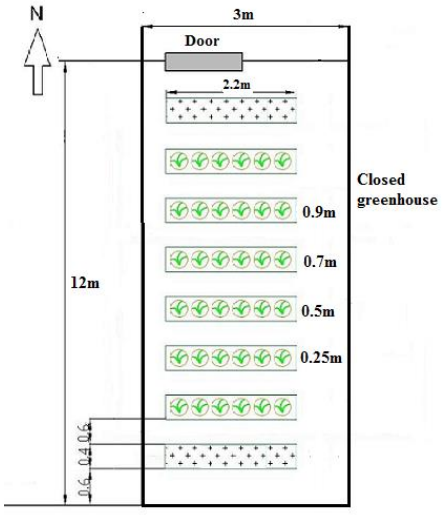


FIGURE 4 Layout of experience

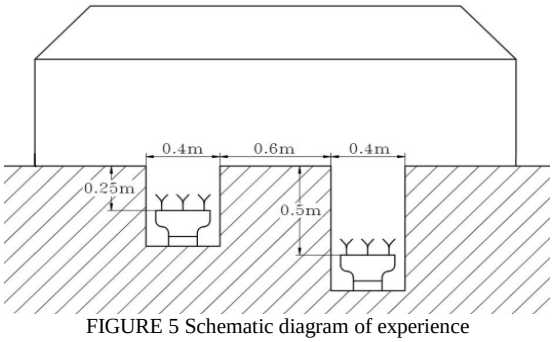


FIGURE 5 Schematic diagram of experience

Multiple rows of ditches with different depth are excavated along the east-west direction in the closed greenhouse. The length and width of each ditch are 2.2m and 0.4m respectively, at the interval of 0.6m, as shown in Figure 4 and Figure 5. Substance surface of the pot is placed in the ditches with the depth of 0.25m, 0.5m, 0.7m and 0.9m respectively. Repeat 5 times for each ditch. To prevent effects of ditch wall desquamation and soil moisture on air humidity, the plastic film is covered on the ditch wall and the bottom of the ditch.

Temperature and humidity in the greenhouse are collected with temperature and humidity collector. The probe is hung at 0.5m in the greenhouse. TES1339 professional illuminometer is used to measure the illumination. Intelligent agriculture controller is adapted to measure soil temperature. The sensor is buried underground at 0.25m in the ditch. The data are stored at the interval of 10min. The data measured at 7:00~20:00 on July 8-13, 2013 are selected as model training and verification data. The main measurement items and test instruments are shown in Table 1.

TABLE 1 List of experimental apparatus

Measurement items	Measuring instrument	Technical parameter
Temperature	Temperature and humidity collector	Measuring range: -30~70℃; Measuring accuracy: 0.01℃
Humidity	Temperature and humidity collector	Measuring range: 0~99.9; Measuring accuracy: 0.1
Illumination	TES 1339 Professional illuminometer	Measuring range: 0.1lux~200000lux; Measuring accuracy: 0.1lux
Soil Temperature	Temperature collector	Measuring range: -40~80℃; Measuring accuracy: 0.2℃
Depth Ditch	Meter	Measuring range: 0~2m; Measuring accuracy: 0.001m

4 Crop canopy temperature model based on ANN

According to the analysis of canopy temperature transmission mechanism, the temperature, humidity, soil temperature and illumination in the greenhouse as well as ditch depth and their change rate have certain mapping relation with crop canopy temperature in the greenhouse. The change of any parameter can impose direct influences on crop canopy temperature in the greenhouse. So, this model predicts crop canopy temperature through monitoring changes in the temperature, humidity, illumination, and soil temperature and ditch depth on July 8-13, 2013. Since the experiment mainly considers effects of high-temperature environment in the closed temperature on ditch-cultivated crop canopy temperature, the data at 7:00～20:00 on each measuring day are selected. 200 groups of data are gained after abnormal data and repeated data are eliminated. The data are divided into 2 sample sets. 170 groups of data as training sample set of the network are used to train the network established. 30 groups of data as inspection sample set of the network are used to inspect accuracy of model prediction after the training.

4.1 CROP CANOPY TEMPERATURE MODEL BASED ON BP

Crop canopy temperature model based on BP network is shown in Figure 6.

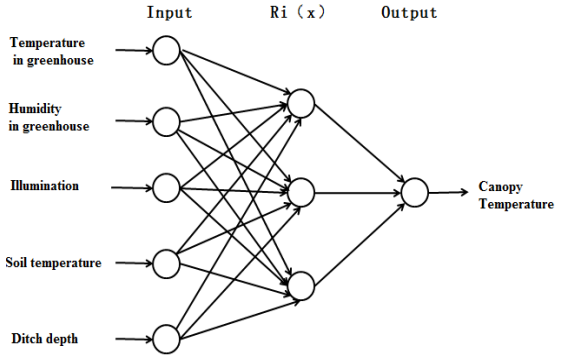


FIGURE 6 BP model of canopy temperature

The number of neurons in the input layer is 5, including the temperature, humidity, illumination, and soil temperature and ditch depth in the greenhouse. The number of neurons in the output layer is 1, indicate as crop canopy temperature. In BP network model, the hidden layer adopts *S* activation function. The output layer adopts linear activation function. It is known from empirical formula and actual training that the number of neurons in the hidden layer is 27.

$$(S_i+1)S_h+(S_h+1)S_o=n_pS_o\,, \tag{6}$$

S_i means the number of neurons in the input layer; S_h is the number of neurons in the hidden layer; S_o is the number of neurons in the output layer; n_p is the number of training samples.

Since the dimension of network input and output is different and the quantity differs greatly, normalization processing of sample data must be conducted before network training to prevent small numerical value information from being submerged by large numerical value information, i.e. normalize input signal to [0, 1]. The normalization equation is:

$$X^*=\frac{X-X_{\min}}{X_{\max}-X_{\min}}\,, \tag{7}$$

X^* refers to normalized data (the value is between 0 and 1); X is the actual value of network input data; $X_{\max}$ and $X_{\min}$ are the maximum value and the minimum value of input data.

To verify effectiveness and reliability of BP neural network model on canopy temperature prediction, MATLAB software is used to train and simulate BP neural network model. Since the length of the paper is limited, simulation results of only 12 groups of experimental data are provided, as shown in Table 2.

In accordance with the simulation results, actual output and expected output of BP network model differ little, with mean error of 0.8408℃ and root-mean-square error of 0.5789℃. It thus can be seen that BP network has relatively accurate prediction result for ditch-cultivated crop canopy temperature.

TABLE 2 Simulation results of BP networks model

NO.	Input					Output	Measured
	Temperature / ℃	*Humidity/%*	*Illumination/klx*	*Soil temperature / ℃*	*Ditch depth/m*	*Canopy Temperature/ ℃*	*Canopy Temperature/ ℃*
1	36.84	41.56	16.27	14.1	0.25	35.478	35.1
2	44.17	32	42.02	14.2	0.25	41.62	42.48
3	48.94	31.78	47.06	14.21	0.25	44.773	45.23
4	36.84	36.44	11.01	13.55	0.5	33.397	32.96
5	44.17	29.78	41.09	13.6	0.5	39.03	41.6
6	48.94	29.11	44.46	13.67	0.5	42.868	45.31
7	44.17	27.78	19.88	12.2	0.7	35.927	34.94
8	48.94	28	41.76	12.22	0.7	38.655	38.61
9	46.35	27.56	32.26	12.25	0.7	37.089	36.88
10	36.84	66.67	5.628	10.9	0.9	29.796	29.44
11	44.17	64	12.61	10.9	0.9	33.226	32.19
12	48.94	57.11	36.58	10.9	0.9	34.523	34.21

4.2 CROP CANOPY TEMPERATURE MODEL BASED ON RBF

Negative gradient descent is adopted to adjust BP network weight. Such method has certain limitations. It has the shortcomings as follows: slow convergence rate; long training time; easy to be caught in local minimum. Radial basis function network is superior to BP network in terms of approximation capability, classification ability and learning rate. Thus, this paper adopts radial basis function network to establish crop network temperature model in order to improve prediction precision of canopy temperature.

In a bid to verify effectiveness and reliability of RBF neural network model on canopy temperature prediction, MATLAB software is used to train and simulate RBF neural network model. The same 200 groups of measured data serve as sample data of the neural network. The input of the neural network includes the temperature, humidity, illumination, soil temperature and ditch depth in the greenhouse. Expected output of the network is crop canopy temperature. The neural network model is as shown in Figure 7.

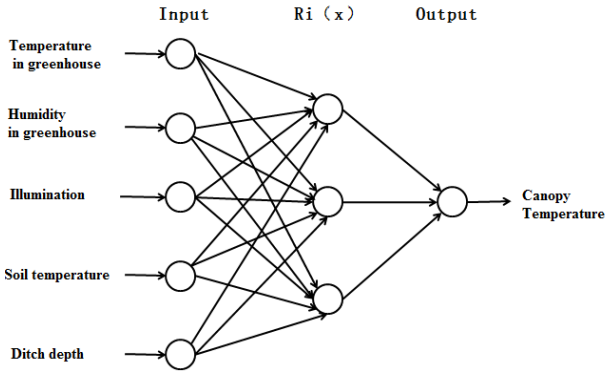


FIGURE 7 RBF model of canopy temperature

To prevent small numerical value information from being submerged by large numerical value information, normalization processing of sample data must be conducted before network training to normalize input signal to [0, 1]. 170 groups of experimental data are taken as the training samples of the network after normalization. During the training, the width of the basis function is 0.15. After the training, the number of neurons in the hidden layer is 92. 30 groups of data serve as inspection sample set of the network. It is known through network training that the model has fast convergence rate and short training time. Table 3 shows RBF network simulation results of 12 groups of experimental data.

TABLE 3 Simulation results of RBF networks model

NO.	Input					Output	Measured
	Temperature / ℃	*Humidity/%*	*Illumination/klx*	*Soil temperature / ℃*	*Ditch depth/m*	*Canopy Temperature/ ℃*	*Canopy Temperature/ ℃*
1	36.84	41.56	16.27	14.1	0.25	35.106	35.1
2	44.17	32	42.02	14.2	0.25	42.645	42.48
3	48.94	31.78	47.06	14.21	0.25	45.082	45.23
4	36.84	36.44	11.01	13.55	0.5	33.05	32.96
5	44.17	29.78	41.09	13.6	0.5	41.518	41.6
6	48.94	29.11	44.46	13.67	0.5	45.026	45.31
7	44.17	27.78	19.88	12.2	0.7	36.32	34.94
8	48.94	28	41.76	12.22	0.7	38.489	38.61
9	46.35	27.56	32.26	12.25	0.7	36.992	36.88
10	36.84	66.67	5.628	10.9	0.9	29.457	29.44
11	44.17	64	12.61	10.9	0.9	32.464	32.19
12	48.94	57.11	36.58	10.9	0.9	34.206	34.21

Based on the simulation results, RBF network model has fast computing speed. Besides, the simulation value and measured value differ little, mean error of 0.2236℃ and root-mean-square error of 0.3496℃. It thus can be seen that

RBF network has more accurate prediction result of ditch-cultivated crop canopy temperature than BP network.

5 Conclusions

Aiming at the mechanism model are influenced by multiple random factors and it is difficult to confirm soil parameters, this paper takes Chinese cabbage as the object of study and establishes canopy temperature simulation models based on BP network and RBF network respectively. The models take the temperature, humidity, illumination, soil temperature and ditch depth in the greenhouse as input neurons and takes canopy temperature as the output neuron. Moreover, experimental verification of the models is carried out through the measured data. The results show that both models can well predict ditch-cultivated crop canopy temperature. The mean error between the simulation value and measured value of BP network model is 0.8408℃, and root-mean-square error of 0.5789℃. Actual output and expected output of RBF network model differ little, mean error of 0.2236℃ and root-mean-square error of 0.3496℃. In contrast, RBF network model can more accurately predict crop canopy temperature than BP network model.

Acknowledgments

This work was supported by National High Technology Research and Development Program of China (Project NO: 2012AA101903; 2011AA100504), and Youth Foundation of Northwest A&F University (Z109021202).

References

[1] Alderfasia A, Nielsen D C 2001 Use of crop water stress index for monitoring water status and scheduling irrigation in wheat *Agricultural Water Management* **47**(1) 69-75
[2] Wang Xiaochan, Ding Weimin, Luo Weihong, Dai Jianfeng 2007 Simulation and analysis of cucumber canopy temperature in greenhouse in Yangtze River delta during summer *Transactions of the CSAE* **23**(4) 196-200
[3] Bristow K L 1987 On solving the surface energy balance equation for surface temperature *Agricultural and Forest Meteorology* **39**(1) 49-54
[4] Zhang Keying, Ma Youxin, Li Yourong, Liu Yuhong 1999 A preliminary study on climatologically simulating model of plant canopy temperature *Acta Meteorologica Sinica* **57**(4) 473-81
[5] Kittas C, Katsoulas N, Baille A 2003 Influence of an aluminized thermal screen on greenhouse microclimate and canopy energy balance *Transactions of the ASAE* **46**(6) 1663
[6] He Fen, Ma Chengwei 2009 Modeling greenhouse plant canopy temperature in north China during winter *Transactions of the Chinese Society for Agriculture Machinery* **40**(5) 169-72
[7] Zhan Y, Jewett T J, Shipp J L 2002 A dynamic model to estimate in-canopy and leaf-surface microclimate of greenhouse cucumber crops *Transactions of the ASAE* **45**(1) 179-92
[8] Wang S, Deltour J 2004 Leaf temperature modeling of greenhouse grown tomato *International Agricultural Engineering Journal* **13**(1) 64-70
[9] Zhu Mingxing, Zhang Delong 2000 Research on the function centre selection algorithm of RBF network *Journal of Anhui University Natural Science Edition* **24**(1) 72-8

Authors

Min Zhang, born on February 2, 1982, China

Current position, grades: candidate in College of Science, Northwest A&F University.
Scientific interests: environmental biophysics and model.

Qiang Fan, born on September 27, 1980, China

Current position, grades: doctor candidate in College of Mechanical and Electronic Engineering, Northwest A&F University
Scientific interests: wireless sensor network and compressed sensing.

Fucang Zhang, born in February, 1962, China

Current position, grades: professor, doctor candidate instructor.
Scientific interests: water saving irrigation theory and technique, agricultural soil and water environment, etc.

Xia Li, born in October, 1978, China

Current position, grades: associate professor.
Scientific interests: environmental biophysics and model.

Xuzhang Xue, born in October, 1967, China

Current position, grades: research fellow.
Scientific interests: water saving agriculture, precision agriculture, etc.

Guodong Wang, born in August, 1957, China

Current position, grades: professor, doctor candidate instructor.
Scientific interests: environmental biophysics, etc.

Analysis on vibration of small-scale hydroelectric generating unit

Liying Wang*, Lisha You

School of Water Conservancy and Hydropower, Hebei University of Engineering, Handan 056021, China

Received 1 June 2014, www.cmnt.lv

Abstract

An analytical model for the vibrations of operating conditions of small-scale hydroelectric generating unit is developed based on frequency domain and time domain. Firstly the vibration of unit 2 in Xida hydropower station is tested by using intelligent data logger, where the DASP10 software is used to collect the data and analyse them; and then the data are analysed by the time domain analysis, shaft centreline orbit analysis and auto-spectrum analysis respectively. Finally some instructive conclusions on the exceedance of shaft degrees and the overweight phenomena are drawn, which may assist an overhaul to raise the operating efficiency and the power output.

Keywords: Hydroelectric Generating Unit, Vibration, Time Domain Analysis, Shaft Centreline Orbit, Auto-spectrum Analysis

1 Introduction

In recent years, a large number of small hydropower stations have sprung up in many urban and rural areas, it has become an indispensable part of residential living electricity and industrial electricity in China, but the small-scale hydroelectric generating units have a poor performance in the quality and efficiency at present due to low investment, the limitation of technical level and there is often lack of the monitoring and analysis of unit vibration, which causes many units damage [1]. At present, the analysis and research methods, which are used in turbine vibration have the following kinds: fluid numerical calculation, model test, real machine test, and system modelling identification study [2-5]. In this paper, according to the reality and other factors, the real machine test method is used to monitor and analyse the vibration of unit 2 in Xida hydropower station, meanwhile the obtained data are analysed by the time domain analysis, shaft centreline orbit analysis and auto-spectrum analysis respectively, some instructive and important conclusions on the exceedance of shaft degrees and the overweight phenomena are drawn.

Xida hydropower station located in She county of Handan was founded in the late 1980s, and it belongs to the typical small hydropower, its rated power is 500kw, its rated speed is 375r/min. In the daily operation process, the engineering technicians find that there exists an abnormal vibration and a higher oil temperature in the bearing of units after excitation.

2 Experimental test systems

2.1 MONITORING POINT ARRANGEMENT

Unit 2 in Xida hydropower station is the test object, the X, Y directions of large shaft flange and water guide bearing all decorate eddy current transducers, and the X, Y, Z directions of upper bracket, lower guide bearing, water guide bearing and the head cover respectively arrange acceleration sensors. Figure 1 is the test site arrangement.

FIGURE 1 Test site arrangement

2.2 WORKING CONDITIONS SET

The test is divided into 12 working conditions which are as follows: the boot, plus exciting ,synchronization, load increasing to 250kw, load increasing to 500kw, load increasing to 600kw, keeping 600kw, load reducing to 500kw, load reducing to 250kw, splitting, reducing excitation, shutdown.

3 The basic analysis method

3.1 TIME DOMAIN ANALYSIS METHOD

Time domain analysis method is most commonly used in vibration signal analysis, the waveform is generated by the original waveform signal, which has a strong intuitive, and can accurately show the change trend of the signal amplitudes with the time [6, 7].

* ***Corresponding author*** e-mail: 2000wangly@163.com

In vibration measurement, there are three basic parameters: displacement, velocity and acceleration, the formulas are given below respectively:

$$x = A\sin(\omega t + \phi), \tag{1}$$

$$\dot{x} = A\omega\cos(\omega t + \phi), \tag{2}$$

$$\ddot{x} = -A\omega^2 \sin(\omega t + \phi). \tag{3}$$

3.2 SHAFT CENTRELINE ORBIT ANALYSIS

Shaft centreline orbit is refers to the axis of rotation in the rotating machinery and the axis is relative to the trajectory of the bearing. In the actual operation process, the abnormal vibration of hydroelectric generating unit will respond on the locus of journal bearing. The Shaft centreline orbit diagram shows the operation situation, which is a simple visual image method, and one can see the change situation from the diagram, and is easier to determine the vibration sources, which influence the stability of unit [8, 9].

3.3 AUTO SPECTRUM ANALYSIS

Auto-spectrum analysis transforms signals from time domain to frequency domain, and each periodic harmonic component corresponding to the spectrum of the signals is obtained. With the change of factor generated by the frequency of the vibration signals, the vibration will be changed., the frequency domain analysis is carried out on the collected signal to better grasp the dynamic law of vibration. The meaning of auto-spectrum analysis is a more complex signal of waveform in a specific period, which is, transformed various harmonic amplitudes the frequency and phase information and they are divided into multiple independent harmonic components [10, 11].

The signals are transformed by fast Fourier, which is defined as follows:

$$x(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt, \tag{4}$$

where: $\omega=2\pi f$, $f(t)$ is time domain data sequence and $x(\omega)$ is the frequency spectrum function of the sequence. Its inverse transformation form is:

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(\omega)e^{j\omega t}dt, \tag{5}$$

$x(\omega)$ is a complex function:

$$x(\omega) = |x(\omega)|e^{j(\omega)}, \tag{6}$$

$|x(\omega)|$ is the amplitude function of $x(\omega)$. In practice, the amplitude-frequency drawing cannot be coming out, the frequency density of $x(\omega)$ usually can be used to make an approximate description:

$$x(\omega) = 2\pi x(f). \tag{7}$$

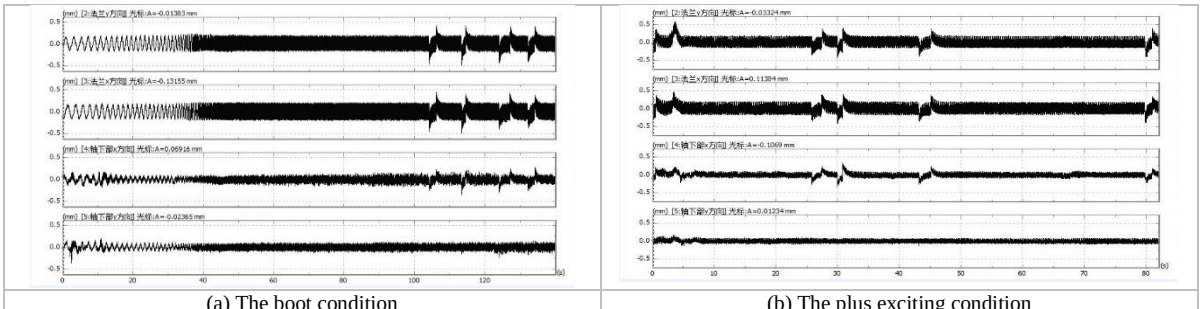
In this test, the spectrum analysis is used to judge the vibration of hydro-generator units. In analysis, FFT is used to make a frequency domain analysis for the vibration signals.

4 The analysis results

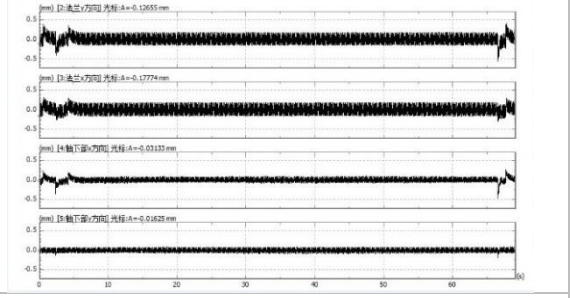
4.1 TIME DOMAIN ANALYSIS RESULTS OF SHAFT SWING

Under the working condition of the above 12, the data are analysed respectively for time domain analysis; the results are shown in Figure 2.

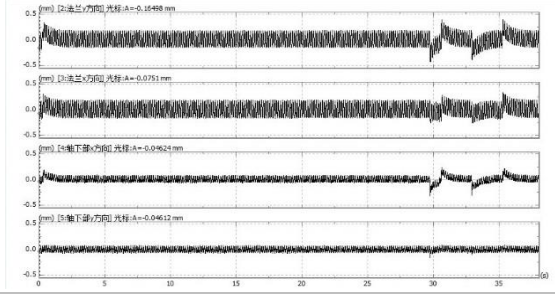
It can be seen that, from Figure 2, when the conditions are in a state of stable work, shaft swing is beyond status, the maximum superscalar is 0.465mm and 0.447mm in X and Y directions, (the allowed values is 0.250mm); under most of the working conditions, the shaft swing at water guide bearing in X and Y direction is in the acceptable range, only in the part of the time period, it is in excessive state, X direction is 0.353mm and 0.263mm, Y direction is 0.575mm and 0.362mm, (the allowed values is 0.250mm). When the working conditions are changing, the shaft swing is also in excessive state, the maximum superscalar is 1.808 mm and 0.974mm in X and Y directions under the reduction excitation condition, (the allowed values is 0.250mm); the shaft swing at water guide bearing in X direction is in excessive state, the biggest superscalar is 0.766mm under the reduction excitation condition, (the allowed values is 0.250mm), the shaft swing in Y direction is in the acceptable range only under the processes of synchronization, load increasing and keep 600kw. Under other conditions, the swing is overweight; the largest superscalar also occurs under reduction excitation condition, the value is 1.206mm, (the allowed values is 0.250mm). Based on the above analysis, we can draw a conclusion that the shaft swing is the biggest in the reduction excitation condition.



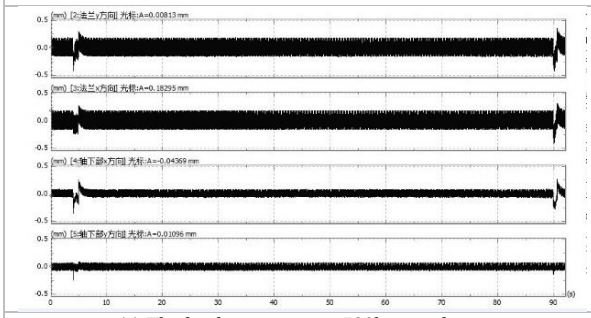
(a) The boot condition (b) The plus exciting condition



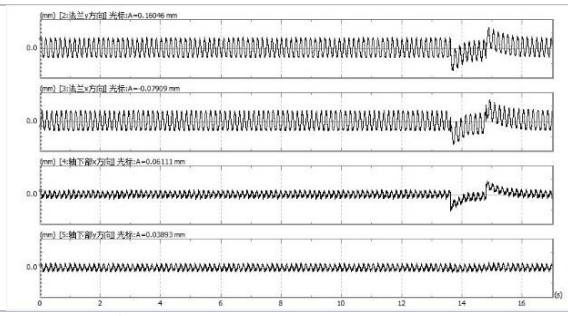
(c) The synchronization condition



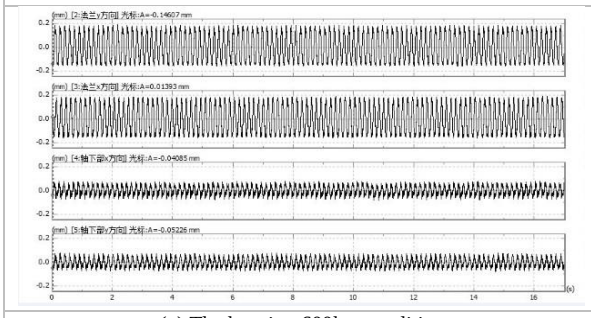
(d) The load increasing to 250kw condition



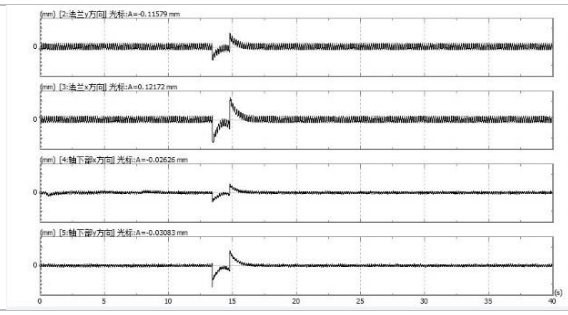
(e) The load increasing to 500kw condition



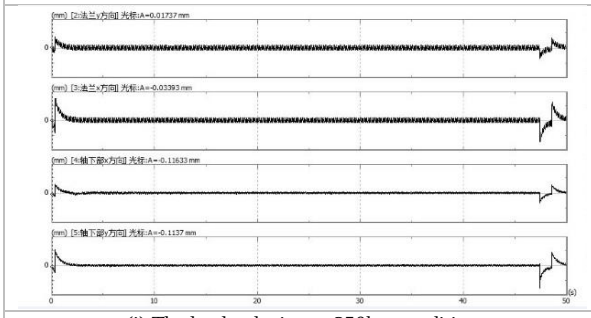
(f) The load increasing to 600kw condition



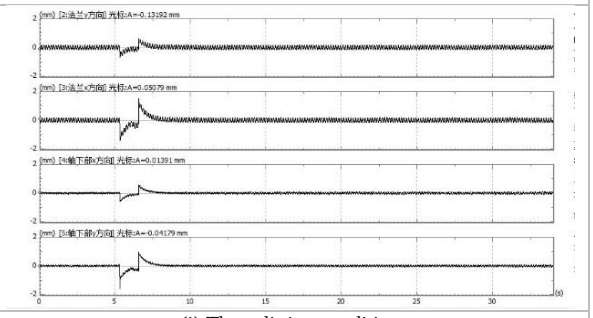
(g) The keeping 600kw condition



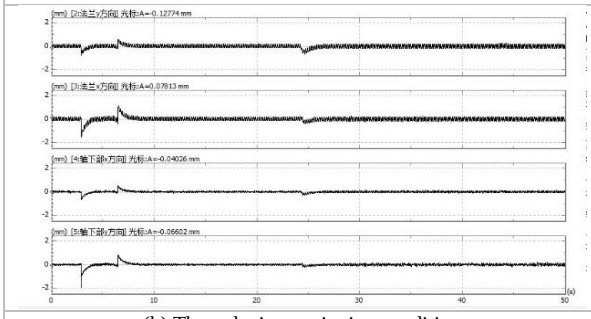
(h) The load reducing to 500kw condition



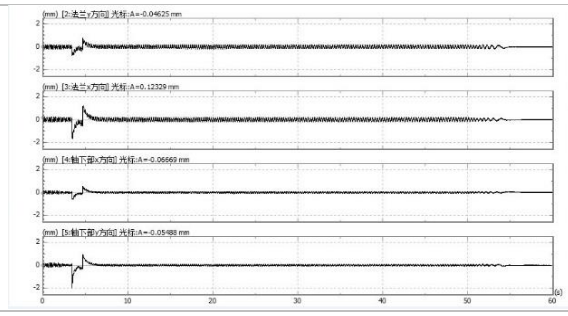
(i) The load reducing to 250kw condition



(j) The splitting condition



(k) The reducing excitation condition

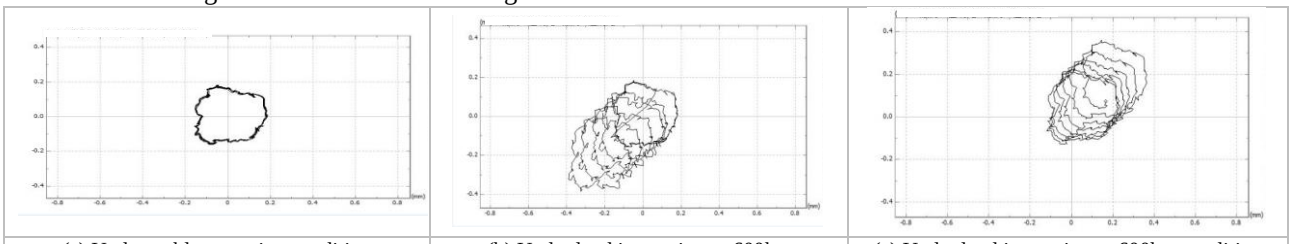


(l) The shutdown condition

FIGURE 2 The time domain of shaft swing at flange and water guide bearing under various conditions

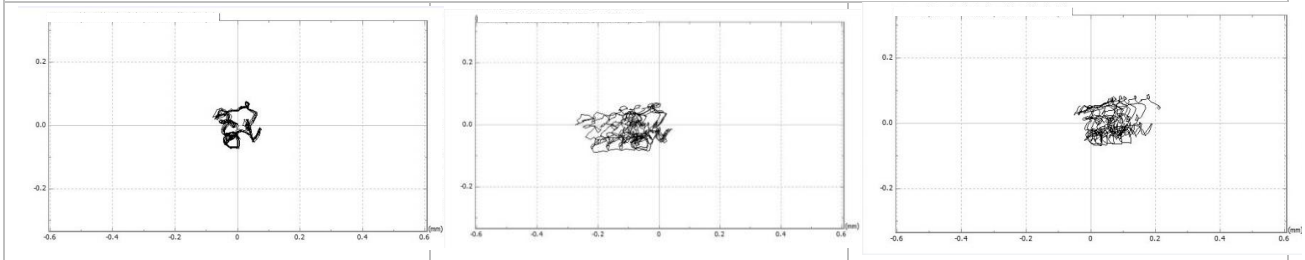
4.2 SHAFT CETRELINE ORBIT ANALYSIS RESULTS OF SHAFT CSCILLATION

The x-y graph analysis is proceeded at flange and water guide bearing under all conditions, due to space problem, we take the working condition of load increasing to 600kw as an example; the axis trajectory deviation is shown in Figure 3 and Figure 4, From the graph, we can see that when conditions change or the load is adjusted, the axis trajectory is unstable and produces deviation, the offset direction is roughly same, and the offset amplitudes are different.



(a) Under stable operation condition | (b) Under load increasing to 600kw condition | (c) Under load increasing to 600kw condition

FIGURE 3 The axis trajectory for shaft flange



(a) Under stable operation condition | (b) Under load increasing to 600kw condition | (c) Under load increasing to 600kw condition

FIGURE 4 The axis trajectory for water guide bearing

4.2 AUTO SPECTRUM ANALYSIS RESULTS OF SHAFT OSCILLATION

The auto-spectrum analysis is used in shaft oscillation; the analytical results under each condition are shown in Table 1. From the Table 1, we can find that the vibration frequencies mainly concentrate in 1HZ, 2HZ, 6HZ, 19HZ, 25HZ, and 50HZ, the turbines rotation frequency is 6.25HZ, the vibration frequency 6HZ is close to rotation frequency 6.25HZ, the frequencies 25HZ and 50HZ respectively are 4 times and 8 times of rotation frequencies, mainly from the different number of guide vanes and runner blades, the frequencies 1HZ and 2HZ are from 0.15 to 0.3 times of rotating frequencies.

TABLE 1 Auto-spectrum analysis (HZ)

Direction	X of flange	Y of flange	X of water guide bearing	Y of water guide bearing
the boot	7,1	7,1	7,1,25,20	7,1
plus exciting	6,1	6,1	6,1,25,19	6,19
synchronization	6,1	6,1	6,1,25,19	6,19
load increasing to 250kw	6	6,1	6,1,25,19	6
load increasing to 500kw	6,1	6,1	6,1,25,19	6
load increasing to 600kw	6,1	6,1	6,1,25,19	6
keeping 600 kw	6	6	6,25,19	6
load reducing to 500kw	6,1	6,1	6,1,25,19	1,6
load reducing to 250kw	6,1	6,1	1,6,19	1,6,19
splitting	1,6	6,1	1,6,19	1,6,19
reducing excitation	6,1	6,1	1,6,25	1,6
shutdown	4,1	4,1	1,4,25	1,4

4.3 THE REASON OF SHAFT SWING

Based on the above analysis, we can discover that the shaft swing is biggest change in the reduction excitation condition, it is because that the generator unbalanced magnetic pull and the bearing clearance are not uniform, moreover, the hydraulic turbine thrust bearing is positioned in the upper guide bearing and the swing of flange is always greater than that of water guide bearing, the shaft swing exists in the inflection point. In runtime, the shaft swing degrees of the main frequency is close to 6.25HZ, the rotation frequency of the turbine indicates that the axis of turbine is not straight or has some mechanical l problems such as bad entering and so on.

When the working condition changes, the shaft centreline orbit has a large range of migration, which is concerned on the electromagnetic factor and hydroelectric factor. The flange shaft orbit is elliptical and the water guide bearing shaft orbit is distorted and has more spikes, the reason is that the runner blades are impacted by water, and the different number of guide vanes and runner blades lead to the hydraulic factors such as frequency difference of water conservancy.

5 Conclusions

In all working conditions, the shaft degrees are beyond the scope of the standard state, when the conditions change, the shaft centreline orbit will appear larger range of deviation. The cause of this abnormal situation is more complex, on

the one hand, the main shaft axis is not straight has an undesirable phenomena, on the other hand, the hydro-generator exists unbalanced magnetic pull and uneven bearing clearance and so on.

In addition, there exists overweight phenomena in upper bracket, lower guide bearing, water guide bearing and head cover; when the load is 500kw, the vibration is the most serious. From the vibration frequency analysis, it also can be reflected that the causes of turbine vibration is more complex, not only the mechanical and electromagnetic aspects, but hydraulic factors are all should be considered.

Acknowledgments

Supports from Xida hydropower station are highly appreciated. The helpful comments from colleagues and classmates are also gratefully acknowledged.

This work is supported by the National Natural Science Foundation of China No.11202062, the Natural Science Foundation of Hebei Province of China No. E2010001026 and the Science and Technology Research Project of University of Hebei Province No. Y2012016.

References

[1] Jian B Q 2007 Analysis on the development of small hydropowers in China *Agricultural Engineering Technology* **3** 47-9
[2] Tao X M, Liu G N 2002 Hydraulic Stability Problem of Francis Turbine *Large Electric Machine and Hydraulic Turbine* **2** 40-9
[3] Pan L P, Gao M 2002 Analysis of Hydraulic Stability of Turbine *Journal of Chang chun Institute of Technology* **3**(4) 41-3
[4] Wang L Y, Shen T Sh, Wei D H, Liu P 2013 Experiment on vibration of aqueduct and analysis on its dynamic characteristics *Water Resources and Hydropower Engineering* **44**(11) 34-9
[5] Škifić J, Radošević A, Brajković Đ, Družeta S, Čavrak M 2013 Numerical simulations of hydraulic transients in hydropower plant Jajce II *Engineering Review* **33**(1) 51-6
[6] Wang L H 2011 Hydroelectric generating set and analysis of vibration *The Yellow River water conservancy press*: Zhengzhou
[7] Li R Y, Li Zh H 2014 Study on the effect of sensor placement on the characteristic parameters of the turbine vibration *Electronic Design Engineering* **22**(5) 161-4
[8] Liang X 2014 An analysis of fault diagnosis of hydraulic turbine-generator units *China Rural Water and Hydropower* **1** 165-8
[9] Hu X 2014 Research on fault diagnosis of turbine based on modal analysis *Water conservancy science and technology and economy* **20**(1) 106-8
[10] Han Y, Song Y H 2003 Condition monitoring techniques for electrical equipment a literature survey *IEEE Trans on Power Delivery* **18**(1) 10-4
[11] Hagenauer Z D 2004 Genomic analysis using methods from information theory *Information Theory Workshop, IEEE* 55–9

Authors

Liying Wang, born on January 6, 1978, in Shijiazhuang, Hebei Province, China

University studies: M.S. degree in Institute of Mechanism, Shijiazhuang Tiedao University, in 2003; she received the Ph.D. degree in vehicle engineering from Beijng Jiaotong University, China in 2014.
Scientific interests: intelligent computing, intelligent control theory, and control systems engineering.

Lisha You, born on October 14, 1989, China

Current position, grades: a graduate student at Hebei University of Engineering, China
University studies: Bachelor degee in thermal energy and power engineering from Hebei University of Engineering, China in 2012.
Scientific interests: hydraulic turbine and mechanical vibration.

The application of fusion structure in the coal mine safety state evaluation

Shisong Zhu[1, 2]*, Haiyan Zhang[1], Cuiyun Zhang[3], Changqing Li[1]

[1]*School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, 454000, P.R. China*

[2]*Key Laboratory of Mine Informatization, Henan Polytechnic University, Jiaozuo, 454000, P.R. China*

[3]*Henan College of Industry & Information Technology, Jiaozuo, 454000, P.R. China*

Received 1 October 2014, www.cmnt.lv

Abstract

Coal mine safety is a very important link in the current coal mine production. Currently, there are a lot of kinds of methods for the evaluation of coal mine safety environment in order to solve the evaluation of the coal mine safety status. However, the complexity and diversity of coal mine environment cause that there is no method can evaluate coal mine safety status. So a new method is proposed, which called two-stage fusion structure according to China's natural environment of the underground coal mine, is used to evaluate the condition of coal mine safety. Firstly, it gets various parameters for affecting the safety of underground using the first level evaluation method; secondly, the second level fusion method is used to realize the coal mine security situation assessment. The experimental results show that this method has better performance more than other methods in evaluating coal mine safety.

Keywords: the first level fusion algorithm, fusion decision, mine safety, better performance

1 Introduction

In recent years, Coal mine disaster accidents frequently happen, which makes people's life lose. It is very important to evaluate the security situation in time, correctly determine the mine safety status to avoid or reduce the coal safety accident [1].

Currently, there are some coal mine safety evaluation methods: expert system, data envelopment analysis method, neural network and support vector machine [2]. Expert evaluation system is mainly based on expert experience and knowledge of coal mine safety, the evaluation results has much to do with expert level, so the evaluation results is difficult to reflect subjectively the degree of coal mine safety; data envelopment analysis is not yet mature, it is difficult to reflect the mapping relation between evaluation index and evaluation results, and at the same time it has certain limitations, the evaluation accuracy is lower. Neural network is a kind of "black box" evaluation method, which needs a lot of sample data for coal mine safety evaluation, but the evaluation results is poor if using the pure quantitative method to evaluate coal mine safety, the same as artificial intelligence [3].

Production of coal mine is a special industry which contains personnel management, machinery operation, mining technology, geological conditions and natural environment, and all the factors have differences [4, 5]. In the coal mine safety evaluation process, firstly, needs to establish a scientific and perfect coal mine safety evaluation, then calculates the index weight, and finally gets the safety level.

Therefore, in order to improve the coal mine safe evaluation accuracy and reliable, the article proposes two-level fusion structure, which is based on improved real-time adaptive weighted fusion algorithm and fuzzy information fusion theory. The method adopts the improved real-time adaptive weighted algorithm to achieve a spatial real-time data fusion which was distributed in different geographical location firstly, and then used Gaussian functions and fuzzy transform information fusion theory to get coal mine safety evaluation. The experimental results show that the method is feasible and effective for multi-sensor mine safety evaluation system.

2 The two-level fusion algorithm structure

Gas, carbon monoxide, oxygen, temperature, wind speed in the air are the fatal factors which affect the safety status under the shaft. This paper puts forward a sensor group which contains gas, carbon monoxide, oxygen, temperature, and wind speed sensors to monitor different positions of a mining area and then using the improved real-time adaptive weighted fusion algorithm to deal with monitoring parameter values to realize the first level data fusion in the spatial domain, and at last adopting the fuzzy information fusion theory to achieve the secondary decision of system. The model of the method is shown in Figure 1.

***Corresponding author** e-mail: zss@hpu.edu.cn

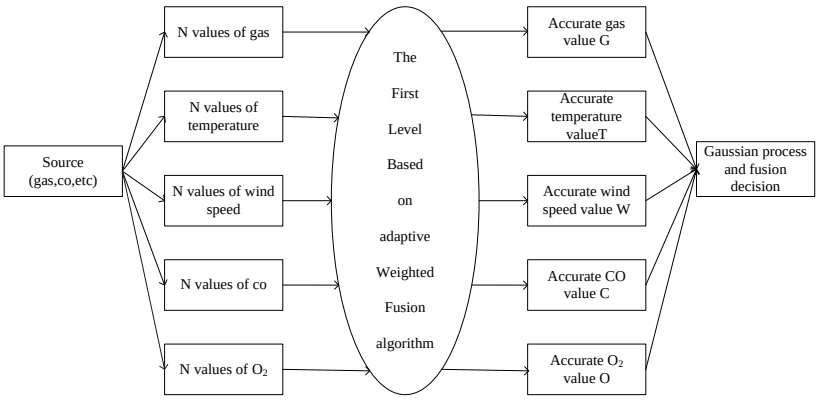


FIGURE 1 The model of two-level fusion structure

3 The first level fuzzy algorithm

The parameters in a mining area are monitored, such as gas, co, and wind speed and so on, and the sensors are distributed in a large scale. To understand the status of the mine safety comprehensively and accurately, multi-sensor groups must be used to complete the task about monitoring data. The parameter, which is monitored by sensor group is more accurate than single sensor. To processing underlying data of system, this paper proposes the real-time adaptive weighted fusion algorithm, which does not need prior knowledge and the programming is easily to realize to get the first level fusion in spatial domain.

3.1 THE FIRST LEVEL FUZZY ALGORITHM

Adaptive weighted fusion algorithm is a common method to fuse the underlying data for multi-sensor. The use of this method is not restricted even in the case of weak or inadequate information [6]. Taking the data from gas sensor as an example, the description of algorithm is as follows:

1) In the coal mine safety evaluation system which contains multi-sensor group, the weight of gas sensor is set as $W_{i1}(\mathrm{i}=1,2,\cdots,n.)$, variance is $\sigma_{i1}^2(\mathrm{i}=1,2,\cdots,n.)$ and W_{i1} is needed to be fulfilled: $\sum_{i=1}^{n} W_{i1}=1$.

2) The gas estimation true value was X, a gas sensor measured value was X_i for a moment, the optimal fusion value $\hat{X}(\mathrm{x})$, which meets the optimal condition that the mean square error was minimum, and the last $\hat{X}(\mathrm{x})$ needs to meet Equation (1):

$$\hat{X}(\mathrm{x})=\sum_{i=1}^{n} W_{i1}X_i \tag{1}$$

The total mean square error of $\hat{X}(\mathrm{x})$ [7]:

$$\sigma^2=E[X-\hat{X}(\mathrm{x})]^2=E[\sum_{i=1}^{n} W_{i1}^2(X-X_i)^2]+ 2\sum_{i=1,j=1,i\neq j}^{n} W_{i1}W_{j1}(X-X_i)(X-X_j)] \tag{2}$$

Because X_1, X_2,…, X_n were independent each other $E[(X-X_i)(X-X_j)]=0\ (i\neq j,i,j=1,2,\ldots\mathrm{n})$:

$$\sigma^2=E[\sum_{i=1}^{n} W_{i1}^2(X-X_i)^2]=\sum_{i=1}^{n} W_{i1}^2\sigma_{i1}^2. \tag{3}$$

According to the multivariate function extreme value theory, we can calculate the corresponding weighted factor when the total mean square error is minimized [8]:

$$W_{i1}^*=1/\sigma_{i1}^2\sum_{i=1}^{n}\frac{1}{\sigma_{j1}^2},\quad i=1,2,\ldots\mathrm{n}. \tag{4}$$

The minimum mean square error is [9]:

$$\sigma_{\min}^2=1/\sum_{i=1}^{n}\frac{1}{\sigma_{i1}^2}. \tag{5}$$

The above estimation is based on the real-time values of the gas sensors in some time, and if the estimation is constant, we can estimate according to average historical data [10]. Estimation and the total mean square error at this time are as follows respectively:

$$\overline{X}=\sum_{i=1}^{n} W_{i1}\overline{X_i}(\mathrm{K}). \tag{6}$$

$\overline{X}$: The gas optimal fusion for the first level fusion.
K: The frequency for historical data measurement.

$$\overline{\sigma^2}=\frac{1}{K}W_{i1}^2\sigma_{i1}^2=\frac{\sigma_{\min}^2}{K} \tag{7}$$

$\overline{\sigma^2}$: The optimum fusion of the minimum variance.

3.2 THE RESULT OF FIRST LEVEL FUSION

The mine gas concentration which comes from a coal mine safety production monitoring and control system is denoted by X_i (%) and its variance is σ_{i1}^2 and weights W_{i1}. The properties of each sensor, as X_i, is displayed in Table 1.

TABLE 1 The properties of the gas sensor σ_{i1}^2

G_i	G_1	G_2	G_3	G_4	G_5	G_6
X_i	0.54	0.55	0.45	0.22	0.57	0.44
σ_{i1}^2	0.53	0.48	0.54	0.53	0.39	0.44
W_{i1}	0.151	0.166	0.147	0.152	0.203	0.181

By Equation (6), we could calculate the optimal fusion value of gas concentration 0.4678.

The system could obtain the optimal fusion about temperature, wind speed, etc. by using the above method.

The improved adaptive weighted fusion method is implemented in the first level of data fusion to get relatively accurate parameter information. As the parameters are independent and it is difficult to assess the condition about mine. So the system is required to use fuzzy information fusion theory to realize the secondary fusion processing.

4 The second level fuzzy algorithm

A precise method to describe the association between parameters is very difficult because underlying parameters from the first level fusion are independent. So the application of fuzzy information fusion theory is helpful to make the coal mine security situation evaluation more accurate.

4.1 THE THEORY OF SECOND LEVEL FUSION

Making the absolute subordinate relations of general set become flexibly and the elements to membership of collection can take any value in [0,1] are the basic idea of fuzzy set. The idea is suitable for describing and processing uncertainty sensor data [10, 11]. The theory, which is based on Gaussian membership function and fuzzy transformation can make decisions in the secondary fusion processing of evaluation system.

(U,V,R) is the fuzzy comprehensive evaluation model [13], U is comment set, and V is factors set. n kinds of coal mine safety status and the status collection: $\{U_1,U_2\cdots U_n\}$; the first level fusion feature parameters constitute a collection $\{V_1,V_2,\cdots V_m\}$. R_{ij} is the possibility of status set from characteristic parameter. It is the degree of membership to U_i and V_j, at last it constitutes a fuzzy relation matrix R. The factors weight vector: $A=(A_1,A_2,\cdots A_m)$ is derived from the different influence degree of characteristic parameters for judging result.

The fuzzy transformation is as follows: $B=A\circ R=(B_1,B_2,\ldots,B_n)$,

$$B=(A_1,A_2,\ldots,A_m)\circ\begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ R_{m1} & R_{m2} & \cdots & R_{mn} \end{pmatrix}, \quad (8)$$

B : The possibility of coal mine safety status,
A : The factor weight vector,
R : The fuzzy matrix,
B_j : The degree of membership of fuzzy subsets about the comprehensive evaluation of down-hole safety status.

4.2 THE RESULT OF SECOND LEVEL FUSION

The gas concentration, temperature, wind speed, co concentration and oxygen concentration are the main factors, which affect the coal mine safety; the first level fusion structure achieves data fusion from different sample point to the same parameters in spatial domain. For secondary decision-making fusion, firstly, the first level fusion parameters are given different weights according to the experimental data and then get weight vector $A^*(A_1^*,A_2^*,A_3^*,A_4^*,A_5^*)$, it is the first step to realize fuzzy information fusion algorithm. Next, it is time to make sure the degree of membership of each coal mine status from the environment parameters.

For continuous analogy parameter variable, Cauchy fuzzy sets and Gaussian pattern matching algorithm are two different mathematical methods to seek membership [12]. Because constructing the process of Cauchy fuzzy sets is more complex and Gaussian membership function is chosen, the basic idea is as follows: $\{N_1,N_2,\ldots,N_n\}$ makes up of n kinds of status in status database, where $N_i(i=1,2\cdots,n)$ represents different status and each status corresponding to a standard joint feature vector containing k characteristic parameters: $\overline{N_i}=\{N_{i1},N_{i2},\cdots,N_{iK}\}$, $N_{ij}(j=1,2,\cdots,K)$ represents the j^{th} standard characteristic parameter of the i^{th} kind state. All kinds of parameter respectively can become a vector containing K characteristic parameters after the first level fusion in a period of time; $\overline{X_i}=(X_{i1},X_{i2},\cdots,X_{iK})$, where $X_{ij}(j=1,2,\cdots,K)$ represents the j^{th} characteristic parameter of the i^{th} kind state.

For the i^{th} kind condition:

Cauchy fuzzy sets and Gaussian pattern matching algorithm are two different mathematical methods to seek membership [12]. Because constructing the process of Cauchy fuzzy sets is more complex and Gaussian membership function is chosen, the basic idea is as follows: $\{N_1,N_2,\ldots,N_n\}$ makes up of n kinds of status in status database, where N_i $(i=1,2\cdots,n)$ represents different status and each status corresponding to a standard joint feature vector containing k parameters: $\overline{N_i}=\{N_{i1},N_{i2},\cdots,N_{iK}\}$, $N_{ij}(j=1,2,\cdots,K)$ represents the j^{th} standard characteristic parameter of the i^{th} kind state. All kinds of parameter respectively can become a vector containing K characteristic parameters after the first level fusion in a period of time; $\overline{X_i}=(X_{i1},X_{i2},\cdots,X_{iK})$, where $X_{ij}(j=1,2,\cdots,K)$ represents the j^{th} characteristic parameter of the i^{th} kind state.

For the i^{th} kind condition:

1) Membership functions of the parameters:

$$\mu_{N_{ij}}(X_{ij}) = \exp\left(-\frac{(X_{ij}-N_{ij})^2}{2\sigma_{ij}^2}\right). \quad (9)$$

X_{ij} : The j^{th} characteristic parameter of the i^{th} kind state.

N_{ij} : It means the j^{th} standard characteristic parameter of the i^{th} kind state.

σ_{ij}^2 : It means the j^{th} standard deviation of the i^{th} kind state.

2) Vector membership function:

$$\mu_{\overline{N_j}}(\overline{X_i}) = \sum_{t=1}^{K} \alpha_t \times \mu_{N_{ij}}(X_{ij}). \quad (10)$$

$$\sum_{t=1}^{K} \alpha_t = 1,\ \alpha_t > 0,\ t = 1,2,\cdots,K. \quad (11)$$

α_t : It represents the weight coefficient of the t^{th} characteristic parameter.

3) The result of fuzzy transform:

$$B = A^* \circ \mu_{\overline{N_j}}(\overline{X_i}) = (A_1^*, A_2^*, A_3^*, A_4^*, A_5^*) \circ \begin{pmatrix} \mu_{\overline{N_1}}(\overline{X_1}) & \mu_{\overline{N_1}}(\overline{X_2}) & \mu_{\overline{N_i}}(\overline{X_3}) \\ \mu_{\overline{N_2}}(\overline{X_1}) & \mu_{\overline{N_2}}(\overline{X_2}) & \mu_{\overline{N_2}}(\overline{X_3}) \\ \mu_{\overline{N_3}}(\overline{X_1}) & \mu_{\overline{N_3}}(\overline{X_1}) & \mu_{\overline{N_3}}(\overline{X_3}) \\ \mu_{\overline{N_4}}(\overline{X_1}) & \mu_{\overline{N_4}}(\overline{X_2}) & \mu_{\overline{N_4}}(\overline{X_3}) \\ \mu_{\overline{N_5}}(\overline{X_1}) & \mu_{\overline{N_5}}(\overline{X_1}) & \mu_{\overline{N_5}}(\overline{X_3}) \end{pmatrix}. \quad (12)$$

5 Results and analysis

The elements in B represent respectively three kinds of status: safety, generally safety and dangerous. The influence degree of each sensor group in system was different, this paper set the gas (G), temperature (T), wind speed (V) and carbon monoxide (C) and oxygen (O) as the weight vector $A = (0.21, 0.13, 0.37, 0.17, 0.12)$ through parameters test and data analysis for many times in one coal mining.

TABLE 2 Fuzzy matrix

Evaluation factors	Safety state	General safety	Dangerous state
G	0.520	0.480	0
T	0.240	0.740	0.020
V	0.123	0.637	0.240
C	0.162	0.420	0.418
O	0.105	0.213	0.682

For a certain moment, the system transmitted the five parameters from the first level fusion to the centre of decision fusion, then the fusion centre could get the fuzzy matrix according to Gaussian membership function and at last the system compounded and normalized processing the weight vector of characteristic parameters and fuzzy matrix to realize safety status evaluation coal mine.

We could get the fuzzy matrix from Equation (9), (10) and (11), fuzzy matrix in Table 2:

We could get the result from Equation (12): $(B_1, B_2, B_3) =$

$$(0.21, 0.13, 0.37, 0.17, 0.12) \circ \begin{pmatrix} 0.520 & 0.480 & 0 \\ 0.240 & 0.740 & 0.020 \\ 0.123 & 0.637 & 0.240 \\ 0.162 & 0.420 & 0.418 \\ 0.105 & 0.213 & 0.682 \end{pmatrix}.$$

$$= (0.274, 0.440, 0.256)$$

The fusion degree of membership of B_2 is the largest; we could determine "general safety" as coal mine status at the moment according to maximum membership judgment rules.

Every sensor group degree of membership value is presented in Table 3. We got down-hole safety determination according to decision rules and at same time obtained comprehensive judgments on the basis of fuzzy subset.

TABLE 3 The contrast between single sensor recognition and fusion recognition

Mine state	Factors	Safety	General safety	Dangerous	Results
Safety	G	1.000	0	0	Safety
	T	0.820	0.180	0	Safety
	V	0.505	0.327	0.168	Safety
	C	0.328	0.473	0.199	General safety
	O	0.256	0.323	0.421	Dangerous
	Fusion	0.422	0.373	0.168	Safety
General safety	G	0.520	0.480	0	Not sure
	T	0.240	0.740	0.020	General safety
	W	0.123	0.637	0.240	General safety
	C	0.162	0.420	0.418	Not sure
	O	0.105	0.213	0.682	Dangerous
	Fusion	0.274	0.440	0.256	General safety
Dangerous	G	0	0.050	0.950	Dangerous
	T	0	0.180	0.820	Dangerous
	W	0.327	0.311	0.362	Dangerous
	C	0.103	0.324	0.573	Dangerous
	O	0.231	0.572	0.197	General safety
	Fusion	0.265	0.345	0.390	Dangerous

Through experimental contrast, it can be seen that one kind sensor for the evaluation of mine safety status is uncertain, but the application of fuzzy fusion increases credibility of state evaluation, enhances and improves the performance of the safety evaluation system.

6 Conclusions

The two-level fusion structure is proposed in this paper, which based on data fusion technology in spatial domain and information fusion algorithm. With the two-stage fusion structure implemented in coal mine safety evaluation system, the contrast experiment shows that the improved adaptive weighted fusion algorithm disposes the data on the space is more reliable, and at same time, fuzzy fusion algorithm in the secondary structure can improve the credibility of the whole evaluation system. Therefore, the practical application of structure in coal mine safety evaluation is feasible. Besides, the time complexity algorithm needs to be part of our future.

Acknowledgment

This work was supported by the Opening Project of Key Laboratory of Mine Informatization, Henan Polytechnic University (KZ2012-02).

References

[1] Chen Z, Yang S, Wu C 2007 Based on entropy and unascertained measure model and the application of coal mine safety evaluation *Journal of mining safety and environmental protection* **34**(1) 757 (*in Chinese*)
[2] Dong C, Cao Z 2006 The coal enterprises evaluation method based on rough set and grey correlation *Science and Technology of Heilongjiang Journal* **16**(6) 400-3 (*in Chinese*)
[3] Zhang T 2012 Study and simulation on coal mine safety assessment *Computer Simulation* **29**(3) 213-7 (*in Chinese*)
[4] Yang C, Zhang M 2005 Fuzzy comprehensive evaluation of mine safety *Journal of coal technology* **4**(1) 37-9 (*in Chinese*)
[5] Qin T, Shang Y 2010 The coal mine safety evaluation index research based on the analytic hierarchy *Journal of modern mining* (5) 70-2
[6] Li Z, Chen R, Zhang B 2006 Study of adaptive weighted estimate algorithm of congeneric multi-sensor data fusion *Journal of Lanzhou University of Technology* **32**(4) 78-82 (*in Chinese*)
[7] Zhai Y, Dai Y 1998 The research of multi-sensor data adaptive weighted fusion estimated algorithm *Journal of measurement* **19**(1) 69-75 (*in Chinese*)
[8] Samarasooriya V N S, Varshney P K.A 2000 Fuzzy modeling approach to decision fusion under uncertainty *Fuzzy Sets and Systems* 59-69
[9] Chair Z, Varshney P K 1986 *IEEE Trans AES* **22**(1) 98-101
[10] Fu H, Gao T, Liu Y 2008 Application of multi-sensor fuzzy information fusion in mine safety *Transducer and Microsystem Technologies* **27**(5) 114-20 (*in Chinese*)
[11] Wang X, Sheng Q 2008 The research of product evaluation model based on fuzzy transform *Modern Shopping Mall* 29-30 (*in Chinese*)
[12] Niu P, Wang S, Ma J 2007 Radar target recognition based on Subordinate Function and D-S Theory *Micro computer information* **23**(11) 218-20 (*in Chinese*)

Authors

Shisong Zhu, born in 1965, Jiao Zuo, China

Current position, grades: an associate professor in Henan Polytechnic University, Jiaozuo, Henan, China.
University studies: M.S. and PhD degree from the Graduate School of Science and Engineering, Yamagata University, Japan, in 2006 and 2009.
Scientific interest: digital image processing and mine informatization.

Haiyan Zhang, born in 1988, Xin Cai, China

Current position, grades: master student in Henan Polytechnic University, China.
University studies: B.S. degree from the Department of Computer Science and Technology, Henan Polytechnic University, China, in 2012.
Scientific interest: graphics.

Cuiyun Zhang, born in 1959, Ji Yuan, China

Current position, grades: Henan College of Industry & Information Technology, Jiaozuo, China.
University studies: M.S. degree from the School of Electrical Engineering and Automation, Henan Polytechnic University, China, in 2008.
Scientific interest: electrical automation and mine informatization.

Changqing Li, born in 1956, Hui Xian, China

Current position, grades: professor in Henan Polytechnic University, Jiaozuo, Henan, China.
Scientific interest: computer application in teaching and research work.

Seismic signals wavelet packet de-noising method based on improved threshold function and adaptive threshold

Liu Shuchong*, Chen Xun

Institute of Disaster Prevention, Sanhe 065201

Received 1 March 2014, www.cmnt.lv

Abstract

Wavelet analysis is one of the effective method to improve the signal to noise ratio and resolution of seismic data, a wavelet packet seismic signal denoising method based on a new threshold function and adaptive threshold was put forward according to the distortion problem of traditional threshold function denoising method, which make up the defects of traditional thresholding method. Wavelet packet decomposition techniques was used for seismic wave signal denoising processing, and the synthetic seismic signals and actual seismic data was done wavelet packet decomposition processing through MATLAB, better removing high frequency random noise to retain the useful signals. Experimental results showed that the method can effectively remove noise and improve seismic resolution, with better denoising effect.

Keywords: seismic signal, adaptive threshold, denoising, resolution

1 Introduction

In seismic exploration, the signals-noise-ratio of seismic data will directly affect the reliability of seismic data, the accuracy of parameter extraction and the resolution enhancement effect etc. Therefore, denoising research plays a very important role in seismic data processing. In the high-resolution seismic exploration, seismic records have very wide frequency band, in the frequency range of significant wave may contain surface wave, random noise interference noise, which makes effective separation of signal to noise more important. Currently in seismic data processing the de-noising method are more, but each de-noising method has its own applicable conditions. Wavelet transform technique as a new approach has been widely applied to the signal denoising because of its time-frequency analysis and multi-resolution and other characteristics will have a good research value and application prospect when applied to the seismic signal noise elimination treatment. Wavelet packet transform [3-4] is the promotion of the wavelet transform, is a more detailed analysis and reconstruction methods. It divides the band by multiple-levels, further divide the high frequency portion which wavelet analysis cannot be subdivided and can adaptively select the appropriate frequency band so as to match the signal spectrum according to the characteristics of the analysed signals, thereby improving time-frequency resolution.

A wavelet packet seismic signal denoising method based on improving threshold function and adaptive threshold was presented for a more sophisticated decomposition to the low-frequency and high-frequency of seismic signals, wavelet packet coefficients derived were processed by the improved refinement threshold function and self-adaptive thresholding, which could effectively remove the noises, reduce the useful signals removal. Synthetic seismogram denoising and real seismic signal de-noising was achieved by simulation. Studies show that the improved threshold function and adaptive threshold wavelet packet seismic signal denoising method is very effective in denoising, have got better denoising effect.

2 Wavelet packet transform

2.1 WAVELET PACKET DEFINITION

In the multi-resolution analysis, $L^2(R) = \bigoplus_{j\in Z} W_j$ show multi-resolution analysis is that the Hilbert space $L^2(R)$ is decomposed to subspaces $W_j (j \in Z)$ orthogonal according to a different scale factor j. In which, W_j is the closures for the wavelet function $\varphi(t)$ (wavelet subspace). The wavelet subspace W_j was done frequency subdivision as binary fractions, in order to achieve the purpose of improving the frequency resolution. The scale subspace V_j and wavelet subspace W_j characterize uniformly with a new subspace U_j^n, and order

$$\begin{cases} U_j^0 = V_j \\ U_j^1 = W_j \end{cases}, (j \in Z) . \tag{1}$$

The Hilbert space orthogonal decomposition $V_{j+1} = V_j \oplus W_j$ could unify using U_j^n decomposition:

**Corresponding author* e-mail: luckyorng1009@gmail.com

$$U_{j+1}^0 = U_j^0 \oplus U_j^1, j \in Z . \quad (2)$$

Define subspace U_j^n is the closure space of function $u_n(t)$, U_j^{2n} is the closure space of $u_{2n}(t)$ and order $u_n(t)$ satisfies the following two-scale equation:

$$\begin{cases} u_{2n}(t) = \sqrt{2}\sum_{k\in Z} h(k)u_n(2t-k) \\ u_{2n+1}(t) = \sqrt{2}\sum_{k\in Z} g(k)u_n(2t-k) \end{cases}, \quad (3)$$

where $g(k) = (-1)^k h(1-k)$, that is two coefficients have orthogonal relationships. When $n = 0$, the above equation becomes

$$\begin{cases} u_0(t) = \sum_{k\in Z} h_k u_0(2t-k) \\ u_1(t) = \sum_{k\in Z} g_k u_0(2t-k) \end{cases}, \quad (4)$$

$u_0(t)$ and $u_1(t)$ are respectively scale function $\phi(t)$ and wavelet function $\varphi(t)$, Equation (4) is the equivalent representation of Equation (2). This equivalent representation were extended to $n \in Z_+$ (non-negative integer), that is equivalent representation of Equation (3)

$$U_{j+1}^n = U_j^{2n} \oplus U_j^{2n+1} \qquad j \in Z, n \in Z_+ . \quad (5)$$

This allow the wavelet subspace further subdivided as binary:

$$\begin{aligned} &W_j = U_j^1 = U_{j-1}^2 \oplus U_{j-1}^3 \\ &U_{j-1}^2 = U_{j-2}^4 \oplus U_{j-2}^5, U_{j-2}^5 = U_{j-2}^6 \oplus U_{j-2}^7 \\ &... \\ &W_j = U_{j-k}^{2^k} \oplus U_{j-k}^{2^k+1} \oplus ... \oplus U_{j-k}^{2^k+m} \oplus ... \oplus U_{j-k}^{2^{k+1}-1} \end{aligned}, \quad (6)$$

in which orthonormal basis $\left\{2^{(j-k)/2}\varphi_{2^k+m}(2^{j-k}t-l); l \in z\right\}$ are called wavelet packet corresponding to $k = 1,2,...j; j = 1,2...$ and subspaces $U_{j-k}^{2^k+m}$.

Assume $g_j^n(t) \in U_j^n$, then $g_j^n(t)$ can be expressed as $g_j^n(t) = \sum_l d_l^{j,n} u_n(2^j t - l)$. Wavelet packet decomposition are that $\left\{d_l^{j,2n}\right\}$ and $\left\{d_l^{j,2n+1}\right\}$ are obtained by $\left\{d_l^{j+1,n}\right\}$, that is

$$\begin{aligned} d_l^{j,2n} &= \sum_k a_{k-2l} d_k^{j+1,n} \\ d_l^{j,2n+1} &= \sum_k b_{k-2l} d_k^{j+1,n} \end{aligned}. \quad (7)$$

Wavelet packet reconstruction are that $\left\{d_l^{j+1,n}\right\}$ are obtain by $\left\{d_l^{j,2n}\right\}$ and $\left\{d_l^{j,2n+1}\right\}$, that is

$$d_l^{j+1,n} = \sum_k [h_{l-2k} d_k^{j,2n} + g_{l-2k} d_k^{j,2n+1}] . \quad (8)$$

2.2 DENOISING

In seismic exploration, the general wave interference include: surface waves, high frequency random noise, side wave, multiples, etc, in which the high-frequency surface wave and random interference waves have more serious impact on the effective wave, therefore it need to reduce noise processing to extract the useful signals. For non-stationary seismic signals, the wavelet packet decomposition technique is an effective noise reduction method.

In the wavelet analysis, the signal is decomposed into low frequency portion and high frequency detail rough. However, only the low-frequency portion is used for the second layer decompose while the high-frequency part without treatment, when the scale increases frequency resolution of wavelet analysis became lower and the resolution in high frequency became poor.

Wavelet packet analysis is further promotion of the wavelet analysis, which provides a more complex and flexible analysis tools. Wavelet packet analysis subdivide the low-frequency part and high-frequency portion of last layer while the high frequency part of the signal can be described more detailed, with a more precise local analysis capabilities. Typically, wavelet packet noise reduction steps [5, 6].

The first was wavelet packet decomposition of the signal: Choose a wavelet and determine a wavelet decomposition level *N*, then do *N*-layer wavelet packet decomposition on the seismic waves signal *S*. Described with a three-tier decomposition, the wavelet packet decomposition tree were shown in Figure 1. In Figure 1 *A* represented a low frequency, *D* represented high frequency, the end serial number represented wavelet packet decomposition layers, namely the scale number. The original signal *S* is equivalent to:

$$\begin{aligned} S = {} & AAA3 + DAA3 + ADA3 + DDA3 + \\ & AAD3 + DAD3 + ADD3 + DDD3 \end{aligned}. \quad (9)$$

The second was to calculate the optimal tree: compute the best tree for a given entropy criteria. Commonly used entropy criteria were Shannon threshold norm log energy sure and user and so on.

The third was thresholding quantization of wavelet packet decomposition coefficiences. Select a threshold and do coefficients thresholding quantization for each wavelet packet decomposition coefficients. Wavelet transform coefficient values are compared with a threshold, it is believed that the values smaller than the threshold value were generated by the noise and set to zero, the values greater than the threshold values were corresponding to the signal mutation point and retained in order to achieve the purpose of denoising. In the process of denoising there were usually three treatments: force denoising the default threshold denoising given soft (or hard) threshold denoising.

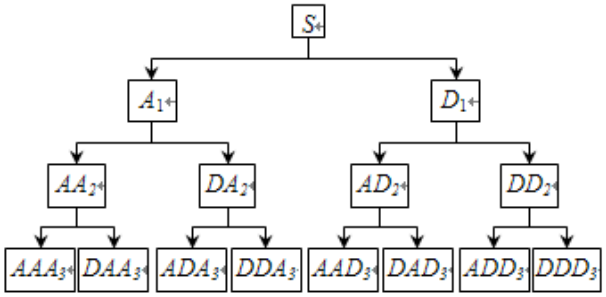


FIGURE 1 Wavelet packet decomposition technique

Soft (or hard) threshold denoising [7, 8] are most commonly used, but there is distortion.

The fourth was signal reconstruction: Do wavelet packet reconstruction on signals according to the L layer wavelet packet decomposition low-frequency coefficients of the original signal and high frequency coefficients after threshold quantization processing.

3 Selection of wavelet and threshold functions

3.1 DETERMINATION OF THE BEST WAVELET PACKET BASIS (CALCULATION OF OPTIMAL TREE)

A signal with length of $L = 2^N$ could have 2^N kinds of different signal decomposition method, while the number of full two forks trees with depth of N are 2^N. This number is too large to enumerate every situation, and the minimum entropy criterion can be obtained by an optimal signal decomposition method.

There are several types of traditional standards based on entropy: Shannon entropy, threshold entropy, norm entropy, logenergy entropy, sure entropy and user entropy and so on. Shannon entropy is defined as follows:

$$El(s_i) = -s_i^2 \log(s_i^2) \quad 0 \log 0 = 0 . \tag{10}$$

The signals were decomposed layer by layer, the entropy of each decomposition node was calculated, entropy values of a node and its child nodes were compared, the base obtaining minimum entropy is the optimal wavelet packet basis.

3.2 SELECTION OF THRESHOLD FUNCTION

Wavelet thresholding denoising was threshold processing of high frequency coefficient according to the characteristics of noise manifests high-frequency signals, if the coefficients less than the threshold it was considered by the noise and set zero, if the coefficients greater than the threshold value it correspond to the useful signals and should be retained, thereby to achieve the purpose of de-noising.

The choice of traditional threshold is the hard threshold processing and soft threshold processing. Soft threshold processing is comparing the absolute value of signals and the threshold, when the absolute value of data is less than or equal to the threshold, set it to zero, when greater than the threshold, it becomes the difference between the point with a threshold value. The hard threshold is comparing the absolute value of signals and the threshold, if less than or equal to the threshold the point was set zero, greater than the threshold value it remains unchanged.

Hard thresholding:

$$\hat{W}_{j,k} = \begin{cases} W_{j,k}, |W_{j,k}| \geq \lambda \\ 0, |W_{j,k}| < \lambda \end{cases} \tag{11}$$

Soft thresholding:

$$\hat{W}_{j,k} = \begin{cases} \text{sgn}(W_{j,k})(|W_{j,k}| - \lambda), |W_{j,k}| \geq \lambda \\ 0, |W_{j,k}| < \lambda \end{cases} . \tag{12}$$

In hard threshold, $\hat{W}_{j,k}$ were discontinuous on the points of $W_{j,k} = \lambda$, which brings oscillationto the reconstructed signal; Despite the $\hat{W}_{j,k}$ calculated in soft thresholding have overall good continuity, $\hat{W}_{j,k}$ and wavelet coefficients of noise signal have constant deviation, reconstructed signal appears to be too smooth that precision declined. Because of some flaws of hard and soft threshold function itself, there is a certain deviation in reconstructed signal, it need to improve the threshold function, to reduce the deviation of the wavelet coefficients, to make it continuous in wavelet space, with higher order derivatives. So an improved threshold function to combine hard and soft thresholding was proposed [9, 10]:

$$\hat{W}_{j,k} = \begin{cases} W_{j,k} - 0.5\dfrac{\lambda^n \cdot k}{(W_{j,k})^{n-1}} + (k-1)\lambda, W_{j,k} > \lambda \\ 0.5\dfrac{|W_{j,k}|^m \cdot k}{(\lambda)^{m-1}} sign(W_{j,k}), |W_{j,k}| \leq \lambda \\ W_{j,k} + 0.5\dfrac{\lambda^n \cdot k}{(W_{j,k})^{n-1}} - (k-1)\lambda, W_{j,k} < -\lambda \end{cases} . \tag{13}$$

In Equation (13), *m*, *n*, *k* are the adjustment factor for improving threshold function, which enhance the flexibility of the threshold function. Parameters *m* and *n* determined the form of threshold function, the parameters *k* values between zero and one, if *k* = *0*, the threshold function is equivalent to the soft threshold function, if *k* = 1, then the threshold function is equivalent to the hard threshold function. Therefore, the adjustable parameter *k* can overcome the discontinuity of hard threshold functions and the constant deviation of soft threshold function when dealing with wavelet coefficients, but also retains the original advantages of the soft and hard threshold. Improved threshold function have infinite order continuous derivative, which provides the foundation for selecting wavelet adaptive threshold.

4 Selection of adaptive threshold

Traditional threshold function will produce the phenomenon of over-kill, perform poor in practical applications. Since the noise has a negative singularity, its magnitude and dense

degree decreases with the increase of scale, but the signal is the opposite. With the increase of the scale, amplitude and dense modulus degree of maxima controlled by the noise quickly reduced and amplitude and dense modulus degree of maxima of the signal will be significantly increased. It can be seen that using the same scale thresholds on the same level are clearly inappropriate, because at a lower scale it will removes useful information and at the largest scale it will leave parts of the noise.

Adaptive threshold is a adaptive threshold selection algorithm using the wavelet transform coefficients of minimum risk mount as threshold. By the Baswarr theorem, the square of wavelet coefficients have energy dimension, therefore the square of decomposed wavelet coefficients are sorted, the risks are calculated according to the given threshold values, which will gets its likelihood estimation. The non-likelihood minimization is done and the selected threshold values are obtained, which is a soft threshold estimator. The specific algorithm is:

1) The square of the wavelet transform coefficient $w_{j,k}$ of each layer are arranged in ascending order, to get a vector $w=[w_1, w_2, ..., w_n]$, $w_1 \le w_2 \le ... \le w_n$, wherein n is the number of wavelet coefficients.

2) Calculate the risk vector $R=[r_1, r_2, ..., r_n]$, then $r_i = \left| n-2i+(n-i)w_i + \sum_{k=1}^{i} w_k \right| / n$, where r_i is the introduced risk vector elements.

The minimum r_i yielded by the above equation multiple iterations are denoted by r_0, and the corresponding w_i are denoted by w_0.

3) Calculate the threshold value

$$\lambda = \sigma (w_0)^{\frac{1}{2}}. \tag{14}$$

5 Simulation

In order to compare the different threshold noise reduction methods, the original signals and the noisy signals are selected as the standard signal $f(n)$ and $s(n)$, the length of signals are L, the signal-noise-ratio (SNR) is defined as the Equation:

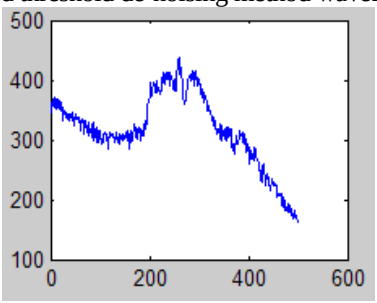
$$SNR = 10 * \log \left[\frac{\sum_{i=1}^{L} f^2(i)}{\sum_{i=1}^{L} (s(i)-f(i))^2} \right] (db). \tag{15}$$

Root mean square error (RMSE) between the original signal and the estimated signals is defined as:

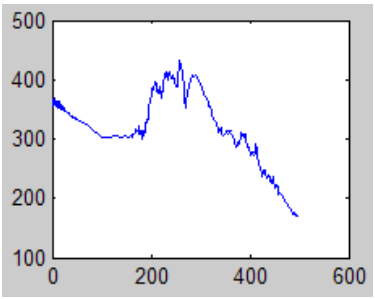
$$RMSE = \sqrt{\frac{1}{L} \sum_{i=1}^{L} (s(i)-f(i))^2}. \tag{16}$$

5.1 DENOISING OF SIMULATION SIGNAL

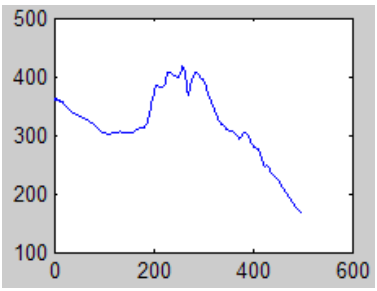
In order to verify the effect of improved threshold denoising method, leleccum signals in matlab are selected in simulation experiments, "sym5" wavelet basis are selected in wavelet packet decomposition, the maximum decomposition scale is three. Figure 2a are the original analog signals, Figure 2b are conventional wavelet thresholding waveforms, Figure 2c are the improved threshold de-noising method waveform.



a) The original signals



b) Signals of conventional wavelet thresholding denoising



c) signals of improved thresholding denoising

FIGURE 2 simulation signal denoising

As can be seen from Figure 2, the two denoising methods improve the signal to noise ratio. Most of the noise have been suppressed, but the use of the improved threshold denoising method get small distortion and better smoothness, high-frequency useful information of signals are relatively well preserved, the SNR is higher than the conventional wavelet threshold de-noising and has a good effect.

4.2 SYNTHETIC SEISMOGRAM DENOISING

Seismic data simulation was done by using MATLAB, common shot gather profiles are synthesized with a low-frequency ricker wavelet and a high-frequency ticker wavelet superimposed, specific parameters: a depth of one hundred meters, upper reflective layer velocity of two thousand me-

ters every second, lower speed of three thousand meters every second, the minimum offset of two meters, the number of channels of sixty, a record length of one hundred and twenty millisecond. Figure 3 is the synthetic common shot gathers, Figure 4 is common shot gathers with 5db random noise, Figure 5 is the common shot gathers after the common wavelet packet denoising. Figure 6 is the common shot gathers after the improved threshold wavelet packet denoising.

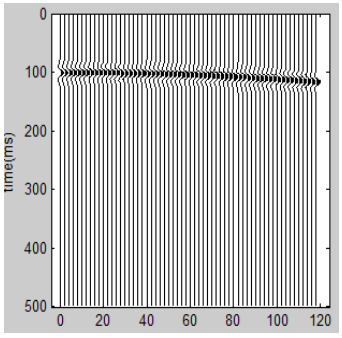


FIGURE 3 Synthetic common shot gathers

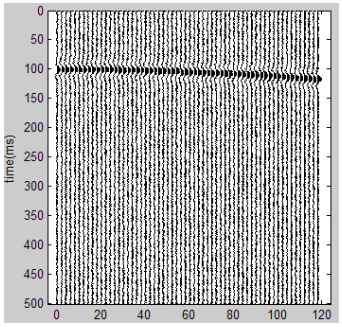


FIGURE 4 Common shot gathers with noise

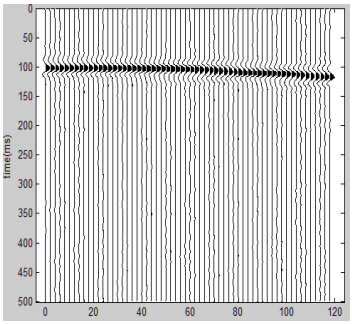


FIGURE 5 The common shot gathers after the common wavelet packet denoising

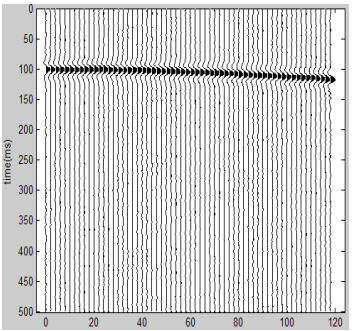


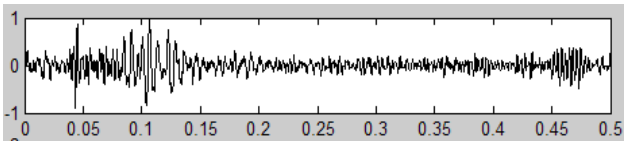
FIGURE 6 The common shot gathers after the improved threshold wavelet packet denoising

It can be seen from Figure 5 and Figure 6, when the valid signal contains high frequency components, the denoising visual effects of improved method is better than conventional wavelet thresholding denoising. The SNR of conventional wavelet thresholding and the new threshold wavelet packet denoising were 25.3018 db and 32.1148 db, SNR have also increased, SNR of improved threshold wavelet packet denoising are higher.

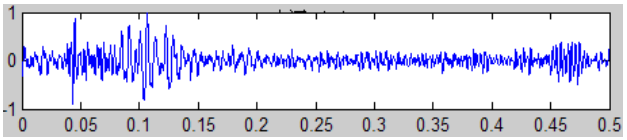
5 Seismic data denoising

The interference noise is inevitable in real seismic data acquisition and the collected data must be denoised. In the test Germany SUMMIT distributed seismograph were used for observation seismic data of a monitoring point, channels were twenty-six, channel spacing were one kilometers, offset distances (two short distances) were ten kilometers, gun dot distances were two kilometers, the detector arrays and source arrays were with one line, recording time was six seconds. Actual seismic data are generally stored in the format of segy, it need to convert, and to remove the channel header and volume header in real data processing. sym4 wavelet were selected for three times decomposition to the acquired seismic wave.

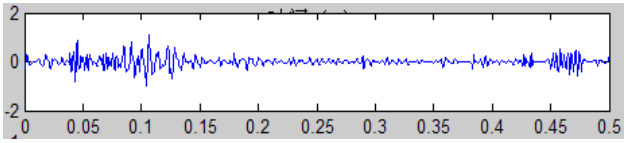
Figure 7a are part data of the original signal in one channel, Figure 7b are signals after conventional wavelet thresholding denoising, and Figure 7c are signals of improved threshold function denoising.



a) The original signals



b) The signals after conventional wavelet thresholding denoising



c) The signals of improved threshold function denoising

FIGURE 7 The common shot gathers after the improved threshold wavelet packet denoising

As can be seen from Figure 7 that most noises of seismic signals by hard threshold function denoising are filtered, which also weaken the details of the original signals, resulting in distortion, which is the results of the coefficients less than the threshold value are set to zero caused by hard threshold value function. The amplitude of soft threshold function denoising and the original seismic signals are of maximum distortion. Instead, the smoothness of seismic signals with the new threshold function denoising are best, the useful signals

components are well reserved reducing the distortion. It can be seen visually seismic signals denoised by improved threshold function are closest to the original signals.

6 Conclusions

An denoising method based on improved wavelet threshold function and adaptive threshold wavelet packet seismic signals was presented. In wavelet packet decomposition process, the improved threshold function and a hierarchical adaptive threshold were selected to apply to the synthetic seismogram and real seismic records. It can be seen from the SNR improvement and simulation results, that this method can effectively remove noises from the sampled signals, keep the useful signal and improve signal to noise ratio of seismic data.

Acknowledgments

This work was financially supported by Teachers' Scientific Research Fund of China Earthquake Administration (20140105).

References

[1] Zhang H, Liu T, Zhang Y 2009 Denoising of seismic data via multi–scale ridgelet transform *Earth q Sci* **22** 493-8 (*in Chinese*)
[2] Yi Z, Lizhi C 2011 A New wavelet denoising method of seismie signals based on a recoursive optimal thresholding *Acta Scientiarum Naturalium Universitatis Nankaiensis* **44**(4) 67-71 (*in Chinese*)
[3] Wei H, Huang S, Xu F 2008 Application of the wavelet packet denoising in pretreatment of seism signal *Seismological Research of northeast China* **24**(2) 45-9 (*in Chinese*)
[4] Zhang R, Lin D, Qiao L 2011 Application of the optimum wavelet packet transform in seismic signals denoising *Journal of Seismological Research* **34**(3) 358-63
[5] Zhou H L, Li L M, Luo S X 2006 A method of improved threshold function based on wavelet analysis *Computing Techniques For Geophysical And Geochemical Exploration* **28**(2) 0173 (*in Chinese*)
[6] Zhou S, Zuo D 2012 Wavelet denoising basaed on improved threshold function and adaptive threshold *Electronic Sci & Tech* **25**(11) 31-4 (*in Chinese*)
[7] Zhang H, Chen X, Yang H 2011 Optimistic wavelet basis selection in seismic signals elimination *OGP* **46**(1) 70-5 (*in Chinese*)
[8] Liu X, Pan H, Gao X, Wu W 2010 Seismic data denoising based on wavelet thresholding *Science Technology and Engineering* (29) 7251-4 (*in Chinese*)

Authors

Liu Shuchong, born in November, 1983, Huhehaote City, Nei Mongolia Province, China

Current position, grades: a lecture in Institute of Disaster Prevention Yanjiao District, San he City, Hebei Province, China.
University studies: institute of Disaster Prevention.
Scientific interest: signal processing.
Publications: 6 papers.

ChenXun, born in 1974, Xi,an City, Shan xi Province, China

Current position, grades: Department of Instrument, Institute of Disaster Prevention, Yanjiao District, San he City, Hebei Province. China.
University studies: Institute of Disaster Prevention Yanjiao District, San he City, Hebei Province, China.
Scientific interest: image processing, signal processing.
Publications: 20 papers.

Study on automatic identification method of X-ray fluorescence spectrum

Jiening Xia[1, 2*], Zhigao Chen[1, 2]

[1]*Hubei Key Laboratory of Earthquake Early Warning, Institute of Seismology, China*

[2]*Wuhan Base of Institute of Crustal Dynamics, CEA, Wuhan, China*

Received 1 July 2014, www.cmnt.lv

Abstract

The way to improve the accuracy and reliability of automatic unscrambling and identification technology on X-ray fluorescence spectrometer spectrum is studied in this essay. Accordingly, two different automatic identification methods based on Fast Fourier Transform and Wavelet Transform are presented. By the tool LabVIEW, such two methods are applied to the qualitative analysis on X-ray fluorescence spectrums, and the features of such two methods are compared. Based on the experiments and analysis on amount of samples, it can be concluded that the automatic identification method based on the Wavelet transform theory is better than the other method for the former has a better local resolution. Therefore, the characteristic values of the singular points are more clearly recognized by the method based on the Wavelet transform. Through the study in this essay, theories on automatic identification are enriched, which set a foundation for further studied in future.

Keywords: X-ray fluorescence spectrometer, spectrum identification, LabVIEW, fast Fourier transform, wavelet transform

1 Introduction

X-ray fluorescence spectroscopy technology as a widely used ray-detection technology, began in the early 1950s. Since the spectrum is less interfered, accurate, precise, fast, can detect multiple elements, has a wide range of measurement, and performs non-destructive testing, it is applied to various fields, which becomes an indispensable and means in the laboratory and no alternatives. And now it becomes an indispensable method for analysis of substance composition. X-ray fluorescence spectroscopy detection technology has overwhelming advantages such as fast analysis speed, wide range of elements, which can be tested, simple pretreatment, free of pollution, low cost and non-destructive testing, etc. Nowadays, there are more and more types of X fluorescence spectrometers. Those X fluorescence spectrometers can be divided into two types according to the method of acquiring and distinguishing the characteristics of the X-ray fluorescence spectroscopy: wavelength dispersive X-ray fluorescence spectrometers (WDXRF) and energy dispersive X-ray fluorescence spectrometers (EDXRF) [1].

An X-ray fluorescence spectrometer is generally made up of three parts in structure: a light source for exciting the sample (X-ray tube), a detection system (for dispersion, detection, and controlling the spectrometer), and a data processing system [2]. Incident X-rays are generated from the X-ray tube for exciting the sample. Accordingly, secondary X-rays can be emitted from the to-be-tested elements of the sample by the excitation, in which different secondary X-rays emitted from excitation of different elements each owns specific energy characteristics and wavelength characteristics. The energy of such secondary X-rays are collected and measured by the detection system, and through a data processing software in the detection system, the collected fluorescence information of such secondary X-rays are calculated into types and contents information of the to-be-tested elements of the sample. Spectrum characteristics of a certain element is not only related to the energy and intensity of the excitation source, but also related to the level of such element in the sample. According to the spectrum characterristics of such element, the content information thereof can be obtained by converting.

There are mainly three types of analysing methods of the X-ray fluorescence spectroscopy detection technology: qualitative analysing method, semi-quantitative analysing method and quantitative analysing method [3]. The qualitative analysing method is discussed in this essay, which indicates the aforementioned solution and identification methods on the spectrum. The Moseley's law tells that the wavelength of the secondary X-ray emitted from the to-be-tested element is one to one related to the atomic number. In qualitative analysis, characteristic peaks and common interference lines of the scanned spectrum figure should be identified, thereby to determine the types of elements contained in the sample, and thus to determine the types of materials of the sample.

Currently, the automatic recognition algorithm of the X-ray fluorescence spectrometer is being studied by various spectrometer manufacturers. According to the algorithm, by determining the peak of the scanned spectrum, calculating the net intensity of the background and peak, as well as pairing with the characteristic spectrum line database, the types of the to-be-tested elements and the types of the spectrum lines are determined. However, the actual working situation is complex and changing. For example, when the

**Corresponding author* e-mail: dr.xjning@gmail.com

content of an element is too little or there exists interference between of spectral lines of different elements, the artificial identification by experiences of the operator is essential. First, the characteristic X-ray lines of the target X-ray tube and the associated lines of the peak are identified. Afterwards, the remaining spectrum lines are marked according to energy. When unknown spectrum lines are analysed, other factors such as the origin and nature of the sample should also be taken into consideration, so as to get comprehensive judgments.

In order to improve the accuracy and reliability of the automatic recognition and reconciliation of spectrums, this essay presents a method for feature extraction and recognition based on fast Fourier transform, and a method for spectrum solution based on wavelet. And such two methods are proved and compared by specific experiments with LabVIEW.

2 RELATED WORK AND BACKGROUND

Spectrum recognition models either domestic or international are essentially divided into two types: a spectrum matching method based on spectrum reconstruction and standard spectrum similarity measurement, and an intelligent recognition method based on sample knowledge. The spectrum matching method is to compare the reconstructed spectrum with the reference spectrum, and to measure the similarity or correlation between them by a certain measurement function, so as to identify the sample. The similarity measurement function can be a distance function (Euclidean distance, Mahalanobis distance), similarity coefficient, correlation coefficient, spectral information divergence or spectrum vector angle, etc.; and the reference spectrum can generally be a standard spectrum in the spectrum database.

According to the data used during the matching process, the spectrum matching method can be divided into two types: direct matching and indirect matching. The direct matching method uses the reconstructed spectrum to match the reference spectrum data directly; whereas the indirect matching method encodes or transforms the spectrum data first, and uses the encoded and transformed data to match, such as encoded spectrum matching, amplitude and phase of the Fourier frequency of the spectrum, wavelet or fractal dimension matching, other than using the raw spectrum data to match directly.

When the X-ray fluorescence spectrum analysis method is carried out, due to the interaction of X-rays substances, such as coherent scattering, incoherent scattering, Compton scattering and other reasons, the Almighty certain peaks of the characteristic X-ray may superimpose on the background. In the qualitative and quantitative analysis, it is very necessary to subtract the background accurately in order to obtain the peaks of the spectrum and the net intensity of peak areas. It is a very critical and difficult technology to subtract the background effectively and accurately. Therefore, various digital filtering and deconvolution algorithms are powerful tools for the spectrum solution.

In recent years, more and more advanced intelligent methods have been used in such field. The intelligent identification method is on the basis of physical spectrum knowledge. According to this method, appropriate diagnostic spectrum features or spectrum parameters with identification ability are selected first, then identification rules are established, and samples are identified accordingly. Characteristics or parameters of the spectrum can be the similarity measure for the whole spectrum, such as the matching degree or matching rate compared to the standard spectrum, or the parameters of specific bands, or both. The identification rules can be generally expressed in 3 ways: uniqueness identification (existence identification), i.e., when some certain conditions are met, it can be considered that some certain element exists; negative identification, i.e., when one or more features occur, it can be considered that some element must not exists; likelyhood identification, i.e., the likelihood whether the material to be identified is determined according to the conditions.

The intelligent identification method combines the full spectrum matching method and the feature spectrum band identification method, which takes the advantages of the two methods, and establishes an identification rule referring to the expert system. As such, the reliability and automation of the identification are greatly improved. For example, someone introduces the genetic algorithm to the solution of X-ray fluorescence spectrum and achieves good results. The genetic algorithm is a kind of probabilistic algorithm based on simulation on the laws of life, which searches for a solution suitable for the environment. Some studies present a method that takes the fractal dimension as the spectrum identification feature. In this method, the fractal dimension of the spectrum signals is obtained by phase-space reconstruction. Different spectrums can be identified through comparing the fractal dimension of the spectrum signals. Someone also proposes another spectrum identification scheme based on multi-spectral features and integrated neural network, and implements system design. However, most of the aforementioned methods and corresponding implementation software are on the basis of human involvement or assistance. This paper attempts to present an automatic spectrum identification method based on traditional signal processing methods and advanced algorithms. Under the current condition, overly complex algorithms do not required for the automatically recognizes by machines, while characteristic values are needed used for the machine's independent judgment.

The basic ideas of both the Fourier transform automatic identification method and the wavelet transform automatic identification method require a standard library, which is used to compare the calculated characteristic values of the samples with the characteristic values in the library obtained by the same method. Therefore, it is very significant to create such a feature library. The general spectrums of an alloy, which includes Stainless steel, cobalt alloy, nickel base alloy, Titanium alloy and middle-low alloy steel are chosen as the spectrum library, wherein each library contains multiple spectrums. The spectrum library is characterized by a processing method based on the fast Fourier transform and characterized by a processing method based on wavelet trans-

form respectively, which will be mentioned below. The feature libraries corresponding to the two methods thus obtainned. The alloy of Stainless steel, cobalt alloy, nickel base alloy, Titanium alloy and middle-low alloy steel are selected as the testing samples. The spectrums of the samples are general random spectrums completely different from the feature libraries.

A suitable experimental environment and proper tools are in need on the basis of aforementioned works, We had chosen the NI LabVIEW environment. LabVIEW (Laboratory Virtual Instrument Engineering Workbench) is a kind of graphical programming language, which uses icons to create applications instead of text lines. In traditional text-based programmming languages, programs are executed in sequences of statements and directives. Whereas, in LabVIEW a dataflow programming mode is used, that VI (abbreviation of virtual instrument, refers to program modules in LabVIEW) and execution sequences of functions are determined by the data flow between nodes of block diagrams. With the LabVIEW programming language, flow charts or diagrams are used instead of program codes. LabVIEW is an end-user oriented tool, which adapts terms, icons and concepts familiar with technical staffs, scientists, and engineers as much as possible. LabVIEW can enhance the ability of building your own science and engineering systems, and can provide a convenient way to implement instrumental programming and data acquisition systems. The working efficiency of the instrument system can be greatly improved with LabVIEW in its principle researching, designing, testing and implementation.

3 The spectrum unscrambling and identification method based on fast Fourier transform

The Fourier transform would be firstly taken into consideration if Mentions signal analysis in transform domain. Through the discrete Fourier transform (DFT), a finite-length sequence *x(n)* can be transformed into a discrete finite-length sequence *x(k)* in its frequency domain [4].

$$x(k)=\sum_{n=0}^{N-1}x(n)W_N^{nk},\ k=0,1,2,\ldots,N-1. \tag{1}$$

Whose inverse discrete Fourier transform (IDFT) is

$$x(n)=\frac{1}{N}\sum_{k=0}^{N-1}x(k)W_N^{nk},\ n=0,1,2,\ldots N-1, \tag{2}$$

wherein $W_N=e^{-j2\pi/N}$, and *x(n)* and *x(k)* may be a real number or a complex number. It can be concluded that to calculate a sample sequence, *N* times of multiplication operation and *N*-1 times of addition operation of complex numbers should be done. Therefore, it is difficult to perform real-time processsing with DFT because of its huge amount of computations.

Since it is very important to do discrete Fourier transform in signal analysis, it is very necessary to find fast algorithms for discrete Fourier transform and corresponding inverse transform. Fast Fourier Transform (FFT), namely fast discrete Fourier transform (FFT) is a kind of fast DFT algorithm that invented by J.W.Cooley and J.W.Tukey in 1965, wherein the periodicity and symmetry of the factor of the W_N is utilized. The Discrete Fourier Transform is equivalent to multiplying a sequence *x* (as a column vector of *n* elements) by a $n\times n$ matrix F_n, which needs n^2 multiplications. Whereas the fast Fourier transform is to change the matrix F_n into a special form, which can reduce the times of multiplications into approximate $5n\log_2 n$ times. If n=1000, then by the fast Fourier Transform, the times of multiplications can be reduced from millions to about 50, 000.

In the later decades, it is further developed with the FFT algorithm, wherein the most commonly used FFT algorithms now are radix-2 algorithm and split-radix algorithm. It is known that the FFT algorithm has been introduced in detail in many other documents, so this essay will not describe it in detail.

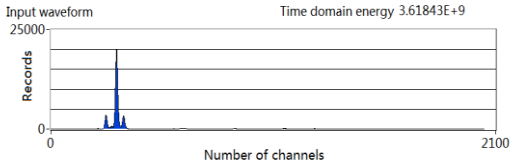


FIGURE 1 A spectrum captured by X-ray fluorescence spectrometer

Referring to Figure 1, a spectrum is shown. In physical, it is the channel value that corresponds to the horizontal axis, rather than a time-sequence. However, in a broad view, it can still be taken as a complete process. The aforementioned fast Fourier transform can be performed using the signal analysis tool in LABVIEW, as shown in Figure 2.

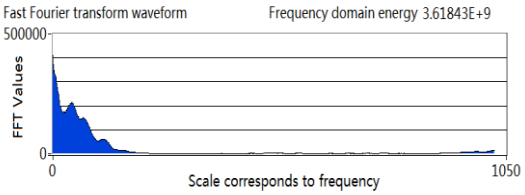


FIGURE 2 The result of fast Fourier transform

It is shown in the Figure 1 and Figure 2 that the energy in time domain calculated from the original pattern before the transformation must be equal to the energy in frequency domain calculated after the transformation, which is in line with Parseval's Theorem. It should be noted that the points in the abscissa only needs half of the original after the Fourier transform, even a quarter or less effective portion thereof with experiences. Accordingly, the amount of computation can be greatly reduced. Unfortunately, it is clear that the characteristic frequencies of almost all the general spectrum samples after the FFT in different spectrum patterns cannot be clearly distinguished by the Fourier transform. In order to make the FFT method continue, a method which combines the energy values and the boundary detection is used, considering the Parseval's Theorem. In this essay, the energy value is selected as the characteristic value, and then transformed by FFT transform, and last detected by its boundary for spectrum pattern recognition.

As shown in Figure 3, the energy value of the to-be-identified spectrum pattern is compared with that of the sample spectrum pattern in Library. The situation should be

very clear if the two energy values are not in the same magnitude. Otherwise, a threshold of range should be defined if the two energy values are close to each other. Additionally, comparison between the boundary values can be performed under the premise that the energy values are as the characteristic values. The pink line and blue line in the Figure 3 represent the top boundary value and bottom boundary value respectively. The identified spectrum pattern is transformed by FFT transform and compared with the boundary values. Points out of the boundaries are marked in red. In this figure, points that do not match occupy 0 percent of the total number of points, in some instances, this value may become large. Therefore, such percentage can also be taken as the threshold value of range for automatic identification.

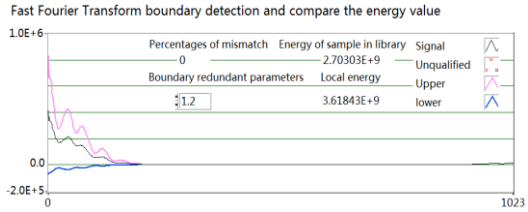


FIGURE 3 The spectrum unscrambling based on Fast Fourier transform

As can be concluded from the identification result, the material as shown in the Figure 3 is probably stainless steel.

4 The spectrum unscrambling and identification method based on wavelet transform

The basic principal of wavelet transformation is: after being dilated and translated, the original signal is decomposed into a series of sub-band signals with good local features in time domain and frequency domain, thus achieving local time and frequency analysis [5]. Wavelet transform is strongly correlated to the data, which make the energy of the signal exist only in some large wavelet coefficients in the wavelet domain, whereas the energy of the noise distributes throughout the whole wavelet domain. Thus, through the wavelet decomposition, the amplitude of the wavelet coefficients of the signal should be greater than that of the noise. It is considered that whose amplitude of the wavelet coefficients is relatively larger should be generally the signal, while whose amplitude of the wavelet coefficients is relatively smaller should probably be the noise. Therefore, wavelet transform shows better performance of detection in non-stationary signals with low SNR. When the scale of a signal to be detected changes, there will be a wavelet with an appropriate scale to match it. On the contrary, by the short time Fourier transform, the profiles of the signal can only be detected when the basic function matches the signal scaling function, while it is difficult to get a satisfactory result in other cases.

The purpose of Wavelet transform and the purpose of Fourier transform are the same: the signal is expressed as a linear combination of basic functions. The difference between the two transformations is: the basic function of the Fourier transform is the harmonic function $e^{-j2\pi/N}$, whereas the basic function of the wavelet transform is the Mother Wavelet Function $\Psi(t)$ with compact support generation. In the wavelet transform, a wavelet sequence is obtained through dilating and translating to the mother function $\Psi(t)$:

$$\psi_{a,b}(x)=\frac{1}{\sqrt{a}}\psi(\frac{x-b}{a}),\quad a,b\in R;a\neq 0 \tag{3}$$

where a is the dilating factor, b is the translating factor.

As to a given function $f(t)\in L^2(R)$, its continuous wavelet transform is:

$$W_f(a,b)=<f,\psi_{a,b}(x)>=\int_{-\infty}^{\infty}f(x)\psi_{a,b}(x)dx =\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}f(x)\psi(\frac{x-b}{a})dx \tag{4}$$

The selection of the basic wavelet function is very important, which often depends on the application [6]. The wavelet function generally has two basic characteristics geometrically: it must be an oscillating function and it must be a function with rapid convergence. The above two criteria should be followed When selecting a wavelet function or constructing a wavelet function. Different dilating factors and translating factors will bring about great changes in geometrical shapes of the wavelet function.

Since the translating factor corresponds to the time t, and the dilating factor corresponds to the frequency *f*, the time-frequency plane (*t, f*) becomes a time-scale plane (*b, a*). The principal of weak signal detection based on Wavelet transform is to put the signal with the noise on a time-scale two-dimensional plane, and then process the signal in time-division and frequency-division modes based on distinct characteristics of the signal and the noise. When the scale parameter a increases, the basic function $\psi_{a,b}(x)$ turns to a broadening wavelet, which corresponds to a low frequency function, whose frequency resolution is improved, but time resolution is reduced. Otherwise, when the scale parameter a decreases, the basic function $\psi_{a,b}(x)$ turns to a compressed wavelet, whose time resolution is improved, but frequency resolution is reduced. When the signal to be analysed has a sudden change, it is better to choose a smaller "a" to improve the resolution in the time domain. Therefore, the wavelet transform is more suitable for studying signals which are non-stationary and contains short-term transient components [7]. There is a special signal processing package for wavelet calculation in LabVIEW. It is very convenient to do experimental verifications on the idea of wavelet-based identification.

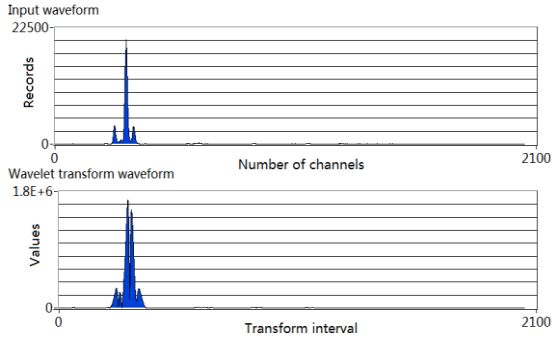


FIGURE 4 The waveform of wavelet transform

As shown in Figure 4, in the upper part, the input signal is a spectrum recorded by the X-ray fluorescence spectrometer. The abscissa of the spectrum physically represents the channel number of the spectrometer, while the ordinate of the spectrum represents values recorded by the instrument in every channel. In the lower part of the figure, there shows a waveform transformed from the input signal by broadly wavelet transform.

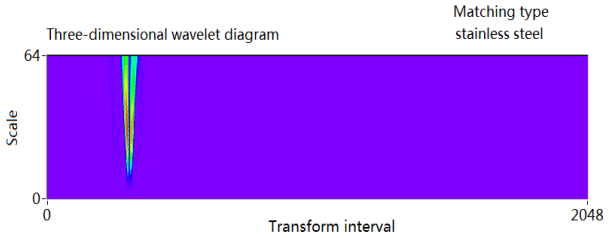


FIGURE 5 Singular points detection in a three dimensinal wavelet digram

As can be seen in Figure 5, it should be noted that the parameter SCALS here controls the scaling ratio of Wavelet transform. And the peaks with different amplitudes of the spectrum in the figure can be considered as singular points of the signal, which can be used as characteristic values of the spectrum. Thus, it can be taken as a detection process of singular points that the solution and identification s process of the spectrum [8]. For a long time, the Fourier transform is a main tool for study the singularity of functions. In detail, by studying the attenuation in the Fourier transform domain of the function of to infer whether there is singularity on the function and how serious the singularity is. However, referring to Fourier transform, its ability to distinguish details is often unsatisfactory. In contrast, wavelet transform is more spatially localized. Accordingly, it is more effective to use wavelet transform to analyse the location of singularity and the singular degree.

A multi-scale peak detection can be carried out toward the data that processed through continuous wavelet transform from the original spectrum, and the detected peaks as shown in Figure 5. Positions of singularity can be considered as characteristic values for identifying the spectrum. The magnitude of the ordinate and the resolution of the abscissa of the wavelet transform data can be used as scaling parameters of the multi-scale peak detection. Peaks less affected can be excluded by adjusting the amplitude threshold, and a proper characteristic value can be accurately found from the overlapped peaks by adjusting the resolution parameter. The characteristic value is corresponding to the position of transform interval, and it can also be corresponding to the channel number of the original input spectrum. Accurate spectrum identification can be performed by comparing that characteristic value with the characteristic library of different materials recorded in the sample library.

Figure 5 gives the automatic identification result, the material shown in the in figure is probably stainless steel.

5 Experiments and results

Two automatic recognition algorithms were designed for testing sample data in LABVIEW environment, which were based on fast Fourier transform and wavelet transform. Sample data contain general spectroscopy spectrum of Stainless steel, cobalt alloy, nickel base alloy, Titanium alloy and middle-low alloy steel five alloys. Before the experiment, we were prepared alloy feature library for each algorithm.

5.1 TESTING FOR THE AUTOMATIC IDENTIFICATION METHOD BASED ON FAST FOURIER TRANSFORM

The FFT method similar to a spectrum matching filtering method (Matched Filtering) is a kind of method to identify a specified target from a background whose spectral information is unknown. The basic idea thereof is to take the material spectrum in a standard spectrum database as a vector in the Q mean space, and to build a filtering vector detector according to the matching degree between the spectrum of the image and the reference spectrum vector, then to filter the input image, so that the spectral response of the target is maximized while the spectral response of other end of the element is suppressed. As such, the contrast between the target and the background is maximized and an optimal matching is achieved, and then an appropriate threshold should be selected, and the target information can be extracted.

Originally, the matching filtering is developed for calculating the relative abundance of the material content of less. Therefore, when the content of the target substance is not few, it should be careful and cautious to apply the method and to interpret the results of the method. The Mixture Tuned Matched Filtering (MTMF) [9] is a composite method, which combines the linear mixed decomposition method and the matched filtering method. It combines the advantages of matched filtering method that do not need other background end elements and the constraint condition that the content of each end element of the image element in the linear mixed decomposition method should be positive and the sum of the content should be 1, thereby the detection limit of the elements is reduced, and the trace elements in the samples that other methods could not detect would be detected.

In Table 1, results by artificial matching identification of 51 samples are excerpted, which are compared with the results by FFT automatic identification algorithm. As can be seen, most of the results are consistent, except that the results in No. 24 and No. 25 by automatic identification identified the stainless steel as the nickel base alloy. After careful analysis, it can be known that by the automatic identification algorithm based on FFT, for both the No. 24 and No. 25 samples, it is difficult to clearly distinguish whether they belong to stainless steel or nickel base alloy, either by energy identification condition or by boundary matching condition. As to the energy identification condition and the boundary matching condition, the libraries of the stainless steel and the nickel base alloy themselves are also very similar. This kind of situation is improved in the automatic identification algorithm based on wavelet.

TABLE 1 FFT Automatic identification test results

Sample number	Manual matching results	Automatic matching results
1	cobalt alloy	cobalt alloy
2	cobalt alloy	cobalt alloy
3	cobalt alloy	cobalt alloy
4	cobalt alloy	cobalt alloy
5	cobalt alloy	cobalt alloy
6	cobalt alloy	cobalt alloy
7	cobalt alloy	cobalt alloy
8	middle-low alloy steel	middle-low alloy steel
9	middle-low alloy steel	middle-low alloy steel
10	middle-low alloy steel	middle-low alloy steel
11	middle-low alloy steel	middle-low alloy steel
12	middle-low alloy steel	middle-low alloy steel
13	nickel base alloy	nickel base alloy
14	nickel base alloy	nickel base alloy
15	nickel base alloy	nickel base alloy
16	nickel base alloy	nickel base alloy
17	nickel base alloy	nickel base alloy
18	nickel base alloy	nickel base alloy
19	nickel base alloy	nickel base alloy
20	nickel base alloy	nickel base alloy
21	nickel base alloy	nickel base alloy
22	nickel base alloy	nickel base alloy
23	nickel base alloy	nickel base alloy
24	stainless steel	nickel base alloy
25	stainless steel	nickel base alloy
26	stainless steel	stainless steel
27	stainless steel	stainless steel
28	stainless steel	stainless steel
29	stainless steel	stainless steel
30	stainless steel	stainless steel
31	stainless steel	stainless steel
32	stainless steel	stainless steel
33	stainless steel	stainless steel
34	stainless steel	stainless steel
35	stainless steel	stainless steel
36	stainless steel	stainless steel
37	stainless steel	stainless steel
38	stainless steel	stainless steel
39	stainless steel	stainless steel
40	stainless steel	stainless steel
41	Titanium alloy	stainless steel
42	Titanium alloy	Titanium alloy
43	Titanium alloy	Titanium alloy
44	Titanium alloy	Titanium alloy
45	Titanium alloy	Titanium alloy
46	Titanium alloy	Titanium alloy
47	Titanium alloy	Titanium alloy
48	Titanium alloy	Titanium alloy
49	Titanium alloy	Titanium alloy
50	Titanium alloy	Titanium alloy
51	Titanium alloy	Titanium alloy

In Table 2, results that comparing and matching a sample with the samples in the library database of FFT algorithm are excerpted, in which the calculation process of such automatic identification algorithm can be reflected. First, it can be seen that for the energy identification condition, the sample is relatively easier to pass cobalt alloy, stainless steel and Titanium alloy. Then the boundary matching condition should be considered, and also the statistical analysis law, through which it can be determined that the sample belongs to Titanium alloy.

TABLE 2 Single sample FFT automatic recognition and matching rsults

Library types	If energy comparison passed?	No matching percentage
cobalt alloy1	T	42.5
cobalt alloy2	T	42.5
cobalt alloy3	T	42.5
cobalt alloy4	T	41
cobalt alloy5	T	66.5
cobalt alloy6	T	52
cobalt alloy7	T	46.5
cobalt alloy8	T	65
cobalt alloy9	T	50.5
cobalt alloy10	T	56.5
cobalt alloy11	T	69
middle-low alloy steel1	F	100
middle-low alloy steel2	F	98.5
middle-low alloy steel3	F	100
middle-low alloy steel4	F	100
middle-low alloy steel5	F	100
middle-low alloy steel6	F	100
middle-low alloy steel7	F	100
middle-low alloy steel8	F	98.5
nickel base alloy1	T	60
nickel base alloy2	F	57.5
nickel base alloy3	F	46
nickel base alloy4	T	44
nickel base alloy5	T	45.5
nickel base alloy6	T	44
nickel base alloy7	F	51
nickel base alloy8	F	72
nickel base alloy9	F	62.5
nickel base alloy10	T	44
nickel base alloy11	T	51
nickel base alloy12	T	42.5
nickel base alloy13	T	40.5
nickel base alloy14	T	43.5
stainless steel1	F	69.5
stainless steel2	T	67.5
stainless steel3	T	70.5
stainless steel4	T	68.5
stainless steel5	T	73.5
stainless steel6	T	72
stainless steel7	T	64
stainless steel8	T	71.5
stainless steel9	T	70.5
stainless steel10	T	71.5
stainless steel11	T	65.5
stainless steel12	T	69
stainless steel13	T	72.5
stainless steel14	T	71.5
stainless steel15	T	69
stainless steel16	T	73
stainless steel17	T	68
Titanium alloy1	T	67.5
Titanium alloy2	T	66.5
Titanium alloy3	T	39
Titanium alloy4	T	74.5
Titanium alloy5	T	73
Titanium alloy6	T	77.5
Titanium alloy7	T	77.5
Titanium alloy8	T	77
Titanium alloy9	T	77
Titanium alloy10	T	72.5
Titanium alloy11	T	77.5
Titanium alloy12	T	78.5
Titanium alloy13	T	77
Titanium alloy14	T	75

To summarize, confirmed by experiments of hundreds of samples, the accuracy rate of identification can reach 90%

by the FFT transform automatic spectrum pattern identification method based on the energy threshold of range and boundary threshold of range. However, there are still some spectrum patterns that could not be clearly distinguished by their characteristics after FFT transform, even the energy value of such spectrum patterns can be easily confused with the sample libraries of different materials. So such spectrum patterns could not effectively be identified through the Fourier transform method. In order to solve such problem and further improve the accuracy of spectrum patterns identification, a wavelet method is introduced here.

5.2 TESTING FOR THE AUTOMATIC IDENTIFICATION METHOD BASED ON WAVELET TRANSFORM

The spectrum is with high spectral resolution, multi-bands, massive data, strong correlation between the bands, high data redundancy, so it is necessary to reduce the dimensions of the data and to weaken the noises before selecting the end elements and identifying the materials. The functions for the wavelet transform are:

1) To separate the information from the noise. After the transformation, the transformation components are arranged in a descending order according to their corresponding characteristic value, whose noise components decreases, while corresponding component information of greater characteristic value is more. Corresponding component information of smaller characteristic value gradually becomes less, while the noise level increases. Corresponding components whose characteristic values are very small and tend to be constant are almost all noises. Therefore, separation the information from the noise can be achieved, and the noise of each component image can be removed on target, or components whose noises are dominant can be abandoned, so as to remove noises.

2) To reduce the dimensions. After the transformation, the correlation between the original spectral bands can be eliminated, so that the information can be concentrated in a small number of components, thus redundancy of data can be reduced. Through observing the characteristic values and corresponding images, the intrinsic dimension of the data can be determined. Selection of end elements and identification of samples can be carried out in transformed low-dimensional space, so as to shorten the duration of data processing and improve the efficiency of processing.

3) To separate features. The spectrum information of samples is isolated in the transformed feature space after wavelet transformation, wherein the spectrum converges on the class feature vector set, and some weak information is enhanced in noise removing transformation. Accordingly, the separability of the data from the sample is increased, and the reliability of the identification is improved.

In Table 3, results by artificial matching identification of 51 samples are excerpted too, which are compared with the results by wavelet automatic identification algorithm. It can be seen that only one sample No.31 cannot be identified by the automatic identification algorithm. As a matter of fact, all the experimental results are not shown in the Table 3 due to the length limitation of the table. There is only this one, which could not be identified in all the hundreds of experiments we have done. Simultaneously, the characteristic values of the samples that calculated through the automatic identification algorithm based on wavelet are shown in Table 3. It can be seen that the numbers of characteristic values that calculated from different samples are different. Value1 and Value2 of each sample (if any) are better consistent. In fact, it can be identified what material the sample belong to basically from the first and second characteristic value. For some samples with more characteristic values, it is better to continue to match with the library database, of course, the premise is that there is similar characteristic value record in the library database, and to prove that the characteristic value is not illegal. For the sample No.31, it can be seen that its characteristic value does not match with stainless steel well. Therefore, it could not get a result by the automatic identification algorithm. The reason may be there some errors occur during the data acquisition process. Interestingly, it can be identified by the automatic identification algorithm based on FFT, which is actually worth further studying.

To summarize, verified by experiments on amount of samples, the identification accuracy of the automatic spectrum identification method based on Wavelet transform can reach 99%. Quite differences were found between the identified spectrum and its corresponding material in the sample library for the unique error, there possibly was something wrong during the data collection process using the spectrometer. It can be concluded that the automatic spectrum identification method based on wavelet transform owns some advantages, but there is still space for further improvement with such method.

TABLE 3 Wavelet automatic identification test results

Sample number	**Manual matching results**	**Characteristic values matching**				**Automatic matching results**
		Value1	*Value2*	*Value3*	*Value4*	
1	cobalt alloy	253	329			cobalt alloy
2	cobalt alloy	253	329			cobalt alloy
3	cobalt alloy	253	329			cobalt alloy
4	cobalt alloy	254	331	376	860	cobalt alloy
5	cobalt alloy	253	329			cobalt alloy
6	cobalt alloy	253	330	860		cobalt alloy
7	cobalt alloy	253	330			cobalt alloy
8	middle-low alloy steel	302				middle-low alloy steel
9	middle-low alloy steel	302				middle-low alloy steel
10	middle-low alloy steel	302				middle-low alloy steel
11	middle-low alloy steel	302				middle-low alloy steel
12	middle-low alloy steel	302				middle-low alloy steel

13	nickel base alloy	357	374			nickel base alloy
14	nickel base alloy	358	375			nickel base alloy
15	nickel base alloy	358	375	860	880	nickel base alloy
16	nickel base alloy	357	374			nickel base alloy
17	nickel base alloy	357	374			nickel base alloy
18	nickel base alloy	357	374			nickel base alloy
19	nickel base alloy	358	374	860	880	nickel base alloy
20	nickel base alloy	357	374			nickel base alloy
21	nickel base alloy	357	374			nickel base alloy
22	nickel base alloy	357	374			nickel base alloy
23	nickel base alloy	357	374			nickel base alloy
24	stainless steel	302				stainless steel
25	stainless steel	302				stainless steel
26	stainless steel	302				stainless steel
27	stainless steel	302				stainless steel
28	stainless steel	252	302			stainless steel
29	stainless steel	302				stainless steel
30	stainless steel	302				stainless steel
31	stainless steel	253	319			No matching type
32	stainless steel	302				stainless steel
33	stainless steel	303	859			stainless steel
34	stainless steel	302				stainless steel
35	stainless steel	303				stainless steel
36	stainless steel	253	303			stainless steel
37	stainless steel	303				stainless steel
38	stainless steel	303				stainless steel
39	stainless steel	303				stainless steel
40	stainless steel	303				stainless steel
41	Titanium alloy	208				Titanium alloy
42	Titanium alloy	208				Titanium alloy
43	Titanium alloy	208	861			Titanium alloy
44	Titanium alloy	209	861			Titanium alloy
45	Titanium alloy	208	817			Titanium alloy
46	Titanium alloy	209	776	861		Titanium alloy
47	Titanium alloy	208				Titanium alloy
48	Titanium alloy	209	818			Titanium alloy
49	Titanium alloy	208				Titanium alloy
50	Titanium alloy	208				Titanium alloy
51	Titanium alloy	209	861			Titanium alloy

6 Conclusions

Methods of automatic unscrambling and identification on spectrums captured by X-ray fluorescence spectrometers are discussed in this essay. It is difficult to accurately distinguish the peaks in the spectrum in the time domain only with experiences, because the peaks in the spectrum are always overlapped and are easily confused. It should be noted that such time domain is a broad concept which is corresponding to the channel number of the abscissa of the spectrum. Therefore, in order to achieve fast and accurate automatic spectrum identification, a method for unscrambling and identifying characteristics based on fast Fourier Transform and a method based on wavelet transform are proposed herein.

It can be concluded from the results of experiments by the aforementioned two methods that: by the method based on fast Fourier transform, the spectrum is transformed from the time domain into the frequency domain (broadly) to find characteristics of the spectrum in the frequency domain, but actually in some spectrums it is difficult to distinguish peaks of different time points (channel numbers) in the frequency domain because the peaks are also overlapped. In contrast, by the method based on wavelet, it can be clearly and accurately to find the characteristic values of the spectrum and thus effectively to identify various elements and materials, wherein the characteristic of "being amplified in part" of non-stationary and short-term transient component accompanied signals is fully utilized.

Accordingly, a theoretical idea for automatic spectrum identification and solution for X-ray fluorescence spectrometers is provided based on the discussion on the aforementioned two methods, and a basis for further development of faster and more accurate qualitative analysis method is built.

References

[1] Zhang L, Dai T 2008 The present status of energy dispersive x-ray fluorescence spectrometry *Modern instrument* **5** 50-5 (*in Chinese*)
[2] An G 2002 Development of wavelength dispersive X-ray fluorescence analysis *Modern instrument* **4** 40-3 (*in Chinese*)
[3] Guo P, Qin D, Hu Z, Du W 2003 A Novel Method for Spectral Signal Pattern Recognition *Spectroscopy and Spectral Analysis* **l23**(4) 811-5 (*in Chinese*)
[4] Yang L, Zhang B, Ye X 2004 Fast Fourier transform and its applications *Opto-Electronic Engineering* **31** 1-7
[5] Xiong Y, Wen Z, Chen G, Huang J 2005 Spectral Signal Recognition Based on Wavelet Packet Analysis *Spectroscopy and Spectral Analysis* **25**(8) 1332-5
[6] Wu B, He L, Cai Z, Wu Y 1993 Feature Extraction and Recognition of Waveform Based on Wavelet Transform *J Huazhong Univ of Sc. & Tech* **21**(1) 77-81

[7] Guo T, Wu C, and Qu D 2004 Wavelet Transform Theory and Its Application Progress: a Review *Information and Control* **33**(1) 67-71
[8] Mallat S, Hwang W L 1992 *IEEE Trans on Information Theory* **38**(2) 618-43
[9] Wang R, Yang S, Yan B 2007 A Review Of Mineral Spectral Identification Methods and Models with Imaging Spectrometer *Remote Sensing for Land& Resources* **1** 1-9 (*in Chinese*)
[10] Song Y, Zeng W-m, Liu G-j, Chang Z-y *2012* Development of X-ray Fluorescence Spectrometer for Determination of Tc *Annual Report of China Institute of Atomic Energy* **164** (*in Chinese*)
[11] Lai MS, Xiang LW, Lin J-M, Li H-F 2013 Quantitative analysis of elements (C,N,O,Al,Si and Fe) in polyamide with wavelength dispersive X-ray fluorescence spectrometry *Science China (Chemistry)* 1164-70 (*in Chinese*)
[12] Yokoi K, Ohashi H, Tochio T, Ito Y, Shoji T, Matsuno T 2006 High resolution X-ray fluorescence spectrometry for water analysis of metals in East China Sea *Chinese Journal of Geochemistry* **S1** 152-3 (*in Chinese*)
[13] Schipp F, Wade W R 1998 Fast Fourier Transforms On Binary Fields *Approximation Theory and Its Applications* **1** 91-100
[14] Novel algorithm for complex bit reversal:employing vector permutation and branch reduction methods *Journal of Zhejiang University(Science A:An International Applied Physics & Engineering Journal)* 2009 **10** 1492-9 (*in Chinese*)
[15] Li S-h, Xian J 2013 A multiscale Galerkin method for the hypersingular integral equation reduced by the harmonic equation *Applied Mathematics:A Journal of Chinese Universities(Series B)* **1** 75-89 (*in Chinese*)
[16] Ma CM, Chen H, Yu JY, Long T 2013 A novel conflict-free parallel memory access scheme for FFT constant geometry architectures *Science China(Information Sciences)* **04** 180-8 (*in Chinese*)
[17] Han WQ, Zhou JY 2013 Acoustic emission characterization methods of damage modes identification on carbon fiber twill weave laminate *Science China(Technological Sciences)* **9** 2228-37 (*in Chinese*)
[18] Arumugam V, Kumar S.B, Santulli C, Stanley A J 2011 Effect of fiber orientation in uni-directional glass epoxy laminate using acoustic emission monitoring *Acta Metallurgica Sinica (English Letters)* **5** 351-64 (*in Chinese*)
[19] Comparison between Windowed FFT and Hilbert-Huang Transform for Analysing Time Series with Poissonian Fluctuations: A Case Study *Chinese Journal of Astronomy and Astrophysics* 2006 **4** 503-12 (*in Chinese*)
[20] Asamwar R S, Bhurchandi K M, Gandhi A S 2010 Interpolation of Images Using Discrete Wavelet Transform to Simulate Image Resizing as in Human Vision *International Journal of Automation & Computing* **1** 9-16
[21] Luo M-J 2007 Application Of Wavelet Transform On Diagnosis And Prediction Of Milling Chatter *Chinese Journal of Mechanical Engineering* **3** 67-70 (*in Chinese*)
[22] Sclabassi R J, Hsu J T, Li C C 2007 A Wavelet Transform Based Pocs Superresolution Algorithm *Journal of Electronics(China)* **5** 642-8 (*in Chinese*)
[23] Wu W 1985 Applications Of Orthogonality Theory Of Discrete Normal Modes To The Identification Of Free Oscillation Eigenfrequencies And Determina Tions Of Spectral Peak Positions And Spectral Line Widths *Science in China* **Ser.B 06** 641-54 (*in Chinese*)
[24] Guo L, Moore J B 1991 Stochastic System Identification Via Adaptive Spectral Factorization *Systems Science and Mathematical Sciences* **03** 275-88
[25] Boumaiza S, Ghannouchi F M 2007 Spectral Methods for Accurate Identification and Quantification of Memory Effects of Wideband RF Power Amplifiers *Chinese Institute of Electronics(CIE). Proceedings of 2007 5th International Conference on Microwave and Millimeter Wave Technology[C].Chinese Institute of Electronics(CIE)* **4**
[26] Kozánek J, Pust L 2011 Spectral properties and identification of aerostatic bearings *Acta Mechanica Sinica* 01 63-7 (*in Chinese*)
[27] Guo Y 2001 Parameter Identification By Mixed Spectral-Pseudospectral Approximations *Acta Mathematicae Applicatae Sinica (English Series)* **02** 218-32 (*in Chinese*)
[28] Balachandar V, Lakshman K B, Sasikala K, Manikantan P, Sangeetha R, Mohana D S 2007 Identification of a high frequency of chromosomal rearrangements in the centromeric regions of prostate cancer patients *Journal of Zhejiang University(Science B:An International Biomedicine & Biotechnology Journal)* **09** 638-46 (*in Chinese*)
[29] Zheng W, Shao J, Wang M, Huang D 2013 A Thin Cloud Removal Method from Remote Sensing Image for Water Body Identification *Chinese Geographical Science* **04** 460-9 (*in Chinese*)
[30] Titulaer M K 2013 Candidate Biomarker Discovery for Angiogenesis by Automatic Integration of Orbitrap MS1 Spectral- and X!Tandem MS2 Sequencing Information *Genomics, Proteomics & Bioinformatics* **03** 182-94
[31] Yao Z, Song Y, Yang J, Gao Q 2012 Spectrum Matching Analysis of Chaotic Binary Modulation Excitation Sequence for Sonar Systems *IEEE Beijing Section.Proceedings of 2012 IEEE 11th International Conference on Signal Processing (ICSP 2012) IEEE Beijing Section* **4**

Authors

Jiening Xia, born in June 27, 1982, NingXia, China

Current position, grades: research assistant, master Institute of Seismology.
University studies: Wuhan University, Institute of Seismology, China Earthquake Administration.
Scientific interest: electronic engineering, photoelectric technology, instrumentation and algorithms.
Publications: 6.

Zhigao Chen, born in April 15, 1975, Wuhan Hubei, China

Current position, grades: the president at the Wuhan Institute of Earthquake Science Instruments.
University studies: associate professor, master.
Scientific interest: measurement and control technology and instrument.
Publications: 5.

Fine detection of mine collapse column by anti-explosive ground penetrating radar

Fan Cui[1]*, Zhiyuan Wu[2], Jia Chen[2], Lei Wang[2]

[1]*State Key Laboratory of Coal Resources and Safe Mining, China University of Mining Technology, Beijing*

[2]*College of Geoscience and Surveying Engineering, China University of Mining Technology, Beijing*

Received 1 October 2014, www.cmnt.lv

Abstract

This work discussed the technique for fine detection of mine collapse column by anti-explosive ground penetrating radar (GPR) and its feasibility. In stage of theoretical research, numerical forward modelling was used to study the response characteristics of collapse column on radar profile. In actual exploration, the spatial distribution pattern of collapse column was accurately stored by installing complex environment observing system in the pit. In processing and explanation, the disturbance types of mine were comprehensively analysed, and various measures were taken in the suppression of interference signals to obtain better exploration effects. The study showed that fine detection of collapse column can be realized by anti-explosive ground penetrating radar under the complex conditions in the pit.

Keywords: Anti-explosive ground penetrating radar (GPR), collapse column, fine detection

1 Introduction

Mine collapse column is one of the key problems for mine production [1]. Collapse column is generally formed by sinking of overlying rock. Such sinking results from the cavity, which is caused by the underground water corrosion of soluble rock underlain in coal measure strata. If the collapse column without solid filling is connected with runoff belt in aquifer of mining area with rich underground water, water burst and even inundation may possibly occur when work surface is directly exposed or gets close [2].

The detection of collapse column in the pit is mainly by drilling and geophysical prospecting. Drilling is perceptual but limited by the layout condition of drill hole, so it cannot completely reflect the distribution of collapse column [3]. With advantages of area detection, geophysical prospecting is widely applied to obtain obvious geological effects. The main methods include seismic detection, direct current electric method, transient electromagnetic, and pit penetration. However, it is difficult to arrange effective observing system in the complex and narrow space of mine when using seismic detection, direct current electric method, etc. Therefore, the exploration completeness and accuracy of collapse column will be affected [4, 5]. Moreover, geophysical prospecting is also limited by the gas environment in the pit. Equipment not reaching anti-explosive safety grade cannot be used in the pit.

Anti-explosive GPR can safely work in the pit. With small volume, light weight, and collection method of point detection or continuous detection, it can flexibly arrange the survey lines according to specific environment. Moreover, the screen-type design can effectively shield interference in the pit [6]. Meanwhile, featured with high resolution and efficiency of common GPR, it can finely detect the spatial distribution of collapse column [7]. This work mainly introduced technology methods and application effects about the detection of mine collapse column by anti-explosive GPR.

2 Theory of GPR

The concept of GPR was proposed by Germany Letmbatch and Lowy in 1910 at the earliest. It can detect the target by emitting high-frequency impulse electromagnetic (in frequency scope of 106~109Hz) to the medium [8]. Mediums with different relative dielectric constants have different reflections of electromagnetic wave, and its factor r can be expressed as follows:

$$r = \frac{\sqrt{\varepsilon_1} - \sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}}, \tag{1}$$

where ε_1 and ε_2 represent relative dielectric constants of medium 1 and medium 2, respectively. The distribution scope of the constant is 1 (air) < ε <81 (water). The larger the difference between relative dielectric constants, the stronger the reflection energy of electromagnetic wave will be.

Figure 1 show the principle that the GPR emits high-frequency electromagnetic pulse to the medium. The travel time of impulse wave can be expressed as follows:

$$t = \frac{4z^2 + x^2}{v}, \tag{2}$$

where, t is the two way travel time of electromagnetic wave, z is reflector depth, v is the wave speed in underground medium, and x is the distance between transmitting antenna and receiving antenna (distance between receiver and transmitter).

**Corresponding author* e-mail: cuifan1984@foxmail.com

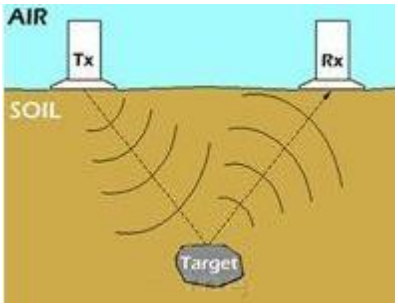


FIGURE 1 Schematic diagram for the principle of ground penetrating radar (GPR) detection

This study used ZRT intrinsically safe GPR developed by China University of Mining and Technology (Beijing). It can be safely used in the gas environment of pit. With central frequency of 180MHz, this radar antenna adopted screened antenna "integrating with receiving and transmission". Compared to the exploration depth *z*, the distance between receiver and transmitter can be deemed as $x \approx 0$ for the close distance between receiving antenna and transmission antenna. Therefore, the two-way travel time of impulse can be expressed as follows:

$$t = \frac{2z}{v}, \tag{3}$$

where wave speed *v* is related to the relative dielectric constant of medium. The larger the relative dielectric constant of medium, the slower the radar wave will transmit in the medium. The relational expression between both is as follows:

$$v = \frac{c}{\sqrt{\varepsilon}}. \tag{4}$$

When the speed of electromagnetic wave in medium *v* is known, the depth of reflection target *z* can be obtained by Equation (3) according to the measured accurate time *t*.

When transmitting in the medium, the properties of electromagnetic wave, including amplitude, phase, and waveform will change with the electromagnetic property and geometrical morphology of medium. Therefore, the structure and nature of medium can be deduced according to the received data about travel time, amplitude, and waveform of radar wave. Radar detection reflects the measured medium with profile, which is the record for the form of reflection wave. The positive and negative energy are represented with black and black, or the gray and color scale, respectively. Compared with other geophysical prospecting, GRP can more intuitively reflect the detection results, and can make preliminary reference and judgment at site.

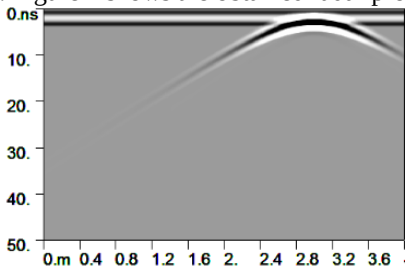
3 Geophysical characteristics of collapse column

Collapse column of coal measures is generally caused by the continuous upward collapse in large-span solution cavity of soluble rock series underlies in coal measures. There is a large difference between electrical property of coal measures and that of the rock mass in collapse column in coal measure strata. Therefore, the radar reflection wave with strong energy can be formed in the boundary of collapse column. In general, the development of collapse can be divided into 4 stages: early development stage, strong development stage, decline stage, and dead stage [9]. For the collapse column in different development stages, the filling degree of inner cement, water transmissibility and abundance, the damage status of surrounding rock distribution, and the associate structure are all completely different. Therefore, their geophysical characteristics are different in electric property.

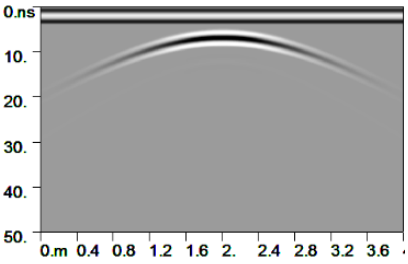
In general, collapse column is elliptic cylinder or oblate cylinder in space, and irregular cylinder in profile [9]. The dielectric constant of collapse column is obviously different with that of surrounding rock stratum and coal bed. Moreover, with water or air in marginal area, the difference of dielectric constant and conductivity may surge to enhance the reflection energy of radar signal. Although the reflection energy is strong, the distribution of wave group is in disorder because of the complex binding form between collapse column and its margin and surrounding rock [10].

In order to study the radar respond characteristics of collapse column, finite difference time domain was adopted to simulate the radar wave generation and the transmission in coal and rock medium.

Firstly, a collapse column with elliptic cross section was designed, and its long axis was 3.2 meters and short axis 1.8 meters in horizontal. Then, the collapse column was detected by simulating GPR penetrating the lateral wall of lane with survey lines parallel to long and short axis were laid in horizontal. Figure 2 shows the obtained radar profile.



a) Simulated effects of radar profile when survey line parallel to the short axis of collapse column in horizontal



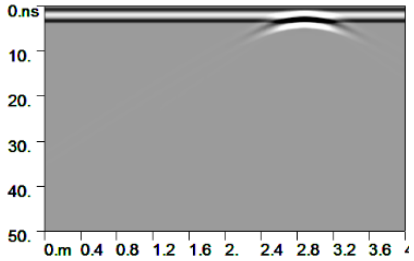
b) Simulated effects of radar profile when survey line parallel to the long axis of collapse column in horizontal

FIGURE 2 Simulated radar profile of collapse column in horizontal section

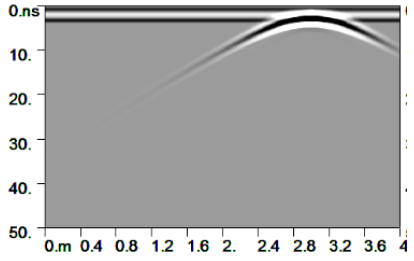
In simulated profile, the response of detected collapse column parallel to long axis should be obviously wider than that to short axis. Therefore, the distribution for the long and short axis of collapse column can be differentiated according to this characteristic.

In addition, the contrast simulation was conducted for the change of collapse column conductivity and dielectric constant caused by different moisture contents. Water may increase the conductivity and dielectric constant of collapse column [11]. Figure 3 shows two respond profiles of collapse column with different conductivities and dielectric constants. In this figure, Figure 3a shows the non-aqueous condition that the conductivity is 10-7S/m and dielectric constant is 7. Figure 3b shows the aqueous condition that the collapse conductivity is 10-2S/m and dielectric constant is 14. According to the radar profile, the response energy of aqueous collapse column is obviously higher than that of non-aqueous collapse column [12].

Furthermore, the contrastively simulations detection of collapse column has been used according to the change of conductivity and permittivity of collapse column for different moisture content as well. The conductivity and permittivity of collapse column will increase because of the water in medium, so in the radar response profile of collapse column of two group different conductivity and permittivity in the Figure 3, Figure 3a shows the conductivity of collapse column is 10-7S/m, and the permittivity is under no water circumstance value 7. Figure 3b shows there contain water in the medium, the conductivity of collapse column is 10-2S/m, the permittivity of it is 14. The response energy of collapse column of containing water is significantly higher than collapse column of no water according to the radar profile [13].



a) Profile of non-aqueous collapse column



b) Profile of aqueous collapse column

FIGURE 3 Comparison chart for simulated profiles of collapse column with different aqueous conditions

According to the forward modelling radar profile, characteristics of collapse column were judged in aspects of layout position, form, water containing, etc.

4 Actual detection

There is a collapse column with 73m long axis and 45m short axis between main rail haulage roadway and air return roadway in the primary mineable coal bed in Shanxi Province. As shown in Figure 4, the exploration of air return roadway keeps away from this collapse column by bypassing up to the mountain. Coal faces 8102 and 8103 are located in the north to collapse column.

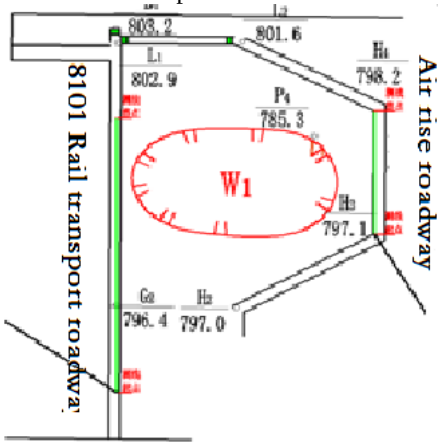


FIGURE 4 Schematic diagram for survey lines of 8101 rail haulage roadway

4.1 DATA COLLECTION

This mine is a newly-established shaft without detailed geological data, so two main radar survey lines were laid for the lateral walls of two roadways neighboring to collapse column after knowing the surrounding rock in transportation roadway and air return way. As shown in Figure 5, two survey lines of lateral walls along crossheading of rail are 110m long, while those along air return roadway are 50m long. As shown in Figure 6, two short survey lines were set in lateral wall and head on work faces 8103 and 8102, which were 18m and 19m long.

This study used ZRT intrinsically safe GPR developed by China University of Mining and Technology (Beijing). Its central frequency is 180MHz. In order to reach 30m detection depth, the time window for collection was set as 700ns, and the sampling point number as 1024. Detection was conducted by continuous collection for 4 times.

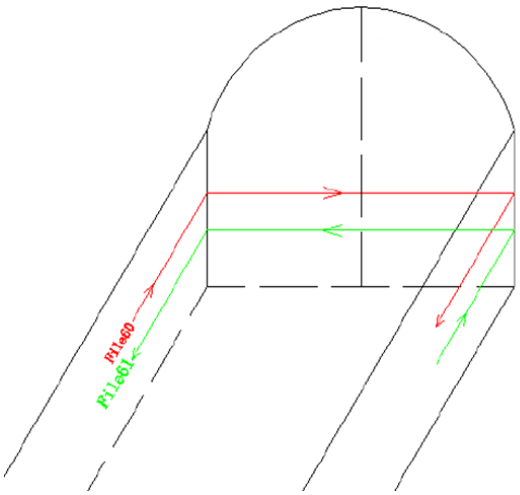


FIGURE 5 Schematic diagram of radar survey line layout

4.2 DATA PROCESSING

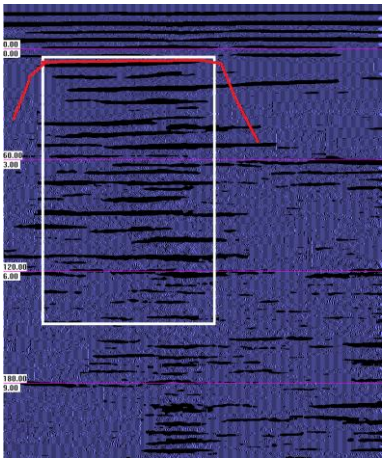
Self-developed GR processing and explaining software was adopted in the processing of collected data. In regular processing, gain adjustment, one-dimensional filtering, FK filtering, and wavelet transform were used. In addition, signal-to-noise ratio can truly reflect the change of detected medium and ensure data quality [14].

4.3 EXPLANATION OF REGULAR DATA

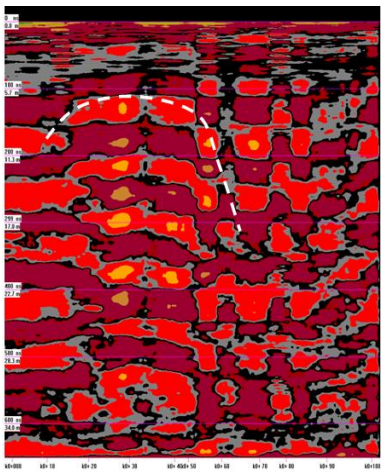
GPR detects anomalous body with the variance of electromagnetic property. GPR analyses the variance between collapse column boundary and surrounding mediums and the reflection of collapse column on radar profile according to the geological data. Then, the distribution scope of collapse column can be determined. The time profile of radar is mainly identified according to the changes of reflection energy on different time axis because there is a large difference of electric property between collapse column boundary and surrounding rock mass. Non-aqueous collapse column may generate disordered reflection waves which are superposed after multiple reflections under the obvious reflecting layer. If the collapse column is aqueous and water-transporting, however, the radar wave energy may be adsorbed and reduced under the reflecting layer of strong margin, while the reflection energy layer may be enhanced. In addition, the depth of abnormal position can be judged by reading the reflection time of abnormality distributing on radar profile according to the relevance between transmission speed of electronic wave in medium and relative dielectric constant of medium [15].

1) Figure 7 shows the detection results of radar in the rail haulage roadway. In the profile detected in rail haulage roadway, there is an obvious lithological variance belt during 100ns-250ns. According to geological data, it can be regarded as the boundary affected by collapse column. The surrounding rocks of haulage roadway are mainly mud rock of Taiyuan Fm in carboniferous system, and their relative dielectric constant is in scope of 8-15. Therefore, the boundary affected by collapse column can be estimated according to the formula of radar wave speed and the equation of time interval during radar wave transmission. This boundary is about 5-12m away from the lateral wall of rail haulage roadway, and its form distribution in plane conforms to the circular arc of collapse column boundary. Moreover, the reflection clutters are still disordered in the lower part of time axis on profile, and the energy does not generate strong absorption and reduction. Therefore, it can be determined that this collapse column is not water-transporting.

2) Figure 7 shows the radar detection results of air return roadway up to the mountain. The air return roadway up to the mountain is located in the east to collapse column. In the radar time profile, strong reflection energy layer can be seen during 120ns-260ns. Therefore, it can be determined that this roadway is in the boundary affected by collapse column. The roadway is about 7-15 meters away from the lateral wall, and has the boundary shape of collapse column. However, the energy on and below reflecting layer obviously enhances than the profile detected in rail haulage roadway before. According to the analysis on radar response characteristics in forward simulation, the moisture content in the collapse column near the air return roadway is higher than that near the rail haulage roadway.



a) Profile of waveform variable area

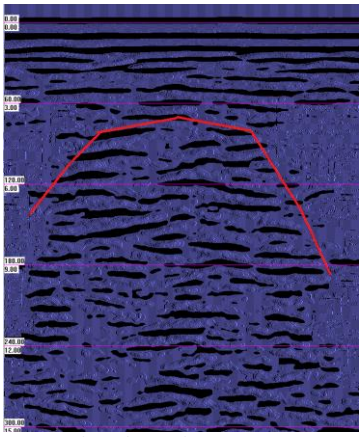


b) Colorful profile

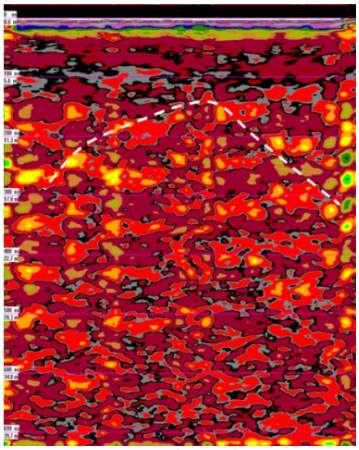
FIGURE 6 Radar detection profile of collapse column in rail haulage roadway

According to the comprehensive analysis on the radar profile of two survey lines, this collapse column has moisture change in partial region, but does not been connected. Therefore, it can be determined that the collapse column is not water-transporting in horizontal.

3) Radar detection results in other regions are as follows. There is anchor network, steel-frame shed, and support plank in the most detection position of lateral wall and frontal plane on 8103 work face. Therefore, the coupling with surrounding rock is bad, and the interference signal is strong during radar detection, which affects the detection results to some extent. After a series of signal processing, the signal-to-noise ratio increases. However, there is no obvious reflecting layer on radar profile when detecting the frontal plane, while an obvious energy reflecting variance belt is found when detecting the lateral wall (See Figure 8). It can be judged as the reflection caused by collapse column. However, the survey line at this section is short, so the judgment of position distribution may have large error.

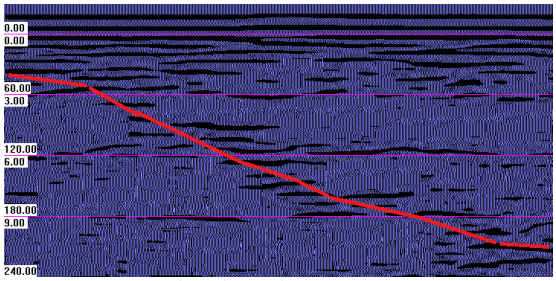


a) Profile of waveform variable area

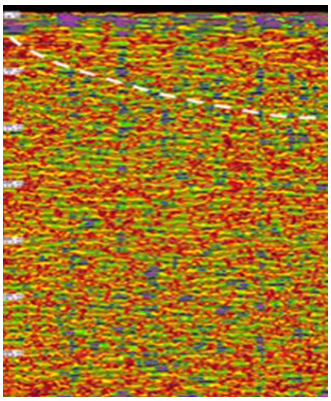


b) Colorful profile

FIGURE 7 Radar detection profile of collapse column in air return roadway up to the mountain



a) Profile of waveform variable area



b) Colorful profile

FIGURE 8 Schematic diagram of reflecting energy variance belt in lateral wall on work face

According to all above results, the basic form of collapse column distribution is determined by radar detection on the horizontal plane surrounded by the lateral wall, rail haulage roadway, and air return roadway up to the mountain on 8103 work face. In order to verify the detection accuracy of GRP in the pit, the explanation achievements are compared with original geological data of tunnel prospecting and drilling. It can be found that the collapse column conforms to the distribution in horizontal. In Figure 9, the area circled with green imaginary line is the distribution scope of collapse column in horizontal determined according to radar profile. And the area circled with red active line is the distribution scope of collapse column provided in the geological data of this mine. These two areas have high superposition if the survey lines are laid long enough to ensure the detection accuracy.

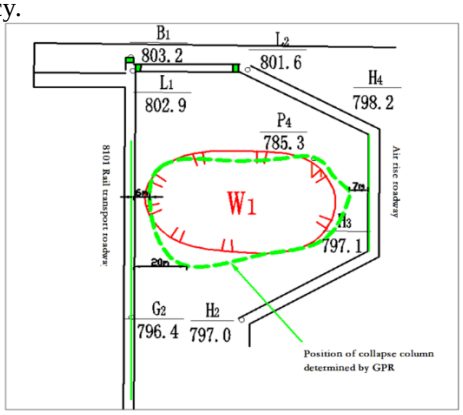


FIGURE 9 Comparison diagram between collapse column positions in radar detection and geological data

5 Conclusions

Theoretical analysis and actual engineering study are conducted on the detection of collapse column in the pit with anti-explosive GPR. Scientific and rational layout of survey lines and utilization of limited space in complex mine environment are significant to obtain radar data with high signal-to-noise ratio.

With rational and sufficient layout of survey lines, the anti-explosive GPR can reflect the spatial distribution of collapse column and moisture and water-transporting of collapse column. Moreover, effective layout of survey lines can be used to accurately determine the distribution scope of collapse column in horizontal.

Practices show that anti-explosive GPR has advantages of high resolution, flexibility, lossless, anti-interference, etc. when detecting the collapse column in the pit. Therefore, it is incomparably advantageous to other geophysical methods. Except collapse column detection, anti-explosive GPR in the pit can also be applied to safety detection of other disaster sources in the mine pit.

Acknowledgement

Funded by independent research project of international key laboratory of coal resources and mining safety (2013CRSMZZ03); Key project of National Natural Science Foundation - Early warning for roof disaster in mine (51234005).

References

[1] Hou M 2013 Simlpe analysed the cause of collapse column, the influence of collapse column on the coal mining and handing practice *Friend of Science Amateurs* **7**(2) 144-5 (*in Chinese*)
[32] Liu G 2003 Geophysical technology applied in the detection of karst collapse columns *Coal Engineering* **8** 4-6 (*in Chinese*)
[33] Liang Q 2011 Underground penetrating radar (GPR) in underground coal mine fast advanced detection and data analysis *Geophysical And Chemical Computing Technology* (5) 531-5 (*in Chinese*)
[34] Li Y, Peng S 2006 The classification and characteristics of karst collapse column in north China coalfields *Coal Geology & Exploration* (4) 53-6 (*in Chinese*)
[35] Song J 2005 The research of detection technology in using underground penetrating radar (GPR) in underground coal mine *Chongqing Branch of China Coal Research Institute* (*in Chinese*)
[36] Yang F, Peng S 2010 The research on principle and method of geological radar (*in Chinese*)
[37] Su H, Zhao Y, Wen J 2005 Application of ground-penetrating radar in Wu Long tunnel's karst and underground river detection *Journal of Shi Jia Zhuang Rail Way Institute* **18**(1) 54-5
[38] Zhang L, Liu C, Wu K 2013 Study on application of ground penetrating radar in Karst Collapse Column detection *Subgrade Engineering* **32**(4) 130-1
[39] Qu S, Zhang J, Zhang H 2014 Application of ground penetrating radar measuring technology in coal mine *Safety In Coal Mines* **45**(2) 60-1 (*in Chinese*)
[40] Lin J 2013 Two-dimensional filtering technology in geological radar detection application of hidden hazard in coal mine *Technological Development of Enterprise* **32**(10) 6-7
[41] Li F, Liao G, Liu X 2013 Application of mine transient electromagnetic method in coal mine collapse area detection *Safety In Coal Mines* **44**(4) 156-7
[42] Yin Q, Pan D, Yu J 2012 Research on seismic identification technique of coal mine collapse column *Progress in Geophysics* **27**(5) 2168-9
[43] Yu Y, Shang Y 2014 Comprehensive detection technology application with TongXin coal mine collapse column *Modern Mining* **31**(4) 63-4
[44] Yang L 2012 Using the ground penetrating radar (GPR) to detect mines FenXi abnormal geological structure of mining area *Mining Safety and Environmental Protection* **39**(5) 83-4
[45] Yang W 2013 Integrated geophysical prospecting principle and method of exploration the water enrichment of coal mine collapse column *Journal of Mining & Safety Engineering* **30**(1) 45-6

Authors

Fan Cui, born on January 3, 1984, Anhui Province of China

Current position, grades: lecture of geophysics of application at China University of Mining and Technology, Beijing.
University studies: Ph.D. degree in geophysics from China University of Mining Technology, Beijing in 2012.
Scientific interest: ground penetrating radar (GPR) in underground detection and other fields.
Publications: 7 papers, 3 patents.

Zhiyuan Wu, born on November 1, 1987, Henan Province of China

Current position, grades: Doctoral students second grade in China University of mining &technology (Beijing).
University studies: master's degree in Mineral resources survey and exploration from China University of Mining Technology, Beijing in 2013.
Scientific interest: ground radar detection of shallow strata of moisture content.
Publications: 1 paper.

Jia Chen, born on January 1, 1989, Sichuan Province of China

Current position, grades: graduate student, grade 2, in China university of mining &technology (Beijing).
University studies: Bachelor's degree from China University of Mining Technology, Beijing in 2013.
Scientific interest: ground penetrating radar detection of shallow strata structure.

Lei Wang, born on October 15, 1992, Shanxi Province of China

Current position, grades: graduate student, grade 2, in China university of mining &technology (Beijing).
University studies: Bachelor's degree from China University of Mining Technology, Beijing in 2013.
Scientific interest: ground penetrating radar detection of shallow strata of moisture content.

Analysis of thermal-mechanical coupling and structural optimization of continuous casting roller bearing

Disi Chen[1], Gongfa Li[1, 2], Honghai Liu[2]*, Guozhang Jiang[1], Jia Liu[1], Ze Liu[1], Weiliang Ding[1], Wei Miao[1], Zhe Li[1]

[1]*College of Machinery and Automation, Wuhan University of Science and Technology, Wuhan 430081, China*

[2]*Intelligent Systems & Robotics Group, School of Computing, University of Portsmouth, Portsmouth, PO1 3HE, United Kingdom*

Received 1 October 2014, www.cmnt.lv

Abstract

The continuous casting roller bearing was one of the important parts of continuous caster [1], its working temperature was relativity high, the working load was also very large and the working condition was complex. Since the external heat of casting roller bearings was much higher than internal, the continuous casting roller bearing was different from the general high-speed rolling bearing with heavy load. If the external heat that the bearing suffered could not dissipate in time, the extreme high working temperature might accelerate the failure of the bearing and severely reduce the productivity of the caster. To optimize the thermal structure of the continuous casting roller bearing, the thermal coupling analysis of the bearing should be conducted. Firstly, the stress field of the working continuous casting roller bearing should be analysed by ANSYS and the three-dimensional geometric CAD model and the CAE model of it should be established. Then find out the location where the bearing suffered the largest force by stress analysis, in the case of bearing block with cooling water, load the working temperatures to the bearing, after analysing, the result showed that the external temperature was the main contributor of internal stress, and the rolling element was the part inside of the bearing which suffered the largest stress. To optimize the structure of the continuous casting roller bearing, the rolling element was turned into axial hollow structure, which could reduce the extreme stress of the rolling element bearing. By analysing all the maximum thermal stress of rolling element in different feasible hollowness, and finally the most suitable hollowness could be found out. The analysis results above showed that the optimization of the rolling element structure of the bearing could effectively reduce the internal stress and improve reliability of bearings in high working temperature. The conclusion of this study was significant in bearing optimizing or designing.

Keywords: thermal stress, coupling, structural optimization, continuous casting roller

1 Introduction

Casting roller bearing [2] was a kind of rolling bearing, the motion of internal rolling bearing was quite complex, since the bearings could be easily damaged in high temperatures, and the temperature analysis and stress analysis of continuous steel billet and continuous caster itself were quite complex, it was too hard to give a theoretically description or analysis of the bearing. There had been many experts and scholars from China or foreign countries who contributed themselves in finding the cause of failure of the bearing on the basis of researching its temperature. But they almost tend to study the complex internal heat which generated by heavy load and high-speed operation, but there was still few studies under the situation of external heat generation higher then internal. The analysis of rolling bearing temperature field is mainly based heat transfer theory [3]. Analyse and calculate the distribution of bearing temperature field during the heat transfer process by using a combination of theory and simulation methods, the parts which had risk to failure and the steady temperature field without lubricating oil could be predicted during the steady process of bearing. The analysis of the rolling bearing temperature field was mainly based on heat transfer theory; researchers could only study it by actual measuring through the site and testing in the laboratory, which largely relies on the experience accumulated by long-term of work, and these methods still had practical significance on today's studies. However, with the development of iron and steel smelting technology and computer science, traditional experimental methods had shown a lot of drawbacks, which unable to give an accurate result in practical; therefore they could not meet the current needs of the development of steel industry. At present, many research institutions and production enterprises, both here and abroad, applied the finite element techniques to metallurgical equipment design and optimization studies, analysed and calculated the distribution of bearing temperature field during the heat transfer process by using a combination of theory and simulation methods, predict the parts which had risk to failure and the steady temperature field without lubricating oil during the steady process of bearing. They had made a lot of phase achievement. Therefore, conducting mathematical simulation by using the finite element analysis software had become an important means of future research work.

**Corresponding author* e-mail: honghai.liu@port.ac.uk

2 The construction of finite element model of the continuous casting Roller

2.1 THREE-DIMENSIONAL MODEL OF THE CONTINUOUS CASTING ROLLER

The relationship feature of casting rolling bearing model was very complex which was not easy to use the self-modelling, so the mesh division should be use finished in professional software. The methods that most researchers used are: Firstly, use certain CAD drawing software (such as SOLIDWORKS, PROE, UG, etc.)to transform two-dimensional engineering drawing into the overall structure three-dimensional drawings, and then export to the file (such as iges, step, etc.) that CAE pre-processing software could recognize, finally introduce the model into CAE pre-processing software and make necessary topologies corrections and mesh division [4], the whole CAE model structure could be eventually finished. Since exporting in STEP format might always lost key feature of the model, in this paper, IGS format was used to export the model after finished it with three-dimensional graphics software, for better topology correction and mesh division in CAE pre-processing software. According to the bearing models and drawings provided by a steel mill, use the popular 3D CAD software to build a three-dimensional CAD assembly model of the whole casting rolling bearing (as shown in Figures 1 and 2).

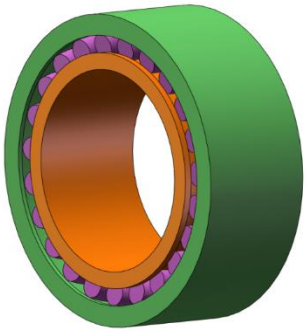


FIGURE 1 Three-dimensional CAD model of the continuous casting roller

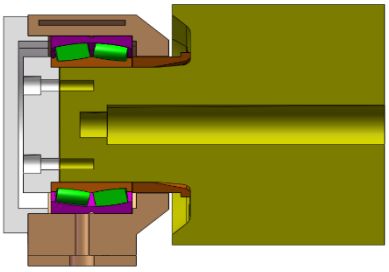


FIGURE 2 Three-dimensional CAD assembly model of the continuous casting roller (section view)

All parts were sat by the way of parametric feature modelling. Firstly, draw sketches of every part, and then generate solid parts according to the shape of each part by stretching, rotating, excision, etc. In sketching and feature modelling, the dimension of three-dimensional geometric model of the part should be set strictly in accordance with the engineering drawings and technical requirements. Every position of parts in the assembly should be located follow the coaxial, overlapping surfaces, parallel with the distance and other different assembly principals. In the assembly process, all the connections in the assembly should be joined in full accordance with requirements of the drawings and specifications of the mechanical coupling.

2.2 BUILD THE FINITE ELEMENT MODEL OF THE CONTINUOUS CASTING ROLLER

In this paper, the mesh model was built by using CAE pre-processing simulation software, in strict accordance with the dimensions and structural characteristics of the three-dimensional CAD geometry model.

This paper mainly considered the impact of load on the bearing, so the commonly used SOLID45 unit should be chosen in the stress analysis. According to the characteristics of the bearing structure, solid elements were used in the meshing process. The model should be reasonable divided to ensure the accuracy of simulation results with the number of grid as less as possible. The model should be divided into mesh after carefully considering, with a large amount of grids, although the computing result could be more accurate, it might also take higher computer resources and a lot more time to finish this simulation. In meshing for bearing stress analysis, the number of grids could be reduced reasonably. As for the density of grid, using the uniform grid, could not only ensure the reasonable amount of meshes, but also avoid the unexpected error caused by the mutations of grids.

Figure 3 was a three-dimensional mesh model of the casting roller, which was used for thermal analysis of the bearing; Figure 4 is a partial mesh structure of bearing caster roller. In order to better simulate the actual working conditions, the assembly mesh model of casting roller bearing was shown in Figure 5. Meshing assembly model was the most difficult task, because of the water cooling pipe inside of it [5], the internal meshing should be divided as well, the tetrahedral and pentahedral mesh were also been used where the shape was extremely irregular. Mesh quality of this site is not as high as hexahedral mesh but the number of tetrahedral mesh or pentahedral mesh grid were little, the impact of the entire model was negligible.

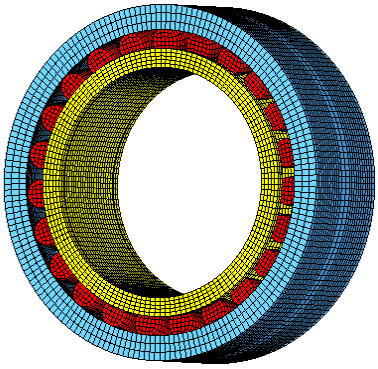


FIGURE 3 The three-dimensional mesh model of the continuous casting roller bearing

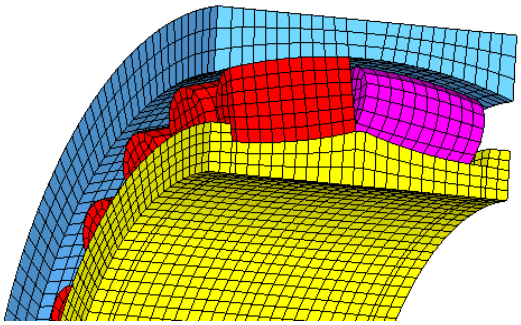


FIGURE 4 The part of three-dimensional mesh model of the continuous casting roller bearing

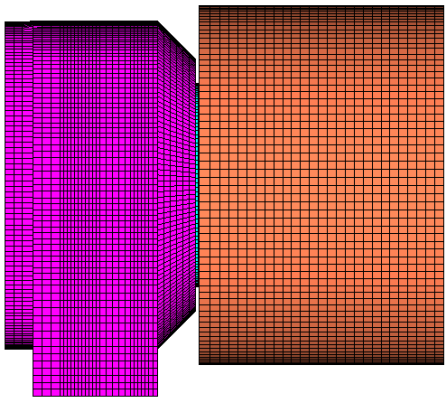


FIGURE 5 The three-dimensional mesh model of the continuous casting roller bearing assembly

3 The continuous casting roller bearing stress field analysis

3.1 MECHANICAL ANALYSIS OF THE CONTINUOUS CASTING ROLLER

3.1.1 The mechanical analysis of the continuous casting roller

In calculating the force [6], the bulging force of steel billet should be considered, the bulging force caused by static pressure of molten steel should be calculated by the mechanical model which based on a certain pair of static pressure between molten and continuous casting roller. As for the force on the nip rolls, it should be calculated follow the theory of material mechanics, such as the equal cross-section beam with both ends fixed and a uniform distributed load (static pressure of molten steel). The calculation of the force on *i*-th roller:

$$F_i = \rho g h_i l_i \left(w - 2k\sqrt{\frac{L}{V_c}} \right), \tag{1}$$

where ρ is the density of liquid steel, the units are kg·m^{-3}. h_i is the relative height of *i*-th pair of rollers to the steel surface, the units are m. l_i is the spacing between *i*-th pairs of roller, the units are mm. w is the width of the billet, the units are mm. L is the arc length of the billet from the mold level, the units are mm. V_c is casting speed, the units are m·min^{-1}. K is the integrated coagulation factor, the units are mm·min$^{-1/2}$.

In the sector segments, when a light dynamic pressure is given to the billet, these segments would get a reacting force caused by the deformation of billet. Therefore, the force could be calculated by flat rolling theory, when the soft-reduction process was conducted on the rollers. As shown in the Equations (2) and (3):

$$P = \sigma 2\delta\sqrt{2R\Delta h - \Delta h^2}\ , \tag{2}$$

$$\Delta h = Ql\ , \tag{3}$$

where σ is the average pressure on unit rolling, the units are t·mm^{-2}. δ is the thickness of casting billet, the units are mm. R is the radius of the pressured rollers on continuous casting roller, the units are mm. Δh is the reduction of casting roller, the units are mm. Q is the reduction rate of casting roller, the units are mm·m^{-1}. l was roller distance, the units are m.

If the production model of steel billet was Q235, the width of billet was 1550mm and the casting speed was 1.2 m/min^{-1}. After calculating, the pressure of No.20 free roller in sector segments (lower roller) was 273KN. All the stress analysis in this paper was based on the 20th free roller (lower roller) casting roller.

3.1.2 The mechanical analysis of the continuous casting roller bearing

Took stress analysis of the No. 20 free roller bearing in sector segment continuous casting roller. Where there were two rollers with total two pairs of bearings, the mechanical analysis of the continuous casting roller was shown in Figure 6.

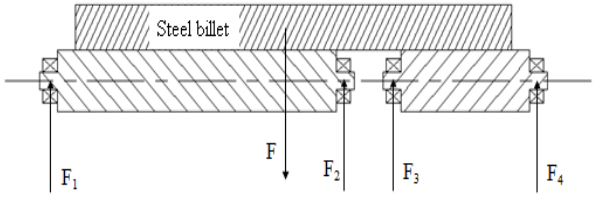


FIGURE 6 The schematic diagram of force on the continuous casting roller

F is the force between roller and casting billets. **F$_1$**, **F$_2$**, **F$_3$**, **F$_4$** are the reaction force given by the bearing.

As shown in the graph, the equation could be drawn according to the mechanical balance [7]: $F-F_1-F_2-F_3-F_4=0$.

By further calculations the following result could be drawn, $F_1=0.27F$, $F_2=0.31F$, $F_3=0.23F$, $F=0.19F$.

Then it could be drawn the maximum stress of second bearing, $F_2=0.31\times273KN=84.63KN$.

The load carrying capacity of the bearing according to the data provided by the manufacturer was shown as follow, $F_{load}=Cr\times e=394KN\times0.32=126.08KN$.

By comparison, it could be known that the supporting bearings in ideal conditions would not be overloaded.

3.2 THE BOUNDARY CONDITIONS OF THE CONTINUOUS CASTING ROLLER THERMAL STRESS ANALYSIS

The coefficient of thermal expansion between different structures that interacted or different parts of the same structure were mismatched. When temperature rose or lowed degree of expansion and contraction was inconsistent with each other, which led to the generation of thermal stress.

3.2.1 The boundary conditions of the continuous casting roller thermal stress analysis

The constraints of continuous casting roller bearing only existed in one place, it was the constraint at the bottom seat of the bearing. That was at the bottom bearing node constraint on its outer surface in radial and tangential and axial displacement of three directions.

3.2.2 The working load of the continuous casting roller thermal stress analysis

Bearing load under steady state was the force from continuous casting roller of bearing. Owing to the model involved only a part of the continuous casting roller, when the applied load needed to bending load, temperature load on the model, using APDL [8] language to read the temperature field calculation results file, the finite element software would automatically obtain the temperature value of each node.

When there was only stress simulation for stress field, the boundary conditions were the same, but it did not need to add the temperature load.

3.3 THE THERMAL STRESS ANALYSIS OF THE BEARING IN THE BEARING SEAT WITH COOLING WATER

The simulation results of maximum stress and thermal stress in each part of the continuous caster roller bearing were recorded, as shown in Table 1.The Table 1 showed that under the temperature load, the continuous casting roller bearing stress almost increased twice and the influence of the temperature on the load bearing was very obvious, and the stress was only 20.4 MPa less than the allowable stress, working under such a high stress the casting roller bearing was easy to failure and it showed that the failure of continuous casting roller was significantly related to the temperature. At the same time, it could be seen that although the temperature of the inner ring was highest, it was still smaller than rolling elements and the maximum stress and the rolling element was the part, which suffered the largest stress in the continuous casting roller bearing.

TABLE 1 The result of each part of the continuous casting roller bearing maximum stress compared with the thermal stress

	Maximum stress (MPa)	**The maximum thermal stress (MPa)**	**The allowable stress of bearing material (MPa)**
Continuous casting roller bearings roller	249.3	411.3	431.7
Outer ring	120.2	245.1	
Inter ring	176.0	321.8	
Rolling element	249.3	411.3	

4 The optimization analysis of the continuous casting roller bearing

Take the inner ring as a benchmark, the temperature of rolling elements was different in different positions of the bearing and the stress was changeable. Because of rotating of the bearing, it could be understood that the temperature changed with the rotating of rolling elements and the thermal stress was also continuously variable. And this change will lead to the fatigue damage of bearing rolling elements. The structure of the bearing rolling body should be changed to reduce the stress, so as to improve the life of the bearing.

The middle part of rolling bearing was changed into a hollow to improve the stress distribution. And the rolling body was changed into a hollow [9], the heat transfer in the heat convection and heat radiation under the same condition was relatively small. Because of the hollow, the rolling body could produce self-deformation due to offset the stress produced by thermal and quality, so, the stress of the overall rolling body decreased. And it prolong the working life of bearing.

Hollowness was used to represent the size of roller hollow. Hollowness refers to the ratio of the radius of hollow roller and actual radius of the hollow, it was an important parameter of hollow roller bearing and it must be moderate. The hollowness was too small to reflect the superiority of hollow roller or too large to reduce the bearing ability. It was not appropriate. According to the convention, the thermal stress was analysed of rolling bodies with hollow degree of 20%, 30%, 40%, 50%, 60%, 70%. The result of simulation analysis was shown in Figures 7-12.

From Figures 7-12,it could be seen that with the increase of rolling body hollow ratio, the stress of the rolling body was constantly decreasing, which once again showed that the finite element analysis was consistent with the actual situation. However, the hollow ratio of rolling body could not increase indefinitely, to ensure the bearing strength and service conditions, there should be an upper limit.

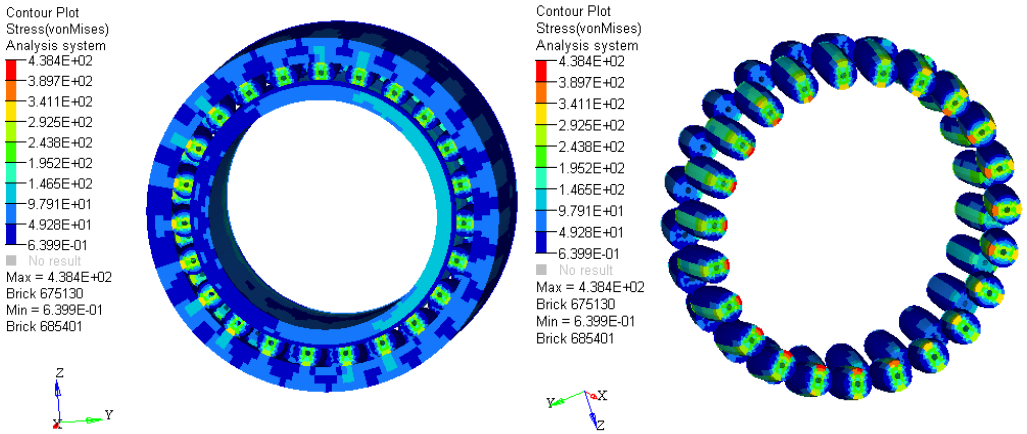


FIGURE 7 The thermal stress field of bearing when the roller hollowness was 20%

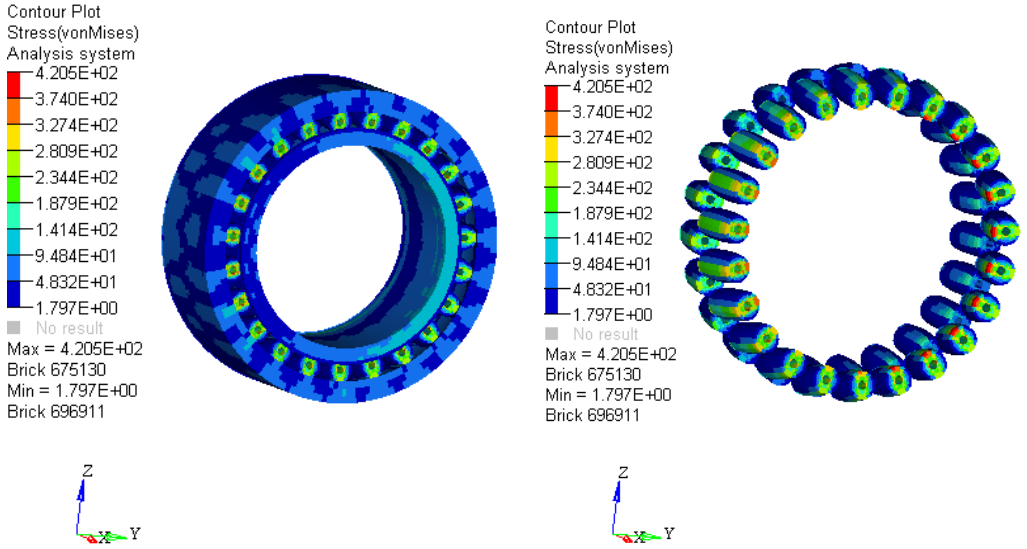


FIGURE 8 The thermal stress field of bearing when the roller hollowness was 30%

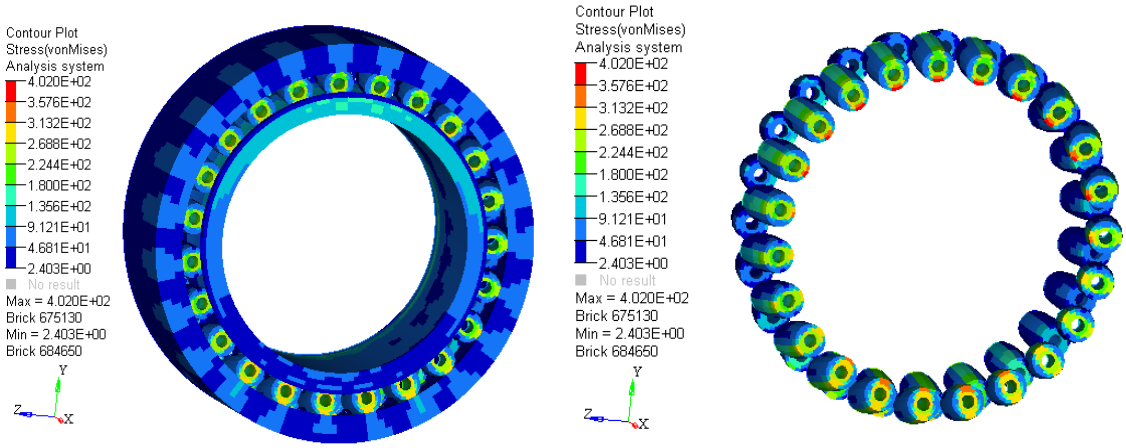


FIGURE 9 The thermal stress field of bearing when the roller hollowness was 40%

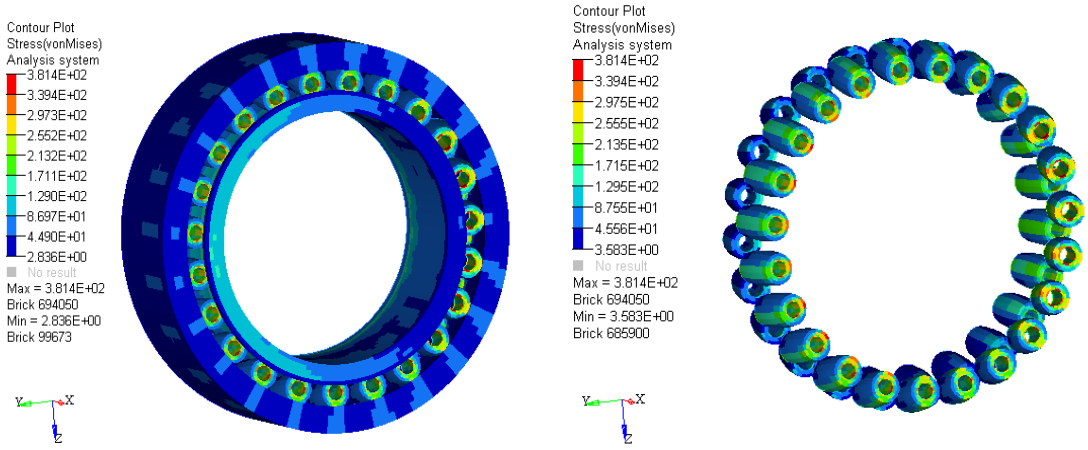


FIGURE 10 The thermal stress field of bearing when the roller hollowness was 50%

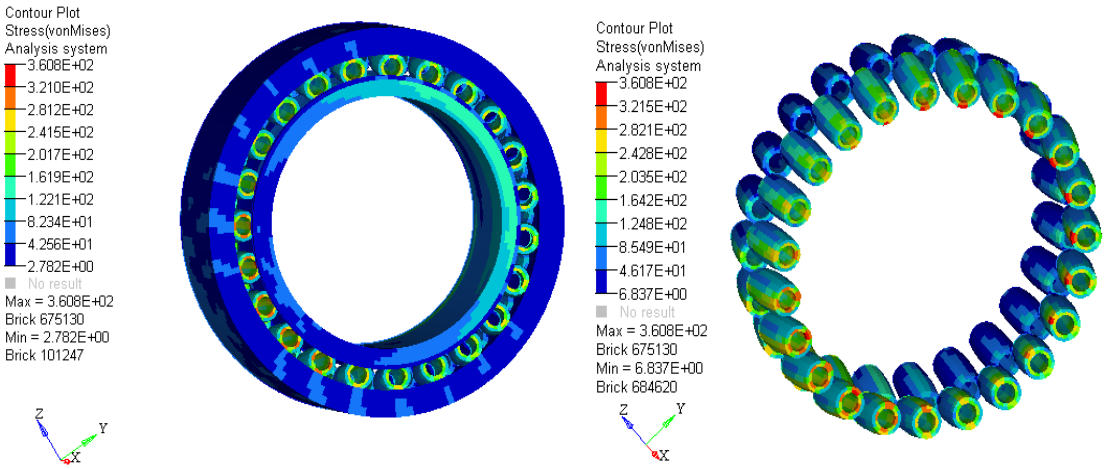


FIGURE 11 The thermal stress field of bearing when the roller hollowness was 60%

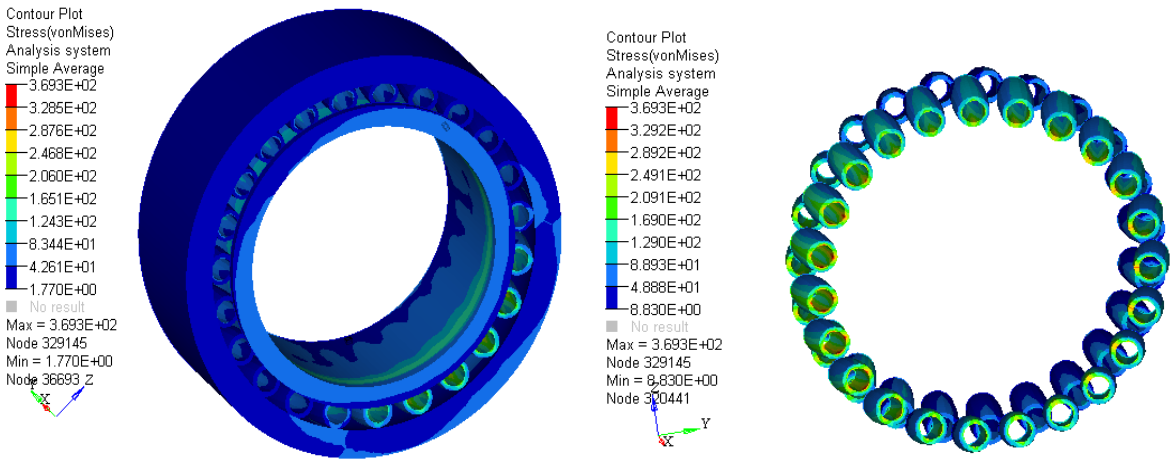


FIGURE 12 The thermal stress field of bearing when the roller hollowness was 70%

By analysing, the maximum thermal stress curve of bearing could be obtained in different roller hollowness. As shown in Figure 13.

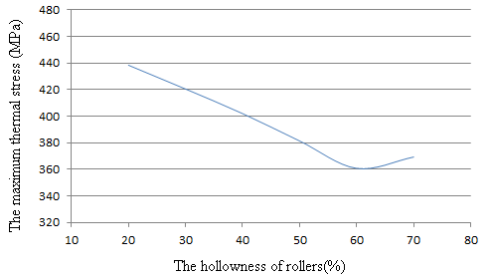


FIGURE 13 The maximum thermal stress curve of bearing roller in different roller hollowness

By observation, it could be seen that the part which the bearing subjected to the maximum stress on the roller. And with the increase of roller hollowness[10], when the roller hollowness was 60%, the bearing subjected to the minimum stress, was 360.8MPa, smaller 50.5MPa than the roller without hollowness. It compared with the allowable stress, in this case the bearing was not easy to failure. Also it could be found that the rolling body stress distribution become uniform, to a certain extent, it also optimized the thermal stress distribution of casting roller bearing and play a role of reference for bearing design.

5 Conclusion

By analysing the stress field of continuous casting roller bearing, it could be found that the maximum stress of bearing was 249.3MPa when the bearing seat was applied cooling water and not applied thermal load, and the coupling stress was 411.3MPa, while the thermal stress caused by temperature was large, the result had been very close to allowable stress of bearing material. At the same time the bearing operating temperature was very high, and the temperature distribution was not uniform, the deformation of bearing different parts caused by different temperature rise of bearing different parts was also inconsistent, affected the bearing accuracy and the rotary bearing capacity of bearing, thus the temperature was the main cause of the failure of continuous casting roller bearing. By altering the hollowness of rolling body from the structure to reduce the stress, and simulated the optimization scheme then the result showed that bearing subjected a minimum stress when the rolling body hollowness was 60%, and the stress distribution was relatively uniform. Thus, it improved the bearing internal heat dissipation and extended the service life of bearing, also played a role of reference for bearing design.

Acknowledgments

This research reported in the paper is supported by National Natural Science Foundation of China (71271160) and China Scholarship Council (CSC). This support is greatly acknowledged.

References

[1] Zyuzin A A, Kaz'min B N, Yurov M D 2012 Roller Reliability in Continuous Casting of Steel Metallurgist **32**(3) 268-273
[2] Tomovic R 2012 Calculation of the boundary values of rolling bearing deflection in relation to the number of active rolling elements Mechanism and Machine Theory (47) 74-88
[3] Yang S, Tao W 2006 Heat Transfer (4th edition) Higher Education Press Beijing Chapter **16** *(in Chinese)*
[4] Guo Y, Parker R G 2012 Stiffness matrix calculation of rolling element bearings using a finite element/contact mechanics model Mechanism and Machine Theory (51) 32-45
[5] Park I S 2011 Fully Numerical Analysis for Effects of Cooling Water Flow Rate and Plate Running Speed on Steel Plate Cooling in Very High Temperature Region ISIJ International **51**(11) 1864-9
[6] Luo Y, Chen X 2014 The optimization of rolling bearing Science and technology research (17) 558-9
[7] Xu H, Cai C, Yan J 2003 Mechanical Design Handbook (2nd edition) China Machine Press Beijing *(in Chinese)*
[8] Lu X, Zhou J, Jiang W 2009 Finite element parametric modeling and analysis of tower crane based on APDL Mechanical &Electrical Engineeering Magazine **26**(7) 34-6
[9] Liu Y, Guan T, Wei Y 2013 Study on the performance of high speed hollow roller bearings Manufacturing Technology & Machine Tool (8) 142-4
[10] Chen S, Li S, Li X 2014 Thermal-Stress Coupled-Field Analysis of Auto Gearbox Test Platform Spindle System Machine Tool & Hydraulics **42**(7) 107-11

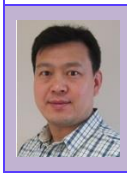
Authors

Disi Chen, born in 1992, Hubei province, P.R. China

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Wuhan Textile University, Wuhan, China, 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Gongfa Li, born on October 7, 1979, Hubei province, P.R. China

Current position, grades: associate professor at Wuhan University of Science and Technology, China.
University studies: PhD degree in mechanical design and theory at Wuhan University of Science and Technology in China.
Scientific interest: intelligent control, modeling and optimal control of complex industrial process.
Publications: 110



Honghai Liu, born in 1973, P.R. China

Current position, grades: professor in Intelligent Systems, Head of Intelligent Systems and Biomedical Robotics, University of Portsmouth.
University studies: PhD in Intelligent Robotics in 2003 from Kings College, University of London, UK.
Scientific interest: approximate computation, pattern recognition, multi-sensor based information fusion and analytics, human machine systems, advanced control, intelligent robotics and their practical applications.
Publications: 320

Guozhang Jiang, born on December 15, 1965, Tianmen, P.R. China

Current position, grades: professor of Industrial Engineering, and the Assistant Dean of the college of machinery and automation, Wuhan University of Science and Technology.
University studies: PhD degree in mechanical design and theory at Wuhan University of Science and Technology, China, in 2007.
Scientific interest: computer aided engineering, mechanical CAD/CAE and industrial engineering and management system.
Publications: 130

Jia Liu, born in 1990, Shanxi, China

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Wuchang institute of Technology, Wuhan, China, 2012.
Scientific interests: mechanical CAD/CAE, signal analysis and processing.

Ze Liu, born in 1989, Hubei province, P.R. China

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Wuhan Institute of Bioengineering, Wuhan, China, 2013.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Weiliang Ding, born in 1991, Hubei province, P.R. China

Current position, grades: student of MS degree in mechanical design and theory at Wuhan University of Science and Technology.
University studies: B.S. degree in measurement and control technology and instrumentation program at Changzhou Institute of Technology, Changzhou, China, 2013.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Wei Miao, born in 1993, Henan province, P.R. China

Current position, grades: M.S. degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Zhengzhou Huaxin College, Zhengzhou, China, 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Zhe Li, born in 1991, Hubei province, P.R.China

Current position, grades: M.S. degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: B.S. degree in mechanical engineering and automation at Wuhan University of Technology Huaxia College, Wuhan, China, in 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Temperature data acquisition and remote monitoring of ladle based on LabVIEW

Fuwei Cheng[1], Gongfa Li[1, 2], Honghai Liu[2]*, Guozhang Jiang[1], Ze Liu[1], Disi Chen[1], Wei Miao[1], Zhe Li[1], Weiliang Ding[1]

[1]*College of Machinery and Automation, Wuhan University of Science and Technology, Wuhan 430081, China*

[2]*Intelligent Systems & Robotics Group, School of Computing, University of Portsmouth, Portsmouth, PO1 3HE, United Kingdom*

Received 1 October 2014, www.cmnt.lv

Abstract

Ladle is an important container parts, plays a vital role to storage and operation of steel smelting. The length of the ladle service life directly affects the efficiency of steel production and the cost of production. The service life of the ladle is mainly affected by thermal mechanical stress of ladle lining, thermal mechanical stress is mainly caused by the severe changes of temperature, in order to fully understand the working state of ladle to ensure the safety of the ladle, the ladle temperature monitoring is particularly important. Ladle design of remote monitoring system uses the structured and modular design thought, on the basis of using the LabVIEW virtual instrument platform; realize the collection of ladle temperature data and remote monitoring. Establish ladle remote client monitoring system by TCP communication protocol, realize the ladle monitoring networked. This system uses the real-time waveform display and friendly human-computer interaction interface, greatly enhanced the real-time and visibility of the ladle remote detection system.

Keywords: the ladle, LabVIEW, remote monitoring list

1 Introduction

Ladle is an important fundamental of steel production in the steel mill, each working procedure in the process of steel-making is required to complete with ladle, and the temperature conditions of ladle directly affect the quality problem of the iron and steel production, especially in high efficiency continuous casting production. Ladle will be impacted by high temperature from molten at work, affecting the service life of the ladle, so the collection of temperature of ladle and precise monitoring is very important. The traditional monitoring equipment for ladle is relatively backward in our country, and its security, reliability and precision and automation degree is low, a lot of testing equipment is in need of upgrading [1]. This paper based on the LabVIEW platform will get the ladle temperature data by sensors, then monitor the collected data with the remote monitoring system based on LabVIEW and determine whether the temperature value is in the normal range, this help the remote workers to rule out the fault in production timely, thus avoid the damage of high temperature to ladle lining, can greatly improve the using efficiency of the ladle, improving the steelmaking efficiency of steel.

2 Structure and life of ladle

The ladle model is as shown in Figure 1. The ladle is mainly composed of trunnion, work layer, permanent layer, ladle wall, impact block, vent brick, nozzle bricks and other parts. Ladle wall has a small slope, lift ladle through the trunnion, and complete the movement of the ladle.

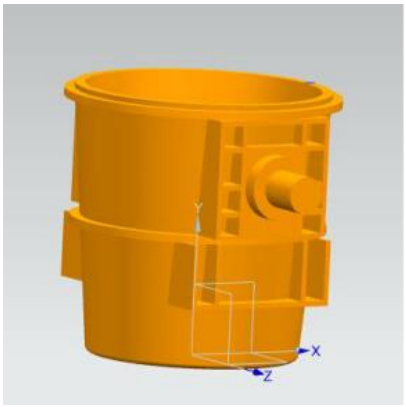


FIGURE 1 Model of the ladle

Figure 2 is the plane view of ladle bottom. There are mainly impact block in the bottom of the ladle, vent brick and runner brick.

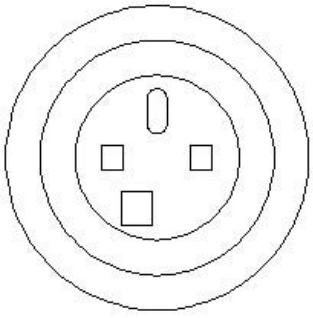


FIGURE 2 Structure figure of ladle bottom

**Corresponding author* e-mail: honghai.liu@port.ac.uk

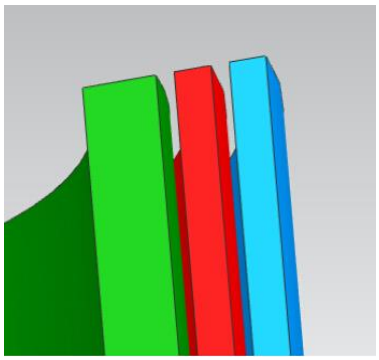


FIGURE 3 Three layer structures of ladle

Ladle generally is consists of three layers, from the inside out are working layer, permanent layer and ladle wall. The material of working layer is commonly alumina and magnesia carbon, material of permanent layer is commonly micro expansion high aluminium, the three layer structure use different materials, so the temperature of ladle is in security.

Ladle life will directly affect the production efficiency of iron and steel and the cost of steel production in steel production. Erosion and burst of refractory lining in the iron and steel production, causing the penetration of liquid steel in the ladle, this is the most common damage in the process of iron and steel production. Thermal mechanical stress and chemical erosion is the main reason of the lining damage. Which the direct reason of resulting the damage of the ladle refractory and the crack of lining layer is the thermal mechanical stress. The damage principle of thermal mechanical stress is that as the temperature of ladle lining layer in a dramatic changes, the produced thermal stress make the micro cracks in the materials gradually spread, eventually led to the fracture and peeling of inner layer material.

The heat transfer phenomenon of temperature in engineering can be roughly divided into three categories: thermal radiation, heat convection and heat conduction. To mainly affect the temperature of ladle are heat convection and heat conduction.

a. Heat convection

Macro relative motion caused by different temperature inside fluid parts caused the heat transfer through mutual movement.

$$q = \lambda \Delta T A, \tag{1}$$

where ΔT – object temperature difference on both sides (K), A – surface area of convective heat transfer $\left(m^2\right)$, λ – surface convective heat change coefficient $\left(W/\left(m^2 * k\right)\right)$.

b. Heat conduction

Each parts with different temperatures or different objects in an object's, through the collision and movement of the microscopic particles in the objects transmitting energy, this way of heat transfer is contacted each other between objects.

$$q = -\lambda A\left(dt/dx\right), \tag{2}$$

where A – area $\left(\left(m^2\right)\right)$, λ – heat transfer coefficient $\left(W/\left(m^2 * k\right)\right)$, t – temperature, x – coordinates on the surface of the heat.

The sharp temperature change of ladle, will make the life of the ladle affect greatly, as the ladle is in the process of working, there happened damage phenomenon, such as lining erosion or burst, ladle needs different levels of maintenance, thus affect the normal production schedule of the steel. Monitoring and displaying every area of the ladle temperature timely, and set the corresponding alarm monitoring, can timely understand the working status of the ladle. Make a corresponding adjustment for the ladle operation according to the monitoring the situation, improving the service life of the ladle.

3 Temperature data collection of ladle

Data collection is a bridge connecting external physical world and the PC, in order to realize the monitoring of ladle temperature signal, first ladle work quantities should be converted into electrical signals of analogy quantities by sensors, then converts processed analogy signals to be the identified digital quantity on ladle remote computer monitoring system, and send to the ladle remote monitoring system [2]. Data acquisition system of ladle obtain needed information from the monitoring objects by sensors, and converts the output signal to the identified digital signal by computer, and then put into the computer to carry on the corresponding processing, get the required data. Ladle data acquisition has a variety of software and hardware combination, can choose appropriate from several combination as the hardware and software of this system.

Good or bad data of acquisition system performance of ladle remote monitoring system is mainly determined by the speed and precision of the ladle data [3]. On the premise of ensuring the accuracy of the collected ladle data, should improve the speed of the ladle data as possible, in order to meet the real-time of ladle data acquisition and processing [4]. Table 1 ladle data acquisition system, generally can be divided into five parts:

TABLE 1 Data acquisition composition of the ladle monitoring system

Sensor	Sensor of monitoring Ladle is a sensor can monitor ladle temperature, and convert the detected signal into the needed signal form to put
Signal	A quantity from a ladle monitoring sensor, signal is transmission medium of ladle data acquisition system, signal usually has two forms, digital and analogy
Signal conditioning	Convert collected temperature signals of ladle to a standard and identifiable signal through a series of transformations
Data acquisition hardware	Data acquisition hardware of ladle monitoring system is the device that convert simulation quantity to digital quantity for ladle monitoring signal
Driver program and application software	The application software of ladle monitoring system is the LabVIEW virtual instrument software and related hardware driver

Data acquisition system of ladle remote monitoring system is shown in Figure 4.

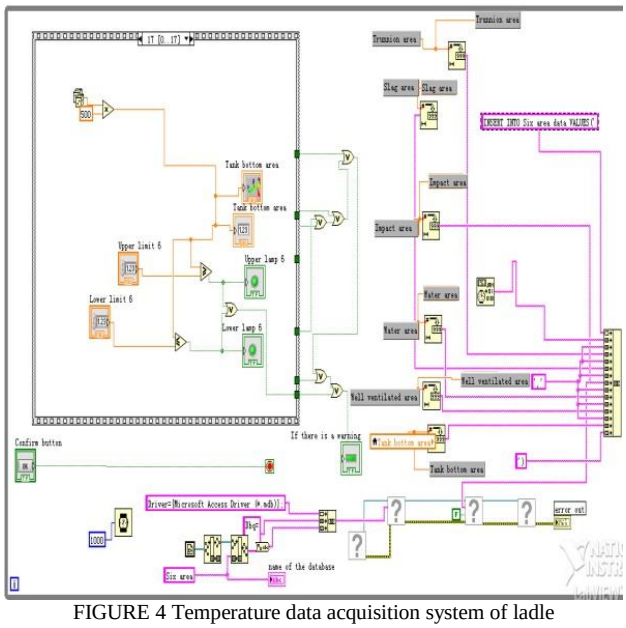


FIGURE 4 Temperature data acquisition system of ladle

The main functions of the data acquisition system are man-machine interface, data acquisition, data display, data storage, printing and output as well as secondary calculating of data shown in the following table.

1) The man-machine interface of ladle data acquisition system.

In the ladle data acquisition, staff dialogue to data acquisition system of the ladle through the mouse, keyboard or touch screen, complete ladle data collection.

2) Data collection of the ladle.

When data acquisition system is working, it will collect for analogy signals into the system or other signals in accordance with the previous ladle collection cycle.

3) Data display of ladle data acquisition system.

The collected temperature data of ladle are displayed in the form of that watchers are convenient for watching, displaying interface is commonly images, forms, trend charts and so on.

4) Storage of the ladle collection data.

Ladle state change is a slow process, set a certain time interval for the storage of collected ladle data, store the collected important data of ladle, and store it to the specified database, convenient for future reference.

5) The ladle data collected printing.

After the ladle data collection, we need to output the data collected to print, in order to understand the performance of the ladle; it will be the required to set a certain time interval to print the data in the form of a graph or table.

6) Secondary calculation of ladle data.

Comparing the collected first data of ladle by the sensor with normal numerical range at work, then making secondary calculation, to identify the running status of the ladle.

4 Ladle temperature remote monitoring

Network communication is an important aspect of Ladle's remote monitoring; the LabVIEW software has a strong network communication function and remote communication function of LabVIEW allow LabVIEW users write remote communication software to realize the remote communication function [5]. Ethernet technology is now widely adopted by the computer network technology, the media access and control mode is adopted by the collision detection of the carrier to listen on multiple accesses. To establish a network publishing of LabVIEW needs to be more than 2 PC through communication lines connected and established communication [6]. So you can through the establishment of communication network to provide resource sharing, communication service, coordination load, data transmission and remote fault diagnosis other functions so on.

From the composition of the entire computer network, which is made up of many WAN and LAN together [7]. Node of the Internet is consisting of the PC, subnet segment, servers and other equipment, which can be established through LAN and telephone line connected to the Internet. Throughout the development of network technology, LAN technology is a focus in the fastest, users can make use of local area network (LAN) in the Internet technology, which brings a lot of convenience [8]. LAN topology structure mainly can be divided into star structure, bus structure, and ring structure and so on. In these several topology types, bus structure is the most simple topological structure, whose local area network (LAN) is based on public information channel [9]. Although this structure is not easy to fault detection, it is easy to be extended and easy to maintenance and installation Each PC machine on the network has an own logo, that is the IP address of the PC, which generally consists of four part numbers. Ladle of network monitoring system is a trend of modern ladle monitoring, combining the Internet and monitoring of the ladle system together, which can make full use of network resources and improve the efficiency of monitoring.

TCP/IP protocol is one of the most basic Internet protocol, TCP and IP forms the core of TCP/IP protocol, both LAN and the Internet are widely adopted, TCP/IP protocol. The application layer, transport layer, network layer, link layer make up the TCP/IP protocol suite. IP is Network protocol, which belongs to Network interconnection layer protocol. In order to transfer the data to the destination, you will need to print data sending into a datagram, and then add the header to datagram and sent to the destination [10]. Because IP layer service is a connectionless service, it is difficult to guarantee the reliability of data in data transferring, which is the reason that seldom directly used IP in LabVIEW programming. TCP is transmission control protocol that is a transport layer protocol, which needs to link first in the data transmission, and most of the network data transmission use TCP transmission control protocol (TCP) now. TCP uses bits of circulation in data transmission and a session needs to be established on the data exchange [11]. In order to guarantee the reliability of data, through the TCP data will be assigned a sequence number. When data in the transmitted is divided

into several small pieces, the receiver can specify whether the data has been received. During the data transfer, the data of the receiver must carry on the confirmation, which realized by sending a reply. When using TCP transmit data, you need to call the IP network protocol, but compared with TCP/IP network protocol, TCP will not repeat or miss data and more reliable in data transmission.

In the use of remote data transmission, TCP client port settings should be consistent with server port, at the same time the client need to specify the right IP address [12]. So as to it can realize correct network interconnection. The design of server-side as shown in Figure 5, TCP server program will first collect the temperature data in the production of the ladle, then write to the TCP server, finally send the data to client.

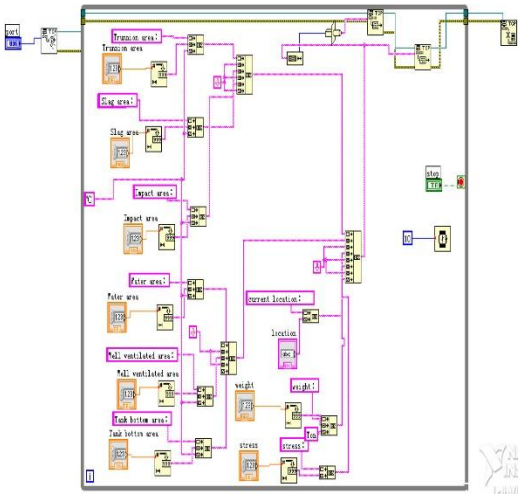


FIGURE 5 The TCP server program

The program diagram of client as shown in Figure 6, The remote monitoring the client program diagram of ladle is programming under the While loop structure, in order to ensure the data writing continuity, with setting the frequency of data writing [13]. In addition to set two attribute node, TextFontSize and Text TextColor, it convenient to observe display window of the front panel data, so that it can read out data transmitting.

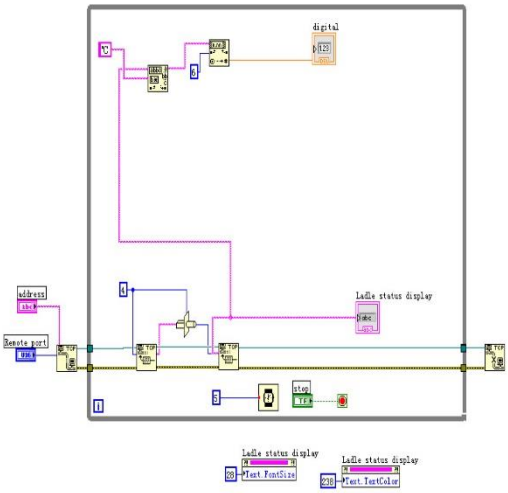


FIGURE 6 The client program of ladle remote monitoring

In the working of ladle, if the temperature status of the ladle is not informed in a timely manner, you will appear many problems, for example:

1) The temperature is too low, so that molten steel solidification needs to be reheated.

2) The temperature of ladle is too low, so the ladle must be heated to a certain temperature before picking up the molten steel.

3) When the ladle is shipping to the continuous casting desktop, too low temperature needs to be reheated.

Because the monitoring object is ladle, system should real-time monitor the ladle temperature data (the temperature of the six, trunnion area, slag zone, under the impact area, the water area, breathable area, and tank bottom area). And count and analysis the acquisition of data, calculate the upper and lower value of collecting data and compare to the setting value, judge whether the working condition of the ladle is normal [14]. It can alert when it is not normal operation, thus reducing the failure rate of the ladle work.

Establish ladle remote monitoring system of remote data transmission module and a remote monitoring server, and then write the data needing to transfer to the remote communication program, which established via TCP communication protocol. To establish a ladle remote monitoring client, you can read the monitoring data of the ladle transmitted from the server, so that to realize remote monitoring [15].

In order to ensure the ladle in production has a good process to produce good quality steel and to study the relationship between temperature field and stress distribution, the need for monitoring the temperature of the ladle six area, trunnion area, slag zone, under the impact area, the water area, breathable area, tank bottom area, is necessary. The temperature monitoring of the ladle front panel diagram as shown in the figure below

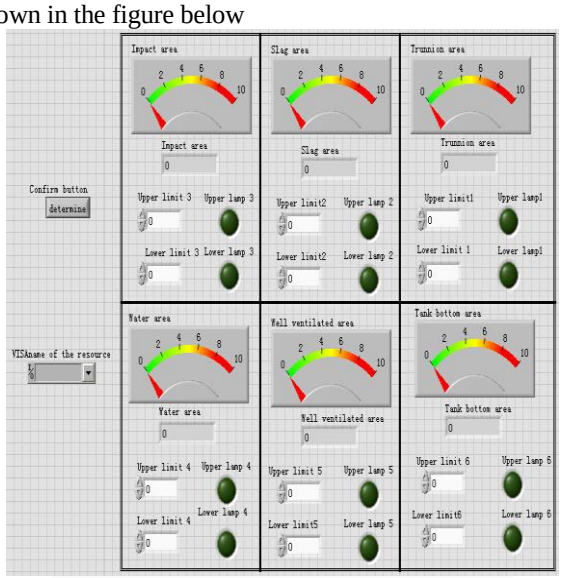


FIGURE 7 The front panel of temperature monitoring

This system is designed to meet the temperature detection, while meet the demand for six area temperature monitoring, in this program using the sequential loop structure to achieve six area through a serial port to read six regional

temperature requirements. Each interval of temperature upper and lower limit value should combine the ladle located automatically to set, which meet the demand of temperature of ladle in any position.

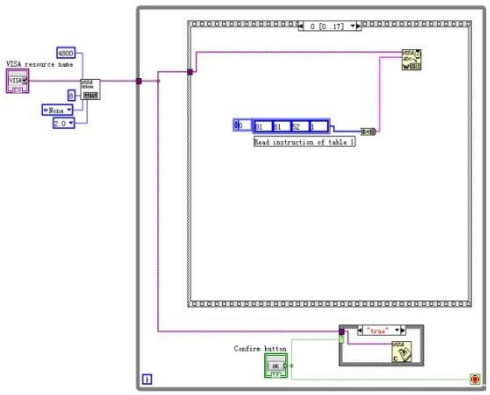


FIGURE 8 Send instruction to the monitoring instrument

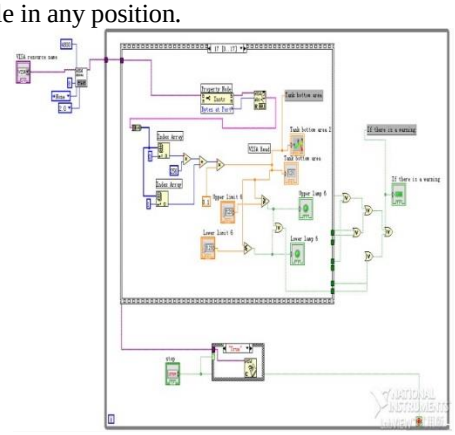


FIGURE 9 Read temperature of six temperature zone

Read the six area temperature value and write temperature data collected through global variables into the database established connection. The temperature monitoring by temperature monitoring module has the six districts, through a serial port of one monitoring module to realize six areas temperature monitoring, that need to be established connection to the temperature acquisition instrument by VISA Write. To judge the collected temperature data with the upper limit and lower limit value, when is greater than the threshold value or less than the lower limit value can change warning light color.

5 Conclusions

The length of the life of the ladle has a vital role on iron and steel production, and the real-time monitoring for ladle production status is conducive to improve the efficiency of steel production. Based on the ladle as monitoring object, this paper starts from virtual instrument technology, then collects and monitories of the temperature parameters of the ladle, next makes a further analysis and research on the data, to ensure that the working temperature of ladle has been in a normal range, thus educes the failure rate, greatly extends the life of the ladle, raises the productivity of the steel.

Acknowledgments

This research reported in the paper is supported by National Natural Science Foundation of China (71271160) and China Scholarship Council (CSC). This support is greatly acknowledged.

References

[11] Jensen M B 2011 Using LabVIEW to demonstrate instrumentation principles *Analytical and Bioanalytical Chemistry* (400) 2673-6
[12] Wang G-m, Jiao S-l, Song H 2009 Mine Pump Comprehensive Performance Testing System Based on Labview *International Conference on Measuring Technology and Mechatronics Automation* (1) 300-3
[13] Cai C-q, Wang F, Dai H 2008 The DAQ system of the vehicle test based on the virtual instruments technology *Proceedings - 2008 International Seminar on Future BioMedical Information Engineering* 202-5
[14] Savu T, Ghercioiu M 2008 Measuring with ultra-low power Wi - Fi sensor tags in LabVIEW *WSEAS Transactions on Computers* (7) 1000-9
[15] Hong L, Huang C-z, Xiao L-y 2011 Research on monitor system of distant coal mine gas based on Labview *Lecture Notes in Electrical Engineering* (3) 457-61
[16] Zhang X-c, Fu Y-XR, Yu J-y, et a1. 2011 A rapid way to build prototypes of Wireless Sensor Network based on virtual instruments technology *Advanced Materials Research* (301-303) 1588-91
[17] Li N, Feng L-p, Zhou B, et a1. 2010 A study on a cRIO-based monitoring and safety eady warning system for large bridge structures *2010 International Conference on Measuring Technology and Mechatronics Automation* 1 361-4
[18] Semakhin V V, Vyalyanov L E 2004 Optimizing the drying and high-temperature heating of the lining of steel-pouring ladles *Metallurgist (S0026-0894)* **48**(5) 275-8
[19] Akselrod L M, Mizin M V G, Filyashin M K, et al. 2003 The steelmaking ladle-ways towards saving heat *Refractories and Industrial Ceramics (S1083-4877)* **44**(3) 123-6
[20] Lcamdali U, Tunc M 2006 Steady state heat transfer of ladle furnace during steel production process *Journal of Iron and Steel Research (S1006-706X)* **13**(3) 18-20
[21] Korvtko N G, Loginov P G, Marin A G, et al. 2005 Monolithic lining for steel-pouring ladles *Metallurgist (S0026-0894)* **49**(3) 91-3
[22] Wen H, Dong X-R, Ma Y-C, et a1. 2010 The research of the databases connection methods in LabVIEW based on ADO *ICCASM2010 - 2010 International Conference on Computer Application and System Modeling, Proceedings* (7) 7229-33
[23] Zhu Q-y, Wen C-l, Xie W-y 2012 General environment integrated monitoring and data management system based on virtual instrument *Lecture Notes in Electrical Engineering* **136** 163-8
[24] Miao M-x 2010 Database visiting method and application in electronic power monitor system by using LabSQL in virtual instrument LabVIEW *Applied Mechanics and Materials* (20-23) 110-5
[25] Bing H, Liu X-j, Shah L 2011 Application and research of data acquisition based on database technology of LabVIEW *Advanced Materials Research* (267) 398-403

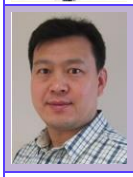
Authors

Fuwei Cheng, born on June 17, 1988, Hubei province, P.R.China

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology, China.
University studies: BS degree in mechanical engineering and automation at Donghu college of Wuhan University, Wuhan, China, 2012.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Gongfa Li, born on October 7, 1979, Hubei province, P.R. China

Current position, grades: an associate professor at Wuhan University of Science and Technology, China.
University studies: PhD degree in mechanical design and theory at Wuhan University of Science and Technology in China.
Scientific interest: intelligent control, modeling and optimal control of complex industrial process.
Publications: 110



Honghai Liu, born in 1973, P.R. China

Current position, grades: professor in Intelligent Systems, Head of Intelligent Systems and Biomedical Robotics, University of Portsmouth.
University studies: PhD in Intelligent Robotics in 2003 at Kings College, University of London, UK.
Scientific interest: approximate computation, pattern recognition, multi-sensor based information fusion and analytics, human machine systems, advanced control, intelligent robotics and their practical applications.
Publications: 320

Guozhang Jiang, born on December 15, 1965, Tianmen, P.R. China

Current position, grades: professor of Industrial Engineering, and the Assistant Dean of the college of machinery and automation, Wuhan University of Science and Technology.
University studies: PhD degree in mechanical design and theory at Wuhan University of Science and Technology, China, in 2007.
Scientific interest: computer aided engineering, mechanical CAD/CAE and industrial engineering and management system.
Publications: 130

Ze Liu, born in 1989, Hubei province, P.R. China

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Wuhan Institute of Bioengineering, Wuhan, China, 2013.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Disi Chen, born in 1992, Hubei province, P.R. China

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Wuhan Textile University, Wuhan, China, 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Wei Miao, born in 1993, Henan province, P.R. China

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Zhengzhou Huaxin College, Zhengzhou, China, 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Zhe Li, born in 1991, Hubei province, P.R. China

Current position, grades: MS degree candidate in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in mechanical engineering and automation at Wuhan University of Technology Huaxia College, Wuhan, China, 2014.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

Weiliang Ding, born in 1991, Hubei province, P.R. China

Current position, grades: student of MS degree in mechanical design and theory at Wuhan University of Science and Technology.
University studies: BS degree in measurement and control technology and instrumentation program at Changzhou Institute Of Technology, Changzhou, China, 2013.
Scientific interest: mechanical CAD/CAE, signal analysis and processing.

A new decoupled sliding mode control approach for the linear motion of a spherical rolling robot

Liang Zhao[1], Tao Yu[2]*

[1]*Modern Educational Technology Centre, Liaoning Medical University, Jinzhou, Liaoning, China*

[2]*Faculty of Mechanical Engineering and Automation, Liaoning University of Technology, Jinzhou, Liaoning, China*

Received 1 March 2014, www.cmnt.lv

Abstract

This paper investigates the dynamics and control aspects of the linear motion of a pendulum-driven spherical rolling robot. The dynamic model is deduced for the linear motion of a spherical robot by using the Euler-Lagrange formulation. By appropriate definitions the equations of motion for the robotic system are transformed into the state space form. A novel decoupled sliding mode control approach is proposed to achieve set-point regulation of the linear motion. This approach consists of the construction of a cascade sliding mode controller and the design of a nonlinear reaching law by using a switching component that dynamically adapts to the variations of the controlled system. The asymptotic stability of the robotic system is verified through Lyapunov analysis, and the validity of the proposed approach is illustrated through numerical simulations.

Keywords: spherical robot, linear motion, dynamic model, decoupled sliding mode control, exponential reaching law

1 Introduction

A spherical robot is a robotic device without wheels or legs, which has a single spherical form that scrolls by itself to conduct missions. The spherical shape of this class of mobile robots offers several advantages over other forms of surface-based locomotion such as wheels, tracks or legs. The sphere is a strong shape providing a high level of robustness with no major weakness points on its surface, whereas wheels, tracks or legs can be damaged, potentially disabling the mobility of the robot. The outer shell can also be resilient and serve as a protective barrier between the outside environment and the inside equipments. A spherical robot is by nature non-invertible further limiting the risk of becoming disabled, while most other mobile robot designs are vulnerable to tipping over or becoming stuck on the terrain where their means of locomotion lose contact with the ground. These advantages indicate that a spherical robot is appropriate for many different applications such as surveillance, reconnaissance, hazardous environment assessment, search and rescue, as well as planetary exploration.

Spherical rolling robots can be categorized into different types according to their internal driving mechanisms [1-9]. Compared with other types of spherical rolling robots [1-6], a pendulum-driven spherical rolling robot [7-9] has a simpler structure further making it easier to be manoeuvred. The schematic diagram of a pendulum-driven spherical rolling robot with dual inputs is illustrated in Figure 1. Linear motion is a basic form of locomotion of pendulum-driven spherical rolling robots, and it is realized by moving a motor-controlled pendulum forwards or backwards. In this paper, a new decoupled sliding mode control approach based on a novel exponential reaching law is presented for stable control of the linear motion. In the proposed controller, a double layer structure is used to guarantee the stability of the whole system, and the sub-sliding surfaces are utilized to drive the tracking errors to zero.

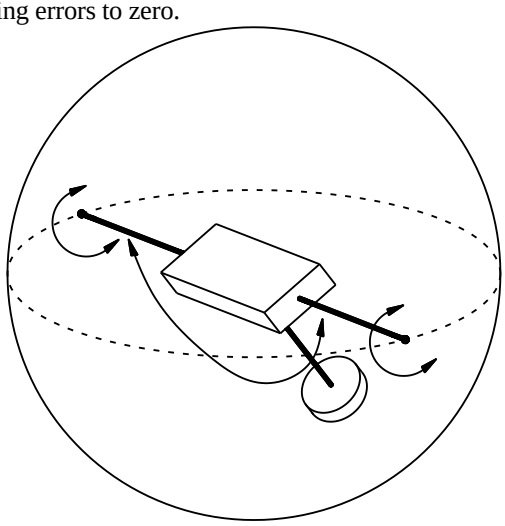


FIGURE 1 Structure of a pendulum-driven spherical robot

2 Dynamic analysis

We start with a simplified planar model, only considering no slip linear motion on flat surfaces. Figure 2 illustrates the simplified model with a side view of a pendulum-driven spherical rolling robot. It represents the spherical shell with its centre of mass *B*, the internal mechanism with its centre of mass *D*, which coincides with that of the spherical shell, and the pendulum (composed of a massless link and a counterweight at its end) with its centre of mass *E* and the axis attached at the centre of the sphere. The definition of the model parameters is listed in Table 1.

***Corresponding author** e-mail: yutaolanjie@163.com

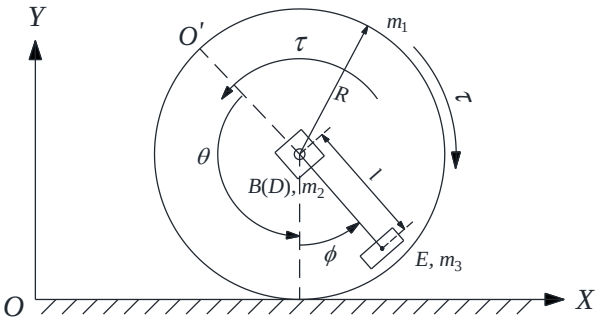


FIGURE 2 Simplified model of the linear motion

TABLE 1 Parameter definition of the planar model

m1, m2, m3	mass of the spherical shell, the internal mechanism and the pendulum, respectively
r, l	radius of the spherical shell and length of the pendulum, respectively
θ, ϕ	roll angle of the spherical shell and sway angle of the pendulum, respectively
I1, I2, I3	moment of inertial of the spherical shell, the internal mechanism and the pendulum, respectively
τ	torque applied to the pendulum

We first choose the roll angle of the sphere θ and the sway angle of the pendulum ϕ as the generalized coordinates of the robotic system, and then we develop the equations of motion by calculating the Lagrangian $L = T - P$ of the system, where T and P are the kinetic energy and potential energy of the system respectively.

The kinetic energy and potential energy of the whole robotic system are given by

$$T = \frac{1}{2} J_1 \dot{\theta}^2 + \frac{1}{2} J_2 \dot{\phi}^2 + m_3 r l \dot{\theta} \dot{\phi} \cos\phi \quad P = -m_3 g l \cos\phi, \tag{1}$$

where g denotes the gravitational acceleration;

$J_1 = M_t r^2 + I_1$, $M_t = m_1 + m_2 + m_3$, $J_2 = m_3 l^2 + I_2 + I_3$.

It is assumed that the viscous friction operates between the sphere and the pendulum. The loss due to the viscous friction is written in an energy dissipation function that depends on the velocities of the system and the damping coefficient ζ associated with the pendulum-sphere bearing.

$$R = \frac{1}{2} \zeta \left(\dot{\theta} + \dot{\phi} \right)^2. \tag{2}$$

Using the Euler-Lagrange Equations [10], the dynamics of the linear motion can be expressed as

$$\boldsymbol{M}(\boldsymbol{q}) \ddot{\boldsymbol{q}} + \boldsymbol{N}(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \boldsymbol{E}(\boldsymbol{q}) \tau, \tag{3}$$

where $\boldsymbol{M}(\boldsymbol{q})$ is the inertia matrix, $\boldsymbol{N}(\boldsymbol{q}, \dot{\boldsymbol{q}})$ is the nonlinear terms, and $\boldsymbol{E}(\boldsymbol{q})$ is the input transformation matrix.

$$\boldsymbol{M}(\boldsymbol{q}) = \begin{bmatrix} J_1 & m_3 r l \cos\phi \\ m_3 r l \cos\phi & J_2 \end{bmatrix}, \quad \boldsymbol{E}(\boldsymbol{q}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\boldsymbol{N}(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \begin{bmatrix} \zeta \left(\dot{\theta} + \dot{\phi} \right) - m_3 r l \sin\phi \dot{\phi}^2 \\ \zeta \left(\dot{\theta} + \dot{\phi} \right) + m_3 g l \sin\phi \end{bmatrix}.$$

Using the control input $u = \tau$ and the state vector $\boldsymbol{X} = \left(\theta, \dot{\theta}, \phi, \dot{\phi} \right)^T$, we can rewrite Equation (3) as follows

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f_1(\boldsymbol{X}) + b_1(\boldsymbol{X}) u \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = f_2(\boldsymbol{X}) + b_2(\boldsymbol{X}) u \end{cases}, \tag{4}$$

where

$$f_1(\boldsymbol{X}) = \frac{m_{22} n_1 - m_{12} n_2}{m_{12} m_{21} - m_{11} m_{22}}, \quad f_2(\boldsymbol{X}) = \frac{m_{11} n_2 - m_{21} n_1}{m_{12} m_{21} - m_{11} m_{22}},$$

$$b_1(\boldsymbol{X}) = \frac{m_{22} - m_{12}}{m_{11} m_{22} - m_{12} m_{21}}, \quad b_2(\boldsymbol{X}) = \frac{m_{11} - m_{21}}{m_{11} m_{22} - m_{12} m_{21}}.$$

Here m_{ij} denotes the element in the i-th row and j-th column of the matrix $\boldsymbol{M}(\boldsymbol{q})$, and n_k represents the k-th component of the vector $\boldsymbol{N}(\boldsymbol{q}, \dot{\boldsymbol{q}})$.

3 Controller design

In this section, we investigate the set-point regulation scheme of the linear motion of a spherical robot, and a decoupled sliding mode controller based on a new exponential reaching law is derived to asymptotically stabilize the robot around its desired equilibrium.

Considering the system represented by Equation (4), we first divide the whole system into two subsystems as

$$A: \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f_1(\boldsymbol{X}) + b_1(\boldsymbol{X}) u \end{cases}, \quad B: \begin{cases} \dot{x}_3 = x_4 \\ \dot{x}_4 = f_2(\boldsymbol{X}) + b_2(\boldsymbol{X}) u \end{cases}.$$

Then we construct the following first layer sliding surfaces for the two subsystems

$$s_1 = \lambda_1 e_1 + \dot{e}_1, \qquad s_2 = \lambda_2 \phi + \dot{\phi}, \tag{5}$$

where λ_1 and λ_2 are positive constants; $e_1 = x_1 - \theta^d$, and θ^d is the desired value of θ.

We define an intermediate variable z which represents the information from subsystem B, and it is incorporated into the sliding surface s_1. Therefore, the second layer sliding surface S is designed as

$$S = \lambda_1 (e_1 - z) + \dot{e}_1 \quad z = z_u \cdot \tanh(s_2), \tag{6}$$

where z_u is the upper bound of $abs(z)$, $0 < z_u < 1$; $\tanh(\cdot)$ is the hyperbolic tangent function defined as follows $\tanh(s_2) = \dfrac{e^{s_2} - e^{-s_2}}{e^{s_2} + e^{-s_2}}$.

Since z_u is less than one, z presents a decaying signal. As s_2 decreases, z decreases too. When $s_2 \to 0$, we have $z \to 0$, $e_1 \to 0$, and then $s_1 \to 0$, and the control objective will be achieved.

Differentiating Equation (6), we can calculate $\dot{z}$ as

$$\dot{z} = \alpha(s_2, z_u)\cdot \dot{s}_2, \tag{7}$$

where $\alpha(s_2, z_u) = z_u \cdot \sec h^2(s_2)$, $\sec h(s_2) = \dfrac{2}{e^{s_2} + e^{-s_2}}$.

Differentiating Equation (6) and using Equation (7) yields

$$\begin{aligned} \dot{S} &= \lambda_1(\dot{e}_1 - \dot{z}) + \ddot{e}_1 = \\ &\lambda_1 x_2 + f_1 - \alpha\lambda_1(\lambda_2 x_4 + f_2) + (b_1 - \alpha\lambda_1 b_2)u \end{aligned}. \tag{8}$$

Then we can obtain the equivalent control as

$$u_{eq} = \frac{\lambda_1 x_2 + f_1 - \alpha\lambda_1(\lambda_2 x_4 + f_2)}{\alpha\lambda_1 b_2 - b_1}. \tag{9}$$

The control input of the system is assumed to take the following form

$$u = u_{eq} + u_{sw}, \tag{10}$$

where u_{sw} is the switching control.

To construct the switching component u_{sw}, we propose the following exponential reaching law

$$\dot{S} = -\frac{\eta}{N(S)}\operatorname{sgn}(S) - \rho S, \tag{11}$$

where η and ρ are positive constants $N(S) = \delta_0 + (1-\delta_0)e^{-\gamma|S|^p}$.

Here δ_0 is a positive constant that is less than one, p is a positive integer, and γ is also a positive constant.

The proposed ERL given by Equation (11) is composed of a variable rate reaching term [12] and an exponential term. Comparing with the conventional exponential reaching law [13], we can see from Equation (11) that if $|S|$ increases, $N(S)$ approaches δ_0, and therefore $\eta/N(S)$ converges to η/δ_0, which is larger than η. This means that $\eta/N(S)$ increases in the reaching phase, and consequently the attraction to the sliding surface S is faster. On the other hand, if $|S|$ decreases, then $N(S)$ approaches one, and $\eta/N(S)$ converges to η. This means that, when the system state approaches the sliding surface S, $\eta/N(S)$ gradually decreases to reduce the chattering. Therefore, the proposed ERL allows the controller to dynamically adapt to the variations of the switching function S by letting $\eta/N(S)$ vary between η and η/δ_0.

Using Equation (8) to Equation (11), we can obtain the switching control as

$$u_{sw} = \frac{N_\eta(S)\operatorname{sgn}(S) + \rho S}{\alpha\lambda_1 b_2 - b_1}, \tag{12}$$

where $N_\eta(S) = \dfrac{\eta}{N(S)}$.

Substituting Equation (12) into Equation (10), we can obtain the following sliding mode control law

$$u = \frac{\lambda_1 x_2 + f_1 - \alpha\lambda_1(\lambda_2 x_4 + f_2) + N_\eta(S)\operatorname{sgn}(S) + \rho S}{\alpha\lambda_1 b_2 - b_1}. \tag{13}$$

Theorem 1: Supposing that the robotic system represented by Equation (4) is controlled by the sliding mode controller given by Equation (13). Then the system defined by Equation (4) is asymptotically stable.

Proof: Considering the Lyapunov function candidate $V = \dfrac{1}{2}S^2$, then $\dot{V}$ can be given by

$$\dot{V} = S\dot{S} = -N_\eta(S)|S| - \rho S^2 \le -\eta|S| - \rho S^2 \le 0. \tag{14}$$

Integrating both sides of Equation (14), we have

$$\begin{aligned} &V(t) = \frac{1}{2}S^2 \le V(0) < \infty \\ &\lim_{t\to\infty}\int_0^t \left(\eta|S| + \rho S^2\right)\mathrm{d}\sigma \le V(0) < \infty \end{aligned}. \tag{15}$$

According to Equation (15), we have $S \in L_\infty$, $S \in L_2$. According to Equation (14), we have $\dot{S} \in L_\infty$. Consequently, by applying Babalat's lemma we can conclude that the sliding surface S is asymptotically stable, i.e. $\lim_{t\to\infty} S = 0$.

Then the system can be guaranteed to be asymptotically stable.

4 Simulation study

In this simulation, the following physical parameters of the spherical mobile robot [14] and design parameters of the sliding mode controller are used.

$m_1 = 1.2\,\mathrm{kg}$, $m_2 = 1.85\,\mathrm{kg}$, $m_3 = 2.05\,\mathrm{kg}$, $R = 0.15\,\mathrm{m}$, $l = 0.12\,\mathrm{m}$, $I_1 = 0.018\,\mathrm{kg\cdot m^2}$, $I_2 = 0.0017\,\mathrm{kg\cdot m^2}$, $I_3 = 0.0006\,\mathrm{kg\cdot m^2}$, $\zeta = 0.03\,\mathrm{N\cdot m\cdot (rad/s)^{-1}}$, $g = 9.81\,\mathrm{m/s^2}$, $z_u = 0.96$, $\lambda_1 = 4$, $\lambda_2 = 0.6$, $\eta = 3$, $\delta_0 = 0.1$, $\gamma = 10$, $p = 1$, $\rho = 7.8$.

In addition, the initial and desired values of the system states are chosen as $x_0 = (0,\ 0,\ 0,\ 0)^T$, $x^d = (\pi,\ 0,\ 0,\ 0)^T$.

The simulation results are depicted in Figure 3 to Figure 6. As it is theoretically expected, we can find that both the roll angle of the sphere and the sway angle of the pendulum are asymptotically stabilized to their desired values, and the anti-sway control is achieved in a rapid manner after only one oscillation.

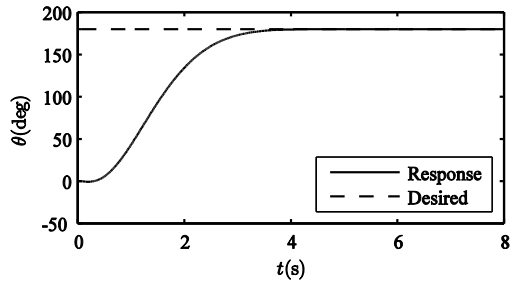


FIGURE 3 Tracking result of the roll angle of the spherical shell

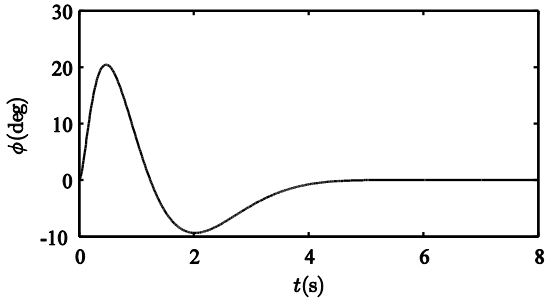


FIGURE 4 Tracking result of the sway angle of the pendulum

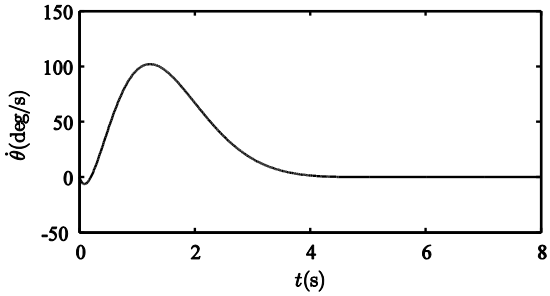


FIGURE 5 Time evolution of the angular rate of the roll angle

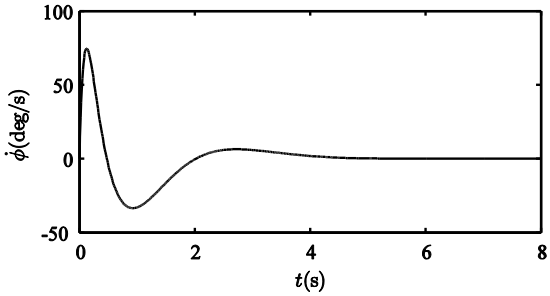


FIGURE 6 Time evolution of the angular rate of the sway angle

5 Conclusions

In this paper, we present a variable structure strategy for set-point regulation of the linear motion of a spherical rolling robot. The control development is based on the construction of a cascade sliding mode controller and a novel exponential reaching law, and the proposed control approach consists of designing a nonlinear reaching law by using a switching term that dynamically adapts to the variations of the system state. The asymptotic stability of the sliding surface of the whole system is theoretically proved, and the simulation results further verify the effectiveness of the proposed controller.

References

[1] Halme A, Schonberg T, Wang Y 1996 *Proceedings of the 4th International Workshop on Advanced Motion Control (Mie) IEEE USA* **1** 259-64
[2] Bicchi A, Balluchi A, Prattichizzo D 1997 Proceedings of the IEEE International Conference on Robotics and Automation (Albuquerque) *IEEE USA* **3** 2620-5
[3] Otani T, Urakubo T, Maekawa S, Tamaki H, Tada Y 2006 *Proceedings of the 9th IEEE International Workshop on Advanced Motion Control (Istanbul) IEEE USA* 416-21
[4] Mukherjee R, Minor M A, Pukrushpan J T 1999 *Proceedings of the 38th IEEE Conference on Decision and Control (Phoenix) IEEE USA* 2132-7
[5] Javadi A H A, Mojabi P 2004 *Journal of Dynamic Systems, Measurement and Control* **126**(3) 678-83
[6] Bhattacharya S, Agrawal S K 2000 *IEEE Transactions on Robotics and Automation* **16**(6) 835-9
[7] Michaud F, Laplante J F, Larouche H, Duquette A, Caron S, Letourneau D, Masson P 2005 *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* **35**(4) 471-80
[8] Zhan Q, Jia C and Ma X 2005 *Chinese Journal of Mechanical Engineering* **18**(4) 542-5
[9] Sun H, Xiao A, Jia Q and Wang L 2005 *Journal of Beijing University of Aeronautics and Astronautics* **31**(7) 735-9
[10] Abbott M S *2001 Kinematics, Dynamics and Control of Single-axle, Two-wheel Vehicles (Biplanar Bicycles)* MS diss. Virginia Polytechnic Institute and State University: Blacksburg
[11] Wang W, Yi J, Zhao D 2005 *Information and Control* **34**(2) 232-5
[12] Fallaha C J, Saad M, Kanaan H Y 2011 *IEEE Transactions on Industrial Electronics* **58**(2) 600-10
[13] Gao W, Hung J C 1993 *IEEE Transactions on Industrial Electronics* **40**(1) 45-55
[14] Yu T 2014 *Study on Control Methodology for the Slope Motion of a Spherical Robot* PhD diss. Beijing University of Posts and Telecommunications Beijing

Authors

Liang Zhao, born in 1979, Jinzhou, Liaoning, China

Current position, grades: lecturer with Modern Educational Technology Centre, Liaoning Medical University, Jinzhou, Liaoning, China.
University studies: BS degree in control science and engineering from Harbin Engineering University, Harbin, Heilongjiang, China 2002, M.Ed. degree in pedagogy from Guangxi Normal University, Guilin, Guangxi, China 2008.
Scientific interest: the dynamics and control of industrial robots and special robots.

Tao Yu, born in 1980, Jinzhou, Liaoning, China

Current position, grades: associate professor with Faculty of Mechanical Engineering and Automation, Liaoning University of Technology, Jinzhou, Liaoning, China.
University studies: PhD degree in mechanical engineering from Beijing University of Posts and Telecommunications, Beijing, China 2014.
Scientific interest: sliding mode control, intelligent control, and robotics.

MEMS model order reduction method based on SPRIM

Xiangjuan Bian[1], Youping Gong[2]*, Liyun Zhen[2]

[1]*School of Faculty & technology, Zhejiang International Studies University, Hangzhou, 310012 China*

[2]*School of Mechanical Engineering Hangzhou Dianzi University, 310018, Hangzhou, China*

Received 1 March 2014, www.cmnt.lv

Abstract

According to large scale MEMS united constraints equations, the paper investigates model order reduction (MOR) techniques based on Structure Preserving Reduced-order Interconnect Macro modelling (SPRIM) method, which can be used to generate computationally efficient solutions for multiphasic MEMS simulation. To united constrained model, the high dimensionality of the original system state space is mapped to a suitable low-dimensional subspace, obtained a low-dimensional state sub-space model. A improved algorithms(SPRIM) from Arnoldi algorithms are implemented to extract low dimensional Krylov subspaces from Unified state subspace models for model order reduction, reduced order electro thermal–mechanical models are generated for a MEMS micro beam using the developed programs. Developed programs automatically generate compact structure preserving models and can be used to significantly improve the computational efficiency without much loss in accuracy and model stability for coupled-field MEMS simulation.

Keywords: MEMS (Micro-Electrical Mechanical System), model order reduction, sub-space model, model reflection

1 Introduction

Micro-Electro-Mechanical Systems (MEMS) are the integration of mechanical elements, sensors, actuators, and electronics at the micron-scale or even at nano-scale through micro fabrication technology, which involves the mechanical, electronic, fluid, thermal, optical, magnetic and other disciplines [1]. Most micromechanical devices have some form of nonlinearity, either geometric nonlinearity due to large amplitude deflections or intrinsic nonlinearities due to governing equations, Multi-energy domain coupling effect of MEMS can hardly be described perfectly, because in a micron-scale or even in a nano-scale working room, physical quantities of energy among the different domains interact with each other. So, in MEMS's design process, the comprehensive effects which have been made by a variety of physical phenomena must be taken into consideration. Most MEMS devices are modelled using fully meshed finite-element based dynamic models, such models often require unacceptably long simulation time. As a result, fully meshed dynamical simulation can be computationally lunation time, and as a result, fully meshed dynamical simulation can be computationally infeasible in a typical workstation environment. Such models need to reduce the computation cost without comprising accuracy, thus allowing hundreds or thousands of dynamical simulations necessary to study how a MEMS device actually functions under various input excitations in a reasonable time.

The paper investigates model order reduction (MOR) techniques based on Structure Preserving Reduced-order Interconnect Macro modelling (SPRIM) method, which can be used to generate computationally efficient solutions for multi-physics MEMS simulation. We organized the paper as 4 parts:

1) the preview of MOR method;
2) the principle of SPRIM;
3) the detailed example of MOR based on SPRIM;
4) an application example.

2 Previous work

The basic idea of MOR method of system state space node variable mapping transform based on the PDES dynamics system is controlled by the spatial discretization, get a great degree of freedom of ODES, then using the coordinate transformation method transform the large-scale ODES system into a system with approximation the low dimension ODES, the dynamic characteristic of the approximate system can depict the original system without sacrificing too much precision condition, so as to achieve the state vector of the aim of reducing the number of degree of freedom. Many existing MOR algorithms are basically the reduction of a linear system model, for nonlinear systems, usually use MOR in the vicinity of the operating point linearization, there are three type of the MOR method in general [2-4]:

1) Based on Krylov approximation subspace or moment matching;

2) Based on balanced truncation method (TBR) with approximate Hankel norm;

3) Combined with singular value decomposition (SVD) and the Krylov subspace iterative method. The transient analysis approximate effect of Krylov subspace methods is good,

**Corresponding author* e-mail:gyp@hdu.edu.cn

and its operation is relatively simple, Krylov methods construct a reduced model based on either explicitly or implicitly approximating the transfer function of a full scale model. Asymptotic waveform evaluation (AWE) is an explicit method that has been used to create reduced models; however, AWE has been shown to be a numerically unstable process. The Arnoldi and Lanczos algorithms are implicit methods that generate Krylov subspaces and reduce models by projection onto these subspaces [5, 6]. The Arnoldi and Lanczos algorithms method, which have been widely used to do model order reduction for VLSI interconnects, is used to generate reduced-order models for a linearized fixed-fixed beam microstructure, the method has been proved to be very efficient to generate reduced-order models when the device is operated in the linear regime. However, when the device undergoes large amplitude deflection, the linearized model deviates from the original nonlinear model significantly, suggesting that nonlinear reduced-order models are needed [7, 8]. To solve this problem, the paper presents a new technique by combining the Taylor series expansion and an improve Arnoldi method - SPRIM method, to develop accurate reduced-order models for nonlinear MEMS devices. The main idea of the approach is to devise a state-variable transformation operation from the linearized device system, and use this state-variable transformation operation to project a high-dimensional device nonlinear state-space model onto a low-dimensional form. Simulation of a fixed-fixed electrostatic actuated beam device shows that the reduced-order nonlinear models can accurately capture the original device nonlinear behavior over a much larger range of device deformation than the conventional linearized model.

3 Structure preserving reduced-order interconnect macro modeling (SPRIM) method

Krylov subspace model order reduction method under the condition of certain can maintain the stability of the original system and passive, so to improve the traditional Krylov subspace methods, makes some important attributes order reduction system can keep the original system. Krylov subspace projection methods extract low dimensional Krylov subspaces from models described by ODEs with the desired model reduction achieved by projection of the models onto the subspaces while dynamic characteristics are maintained through a property called moment matching. To illustrate the MOR formulation procedure and moment matching we will take the first order system shown in Equation (1) as an example and rewrite it as follows:

$$\begin{cases} E\dfrac{dx(t)}{dt} = Ax(t) + Bu(t) \\ y(t) = B^T x(t) \end{cases}, \tag{1}$$

where $E = \begin{bmatrix} M & 0 \\ 0 & H \end{bmatrix} \in R^{n\times n}$, $A = \begin{bmatrix} N & F \\ -F^T & 0 \end{bmatrix} \in R^{n\times n}$, $B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \in R^{n\times P}$, M, N and B_1 has the same number of rows, M, H and N is positive semi-definite matrix, at the same time, a given point, as well as the positive integer, specific algorithm is:

Step 1: let: $G = (S_0E - A)^{-1}$, $Q = (s_0E - B)^{-1}$, use Arnoldi block method compute matrix $\hat{V}$, which satisfies: $\text{colspan}\{\hat{V}\} = K_r(G:Q)$;

a) given matrix $A \in R^n, B \in R^{n\times p}$ and QR decomposition of matrix B, let $B = V_0T$;

b) Compute: $\hat{V}_1 = AV_0 - V_0H_{00}$, where, $H_{00} = V_0^TAV_0$, then QR decomposition $\hat{V}_1$, we can get: $\hat{V}_1 = V_1H_{10}$;

c) Compute: $\hat{V}_2 = A_0V_1 - V_1H_{11} - V_0H_{01}$, where: $H_{11} = V_1^TAV_1$, $H_{01} = V_0^TAV_1$, then QR decomposition $\hat{V}_2$, we can get $\hat{V}_2 = V_2H_{21}$;

d) Then compute following parameter intron: $\hat{V}_r = AV_{r-1} - V_{r-1}H_{r-1,r-1} - ... - V_1H_{1,r-1} - V_0H_{0,r-1}$, where: $H_{r-1,r-1} = V_{r-1}^TAV_{r-1}, ..., H_{1,r-1} = V_1^TAV_{r-1}$, $H_{0,r-1} = V_0^TAV_{r-1}$ and QR decomposition $\hat{V}_r$, we can get $\hat{V}_r = V_rH_{r,r-1}$.

Step 2: According to the characteristics of block structure matrix E and A matrix to do division $\hat{V}$: $\hat{V} = \left[V_1^T, V_2^T\right]$, and constructing the corresponding matrix $V = diag\{V_1, V_2\}$.

Step 3: let $\tilde{M} = V_1^TMV_1$, $\tilde{H} = V_2^THV_2$, $\tilde{N} = V_1^TNV_1$, $\tilde{F} = V_2^TFV_2$, $\tilde{B} = V_1^TB_1$, we get the reduced order system:

$$\begin{cases} \tilde{E}\dfrac{d\tilde{x}(t)}{dt} = \tilde{A}\tilde{x}(t) + \tilde{B}u(t) \\ \tilde{y}(t) = \tilde{B}^T\tilde{x}(t) \end{cases}, \tag{2}$$

where: $\tilde{E} = \begin{bmatrix} \tilde{M} & 0 \\ 0 & \hat{H} \end{bmatrix} \in R^{n\times n}$, $\tilde{A} = \begin{bmatrix} \tilde{N} & \tilde{F} \\ -\tilde{F}^T & 0 \end{bmatrix} \in R^{n\times n}$, $\tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ 0 \end{bmatrix} \in R^{n\times P}$.

4 Application example

In order to illustrate the model reduction technique, we give an example of the well-known beam structure in a electrical environment. Figure 1 shows the beam structure, When a voltage is applied, the top plate of the structure bends downward due to the resultant electrostatic force. Also when then beam bends, the pressure distribution of the ambient air under the beam increases. This pressure increase produces a backward pressure force that damps the beam motion. The beam structure has been used in many sensor applications.

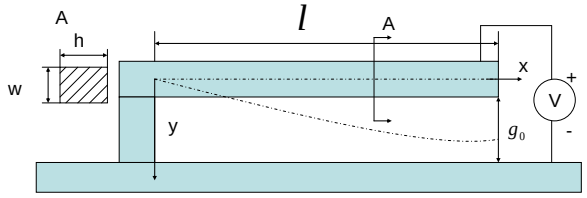


FIGURE 1 Single-ended fixed electrostatic actuating micro beam structure

The beam can be modeled by coupling the Euler beam equation with the electrostatic force and the Reynolds squeeze-film damping equation as:

$$\begin{cases} EI\dfrac{\partial^4 y}{\partial x^4} - s\dfrac{\partial^2 y}{\partial x^2} = -\dfrac{\xi_0 \omega V^2}{2y^2} + \int_0^{\infty}\left(p - p_a\right)dy - \rho\dfrac{\partial^2 y}{\partial t^2} \\ \nabla\cdot\left(y^3 p\nabla p\right) = \dfrac{12\mu}{1+6k}\dfrac{\partial(py)}{\partial t} \end{cases}, \quad (3)$$

where: $y(x,t)$ is the immunity of the z direction, E is young's modulus, $I = \dfrac{wh^3}{12}$ is the moment, ρ is the density, ε_0 is dielectric constant, p is the air damping of the pressure, p_0 is the ambient pressure, μ is the air viscosity coefficient, g is the distance from the plate under initial state, λ is air mean free path (Table 1).

TABLE 1 The parameters of Single-ended fixed electrostatic actuating micro beam structure

Parameters	Value
Length	610 μm
Width	40μm
Thickness	2.2μm
Gap	2.3μm
Young's modulus	149GPa
Density	2330kg/m3
Air viscosity	1.82×10^{-5}kg/(m·s)
Mean-free path of air	0.0064

Let: $\tilde{y} = \dfrac{y(x,t)-g}{g}$, $\tilde{p}(x,t) = \dfrac{p(x,t)-p_0}{p_0}$, substitute them into Equation (3), and in g and p_0 using Taylor series expansion, and take position, speed and pressure as state variables, like: $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, we can get the micro beam for original system state equation:

$$\dot{x} = \begin{pmatrix} \dot{y} \\ \ddot{y} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} x_2 \\ \left(-\dfrac{EI}{\rho}\cdot\dfrac{\partial^4 y_i}{\partial x^4} + \dfrac{s}{\rho}\dfrac{\partial^2 y_i}{\partial x^2} + \dfrac{\varepsilon_0\omega V^2}{g^3\rho}\right)x_1 + \dfrac{p_0}{g\rho}\int_0^{\omega} x_3 dy - \dfrac{3\varepsilon_0\omega V^2}{2g^3\rho}x_1^2 - \dfrac{\varepsilon_0\omega V^2}{2g^3\rho} \\ \left[\left(1+\dfrac{6\lambda}{g}\right)+\left(1+\dfrac{6\lambda}{g}\right)x^3 + \left(2+\dfrac{6\lambda}{g}\right)x_1\right]\dfrac{g^2 p_a}{12\mu}\nabla^2 x^3 - (1+x_3+x_1) \end{pmatrix}. \quad (4)$$

To solve the linearized equation numerically, the Equation (4) should be discretized in space, and then converted into state-space model. We use $(N+1)\times(M+1)$ mesh as shown in Figure 2, where N represents the number of inner grid points in the x direction and M is the number of inner grid points in the y direction. After the mesh is generated, we then project the unknowns $u(x,t)$ and $p(x,y,t)$ onto the mesh points and apply the trapezoidal rule to discretize the special integral operator and the central difference method to discretize the special derivative operators as follows:

$$\frac{\partial^2 y_i}{\partial x^2} = \frac{y_{i+1} - 2y_i + y_{i-1}}{\Delta x^2}, \quad (5)$$

$$\frac{\partial^4 y_i}{\partial x^4} = \frac{y_{i+2} - 4y_{i+1} + 6y_i + 4y_{i-1} + y_{i+2}}{\Delta x^4}. \quad (6)$$

Accelerometer beam clamped at one end, so the boundary conditions are: $y_0 = 0, y_{-1} = y_1$, substitute Equations (5) and (6) into Equation (4) to discrete, then map the n state vector:

$$X = \left[\, y_1 \ \ldots\ y_N \ \frac{\partial y_i}{\partial t} \ \ldots\ \frac{\partial y_N}{t} \ p_{11} \ \ldots\ p_{mn} \,\right]^T$$

to the various nodes, we can convert discrete equations into n dimensional state space Equation (7), where: $n = 2N + NM$.

$$\begin{cases} X(t) = Ax(t) + Bv(t) \\ y(t) = B^T X(T) \end{cases}, \quad (7)$$

where: $E = \begin{bmatrix} M & 0 \\ 0 & H \end{bmatrix} \in R^{n\times n}$, $A = \begin{bmatrix} N & F \\ -F^T & 0 \end{bmatrix} \in R^{n\times n}$, $B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \in R^{n\times P}$.

$y(t)$ is the micro displacement of the beam end, $v(t)$ is the input voltage of the system.

According to the two-dimensional model establishment of beam parameters as shown in the Figure 2. The beam was divided by with 2D quadrilateral element PLANE82, and the air medium was divided with PLANE121, displacement constraints were applied to fix the micro beam anchors in ANSYS, each degree of freedom beam base and left side of beam were restricted, the finite element model as shown in

the Figure 3 then applied effect of voltage on the beam structure. Figure 4 is the displacement of micro-beam.

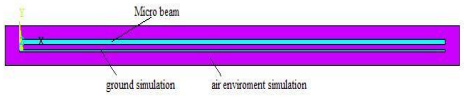


FIGURE 2 Structure Model of Micro Beam

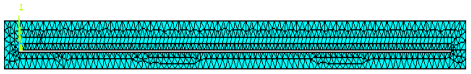


FIGURE 3 Finite Element Model of Micro Beam

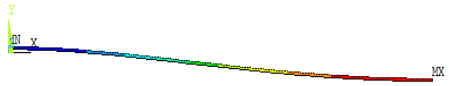


FIGURE 4 The displacement of Micro Beam

Applying the boundary conditions, the finite element model is solved can change the endpoint micro beam under different voltage conditions. The voltage load vector along with the mass and stiffness matrices were extracted and imported into Matlab. For problem simplification, viscous damping was not considered and Rayleigh damping was employed with both proportional damping constants equal to 0.25. Reduced models of size 10 and 30 were generated via the developed SPRIM reduction program. Reduced order models were numerically integrated in Matlab and reduced model solutions were expanded to full scale using Equation (8) to determine the micro beam displacement states. Algorithm stopping criteria were applied similar to those described in the electro structure results section. Figure shows the transient vertical displacement at the lower tip of the micro beam for the reduced and full models. As shown in [5], the simulation results of pull in voltages for the reduced and full models were both nearly in 8.7 V~8.8V, corresponding to relative percent errors of 2.2%.

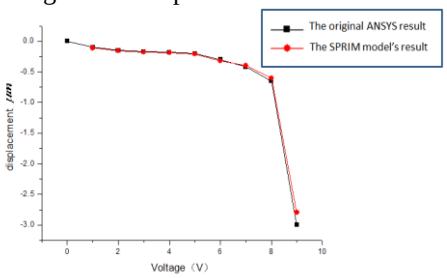


FIGURE 5 the transient vertical displacement of the micro beam for the reduced and full models

The reduced model is applied step signal under the condition of dynamic simulation, the original simulation model and reduced the response curve of the model, the simulation results are shown in Figure 6.

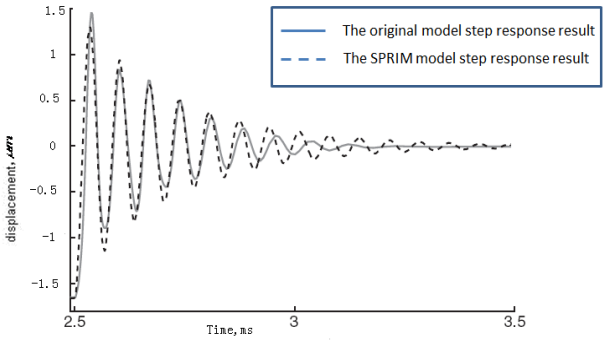


FIGURE 6 dynamic simulation comparison between the SPRIM model and original model

5 Conclusion

Model order reduction methods for coupled metaphysics FEA simulation of a common MEMS device are investigated in this paper. Simulation is performed for various sized reduced models of an electro structure actuated micro beam and compared with the full model simulation. Reduced model comparison illustrates that the reduction methods greatly increase the efficiency of FEA simulation with SPRIM models achieving simulation time reductions 84%, respectively, while SPRIM efficiency improvement was slower than PRIM model, but models experiencing relative errors of 2.2% is least. The relationship between accuracy and reduced model size is demonstrated by an increase in accuracy as model size increases.

Acknowledgments

The paper was support by the National Natural Science Foundation of China "MEMS Multi-fields Uniform Simulation Model Construction and Optimism Research" (61100101), the paper was also supported by Nature Science Foundation of Zhejiang Proviece "MEMS united constraints model order reduction method and solution research (LY14E050026)".

References

[1] Feng L 2005 Parameter independent model order reduction. *Mathematics and Computers in Simulation* **68**(3) 221-34
[2] Hu W, Hu G, Wei X, Xie X 2010 MEMS CAD Research Development and the Present Situation *Piezoelectrics and Acoustooptics* **32**(4) 682-91
[3] Cigada A, Leo E, Vanali M 2007 Electrical method to measure the dynamic behaviour and the quadrature error of a MEMS gyroscope sensor *Sensors and Actuators A* **134** 88-97
[4] Rudnyi E B, Korvink J G 2002 Review: Automatic model reduction for transient simulation of MEMS-based devices *Sensors Update* **11**(1) 3-33
[5] Niessnera M, Schraga G, Iannacci J, Wachutka G 2011 Macromodel-based simulation and measurement of the dynamic pull-in of viscously damped RF-MEMS switches *Sensors and Actuators A* **172** 269-79
[6] Hannot S D A, Rixen D J 2011 Building and Reducing a Three-Field Finite-Element Model of a Damped Electromechanical Actuator. Journal of Microelectromechanical Systems **20**(3) 665-76
[7] Kohler A, Reitz S, Schneider P 2012 Sensitivity analysis and adaptive multi-point multi-moment model order reduction in MEMS design *Analog Integr Circ Sig Process* **71** 49-58

[8] Kohler A, Reitz S, Schneider P 2011 Sensitivity Analysis and Adaptive Multi-point Multimoment Model Order Reduction in MEMS Design *Design, Test, Integration and Packaging of MEMS/MOEMS* 64-71

Authors

Xiangjuan Bian, born in May, 1979, Shandong, China

Current position, grades: lecturer
University studies: Kunming University of Science and Technology
Scientific interests: MEMS design, simulation, data base, design and manufacturing of microsensors and microactuators.

Youping Gong, born in July, 1978, Chongqing, China

Current position, grades: associate professor at Hangzhou Dianzi University. Visiting scholar at the University of Florida in 2014.
University studies: graduated at Zhejiang University in 2007.
Scientific interests: MEMS design, simulation, design and manufacturing of microsensors and microactuators.

Liyun Zheng, born in 1989, Zhejiang provience, China

Current position, grades: master's degree student at the School of Hangzhou Dianzi University.
University studies: bachelor degree at the School of Mechanical Engineering Hangzhou Dianzi University in 2012.
Scientific interests: modeling and simulation optimization design research of MEMS devices.

Wireless monitoring system for temperature and humidity based on ZigBee

Jianjun Chen*

Shaoxing University Yuanpei College, Shaoxing, Zhejiang, 312000, China

Received 1 March 2014, www.cmnt.lv

Abstract

Traditional methods of environmental monitoring have the shortages including difficult network layout and low intelligent of node. A monitoring system for temperature and humidity was designed based on ZigBee (wireless network), with SHT11 (temperature and humidity integrated sensor) and CC2430 (wireless radio frequency module integrated with MCU). Data of temperature and humidity, collected at some acquisition terminals in SHT11 and CC2430, were displayed on monitoring host via centre node. By using solution of single chip ZigBee, the system was optimized with high precision, low power dissipation and simple equipment.

Keywords: CC2430 wireless network; monitoring temperature and humidity

1 Introduction

Environmental temperature and humidity are monitored in fields of biopharmaceutics, food processing, paper making and so on. Functions of current integrated intelligent sensors including such as linearization treatment, automatic temperature compensation and humidity calibration, provide favourable conditions for development of temperature and humidity monitoring system [1]. Wireless network can be easily built with ZigBee technology. Data are secure and stable during transmission with little influence from the terrible circumstance such as outside noise and strong magnetic field. In this work a wireless monitoring system for temperature and humidity was built with CC1010 (ZigBee wireless radio frequency module) and SHT11 (temperature and humidity sensor module). Measuring signal of local sensor is sent to monitoring terminal without setting quantities of temperature cables in conservatory. The system was optimized with real-time performance, high precision and overall treatment of nodes' temperature, as well as simplified installation and cost reduction.

2 Overall plan of temperature and humidity monitoring system

A star-like wireless sensor network was built with several ZigBee acquisition terminals and a ZigBee centre node (coordinator). Temperature and humidity sensor in the acquisition terminals collected environmental temperature and humidity data. The data were corrected by 8051 MCU embedded in ZigBee wireless radio frequency module [2]. Then, the corrected data were sent to centre node by wireless chips via ZigBee wireless network. All acquisition terminals have the same functions and implementation methods. Centre node was composed of a ZigBee wireless radio frequency chip and a USB module. As a network coordinator, the chip was responsible for generation, management and maintenance of the network. In detail, its duty included network address distribution for new equipment, access and departure of nodes, distribution and update of network security keys, etc. Furthermore, the chip should upload collected data to monitoring host, and send out orders from the host.

Monitoring host accepted all data of temperature and humidity and displayed them on PC screen for monitoring. The host periodically read serial port buffer and put data of from centre node into cache. Then, temperature and humidity data were extracted and decomposed by host. ID in each data packet was used to confirm data source. After that, the host displayed temperature and humidity information in real time. The host needed to judge residual energy through working voltage of acquisition terminal nodes. If the voltage was too low, the reliability of sensor data would be greatly reduced. And in that case there would be an alarm of battery replacement. Figure 1 shows structure of overall network.

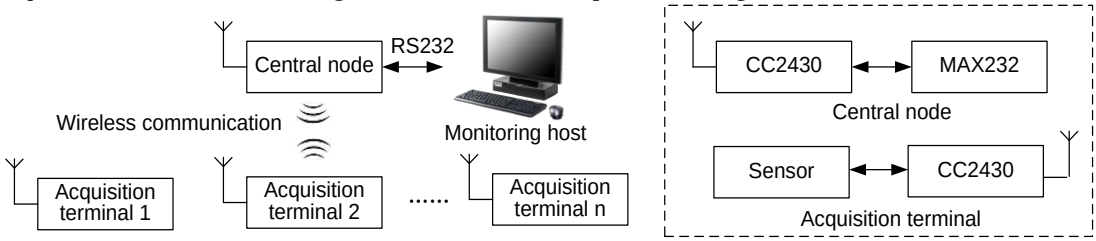


FIGURE 1 Overall design of the monitoring system

***Corresponding author** e-mail: sxchenjianjun@163.com

3 Hardware of temperature and humidity monitoring system

3.1 DESIGN OF ACQUISITION TERMINAL HARDWARE

Acquisition terminal is composed of microprocessor module (with internally-integrated wireless radio frequency module), temperature and humidity data acquisition module and power module. Temperature and humidity data measured by sensor are processed by microprocessor chip and sent to centre node via wireless radio frequency. Figure 2 shows the structure of acquisition terminal. Instead of utilizing the combination of microcontroller and ZigBee, terminal with single ZigBee chip is optimized with high precision, low power dissipation and simple equipment.

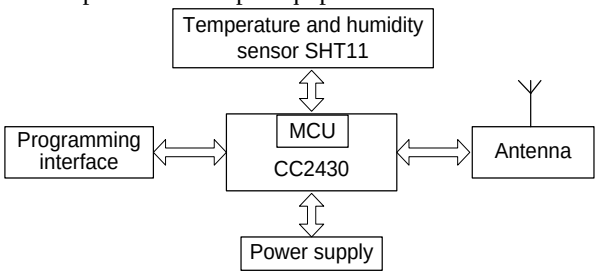


FIGURE 2 Overall block diagram of the acquisition terminal

3.1.1 Wireless RF transceiver chip CC2430 [3]

As core of acquisition terminal, CC2430 is a radio frequency chip produced by Chipcon, with low power dissipation and multiband. It is internally integrated with a 2.4 GHz RF transceiver of direct sequence spread spectrum, as well as an enhanced industrial microprocessor 8051. Microprocessor 8051 has an 8 KB RAM, a programmable flash memory, several timers (including a watchdog timer), a collaborative processor, a 32 KHz crystal oscillator sleep timer, 21 I/O connectors, power detection circuit and power on reset circuit.

CC2430 is also used in wireless radio frequency chip of centre node. Register RFMAIN determines receiving or transmission mode of chip. Seventh RXTX of 8-RXTX RAMAIN can be used to set modes: 0 as receiving mode and 1 as transmission mode.

3.1.2 Temperature and humidity sensor SHT11

SHT11, designed by Swiss Sensirion, is an integrative temperature and humidity sensor with two-wire serial interface. A capacitive polymer humidity sensor can detect the humidity, and energy gap temperature device is for temperature detection. Devices transform temperature and humidity information into weak electrical signals. Signals are amplified by operational amplifier and then directly sent into A/D converter, outputting digital signals via two-wire serial interface bus.

For convenience, debugging and calibration of data are implemented within chip. Besides, users can set measurement accuracy with resolution up to 8, 12 or 14 bits. CRC checking in chip ensures the accuracy of data transmission. Figure 3 shows the hardware connection diagram of SHT11 and MCU. Besides, DATA cable should be externally connected with pull-up resistor. Clock line SCK is used for communication between microprocessor and SHT11. SCK has no limitation of minimum frequency because its interface contains a whole static logic. If working voltage was higher than 4.5 V, frequency of SCK will be up to 10 MHz; else, maximum frequency of SCK will be only 1 MHz.

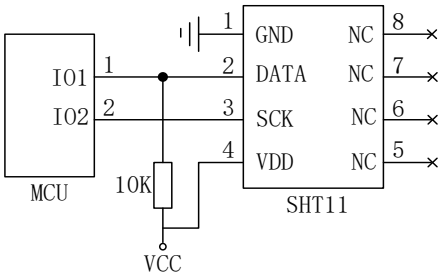


FIGURE 3 MCU connected with SHT11

3.2 HARDWARE OF CENTRE NODE

Centre node is composed of a ZigBee wireless radio frequency CC2430 and a USB MAX232. As a network coordinator, CC2430 can build, maintain and cease ZigBee wireless network [4]. CC2430 is connected with thirteenth pin $R1_{IN}$ and fourteenth pin $T1_{OUT}$ of MAX232 serial module with two serial lines - TXD0 and RXD0. Data are periodically sent to monitoring host via MAX232 USB (see Figure 4).

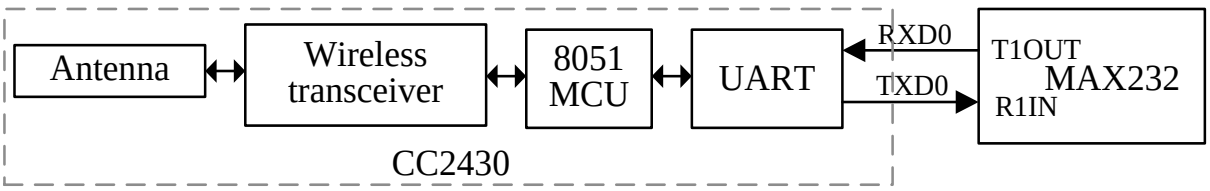


FIGURE 4 Block diagram of central node

CC2430 is regarded as the wireless module for acquisition terminal and centre node because it has four low power modes - PM0, PM1, PM2 and PM3. Largest power dissipation is on PM0 mode, while least power dissipation on PM3 mode. On normal state (PM0), wireless communication module will keep monitoring wireless channel to checks whether data have came. Wireless communication module is closed on other low power modes. Largest energy dissipation emerges on receiving state in normal work, while

slightly less on delivery state. Least energy dissipation results from dormant state. Therefore, system can enter dormancy as soon as possible for low power dissipation, with needless communication [5].

4 Design of system software

Software of system should have functions including building a ZigBee wireless LAN, collection, revision and delivery of temperature and humidity data, packaging, transport and display of each acquisition terminal data. Software design includes three parts, namely the design of acquisition terminal software, centre node software and upper computer system software.

4.1 SOFTWARE OF TEMPERATURE AND HUMIDITY DATA COLLECTION

After acquisition terminal is power on, it will initialize hardware module and ZigBee protocol stack. Then, node searches ZigBee network nearby for accession. Once accessed, terminal turns into low power mode.

Collection of temperature and humidity information included two types – automatic and manual collection. Automatic collection was based on timer at terminal. Timer started after setting an initial value. It would cause interruption when time came to the set value, and the measurement of temperature and humidity began. After measuring, data were sent out via radio frequency, with terminal entering low power mode. Timer was reset for next timing. Thus, automatic collection can be achieved with such circulation.

In manual collection, monitoring host will send out query demand to appointed terminal. Sensor started to collect and process temperature and humidity data once terminal identified the instruction. Then, data were immediately uploaded to monitor host, achieving manual collection. Figure 5 shows the collection flow.

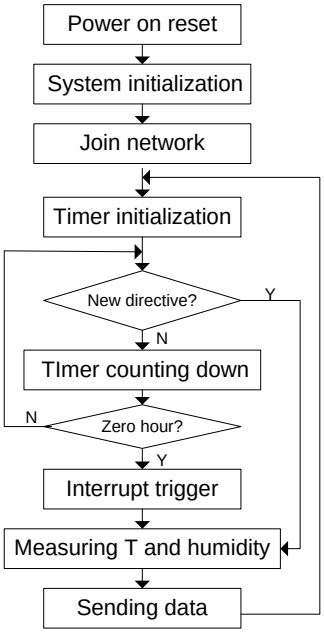


FIGURE 5 Flow chart of acquisition

SHT11 is a temperature and humidity integrated sensor with two measurement commands: 03H for temperature and 05H for humidity measurement. Temperature and humidity data, transferred from SHT11 to CC2430, were nonlinear relative value. The data was sent to centre node after linear compensation by using 8051 MCU embedded in CC2430 chip. Without correction and compensation, the data can be directly used by centre node, saving time on the correction of relative value. Humidity data of 12 bit was converted to actual humidity value. Relative humidity data N can be compensated with Equation (1).

$$RH_L = (-4 + 0.0405N - 2.8*10^{-6} N^2)\%, \quad (1)$$

where N is collected humidity data, and RH_L the transferred humidity value. For temperature's influence on humidity sensor, temperature correction for humidity sensor should be considered when actual temperature was far from 25°C (-77°F). Equation (2) was used for calculation of actual humidity value (represented as RH_T) with temperature compensation.

$$RH_T = RH_L + (T - 25)(0.01 + 0.00008N)\%. \quad (2)$$

Environment or other factors have little effect on SHT11 for its excellent linearity. Digital output of SHT11 can be directly transferred to temperature value with Equation (3). Value of d_1 was determined by working voltage of sensor, while value of d_2 was determined by resolution of sensor [6].

$$Temperature = d_1 + d_2 * SO_T. \quad (3)$$

4.2 SOFTWARE DESIGN OF CENTRE NODE

Software with star-like network topology was designed for centre node. As a coordinator, centre node had two major functions. One was to build a ZigBee wireless network. Web beacons were sent to domain spaces by CC2430 via antenna, so as to inform bound terminal nodes of establishment of wireless network.

Besides, acquisition terminal nodes should receive and reply requests for accession of network. PANid of request equipment was judged. If PANid was concordant, entrance of acquisition terminal was authorized, with an assigned unique 16 bit short address in network.

After ZigBee wireless network was built, acquisition terminal periodically sent temperature and humidity data to centre node via ZigBee wireless network. Useful information, such as temperature and humidity data and the unique short address, were extracted by centre node with accession of data packet. All the information from terminal nodes was processed to be a data packet with structure form. Then, data were then sent to monitoring host through MAX232 serial module. Figure 6 shows the process.

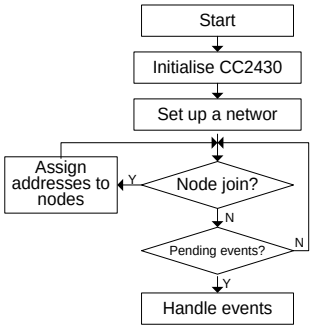


FIGURE 6 Flow chart of the main program in coordinator

4.3 SOFTWARE OF UPPER COMPUTER

PC was used as upper computer, mainly accepting data from centre node via serial port. Moreover, friendly human-machine interface should also be provided for real-time display of data, as well as storage management. Software of upper computer was developed by using VC++6.0 with Mscomm communication control. This control provided all functions of serial communication and achieved data reading or writing via serial port.

With simple programming and convenient debugging, the control encapsulated underlying operation procedure in the process of communication. Users only needed to set up and monitor properties and events of control. Asynchronous serial communication between users and applications can be easily achieved. Figure 7 shows the partial interfaces of monitoring software.

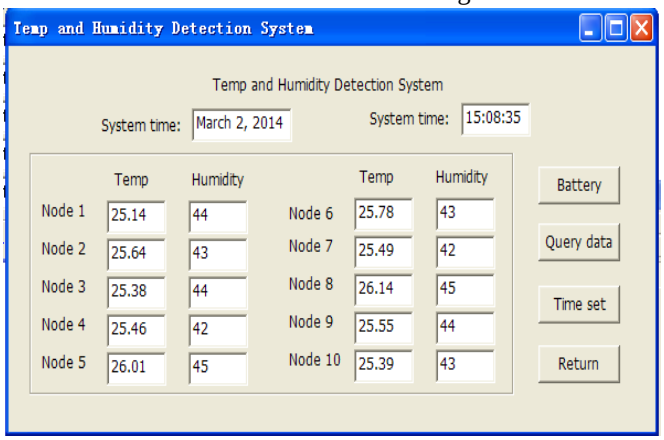


FIGURE 7 Temperature and humidity detection system

Temperature and humidity data management database was established with Microsoft Access 2003 in background. The database covered information of nodes number, temperature, humidity, acquisition time, etc. Table 1 was established in Access for the record and management of nodes formation.

TABLE 1 The node's information

Node number	Temp	Humidity	Battery power	Acquisition time
01	21	65%	48%	10:18:30

Data management of temperature and humidity acquisition terminal became convenient with the database design in this system. With friendly interface display, all databases of this system had functions such as adding, deleting and inquiry.

5 Experimental results

Two 1.5V batteries supplied power for centre node and ZigBee in the experiment. Different programs for centre node and ZigBee acquisition terminal nodes were downloaded by simulator before experiment. Ten acquisition terminals located in 10 different sites, with direct communication to centre node. Centre node was connected with computer via serial port, uploading received data to the computer.

Table 2 shows experimental data of 10 nodes. Test results showed that effective communication distance was about 50 meters in outdoors without obstacle. If there was an obstacle, communication in indoor was stable in 20 meters.

TABLE 2 Experimental values of temperature and humidity

	9:00		10:00		11:00		12:00		13:00		14:00	
Time	*T*	*H*	*T*	*H*	*T*	*H*	*T*	*H*	*T*	*H*	*T*	*H*
Node 1	25.0`4	45	25.45	43	25.87	44	26.01	45	26.31	44	26.28	45
Node 2	25.42	44	25.56	44	25.58	45	26.03	43	26.08	44	26.14	43
Node 3	26.01	43	26.13	42	26.42	43	26.55	44	26.64	43	26.58	43
Node 4	25.40	44	25.38	44	25.58	42	25.94	42	26.12	44	26.20	44
Node 5	25.14	45	25.28	44	25.31	43	25.33	43	25.39	45	25.48	44
Node 6	25.10	43	25.24	44	25.33	45	25.29	43	25.48	44	25.60	43
Node 7	26.14	43	26.18	43	26.20	43	26.33	43	26.38	44	26.46	43
Node 8	26.21	42	26.33	44	26.40	44	26.46	44	26.40	43	26.38	45
Node9	25.04	43	25.14	45	25.20	42	25.36	45	25.44	43	25.44	43
Node10	25.00	45	25.14	43	25.26	43	25.38	44	25.31	43	25.42	44

6 Conclusions

A ZigBee wireless network was built with ZigBee radio frequency chip CC2430. Data are sent and received on wireless, achieving point-to-multipoint data transmission and receiving. System can be utilized in many occasions to ensure the accuracy of data. With the same functions of wired temperature and humidity monitoring system, wireless system avoids the limitation of wiring, with low power consumption and high reliability. In empty place without obstacles, communication distance can reach as far as 50 meters, satisfying temperature and humidity monitoring.

Acknowledgement

The work was funded by Science and Technology Project of Zhejiang (No 2012C21041), "R & D of Efficient Wireless Router of Dual-Band Energy Based on 6LoWPAN".

References

[1] Kinney P 2003 ZigBee Technology: Wireless Control that Simply Works *IEEE Journal of Solid-State Circuits* 190-5
[2] Drew G, Tim G 2004 ZigBee wireless sensor networks-ZigBee is an emerging wireless protocol designed for low-cost, high-reliability sensor networks *Software Tools for the Professional Programmer* (29) 40-3
[3] Polastre J, Szewczyk R, Mainwaring A, Culler D, Anderson J 2004 Analysis of wireless sensor networks for habitat monitoring *Wireless Sensor Networks* 399-423
[4] Mittal, Ruchi, M.P.S Bhatia 2010 Wireless Sensor Networks for Monitoring the Environmental Activities *IEEE International Conference on Computational Intelligence and Computing Research ICCIC* 348-52
[5] TI, CC2430-A True System-on-Chip solution for 2.4 GHz IEEE 802.15.4/ZigBee [EB/OL] REV2.1 2007 http://focus.ti.com/docs/prod/folders/print/cc2430.html
[6] SENSIRION Inc Datasheet SHT1x (SHT10, SHT11, SHT15) Humidity and Temperature Sensor [EB/OL] REV4.0 2008 http://www.sensirion.com.cn

Author

Jianjun Chen, born in September, 1966, Shaoxing, Zhejiang Province, China

Current position, grades: associate professor, master in Shaoxing University Yuanpei College, China.
University studies: computer science and technology.
Scientific interest: computer network and graphics and image processing.
Publications: 25 papers.

Analyses on flow and heat transfer performance and of heat exchanger with continuous helical baffles

Jiazhu Zou, Fengwei Yuan*, Qian Deng

College of Mechanical Engineering, University of South China, Hengyang 421001, Hunan, China

Received 1 October 2014, www.cmnt.lv

Abstract

A numerical simulation for heat exchanger with continuous helical baffles was carried out. The study focuses on the effects of helix angle on flow and heat transfer characteristics, and heat exchanger performance is evaluated by entropy generation number based on the analysis of the second law of thermodynamics. The results show that both the shell-side heat transfer coefficient and pressure drop decrease with the increase of the helix angle at certain mass flow rate. The latter decreases more quickly than the former. The tangential velocity distribution on shell-side cross section is more uniform with continuous helical baffles than with segmental baffles. The axial velocity at certain radial position decreases as the helix angle increases in the inner region near the central dummy tube, whereas it increases as the helix angle increases in the outer region near the shell. The heat exchange quantity distribution in tubes at different radial positions is more uniform at larger helix angel.

Keywords: continuous helical baffles, heat exchanger, helix angle, entropy generation, numerical simulation

1 Introduction

The conventional shell and tube beat exchange most used in the arch baffles, and the shell side fluid flows along the side with large pressure. There are the flow and heat transfer dead regions in the shell side. It is not only easy scaling, but that the high velocity easily induced vibration of heat exchange tube and shorten the service life. For this kind of situation, from the Angle of the change of baffle plate arrangement, put forward the thought of the spiral baffle plate heat exchanger, the application in the shell side of the heat exchanger along the axial of the spiral baffle plate structure, made fluid helical plunger flow. Flow uniform stability, thus overcome the bow baffle plate heat exchanger of the above shortcomings. Existing research shows that compared with the bow baffle plate heat exchanger, spiral baffle plate heat exchanger when the shell side of the same flow unit pressure drop of the heat transfer coefficient is higher, especially for the high degree of double fluid, its advantage is more outstanding.

Most of current research of the spiral baffle plate heat exchanger to apply a set of method and the heat exchange tube axial Angle of fan-shaped certain plane board to lap, approximate spiral surface formation, called the continuous helical baffle plate. Although it is easy to manufacture installation, but this kind of structure because of the discontinuity of the baffle plate and the triangular space is formed between two adjacent frames. The shell side fluid in the region will form a short circuit leakage flow, lead to flow deviation from the ideal spiral flow, thus affecting the performance of heat transfer in the heat exchanger.

To make the shell side fluid to achieve relatively smooth continuous spiral flow, baffle plate should be continuous spiral curved surface, called continuous helical baffles. Baffle plate in the shell on the formation of the spiral curve tangent with the angle β between the cross section is defined as the continuous helical baffle helical angle.

$$\beta = \arctan\left(\frac{H_s}{\pi D_s}\right)_{\varphi}. \tag{1}$$

The shell side of the helical angle determines the Angle between the direction of fluid flow and heat exchange tube bundle, boundary layer thickness and the flow characteristics and so on, in the shell inside diameter must, helical angle determines the cross-section of the shell side of the pitch and circulation, directly affect the shell side heat transfer and resistance properties. So, helical angle is reflecting the continuous helical baffle heat exchanger thermal performance of the important parameters.

Experimental study of continuous helical baffle heat exchanger is made the high installation costs, cycle is long, and not easy to get the distribution of the shell side quantities, this paper established the continuous helical baffle heat exchanger of three-dimensional physical model. A numerical simulation for heat exchanger with continuous helical baffles was carried out by using commercial codes of ANSYS CFD 12.0. The study focuses on the effects of helix angle on flow and heat transfer characteristics, and heat exchanger performance is evaluated by entropy generation number based on the analysis of the second law of thermodynamics.

2 The physical model and geometric dimensions

The continuous helical baffle heat exchanger shell side application, because in the actual processing baffle plate screw

***Corresponding author** e-mail address: 1057487156@qq.com

surface tangent direction of the centre and the shell axial Angle is very small, drilling difficulties, usually equipped with centre false tube helical surface centre position, no false tube fluid flow, the numerical simulation model of heat exchanger is applied with centre of false tube structure. Continuous helical baffle heat exchanger the calculation area physical model as shown in Figure 1, the basic structure size reference GB 151-151, as shown in Table 1.

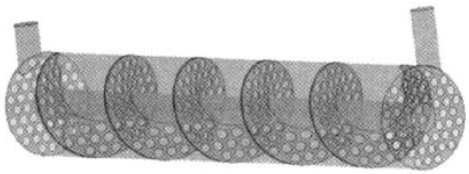


FIGURE 1 Physical model calculation area

TABLE 1 Geometric parameters of heat exchanger

Parameter	Value
inner diameter of shell/mm	207
diameter of central dummy tube/mm	60
diameter of tube/mm	19
central distance of tubes/mm	25
arrangement of tubes	regular triangle
number of tubes	48
helix angle/(°)	10,20,30,40,45
helical pitch/mm	115,237,376,546,650
tube length/mm	1750

3 Mathematical model and numerical simulation method

3.1 TURBULENCE MODEL AND THE CONTROL EQUATION

Continuous helical baffle heat exchanger shell side turbulent pulsating flow with spiral flow interaction, to show high levels of anisotropy and sensitivity to the streamline curvature, so numerical simulation using RNG k model, its strong streamline curvature, the vortex and has good performance of rotating calculation, at the same time in close to the constant application of standard wall function method.

General control equation for related physical quantities

$$div(\rho U\varphi) = div\left(\Gamma_{\varphi} grad\varphi\right) + S_{\varphi} . \quad (2)$$

Substituting different general variables ϕ, Equation (2) to show the continuity equation, momentum equation and energy equation and the equation of $k-\varepsilon$, the generalized diffusion coefficient Γ_{ϕ} and generalized the values S_{ϕ} in the source term is corresponding to each equation.

3.2 WORKING MEDIUM FLOW AND BOUNDARY CONDITIONS

Shell side working medium for heat conduction oil, its steady turbulent flow in the shell side channel. Because of heat conduction oil, the degree of value changing with temperature is bigger, it is a function of temperature to the fitting $\mu = 5.18125\times10^{-6}T^2 - 3.69\times10^{-3}T + 0.66233$, the rest of the physical parameters of constant. In the working medium is water passes and physical parameters are constant. Both gravity, thermal buoyancy lift and clear heat dissipation.

To determine the appropriate boundary conditions on the physical model of simplified. A given shell side inlet mass flow rate M_s, temperature $T_{s,i} = 353.15k$ and turbulence intensity I = 5%; Tube side population rate constant, $\upsilon = 1.5m\cdot S^{-1}$ define temperature $T_{s,i} = 353.15k$ and turbulence intensity I = 5%. The shell side and tube side exports are pressure outlet boundary, given the recycling of static pressure and the appropriate conditions. No sliding surface is defined as the heat exchange tube impermeable solid wall, the application of coupling boundary conditions of cold and hot fluid heat transfer FGC calculation; Baffle plate and shell wall impermeable solid wall is defined as adiabatic without sliding.

3.3 MESHING WITH NUMERICAL SIMULATION METHOD

The shell side of the continuous helical baffle heat exchanger internal structure is complex, so the application of tetrahedron and pyramidal unstructured meshing; in the application passes hexahedron mesh. Selected unit number is respectively 11.82 million, 14.92 million, 17.78 million, 3 sets of grid independence test, when $M_s = 10Kg\cdot S^{-1}$, the spiral Angle of 40°of continuous helical baffle heat exchanger by the 3 sets of the grid to calculate the shell side of heat transfer coefficient and pressure drop, respectively in the import and export $446.1W\cdot m^{-2}\cdot K^{-1}$, $70164.1Pa$,$452.2W\cdot m^{-2}\cdot K^{-1}$ and $65891Pa$,$456.5W\cdot m^{-2}\cdot K^{-1}$ and 65035. 9Pa, after the 2 sets of grid computing results are within 2%. Considering calculation accuracy and efficiency, choose 2 sets of grid computing.

Finite volume method was applied to calculate area and control equations are dispersed and solving, define the convergence condition of the equation of average residual error absolute value is less than 1.0×10^{-5}.

3.4 NUMERICAL SIMULATION VALIDATION

Applying the method of continuous helical baffle heat exchanger are simulated, and the results were compared with the experimental data. Figures 2 and 3 shows the shell side pressure drop simulation value and experiment value deviation for 13. 800-21.2%, the shell side heat transfer coefficient of deviation for the simulation value and experiment value 7.8% - 8%, deviation in a reasonable scope, the validity of the numerical simulation method presented in this paper.

Besides inevitable error of measurement, numerical simulation of the heat exchanger of simplified model and boundary conditions may cause the deviation between the numerical simulation results with the experimental data, for example, numerical simulation of the model ignores the baffle plate and shell and the leakage flow between heat exchange tube, baffle plate and shell wall is simplified to adiabatic boundary condition, etc.

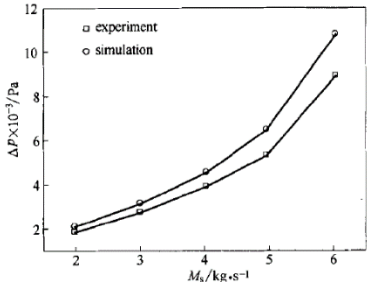


FIGURE 2 Shell side pressure drop experimental value compared with the simulation values

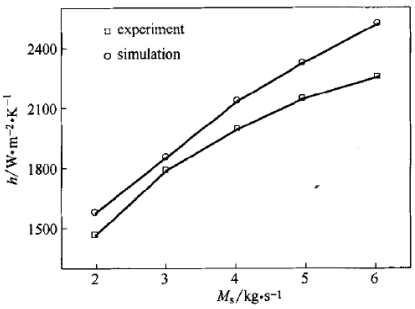


FIGURE 3 Shell side heat transfer coefficient of the experimental value and simulation value comparison

The effect of leakage flow and significantly with the increase of shell side flow, so the deviation between the simulated results with the experimental data with the flow rate increases.

4 The results of numerical simulation and analysis

4.1 SHELL SIDE SCREW CHANNEL AXIAL DISTRIBUTION OF CONVECTIVE HEAT TRANSFER

Continuous helical baffle heat exchanger shell side fluid in the shell, baffle plate, centre pipe and heat exchange tube wall of channel flow, heat transfer along the flow direction of the development process, with the development of the fluid mechanics was conducted at the same time. To high pr, such as heat conduction oil fluid, in fully developed section, heat transfer rate have also been fully developed. So, the shell side of the can from cross section along the axial distribution of local heat transfer coefficient on the shell side of the judgment convective heat transfer full development period of spiral channels.

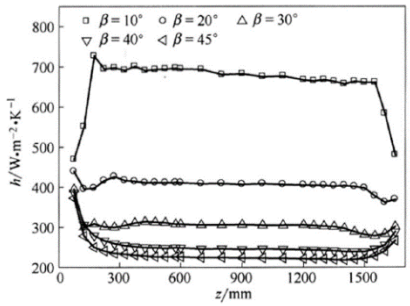


FIGURE 4 Screw channel and the distribution of local heat transfer coefficient along the axial direction

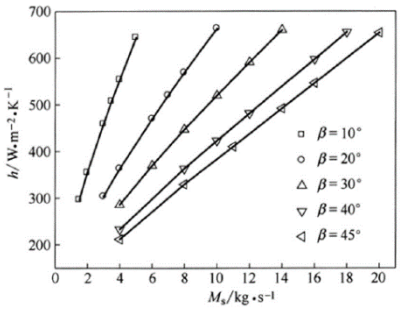


FIGURE 5 Shell side heat transfer coefficient changing with the flow

To decorate with continuous helical baffle and processes do not include the shell The shell side of the outlet pipe of the channel for the screw channel, can see from Figure 4, the shell side mass flow rate is equal to ($M_s = 4Kg \cdot S^{-1}$). The screw channel and local heat transfer coefficient *h* and decreased with the increase of helical Angle. This is because the spiral Angle increases means that the circulation of the shell side of the pitch increases, sectional area increase, results in the decrease of flow velocity, *h* is reduced.

Helix Angle is not at the same time, the axial distribution of *h* is very different: the spiral Angle of 10°, *h* decreased slightly after a sharp increase is seen at the inlet section of the first, and then its value changed little, marked the convective heat transfer into the full development stage, in the export section h drastically reduced; Spiral Angle of 20° and 30°, h in the inlet and outlet section shows the trend of increase with the decrease of, after first; Spiral Angle of 40° and 45°, *h* in inlet and outlet, respectively is the decrease and increase the monotonicity of the trend. This is because of the inlet and outlet of the spiral channel flow at the inlet and outlet of the cross-sectional area by the baffle plate and tube plate of the distance between the decision, and fully development period of the circulation of sectional area is determined by the baffle plate pitch, when the helix Angle is small, the former is greater than the latter, made of import and export as full development period of high flow velocity, and can lead to the development of import and export of *h* is less than fully; with the increase of helical Angle, the situation is the opposite.

In the comprehensive analysis, fluid flow people screw channel after about 1 times the length of the pitch of the part is import section, flow channel is about 0.75 times the length of the pitch before the part of the period for export, import and export for the part between the convective heat transfer full development period.

4.2 FULL DEVELOPMENT PERIOD OF HEAT CONVECTION OF SHELL SIDE HEAT TRANSFER AND RESISTANCE PERFORMANCE

As shown in Figures 5 and 6, the shell side of the same mass flow, the shell side heat transfer coefficient of convective heat transfer full development period of h and pressure drop per unit length Δp_m are reduced with the increase of helical Angle. When $M_s = 4Kg \cdot S^{-1}$ compared with the spiral

Angle of 10°, h and Δp_m decrease respectively 34.4% and 87.9% at 20° and respectively lower 48.3% and 96.5% at 30°, respectively reduce 57.9% and 98.6% at 40°, respectively 61.7% and 99.1% at 45°, the lower the trend will be increased with the increase of flow rate is more and more significant.

This is because, on the one hand, the shell side of the pitch and circulation area increases with the increase of helical Angle, reduce with traffic flow and h and Δp_m so reduce. The shell side of the small spiral Angle, on the other hand, closer to the fluid flow form horizontal tube bundle, same velocity when the heat transfer effect is better. But also inevitably bring large pressure loss. Because along with the increase of helical Angle, Δp_m lower margin than h, significantly, Figure 7 shows the shell side heat transfer and resistance performance will increase along with the increase of helical Angle.

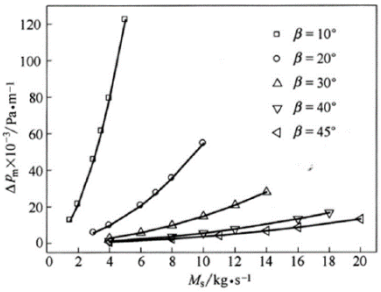


FIGURE 6 Shell side pressure drop per unit length along with the change of flow rate

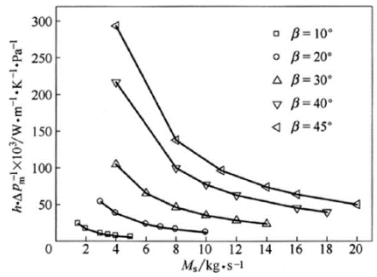


FIGURE 7 Shell side comprehensive performance along with the change of flow

4.3 THE SHELL SIDE VELOCITY DISTRIBUTION

Fluid in the shell side channel spiral flow, the velocity can be broken down for the axial velocity υ_a and tangential velocity υ_θ and radial velocity υ_r. υ_a produce longitudinal from the effect of heat exchange tube bundle, υ_θ produce the effect of horizontal sweep bundles, υ_r that secondary flow is generated by the bundle disturbance. the shell side of the three component makes the function of flow between the longitudinal flow and transverse flow, both of the two characteristics: compared with the longitudinal flow, spiral flow of the tangential component of the shell side fluid velocity gradient along the shell radius direction, destroyed the boundary layer on the surface of the heat exchange tube bundle, enhanced heat transfer; Compared with horizontal flow, spiral flow axial component of the plunger shape is closer to the fluid flow, improve the heat transfer temperature difference.

The shell side of the Figure 8 for the spiral Angle of 40° when convective heat transfer full development period of the velocity vector distribution on cross section. Can see that, because of the continuous helical baffle diversion, formed on the shell side of the fluid in the cross section of regular rotation flow, the distribution of υ_θ is evener, compared with the traditional bow baffle plate, eliminates the flow dead zone, and to reduce the impact of the fluid exchange heat pipe bundle.

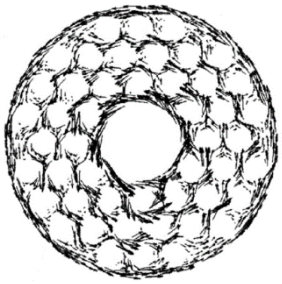


FIGURE 8 The velocity vector distribution on cross section

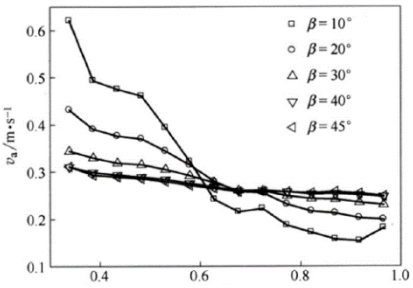


FIGURE 9 Radial distribution of axial velocity along the shell

Bundle between the medial and centre false wall surface and bundle the wall between the lateral and shell because of the heat exchange tube bundle of turbulence effect influence is weak, low flow resistance, fluid showed more significantly than control internal rotation, formed by the spiral flow in two places.

Figure 9 shows the shell side mass flow rate is equal to ($M_s = 4Kg \cdot S^{-1}$), the different spiral Angle υ_a. And decreased with the increase of the radial position δ ($\delta = 2r / D_s$), centre near false tube υ_a is greater than other location. Spiral Angle at 10, υ_a radial along the shell to reduce the sharpest, centre near false tube of shell wall near 4 times, poor distribution uniformity. Gradually, with the increase of helical Angle υ_a along the radial distribution even, $\beta \geq 40^\circ$, the radial distribution uniformity along the shell is better. Shell side spiral channels of distribution can be divided into two areas, near the centre of false tube inner regions, the same radial location along with the increase of helical Angle, and the area near the wall of the outer shell, the same radial position is increased along with the increase of helical Angle.

4.4 SURFACE OF HEAT EXCHANGE TUBE HEAT TRANSFER PERFORMANCE OF THE DISTRIBUTION

Equals the Figure 10 for shell side mass flow rate ($M_s = 4Kg \cdot S^{-1}$), different radial position of the surface of the heat exchange tube Q^* , the change rule of dimension 1 in heat $Q^* = Q_\delta / Q_{\delta=0.42}$. Can see, the spiral Angle is not the change rule of similar at the same time, in addition to the false tube closest to the centre of heat exchange tube, the rest of the radial position of the heat exchange tube are less than 1, indicates that the most close to the centre of the fake tubes due to heat exchange tube bundle inside between the centre of false wall surface and side stream, the influence of surface heat exchange tube in the heat more.

With the increase of radial distance, Q^* downward trend, close to the shell by heat exchange tube bundle of the lateral wall and the influence of the side stream of the shell wall, there will be a certain degree of rise. The smaller spiral Angle, change the magnitude of the more severe, the radial heat exchange tube in different radial positions at this time in exchange for heat distribution inhomogeneity between the stronger, the thermal stress is, the greater the this is the problems that should be paid attention to in the design and use of heat exchanger. When $\beta \geq 30^\circ$ the heat transfer between the heat exchange tube in different radial positions distribution uniformity is better.

4.5 ENTROPY GENERATION ANALYSIS

The second law of thermodynamics, points out that all the actual process in the nature are not reversible. Continuous helical baffle heat exchanger as energy transfer device, the heat transfer process is a typical irreversible process, the energy of the irreversible process will inevitably lead to the loss of available energy. The irreversible loss in the heat exchanger is divided into the irreversible loss caused by finite temperature difference heat transfer and fluid flow in the process of overcoming friction resistance caused by irreversible loss, the sum of the total irreversible loss in the heat exchanger.

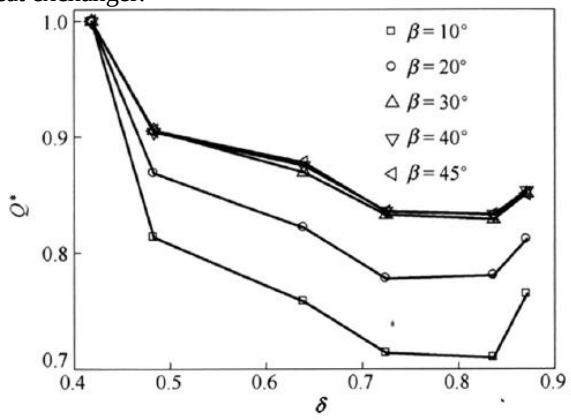


FIGURE 10 Surface of heat exchange tube dimension 1 heat transfer along the radial distribution

The irreversible loss in the heat exchanger can use entropy generation, production rate of heat transfer caused by the entropy is:

$$\dot{S}_{gen,\Delta T} = (Mc_p)_s \mathrm{In}\frac{T_{s,o}}{T_{s,i}} + (Mc_p)_t \mathrm{In}\frac{T_{t,o}}{T_{t,i}}. \quad (3)$$

For incompressible fluid, Resistance to flow caused by the production rate for entropy is:

$$\dot{S}_{gen,\Delta p} = M_s \frac{\Delta p_s}{\rho_s}\frac{\mathrm{In}(T_{s,o}/T_{s,i})}{T_{s,o}-T_{s,i}} + M_t \frac{\Delta p_t}{\rho_t}\frac{\mathrm{In}(T_{s,o}/T_{s,i})}{T_{s,o}-T_{s,i}}. \quad (4)$$

Total entropy yield:

$$\dot{S}_{gen} = \dot{S}_{gen,\Delta T} + \dot{S}_{gen,\Delta p}. \quad (5)$$

Be-Heat transfer caused by the entropy generation and the ratio of the total entropy generation Can be said in the heat exchanger caused irreversible loss caused by heat transfer and flow resistance of the relative importance of the irreversible loss:

$$Be = \frac{\dot{S}_{gen,\Delta T}}{\dot{S}_{gen}}. \quad (6)$$

As shown in Figure 11, when the different helix angle, Be increased with the increase of shell side mass flow is reduced, show that heat transfer caused by irreversible loss percentage of total irreversible loss. In this paper the simulation range, Be less than 0.75, in most cases in 0.9 above, that irreversible loss in the total heat transfer dominates in the irreversible loss. Under small flow rate, the five kinds of helix angle Be were similar, with the increase of flow rate, $\beta = 10^\circ$ when Be falling fastest, $\beta = 20^\circ$ comes second, $\beta \geq 30^\circ$ when Be Dropping lowest. Equals the visible shell side mass flow rate, the smaller helix angle make less of an effect the irreversibility of heat transfer.

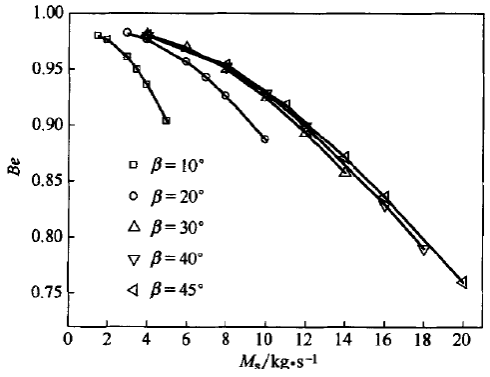


FIGURE 11 Be along with the change of flow rate

For the convenience of comparison of different type of heat exchanger performance under different working conditions, the application needs to be frequently during entropy generation outline 1, entropy generation by the outline 1 is:

$$N_s = \frac{\dot{S}_{gen}}{(Mc_p)_{\min}}. \quad (7)$$

Entropy generation number is smaller, which indicates that irreversible loss is smaller, the better the performance of heat exchanger.

5 Conclusion

1) The helical Angle is not at the same time, the shell side convective heat transfer in helical channel along the shell axial distribution of different. helical channel after about 1 times the pitch to import export before about 0.75 times as part of the pitch, for the full development period of heat convection.

2) The shell side mass flow rate is equal, the shell side heat transfer coefficient and pressure drop per unit length are along with the increase of helical Angle decreases, and the decrease of the amplitude is greater than the former, the shell side heat transfer and resistance along with the increase of helical Angle and improve the comprehensive performance.

3) Shell side cross-sectional tangential flow distribution more uniform, but in the centre of the heat exchange tube bundle and false wall between the tube and shell will be formed by the spiral flow in two places.

Acknowledgements

This work was financially supported by the Hunan Science and Technology Foundation Project (2012GK3130).

References

[1] Bell K 2004 *Journal of Heat Transfer* **6**(126) 877-85
[2] Wang Q W, Chen G D, Chen Q Y 2010 *Heat Transfer Engineering* 10(31) 836-53 *(in Chinese)*
[3] Lutcha J, Nemcansky J 1990 *Chemical Engineering Research and Design* **A**(68) 263-70
[4] Ji S, Du W J, Cheng L 2009 *Proceedings of the CSEE* 32(29) 60-70 *(in Chinese)*
[5] Lei Y G, He Y L, Li R 2008 *Chemical Engineering and Processing* **12**(47) 2336-45
[6] Guo Z-Y, Zhou S-Q, Li Z-X, Chen L-G 2002 Theoretical analysis and experimental confirmation of the uniformity principle of temperature difference field in heat exchanger. International *Journal of Heat and Mass Transfer* **45**(10) 2119-27
[7] Zhao T S 2001 Forced Convection in a Porous Medium Heated by a Permeable Wall Perpendicular to Flow Direction *Journal of Enhanced Heat Transfer* **44** 1031-7
[8] Pesteei S M, Subbarao P M V, Agarwal R S 2005 Experimental study of the effect of winglet location on heat transfer enhancement and pressure drop in fin-tube heat transfer *Applied Thermal Engineering* **25**(11-12) 1684-96

Authors

Jiazhu Zou, born in June, 1977, Hunan, China

Current position, grades: lecturer in Mechanical Engineering, University of South China.
University studies: master of engineering in Mechanical Engineering (2006, University of South China, China).
Scientific interests: fluid control and computer modeling in mechanical engineering.
Publications: 7 papers.

Fengwei Yuan, born in June, 1977, Hunan, China

Current position, grades: associate professor in Mechanical Engineering, University of South China.
University studies: master of engineering in Mechanical Engineering (2001, Central South University of Forestry and Technology, China).
Scientific interests: virtual reality and computer modeling in mechanical engineering.
Publications: 11 papers.

Qian Deng, born in October, 1987, Hunan, China

Current position, grades: lecturer in Mechanical Engineering, University of South China.
University studies: master of engineering in Mechanical Engineering (2012, University of South China, China).
Scientific interest: computer modeling in Mechanical Engineering.
Publications: 4 papers.

Numerical simulation on winding CFRP pipe axial compression stability

Bo Nan*, Yue Wu

School of Civil Engineering, Harbin Institute of Technology, Harbin 150090, China

Received 1 October 2014, www.cmnt.lv

Abstract

Because of the requirements of strict weight, airship structures are usually built by high strength and light-weight CFRP composite truss, CFRP composite tube is the basic unit of it. Winding CFRP tube ultimate bearing capacity has been further improved by combinations of various fiber directions, while stable bearing capacity is an important indicator that could influence the bearing characteristics. It concludes the destruction and deformation characteristics, as well as the relationship between stability factor Φ and slenderness λ, through analysing different CFRP tubes of slenderness ratio with The arc-length method, and drawing the Load-displacement curve, which provided a theoretical basis to better the truss design, and discussed the laws of how component defects affect its stability capacity via a large number of parameters.

Keywords: carbon fiber-reinforced plastic (CFRP), the arc-length method, stability coefficient, slenderness ratio

1 Introduction

Truss structure is a traditional form of large support structure, which is using rod or pipe to transfer load statically, because of big stiffness and saving material, which is a highly efficient structure form of supporting. From the use of material types, the existing truss structures are made of steel, aluminum alloy and CFRP composites, etc. Steel truss by the weight of material density, which cannot meet the requirements of aerospace lightweight structure; Quality of aluminum alloy truss structure is lighter and has been widely used in various kinds of large support structure in the field of aerospace. Along with the development of the large airship and space technology, advanced CFRP composite truss structure because of its significant weight loss effect and excellent mechanical properties have also been more and more applied to the satellite, airship, etc all kinds of aerospace structure (as shown in Figures 1 and 2).

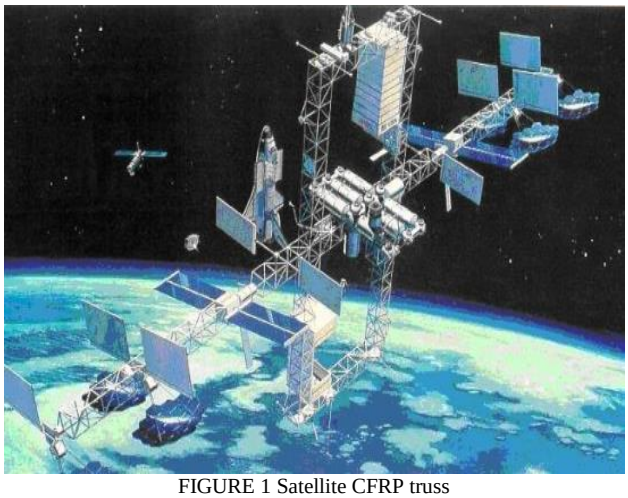


FIGURE 1 Satellite CFRP truss

FIGURE 2 Airship CFRP skeleton

In the composite truss structure, CFRP composites pipe components is the basic composition, and the stability of axial compression bearing capacity is the key to affect the whole structure stability. However, at present the study on stable of long tube CFRP are rare, and the research on bigger slenderness CFRP axis compression member is more rare. But the stability of the long tube from other materials research methods mainly can be divided into the following kinds:

Theoretical derivation:

The earliest Study of aluminum bending member overall stability is researched by Hill and Clark [1, 2], the stable theory research is Mainly done in solid rectangular section, box and i-section section aluminum alloy extrusion profiles component. And the related formula of axial force and bending moment was put forward.

Experimental study:

Aluminum alloy axial compression member is researched by Shen Zuyan, Guo Xiaonong [3], Qian Ruojun, Li Ming [4], etc. Cross section forms include circular pipe and "H" type of section, the test results show that members are bent instability. And for glass fiber reinforced composites (GFRP) pipe compression members, Qian Peng, Ye Lieping [5,6] get the basic parameters of the mechanical properties of GFRP pipe through short pipe compression test firstly,

***Corresponding author** e-mail: nb2003ccc@163.com

then select 4 different slenderness ratio of 12 GFRP long pipe to do experiment to study its stability.

Numerical analysis:

Compared to these two materials, Steel constitutive relationship has been relatively mature, Shi Yongjiu, Wang Yuanqing et al. [7, 8] using ANSYS finite element software, calculate overall stability capacity of High-strength steel welded box-section axial compression column, accurately simulate residual stress and geometric components of the initial defect. The results are in good agreement with the experimental results.

The above method has the following disadvantages: if the material model is simple, Stability capacity of members can be obtained accurately both the theoretical analysis and simulation; if the material model is complex (FRP composite material), results are obtained through the test method, and can only get a small amount of results.

Different angle of winding short CFRP pipe was analysed, then choose a better layer, and establish effective stress strain model, get its nominal yield strength according to Ramberg-Osgood equation; then through different slenderness ratio CFRP components axial compression, derived relationship Stability factor between the slenderness ratio, Finally, large number of parameters was analysed to discuss the law for components of the initial defect affecting its stability capacity.

2 Numerical simulation for short winding CFRP pipe axial compression

2.1 HASHIN PROGRESSIVE FAILURE CRITERIA

CFRP composite laminated structure mainly has 5 kinds of common failure criteria: yhe maximum stress theory, the maximum strain theory, Tsai-Hill criteria, Hoffman criteria, Tsai-Wu criteria. Failure criterion is single directional plate fiber by using a variety of the strength of the composite material under different load conditions; these theories do not consider the injury accumulation before the damage completely. The Hashin failure criterion [9] is a material consideration cumulative injury, damage evolution to achieve the criteria.

The basic principle of this theory includes fiber tensile fracture, compression buckling under transverse tensile and shear fracture when matrix under transverse compression and shear failure such as crushing, mainly includes the following forms:

Fiber tension ($\hat{\sigma}_{11} \geq 0$):

$$F_f^t = \left(\frac{\hat{\sigma}_{11}}{X^T}\right)^2 + \alpha\left(\frac{\hat{\tau}_{12}}{S^L}\right)^2. \tag{1}$$

Fiber compression ($\hat{\sigma}_{11} \leq 0$):

$$F_f^c = \left(\frac{\hat{\sigma}_{11}}{X^C}\right)^2. \tag{2}$$

Matrix tension ($\hat{\sigma}_{22} \geq 0$):

$$F_m^t = \left(\frac{\hat{\sigma}_{22}}{Y^T}\right)^2 + \left(\frac{\hat{\tau}_{12}}{S^L}\right)^2. \tag{3}$$

Matrix compression ($\hat{\sigma}_{22} \leq 0$):

$$F_m^c = \left(\frac{\hat{\sigma}_{22}}{2S^T}\right)^2 + \left[\left(\frac{Y^C}{2S^T}\right)^2 - 1\right]\frac{\hat{\sigma}_{22}}{Y^C} + \left(\frac{\hat{\tau}_{12}}{S^L}\right)^2, \tag{4}$$

X^T is axial tensile strength. X^C is axial compressive strength. Y^C is transverse compression strength. Y^T is transverse tension strength. S^L is axial shear strength. S^T is transverse shear strength. α is the coefficient of shear stress for fiber tensile failure $\hat{\sigma}_{11}$, $\hat{\sigma}_{22}$, $\hat{\tau}_{12}$ are the component effective stress $\hat{\sigma}$. $\hat{\sigma}$, which assess the initial damage equation:

$$\hat{\sigma} = M\sigma, \tag{5}$$

where σ is the nominal stress, M is the damage factor:

$$M = \begin{bmatrix} \frac{1}{(1-d_f)} & 0 & 0 \\ 0 & \frac{1}{(1-d_m)} & 0 \\ 0 & 0 & \frac{1}{(1-d_s)} \end{bmatrix} \times d, \tag{6}$$

d_f, d_m, d_s for the internal damage variable of shear characterization of fiber, matrix. They are derived from the previously mentioned four damage model related damage variable d_f^t, d_f^c, d_m^t, d_m^c:

$$d_f = \begin{cases} d_f^t \_if\_\hat{\sigma}_{11} > 0 \\ d_f^c \_if\_\hat{\sigma} < 0 \end{cases}, \tag{7}$$

$$d_m = \begin{cases} d_m^t \_if\_\hat{\sigma}_{22} > 0 \\ d_m^c \_if\_\hat{\sigma}_{22} < 0 \end{cases}, \tag{8}$$

$$d_s = 1-\left(1-d_f^t\right)\left(1-d_f^c\right)\left(1-d_m^t\right)\left(1-d_m^c\right). \tag{9}$$

Before the injury occurred, M is unit matrix, so $\hat{\sigma} = \sigma$. Once the injury began, it evolved at least one kind of mode above, injury began to occur, $\hat{\sigma}$ represents the real bearing capacity.

2.2 NUMERICAL SIMULATION OF SHORT PIPE AXIAL COMPRESSION BASED ON LAYERED SHELL ELEMENT

2.2.1 Layer Angle

Fiber is called the longitudinal direction, expressed with an "*x*", in vertical direction (sometimes have woven fiber, fiber content less) called lateral, expressed with "*y*", Thickness direction with "*z*", *tk is* the thickness of single layer. θk is angle between Fiber and principal axis direction (as shown in Figures 3-5).

Size	L=150mm, D=50mm
thickness	4mm
Number of layer	8
Element type	S4R
Element number	2400
Boundary conditions	One end fixed, one end free
Load	2mm displacement
Diameter-thick ratio	12.5

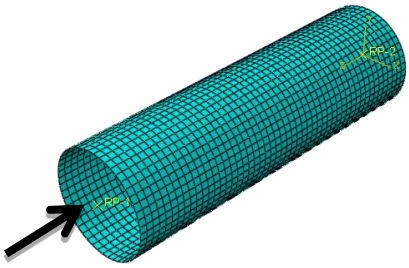


FIGURE 3 Model grid partition map

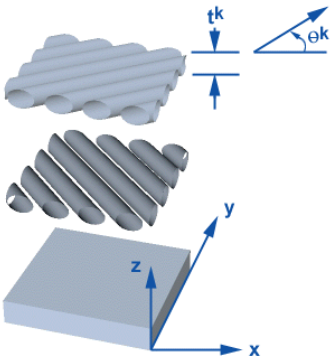


FIGURE 4 Fiber layer directions

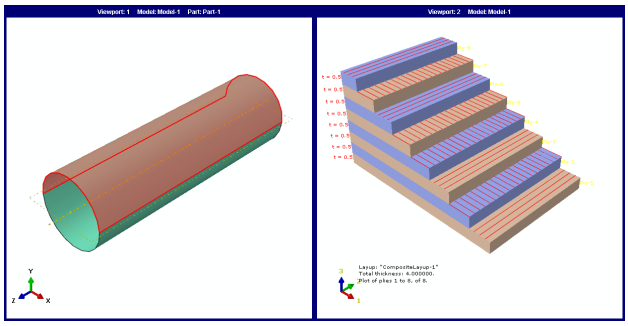


FIGURE 5 90/0/90/0/90/0/90/0 layer angle

2.2.2 The calculation results

Through to the mechanical properties of layer angle analysis of CFRP pipe, the strength of 90/0/90/0/90/0/90/0 orthogonal layer is high, but it belongs to brittle failure, if accompanied by plus or minus 45° layer can achieve very good effect (Figure 6).

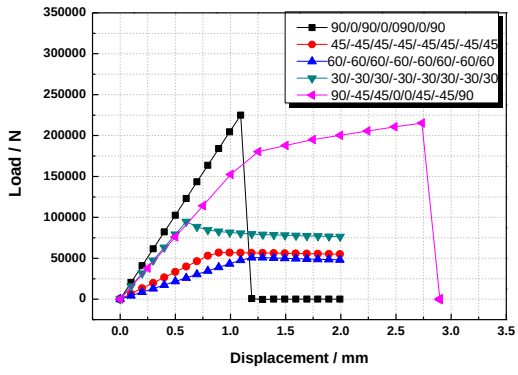


FIGURE 6 Load - displacement curve for each layer angle

2.3 THE EQUIVALENT STRESS AND STRAIN CURVE

According to the load-displacement curve, we found that 90/-45/45/0/0/45/-45/90 angle is the best. Select the layer mode as the research object, using the homogenization thought the macro components of load-displacement curve into material such as microscopic stress-strain curve, the following equation:

$$\sigma = \frac{F}{A} = \frac{F}{\frac{\pi}{4}(D^2 - d^2)}, \tag{10}$$

$$\varepsilon = \frac{\Delta l}{l}. \tag{11}$$

The equivalent of the stress-strain curve as follows:

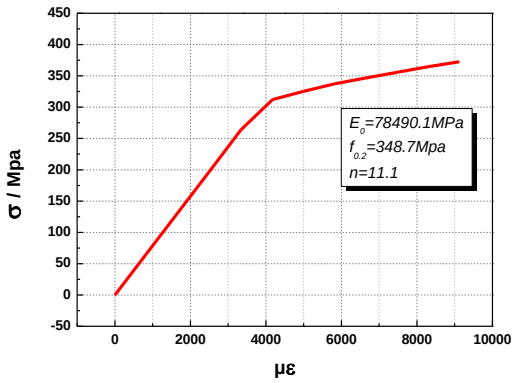


FIGURE 7 Equivalent stress - strain curve

Stress-strain curve of CFRP has no obvious yield platform, Ramberg - Osgood [10] formula was used to describe that:

$$\varepsilon = \frac{\sigma}{E_0} + 0.002\left(\frac{\sigma}{f_{0.2}}\right)^n, \tag{12}$$

E is material elastic modulus. f0.2 is nominal yield strength. *n* is used to describe the strain hardening material, Its value can be calculated according to 0.1% and 0.2% nominal yield strength.

$$n = \frac{\ln(2)}{\ln(f_{0.2} / f_{0.1})}. \tag{13}$$

The equivalent material parameters such as Table 1.

TABLE 1 Material parameters

Material	E0 (MPa)	f0.1 (MPa)	f0.2 (MPa)	fu (MPa)	n
CFRP	78490.1	327.5	348.7	372.4	11.1

3 Research on winding CFRP long pipe stable performance

Calculating ideas of long pipe stable are as follows. The real component is modeled using finite element software. First linear modal analysis was done, a first-order modal L/1000 was extracted as the initial defect component to be applied to the original member. Then the arc method was used to do nonlinear finite element calculation considering the effect of large deformation (Figures 8-10).

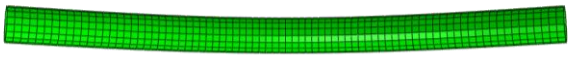


FIGURE 8 Modal analysis

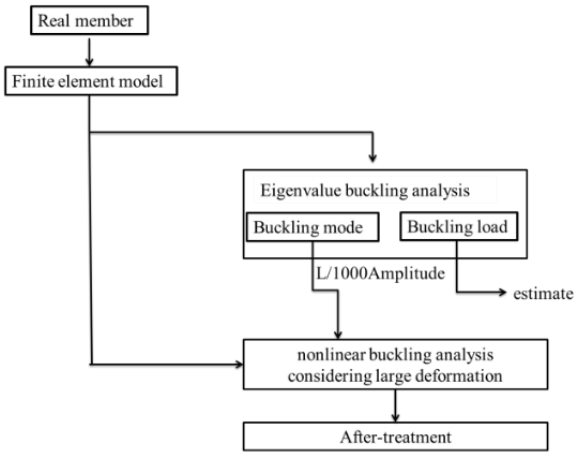


FIGURE 9 Calculation process

FIGURE 10 Nonlinear finite element calculation

3.1 LOAD - DISPLACEMENT ANALYSIS

In order to avoid local buckling, select section for Φ70×10 pipe [11], and calculate the pipe of slenderness ratio of 40, 60, 80, 100, 120, get the whole process of load – displacement curve (as shown in Figure 11) and load – mid-span deflection curve (as shown in Figure 12).

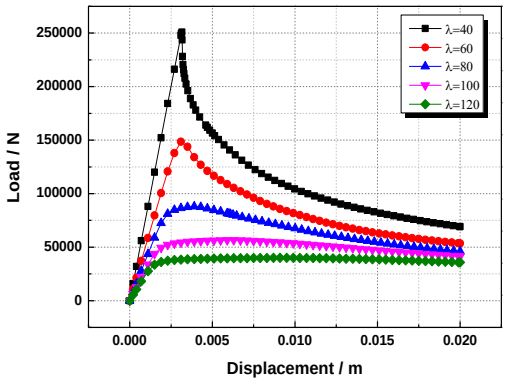


FIGURE 11 Load 3 – displacement curve

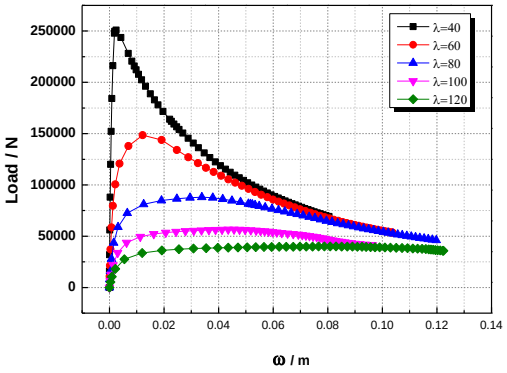


FIGURE 12 Load - deflection curve

As shown in Figure 11 along with the increase of slenderness ratio, compression stability bearing capacity of pipe falling; Slenderness ratio λ of 40 and 60, the shape of load – displacement curve is sharp, brittle failure mode, Mainly because CFRP is brittle material When λ is 80, 100, 120, the shape corners disappear into a smooth gradient inflection point, it shows overall instability of the structure.

As shown in Figure 12, for slenderness smaller components, load - mid span deflection curves presents "tip" is brittle failure, less deformation span; for slenderness relatively large components, load - mid span deflection curve is smooth, elastic deformation, the larger the amount of deformation span.

3.2 STABLE FACTOR – SLENDERNESS RATIO CURVE

The Figure 11 shows the extraction the stability bearing capacity of the whole process load - displacement curve and the stability bearing capacity of the corresponding displacement are listed in the table below.

TABLE 2 List of stable bearing capacity - slenderness ratio

Slenderness ratio λ	Stable bearing capacity (N)	The stability bearing capacity of displacement (mm)
40	250842	2.8
60	148483	3.1
80	88027	3.9
100	56520	5.9
120	39975	9.5

In order to facilitate the comparison of component stability bearing capacity of different sizes, stable factor is introduced; stable factor φ can be calculated through the following equation:

$$\varphi = \frac{F_u}{F_{0.2}}, \tag{14}$$

where F_u for buckling load, $F_{0.2}$ for the product of nominal stress and the cross section (Figure 13).

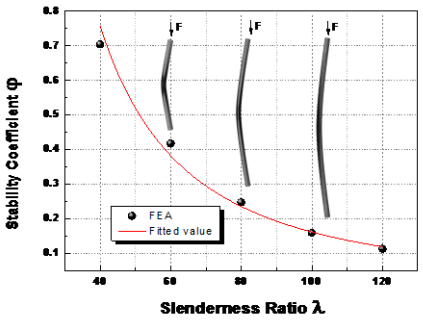


FIGURE 13 $\bar{\lambda} - \varphi$ curve

According to $\lambda - \varphi$ curve, stability coefficient decreases with the increase of slenderness ratio, by fitting curve can be ensure stable factor of slenderness ratio of 50 at 0.5.

3.3 INFLUENCE OF INITIAL DEFECTS

In order to investigate the initial defects of different cross section and the influence of different elastic modulus components, regularization slenderness ratio could be calculated to achieve it, specific methods are as follows:

$$\bar{\lambda} = \sqrt{\frac{F_{0.2}}{F_E}}, \tag{15}$$

$$F_{0.2} = Af_{0.2}, \tag{16}$$

$$F_E = \frac{\pi^2 E_0 I}{L_e^2}, \tag{17}$$

where A is term cross-sectional area, $f_{0.2}$ is nominal buckling strength, E_0 is initial elastic modulus, I is moment of inertia, L_e is calculated length, $F_{0.2}$ is load yield, F_E is Euler critical force (Figure 14).

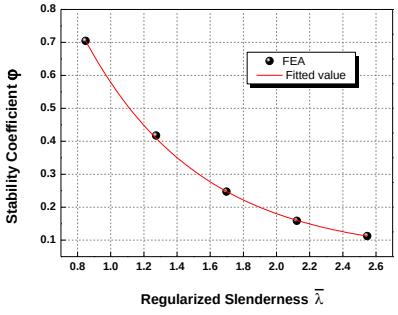


FIGURE 14 $\bar{\lambda} - \varphi$ curve

In this paper, the overall magnitude of bending the initial defect was analysed, includes 6 kinds of conditions, L_e/500, L_e/800, L_e/1000, L_e/1500, L_e/2000, L_e/3000, the results as shown below (Figure 15):

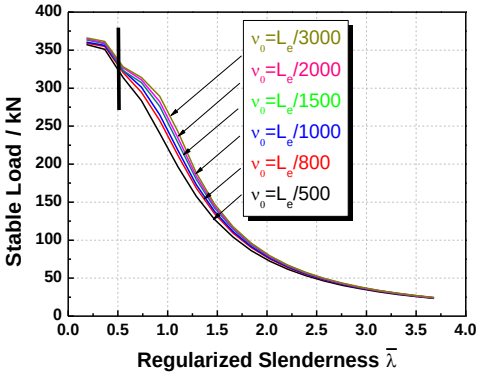


FIGURE 15 Sensitivity analysis of initial defect

The calculation results show that, the length of the smaller member ($\bar{\lambda} < 0.5$) is the intensity of damage, others are overall instability. in the whole range, buckling load of member decreases with the increase of the amplitude of the initial defect, initial defect has the larger the effect on member of the regularization slenderness ratio ($0.5 \le \bar{\lambda} \le 2.0$), when slenderness ratio ($\bar{\lambda} = 1$), Buckling load is most sensitive to changes in the initial defect, has Little effect even negligible when $\bar{\lambda} > 2.0$.

4 Conclusion

1) The compression of short tube is analysed based on the progressive failure criteria (Hashin Criterion) to get the optimal layer angle 90/-45/45/0/0/45/-45/90, and equivalent relationship of stress and strain is obtained to define this layer material.

2) Before instability, CFRP winding pipe appears linear elastic characteristics, brittle characteristics when destroyed, and the length of the smaller member ($\bar{\lambda} < 0.5$) is the intensity of damage, length of middle and long pipe ($\bar{\lambda} > 0.5$) is overall bending instability.

3) According to the $\lambda - \varphi$ relationship curve, stability factor decreases with the increase of slenderness ratio.

4) Initial defect has the larger the effect on member of the regularization slenderness ratio ($0.5 \le \bar{\lambda} \le 2.0$), when slenderness ratio ($\bar{\lambda} = 1$), Buckling load is most sensitive to changes in the initial defect, has Little effect even negligible when $\bar{\lambda} > 2.0$.

Acknowledgments

The work described in this paper was jointly funded by the National Natural Science Foundation of China under Grant No. 51378150.

References

[1] Hill H N, Hartmann E C, Clark J W 1956 Design of aluminum alloy beam-columns *Transactions of the American Society of Civil Engineers* **121**(1) 19-21

[2] Clark J W 1955 Eccentrically loaded aluminum columns *Transactions of the American Society of Civil Engineers* **120**(1) 1116-1132

[3] Guo X-N, Shen Z-Y, Li Y-Q 2007 Theoretical and experimental research on aluminum alloy members under axial compression *Journal of Building Structures* **28**(6) 118-28 *(in Chinese)*

[4] Li M, Chen Y-J, Qian R-J 2000 Experimental research on the stability parameters of aluminum pipes under axially *Compression Load Spatial Structures* **6**(3) 59-64 *(in Chinese)*

[5] Qian P, Feng P Ye L-P 2007 Experimental study on GFRP pipes under axial compression *Journal of Tianjin University* **1**(1) 19-23 *(in Chinese)*

[6] Qian P, Feng P, Ye L-P 2008 Experimental study on GFRP pipes under axial compression *Frontiers of Architecture and Civil Engineering in China* **2**(1) 73-8

[7] Shi G, Wang Y-Q, Shi Y-J 2009 Behavior of high strength steel columns under axial compression *Journal of Building Structures* **30**(2) 92-7 *(in Chinese)*

[8] Shi G, Wang Y-Q, Shi Y 2009 Finite element analysis on overall bucking of ultra-high strength steel welded box section columns under axial compression **25**(2) 255-61 *(in Chinese)*

[9] Hashin Z, Rotem A 1973 A fatigue failure criterion for fiber reinforced materials *Journal of Composite Materials* **7** 448-64

[10] Ramberg W, Osgood W R 1943 Description of stress-strain curves by three parameters Washington D.C. *National Advisory Committee for Aeronautics* TN-902

[11] Hou W, Zhang X-H, Feng H-C 2010 Stability of axial compression members of Glass Fiber-Reinforced Plastic (GFRP) *Journal of Building Materials* **13**(4) 441-5 *(in Chinese)*

Authors

Bo Nan, born in April, 1984, Liaoning Province, Shenyang City, China

Current position, grades: Doctor at Harbin Institute of Technology.
University studies: Master's degree at Shenyang Jianzhu University in Shenyang City in 2010.
Scientific interests: space structure in civil engineering

Yue Wu, born in December, 1972, Heilongjiang Province, Harbin City, China

Current position, grades: professor at Harbin Institute of University.
University studies: Master's degree at Harbin Institute University in Harbin City in China in 1998. Doctor's degree at Harbin Institute University in Harbin City in China in 2003.
Scientific interests: space structure.

Exploratory development of ancient village based on the low carbon travel development model

Jifei Zhang*

School of Business and Management, Hebei Normal University of Science & Technology, Hebei, 066000, China

Received 2 September 2014, www.cmnt.lv

Abstract

This paper summarized basic theory of low carbon travel and the functional mechanism of the low carbon economy to improve the village tourism development by analyzing its conception, and development status and existing problem of the ancient village tourism in the low-carbon tourism development mode of our country nowadays is put forwarded. Combining with actual situation, this paper aims to propose the emission reduction countermeasures in allusion to these problems above, in order to provide certain reference to the development of ancient villages low carbon tourism.

Keywords: Low carbon concept; Low carbon travel; Ancient village travel; Emission reduction countermeasure

1 Introduction

With the increasing attention to the climatic variation worldwide, "low carbon" enters into view and is brought to the forefront. At the same time, with the voice of developing the low carbon economy, energy conservation and emission reduction, the low carbon travel becomes the fashionable tourism rapidly. Ancient village is a kind of traditional human settlement space, and it unites the most traditional life custom, social structure and thousands of years of farming culture of our country. A new tourism development mode has been built by combining the low carbon travel with ancient village travel development. The spiritual enjoyment of people can be satisfied when they are traveling, and at the same time, the healthy ecosystem can be guaranteed. The environmental friendly and energy-saving travel has a positive meaning to promote the development of the ancient village and protect the integrity of the ancient village [1-4].

2 Overview of the low carbon travel

2.1 CONCEPTION OF THE LOW CARBON TRAVEL

The low carbon travel in our country is not clearly defined because it starts late, and scholars have put forward their own viewpoint of the low carbon travel: Cai Meng et al [1] defined the low carbon travel as the carbon sequestration mechanism and low carbon consumption patterns through the low-carbon technology, in order to get higher quality of tourist experience and tourism economy. The core idea of it was to obtain greater economic, social and environmental benefits with less carbon emissions of tourism development; Qu Bao [2] thought that the low carbon travel was a deeper green tourism with low power consumption, low pollution, lower carbon footprint and carbon dioxide emissions; Hou Wenliang et al [3] thought that fundamental purpose of low carbon travel was to reduce greenhouse gas emissions with the premise of guaranteed satisfaction, and realized the new travel mode and management concept of low carbon economy development through the different ways and the humanized system; this was a new kind of management philosophy, the essence of it was a new tourism consumption with low carbon intensity and its objects were all categories of tourist activities. Making a general survey of each scholars' viewpoint, although the low carbon travel is not clearly defined, all of them agree that low carbon tourism is a new type of tourism, and the traveling experience can be higher in the case of reducing carbon emissions through the new type "energy conservation and emissions reduction" technology and tourist's "low carbon consciousness" [5].

2.2 THE THEORETICAL BASIS OF LOW CARBON TRAVEL

The theoretical basis of low carbon travel is mainly about low carbon economy theory, tourism sustainable development theory and ecological tourism theory [5]. The low carbon economy refers to change development pattern from extensive development pattern to new development model with low pollution, low energy consumption, low emissions through a series of means, including technology, institutional innovation, new energy development, etc; innovation and change are the core essentials, and final goal is to achieve sustainable development. Sustainable tourism development follows its basic concept, that is, meet the need of contemporaries and do not threaten the capacity of future generations to their needs. It balances and coordinates the relationship between travel and nature, economy, culture and harmonious development. Besides respecting the integrity of local natural travel and protecting traveling, ecological tourism also involves the interpretation and education of the natural.

* ***Corresponding author*** e-mail: zjfhbqhd@163.com

3 The development mechanism of ancient village travel based on the low carbon travel development model

The low carbon economy promotes the development of ancient village travel in aspects of low carbon concept, low carbon technology and low carbon policy, and it is a comprehensive function mechanism. The function mechanism of the low carbon economy to promote the development of ancient village travel [5] is shown in the Figure 1.

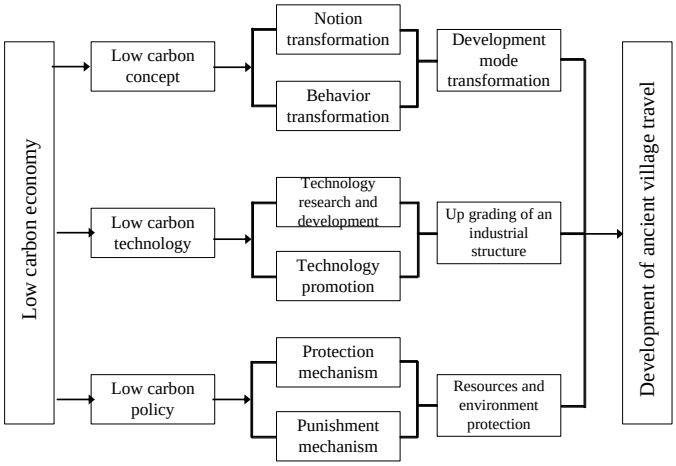


FIGURE 1 Action mechanism of the low carbon economy to promote the development of ancient village

Low carbon concept changes the concept and behaviours of participant and promotes the transformation of development pattern into low pollution, low energy consumption and low emissions in pursuing travel [5]. Most of ancient villages are in the ecological sensitive zone or reserves. The production of the low carbon concept can realize the transformation of development pattern by changing the ecology and resource utilization pattern, and the tourists are boosted to produce low carbon behaviours, and at the same time the ancient villages operators, enterprise, etc can receive self-discipline by calculating the carbon emissions; the low carbon technology provides the support to the low carbon travel, and then upgrade of tourism industry structure of ancient village is promoted. The low carbon technology steady or reduce the demand of non-renewable high carbon fossil energy by improving energy efficiency in each link of each industry in the process of ancient village travel such as eating (food tourism), living (Tourist hotel industry), behaviours (Tourist transportation), travelling (Tourism landscape), purchase (Tourism businesses), amusement (Tourist entertainment industry), etc. The cleaner production and thrift is proposed, and the virtuous cycle and harmonious development of ecology and economy of ancient village is realized by strict control of wastes and emissions; the cleaner production and thrift is proposed, and the virtuous cycle and harmonious development of ecology and economy of ancient village is realized by strict control of wastes and emissions. The low carbon policy provides the safeguard mechanism to low carbon travel by protection mechanism and punishment mechanism in the process of the development of the ancient village travel. It is easy to produce tour order disorder because of the continuously extended ancient village tourism industry scale, which will produce adverse impact to sensitive local ecological environment and resource utilization. Effective utilization of the environment and resources is realized by policy guidance and protection mechanism. The establishment of punishment mechanism can hit the behaviours of environment damage and resources waste to some extent. These two supplement each other, and the objective of ancient village tourism environment protection can be realized under the action of the dual mechanism.

4 The development status of ancient village travel based on the low carbon travel development model

On account of the different national conditions, there has a significant difference between ancient travel development at home and abroad based on the low carbon travel development model [3, 5-7]. First, the research contents of low carbon travel abroad concentrate on tourists' behaviours choice, tourism carbon emissions and reducing travel emissions. The research perspectives of domestic scholars of low carbon travel concentrate on the conceptual framework of low carbon travel, development research and systematic study of low carbon travel. Research focuses at home and abroad have big difference. The research key point at home is low carbon tourism basic theory and gradually tend to study evaluation system of low carbon tourism and the development strategy; abroad study is more tend to measure the tourist traffic emissions through quantitative calculation, at the meanwhile, the corresponding measures of carbon reduction are put forward based on it. By comparison, the value of research results at abroad is generally higher than that at home in terms of science and credibility. The study of travel and protection of ancient village at foreign countries has certain reference significance to our country. Secondly, the history of tourism development research abroad is longer; therefore they have larger advantages compare with our country in the involved field and results. The study of ancient village travel abroad unfold in three aspects, ancient village tourism stakeholders, sustainable development and community participation in tourism, in order to form a virtuous circle of travel development in multiaspect. The study in this aspect at home is still on the exploratory stage, and the points of view in

our study are the problems in the process of tourism, protection and development of ancient villages and stakeholders. Although the involved fields of the theoretical research at home are being extended, lots of them are still rest on the principle and strategy, especially the study of new travel development mode of ancient village travel is rare. Therefore, guidance of the theory and practice is needed in the sustainable development of the ancient village.

The sustainable development of the ancient village is discussed starting from the view of low carbon economy. This paper put forward a technical route framework based on the research and analysis of the low carbon tourism development at home and abroad, in order to provide a certain reference to the study in this aspect. Research on ancient villages' technology roadmap under low carbon economy is shown on Figure 2.

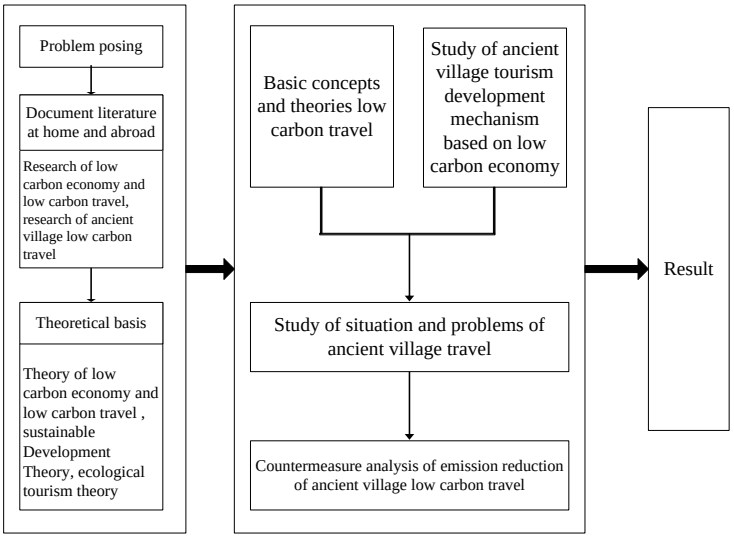


FIGURE 2 Research on ancient villages' technology roadmap under low carbon economy

5 Emissions reduction countermeasures of ancient village tourism development based on low carbon tourism development pattern

The ancient village travel in our country is still in the initial exploration stage; therefore, we should draw on the advanced experience abroad. The environment at home and abroad is different, therefore, how to use the experience for reference is still an issue needed to be researched. The consciousness of low carbon concept is built up from the aspects of "carbon source", "carbon sink" and industrial restructuring. Industry system of ancient village low carbon travel is supported by low carbon technology and guaranteed by low carbon policy, and then the ancient village low carbon travel can realize. The relative countermeasures are shown below:

5.1 ENHANCE THE LOW CARBON CONSCIOUSNESS, AND STRUCTURE INDUSTRIAL SYSTEM OF ANCIENT VILLAGE LOW CARBON TRAVEL

The economy of ancient village is lag behind relatively and the development is slow, which are influenced by its sealing and conservatism. Another study found that domestic tourists did not know completely about the low carbon travel. New knowledge should be increased and the concept should be changed in order to change this phenomenon [5, 7]. We must increase publicity of low carbon tourism consciousness, and the tourism environmental protection concept of "low carbon tourism" should be promoted to the village residents and tourists actively, in order to cultivate and establish a public awareness culture of environmental protection. At the same time, the relative tourist enterprises should devote themselves to forge the low carbon tourism products with local characteristics and guide green and energy-saving consumption. The covers of the tourist industry are very wide, and the tourist industry of low carbon travel is structured by vertical and horizontal industry chain.

5.2 POPULARIZE THE LOW CARBON TECHNOLOGY, AND STRUCTURE THE CLEAN AND EFFICIENT ENERGY SUPPLY SYSTEM

Promotion and application of the low carbon technology is the key to develop the ancient village travel. Comprehensive benefits of ancient village travel can be realized through measures including increasing the invest capital to low carbon travel, completing supportive polices of low carbon travel development, popularizing the development patterns of local ecological agriculture and recycling economy. The commonly used methods for further refinement include many new technologies like carbon capture and sequestration technologies, reduction, recycling, energy utilization technology and so on, and we should promote using proportion of clean energy and process contaminant, which is generated in the production, and live.

5.3 CONSUMMATE THE LOW CARBON POLICY, AND BUILD THE LONG-TERM MECHANISM OF ANCIENT VILLAGE LOW CARBON TRAVEL

The local government should play a leading role to guide the behavioral pattern of enterprise, resident, tourist under the

guidance of sustainable development and low carbon concept. We should formulate the target responsibility, appraise check-up system, develop strategic step combine with the local practical, and make scientific plannings. The rational distribution of local low carbon travel should be bulit. In addition, the experience and lessons of low carbon enconomy and low carbon travel development abroad can be used for reference in building the suitable marketing system and policy system for the development of ancient village low carbon travel. The low carbon travel is regarded as the capital projects of ancient village, and then a sound long-term mechanism will be built.

5.4 ENCOURAGE THE ENTERPRISE TO TAKE PART IN ACTIVELY, AND FORM THE ECONOMICAL PRODUCTION SYSTEM

One of the basic requirements of low carbon travel is to reduce the energy consumption in the production processses and use the clear energy. "Green operation" needs the cooperation of enterprise. Nowadays, the usage rate of ancient village travel enterprises to use low carbon technology and clear energy is not ideal. The enterprise should be initiated to do green consumption and cultivate the low carbon tourism awareness of employees, and production system with scientific, forward-looking and environment friendly should be chose. We can make full use of natural high carbon sinks tourism resources, and the low carbon resources of culture and intangible cultural heritage resources should be excavated and protected in order to realize the ancient village tourism low carbonization.

5.5 EXCHANGE THE CONSUMPTION MODEL OF TOURIST, AND BUILD LOW CARBON CONSUMPTION PATTERNS

Tourists are the subject in the tourist activity. The common consumption model of tourist with resources wasting and increased emissions pollution should be translated into low carbon consumption model, from the aspects of eating, eating, living, behaviours, purchase. The tourists can enjoy a healthy ancient village tourism environment and protect the healthy environment of ancient village at the same time, and the benign development of ancient village travel can be promoted.

5.6 EXERT THE USE OF CARBON SINK, AND STRUCTURE THE CLEAN DEVELOPMENT MECHANISM

Definition of carbon sink is: the course, activity and mechanism of the forest vegetation fix the carbon dioxide in organisms by photosynthesis and reduce the concentration of carbon dioxide in the atmosphere [5]. Ancient villages reduce carbon emissions from the carbon source and pay attention to the carbon sink function of forest, grassland, wetland ecological land etc by using local ecological resources. Therefore, the utilization efficiency of agrarian and the protection of ecologically fragile area should be intensified in the ancient village travel.

6 Conclusion

The ancient village low carbon travel is a new type of tourism sustainable development facing to the increasing climatic variation and environmental resources pressure. It can not only realize the objective of energy conservation and emission reduction but also can quicken the upgrade industries of the tourist industry, and the energy consumption and operation cost of the tourist industry can be reduced, at the same time, the returns can be increased. The tourist industry can develop with dimension and collectivization [3, 5]. At present, it is in the booming period. Based on the further analysis of development mode of ancient village low carbon tourism, this paper started with the research on this aspect of the relevant literature and development mechanism research based on low carbon tourism development at home and abroad, comprehended the research status of ancient village low carbon travel at home and abroad and proposed relative research technology roadmap. On the above basis, we put forward countermeasure of emission reduction from the aspects of "carbon source", "carbon sink" and industrial restructuring, and positively promote the destination construction of low carbon travel.

Acknowledgement

Hebei social science fund project: HB13JJ052.

References

[1] Cai Meng, Wang Yuming 2010 Low-carbon Tourism: A New Mode of Tourism Development *Tourism Tribune* **25**(1) 13-7
[2] Qu Bao 2010 First Exploration of Tourism Development Based on the Low Carbon *Northern Economy* (5) 31-2
[3] Hou Wenliang, Liang Like, Si Dongge 2010 The Study on Concept System of Low-carbon Tourism *Journal of Aayang Teachers College* (2) 86-9
[4] Ren Junying, Wu Jiang 2010 Research on Ancient Vilage Tourism in China *Journal of Anhui Agricultural Sciences* **38**(19) 10292-5
[5] Yu Han 2012 The Research of Ancient Town Low-carbon Tourism-A Case Study of Xiangxi Area *Hunan: Jishou University*
[6] Niu Dandan 2012 Ancient Village Landscape Protection and Tourism Development *Shanxi: Northwest A&F University*
[7] Liang Chao 2012 The Design Research Based on the Idea of Low Carbon Tourism Planning-Take Hengyang Yanfeng District Makina and Bamboo Village as Example *Hunan: Central South University of Forestry and Technology*

Author

Jifei Zhang, born in 1977, Hebei Province of China

Current position, grades: lecturer
University studies: Master's degree was earned in major of software engineering, Beijing Institute of Technology in 2009.
Scientific interest: Tourism, hotels

Content-based retrieval of music using mel frequency cepstral coefficient (MFCC)

Xin Luo*, Xuezheng Liu, Ran Tao, Youqun Shi

School of Computer Science and Technology, Donghua University, Songjiang Districe, Shanghai, 201620, China

Received 1 March 2014, www.cmnt.lv

Abstract

In the last few years, with the growing of multimedia in Internet, MP3 music become one of the most popular types. Some of the MP3 music collections available are approaching the scale of million tracks and this has posed a major challenge for searching, retrieving, and organizing music content. In this paper, we proposed a method to retrieve the MP3 lossy compression format music by using MFCC features. The Kullback-Leibler Divergence and Earth Mover's Distance (EMD) are used to compute music similarity. Experiments show that the retrieving probability of our design can achieve high recall values of 95% out of a total of 1951 tracks in the database.

Keywords: content-based music retrieval, MP3, MFCC, kullback-leibler divergence, EMD

1 Introduction

In the last few years, with the growing of multimedia in Internet, MP3 music become one of the most popular types. Some of the MP3 music collections available are approaching the scale of million tracks. Therefore, it's become increasingly difficult when people classify and manage these massive musical date only by his own hand. We need to query the music information quickly, accurately and efficiently, the traditional approach that query by the text has been far from satisfying the user's search request for such resources. Besides the traditional way, which is using keyword-based mode to retrieve the song information to obtain the desired song, people also want to query it just by the characteristics of the music itself. The CBMR (Content-based Music Retrieval) is an important technique in recent years in handling massive multimedia data in a network environment, together with image retrieval and video retrieval it is today's hotspot in content-based retrieval research field [1].

The work of CBRM research began in the 1990s, the initial research is mostly about the humming retrieval system. Ghiaset [2] carried out pioneering research in Single-track MIDI music humming retrieval system, they use time-domain autocorrelation algorithm to extract pitch information, then by using a fast approximate string matching algorithm they implemented monophonic music retrieval. Tomonariet [3] using span and pitch information as search clues, adopt both dynamic threshold determination and coarse-to-fine matching for matching. Kosugi [4] proposed using pitch direction and distribution at the same time to improve the system performance. They developed a Sound Compass system, it contains 10086 audios, and people must humming coordinate with musical beats while using this system. Seungminet [5, 6] improved the pitch-tracking algorithm, they added the fundamental frequency index function to the traditional tracking algorithm, and by using genetic algorithm-based relevance feedback schemes they signify-cantly improved the retrieval accuracy.

In recent years the audio fingerprint-based fast retrieval technique, which is using binary number to represent the audio features, has becoming the key technology of audio retrieval engine [7-9]. Audio fingerprint extraction first proposed by Philps Institute in Holland, they proposed a global information based audio fingerprints extraction method. This method segment the audio spectrum into many frames, each frame represented by a 0 or 1, thus the whole audio spectrum can be expressed by a binary numbers sequence, the advantage of this method is the whole audio spectrum can be recorded. Another way is proposed by Shazam Company, this UK company try to find some feature points from the audio spectrum, they usually are peak points, and make them be Peak-Pairs, then use the sequence of the Peak-Pairs as the audio fingerprint of this frame. This method centralize the effective messages and got a good noise immunity.

Query by Humming allows the users to find a song by humming part of the tune, use the melody and rhythm to retrieval, make it convenient for the user to find a song compare to the traditional way which use the song title, artist and other text information to retrieval. However, the key part of this way is humming, this make it little meaningful to the user. Audio fingerprints is a better way to retrieval, classify and sorting music for users, but its shortcomings are the feature information is not typical and the noise immunity is also not good enough. Since the general users prefer MP3 format and MP3 as a lossy compression format is designed to significantly reduce the data size of an audio, when we use the audio fingerprint to retrieval a MP3 music it's hard to achieve high accuracy, but for most user's listening experience the reproduction acoustics and the original sound almost in the same level.

This passage proposed a MP3 retrieval system which is using MFCC (Mel-frequency Cepstral Coefficients) to characterization the audio signal. MFCC is based on the human

***Corresponding author** e-mail: xluo@dhu.edu.cn

auditory characteristics, it's a non-liner relationship with the Hz frequency. Through the 1951 Chinese and other language songs retrieval experiment we proved the effectiveness of the proposed method.

2 Proposed audio retrieval

Our proposed way is use MFCC as the audio features. MFCC is the determinants of the spectrum modeling, therefore it can reflects the tone of an audio for a certain degree. We extract MFCC vector sets from each audio and classify them as cluster sets, then calculate the distance between each cluster, use the nearest result as the similar retrieval result.

2.1 RETRIEVAL SYSTEM FRAMEWORK

In this section we review the audio retrieval system flow. The search processes, as shown in Figure 1, can be summarized in the following phases: first, extract MFCC features from all songs in MP3 music database and establish a MFCC feature database. The length of the MFCC feature is the 30 seconds in the middle of the front half of the corresponding song. Next, compress the extracted MFCC features by LBG Design Algorithm, then use k-means cluster them. When to retrieve a song, first extract the MFCC feature MF from query music, and then calculate the similarity distance between MF and the MFCC feature database using KL divergence function or EMD distance function, finally sort the distance in ascending order, output the top 100 songs as final retrieval results.

Details about the feature extraction and distance calculation are given in Figure 1.

2.2 MFCC FEATURE EXTRACTION

Mel frequency cepstrum coefficient just like the cepstrum is the feature, which is used to represent the channel characteristics. Cepstral analysis is mainly to get the audio spectral envelope by Fourier transformation, but, the human auditory perception experiments showed that human auditory perception only focus on certain areas, rather than the whole spectral envelope. The human auditory system is a special non-linear system, MFCC consider the characteristics of human hearing, first, it mapped a linear spectrum to Mel nonlinear spectrum based on auditory perception, and then convert it to cepstrum. Logan et al [10] through music modeling verified the effectiveness of MFCC feature in speech and music field.

The MFCC extraction algorithm performs as follow:

Step 1: Pre-emphasis the voice signal with a pre-emphasis filter, magnify the high frequency part. By doing so, make the channel characteristics more clearly. The pre-emphasis filter is determined as:

$$y(n) = x(n) - px(n-1), \tag{1}$$

$x(n)$ is the waveform of the voice, p is the pre-emphasis coefficient.

Step 2: To reduce the edge effect, apply Hamming window to each frame whose waveform have already attached pre-emphasis filter, calculate the amplitude of the vector by FFT(Fast Fourier Transform).

Step 3: Use Mel Filter Bank compress the Amplitude Spectrum. The filter bank consists of 33 triangular shaped band-pass filters. The focus here is to generate Mel-Frequency. The dimension of Mel is the horizontal axis which is used to reflect the human auditory characteristics, its unit is mel. The lower the frequency, the narrower the interval, the opposite is also true, the higher, the wider. The human ear in the subtle low frequency sound could feel the pitch different, but in the high frequency sound, its goes harder.

Step 4: Transform the compressed values by DCT to remove the correlation between the signals in different dimension, map the signal into low dimensional space.

Step 5: Using the low dimensional composition of the obtained cepstrum as MFCC feature values.

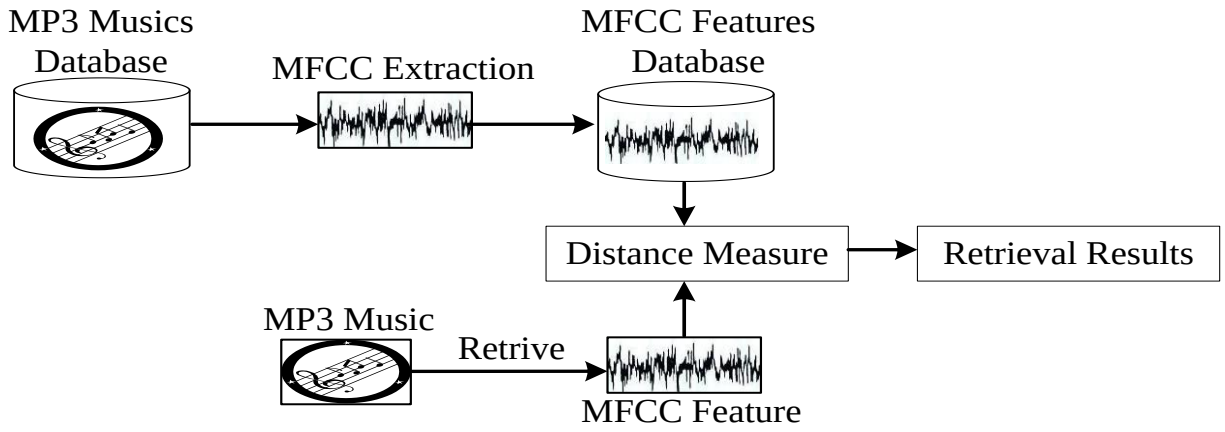


FIGURE 1 Flow of proposed method

This way, a music can be described by a series of spectrum vectors, and each vector is the MFCC feature vector of each frame.

The block diagrams for calculating MFCCs is given as Figure 2.

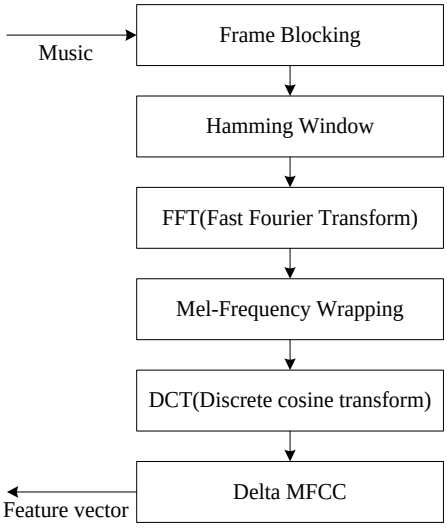


FIGURE 2 MFCC feature extraction processing

2.3 METHOD FOR SIMILARITY DISTANCE CALCULATION

In pure feature vector, the distance between each feature values can be expressed by Euclidean distance. But in audio retrieval, the feature value is not the vector but the normal distribution, that's why we cannot use a simple Euclidean distance.

2.3.1 KL Divergence

KL Divergence, a distance measure index of normal distribution, is the abbreviation of Kullback-Leibler Divergence, also known as Relative Entropy. It measures the difference between two probability distributions in the same sample space. According to the KLD, the definition of Kullback-Leibler Divergence between Normal distribution *N*1 and *N*2 (respectively, the mean vector, variance – covariance matrix) is:

$$D_{KL}(N_1 \| N_2) = \frac{1}{2}\left(tr(\sum_2^{-1}\sum_1) + (\mu_2 - \mu_1)^T \sum_2^{-1}(\mu_2 - \mu_1) - \ln\left(\frac{|\sum_1|}{|\sum_2|}\right) - k \right), \quad (2)$$

where *k* is the dimension of the mean vector. Usually the Kullback-Leibler Divergence is not symmetry, the value of $D_{KL}(N_1 \| N_2)$ and $D_{KL}(N_2 \| N_1)$ are not the same. However, this will not achieve our goals, so we use the Kullback-Leibler Divergence in symmetry case, that is:

$$0.5(D_{KL}(N_1 \| N_2) + D_{KL}(N_2 \| N_1)), \quad (3)$$

in this way we can ensure its symmetry.

2.3.2 The Earth Mover's Distance (EMD)

In computer science, the Earth Mover's Distance (EMD) is a measure of the distance between two multi-dimensional distributions in some feature space where a distance measure between single features. These distributions can be summarized with clustering algorithms, which reduce the feature space in a fixed number of bins. Each cluster c_j is associated with a weight w_j that indicates the size of the cluster (e.g. the occurrences of features in each cluster). The EMD describes the cost that must be paid to transform one distribution (considered as a mass of earth spread in space) into the other (considered as a collection of holes in the same space). The EMD measures the least amount of work needed to fill the holes with earth, considering that a unit of work corresponds to transporting a unit of earth by a unit of distance (ground distance). The EMD evaluation is based on the solution of the transportation problem, which consists in finding the least expensive flow from one distribution to another according to some constraints.

In a more formal way we can express the transportation problem as follows[11].

Let $P = \{(p_1, w_{p1}), ..., (p_m, w_{pm})\}$ be the first distribution with m clusters, where p_i is the cluster representative and w_{ij} is the weight of the cluster. $Q = \{(q_1, w_{q1}), ..., (q_n, w_{qn})\}$ the second distribution with n clusters; and $D = [d_{ij}]$ the ground distance matrix where d_{ij} is the ground distance between clusters p_i and q_j.

The flow $F = [f_{ij}]$ that minimizes the overall cost is computed by Eq. (4) where f_{ij} is the flow between p_i and q_j.

$$WORK(P, Q, F) = \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}, \quad (4)$$

and it is subject to the following constraints:

$$f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n, \quad (5)$$

$$\sum_{j=1}^{n} f_{ij} \leq w_{pi}, 1 \leq i \leq m, \quad (6)$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{qj}, 1 \leq j \leq n, \quad (7)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{pi}, \sum_{j=1}^{n} w_{qj}\right). \quad (8)$$

The first constraint indicates that the items can be moved only from *P* to *Q* and not vice versa. The next two constraints are related to the amount of mass which can be sent from the elements in *P* (it must not exceed the weight values) and to the amount which can be received by elements in *Q* (again limited by the weights). The last constraint forces to move the maximum amount of mass as possible. After solving the transportation problem and computing the total flow *F*, the EMD is defined as the work normalized by the total flow:

$$EMD(P,Q)=\frac{WORK(P,Q)}{\sum_{i=1}^{m}\sum_{j=1}^{n}f_{ij}}. \quad (9)$$

The EMD is a robust method to compare multidimensional distributions of features. It is a true metric if the ground distance is metric and if the total weights of the two signatures are equals

3 Experiment and Evaluation

3.1 CONDITION FOR MFCC FEATURE VALUES EXTRATION

Extraction feature values from the entire song will lead the data amount being too large, to reduce calculation, Pampalket [12] proposed choosing the centred 30 seconds of each song as a proposing target. What's more, according to the characteristics of the song, usually the first half of a song can reflect its characteristic most, and for a considerable part of the song, the second half of them is usually repeating the first half. For this reason, in our experiment we use the MFCC feature extracted from the middle 30 seconds in the first half of a song as its processing target.

Table 1 shows the 30 seconds MFCC feature vectors of song A, sampling frequency 16Hz, shift width 160 samples.

One row in Table 1 represent 20 dimensional vectors of a frame, and a 30 seconds audio include 3000 frames, that's 6000 dimensional vectors in total, the calculation can be pretty big. To reduce this, we use LBG Design Algorithm which proposed by Lindeet [13] to compress it. After clustered 6000 dimensional vectors by this algorithm, we classify the similar vectors to the corresponding class, and assume all classes are normal distribution to calculate its mean vector and covariance matrix, using the result as the new feature value, therefore the feature value of audio *A* can be described as:

$$A=\left\{\left(\mu_{A1},\Sigma_{A1},w_{A1}\right),...,\left(\mu_{Am},\Sigma_{Am},w_{Am}\right)\right\}, \quad (10)$$

where *m* is the number of classes.

Using *k*-means as cluster algorithm. By clustering, we can get the reduced-dimensional vector:

$$A_{MFCC}=CA_i\sigma_i, \quad (11)$$

A_i is the mean vector of each class, σ_i is the covariance matrix, *C* is the number of classes. In this experiment *C* taken 16, 32 and 64 respectively. By using this equation, when the feature vector of audio A using 16 as the number of classes, a 30 seconds MFCC feature vector can be compressed to 6720 from 60000 dimensions.

3.2 DATASETS

To verify retrieval results, our experiments collected 1951 MP3 audios and songs in Chinese and other languages. We manually classified these audio files by its types and kept them in different folders, details are shown in Table 2.

3.3 RESULT AND EVALUATION

Our experiment evaluate the retrieval system from two aspects.

Let us do a brief review of this experiment at first. First choose 10 audios from each class, there are 6 classes, so its 60 audios in total, use these audios for querying and calculate the effectiveness of the retrieval system. While retrieving, extract the MFCC feature MF from the middle 30 seconds in the first half of the query music, calculate the EMD distance between MF and the MFCC feature database, outputs the top 100 songs as final results.

TABLE 1 MFCC Features of Song A

–5.36	–0.20	–0.92	...	–1.40	–1.78	1.40	–0.66	22.44
–1.17	–2.34	5.57	...	3.82	2.54	–0.87	1.42	22.75
–0.77	–5.59	5.56	...	4.30	4.36	–2.55	1.47	22.78
0.17	–6.48	7.63	...	3.76	5.45	–3.11	4.00	22.57
...	...	...	...	...	...	...	...	...

TABLE 2 Mp3 music datasets of experiment

Folder	Audio amount	Size
Encouragement	161	659.4MB
Fervor	334	1.4GB
Gladness	341	1.3GB
Quiet	301	1.2GB
Romantic	332	1.3GB
Sentimental	482	2.1GB
Total	1951	7.96GB

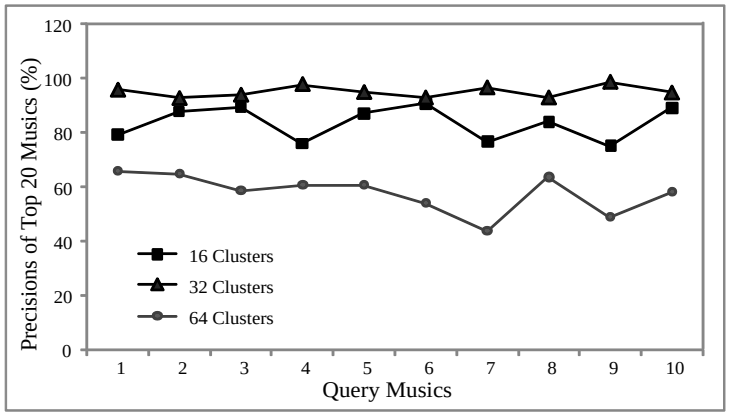


FIGURE3 The retrieval accuracy in using datasets with different classes cases.

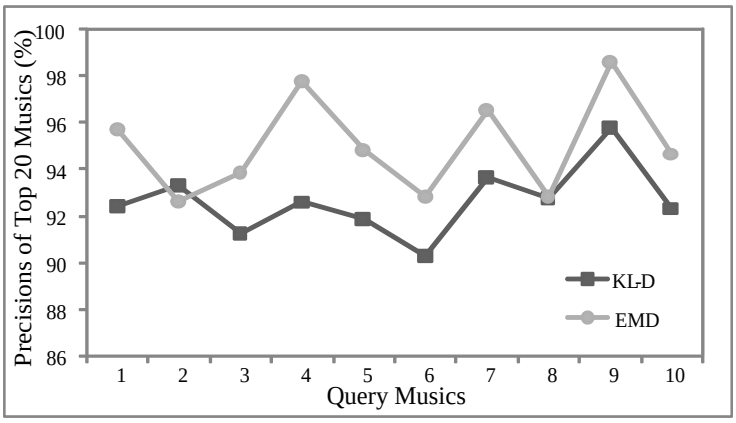


FIGURE 4 The retrieval accuracy by using different distance calculation method

We chose 10 most representative music to do this experiments, each retrieved 3 MFCC datasets (including 16, 32 or 64 classes) respectively. According to the retrieval results, select top 20s to compare with the corresponding class, calculate the precision to compare their accuracy. Our retrieval results show that in a MFCC datasets have 16 classes the accuracy is 83.28%, the accuracy is 95.05% in a 32 classes datasets and 57.85% in a 64 classes datasets. Therefore, the MFCC datasets which have 32 classes can get a better retrieval precision, the larger the number of classes, the lower the accuracy of query and the longer the calculation time. The results are shown in Figure 3.

And after that, for these 10 query music, we use KL divergence and EMD distance respectively to retrieve the MFCC datasets in 32 classes. Results shows as Figure 4.

From the experimental results, we can see that EMD is better than KL divergence under the same conditions.

4 Conclusion and prospects

This passage proposed an audio retrieval system, which is based on MP3 audio compression format, the audio feature values use MFCC and the retrieval distance use KL-D and EMD respectively. The proposed method are mostly focused on the first half of an audio, because the latter part usually repeat the previous section. Therefore we extract MFCC features from the middle 30 seconds of the first half to construct the feature vector datasets, and use LBG design algorithm compress the extracted features, then reduce dimension by using cluster algorithm.

In order to test and verify the effectiveness of our experiments, we collected 1951 different types MP3 music and songs in all kinds of languages. We tested MFCC feature datasets in 16, 32 and 64 classes. Experiment results show that when the class number is 32 the retrieval result is the best. When the number of class goes to 64, the retrieval amount becoming bigger, time consuming more, even the accuracy goes worse.

The proposed method only applies to MFCC feature, and because MFCC is based on human auditory characteristics, it is worked not very well on the low frequency area, which is sensitive to human ear. Our future work will focus on how to extract audio features more efficient. In addition, because the music features are usually got a large amount, in the future we will do more works on how to enhance the retrieval speed and reduce the calculation.

Finally, this experiment using a small MP3 music database, the future we will do further experiments in high-speed retrieval effectiveness aspect on the use of a large-scale database.

Acknowledgements

This research was partially supported by “the Fundamental Research Funds for the Central Universities (No. 13D11205)”.

Reference

[1] Casey M A, Veltkamp R, Goto M, Leman M, Rhodes C, Slaney M 2008 *Proceedings of the IEEE* **96**(4) 668-96
[2] Ghias A, Logan J, Chamberlin D, Smith B C 1995 Query by humming: musical information retrieval in an audio database *Proceedings of the third ACM international conference on Multimedia* ACM 231-6
[3] Sonoda T, Goto M, Muraoka Y 2002 A WWW-based melody retrieval system *Electronics and Communications in Haman (Part II: Electronics)* **85**(9) 63-74
[4] Kosugi N, Nishihara Y, Sakata T, Yamamuro M, Kushima K 2000 A practical query-by-humming system for a large music database *Proceedings of the eighth ACM international conference on Multimedia* ACM 333-42
[5] Rho S, Hwang E 2006 FMF: Query adaptive melody retrieval system *Journal of Systems and Software* **79**(1) 43-56
[6] Rho S, Han B-j, Hwang E, Kim M 2008 MUSEMBLE: A novel music retrieval system with automatic voice query transcription and reformulation *Journal of Systems and Software* **81**(7) 1065-80
[7] Haitsma J, Kalker T 2002 A highly robust audio fingerprinting system *ISMIR 2002* 107-15
[8] Bellettini C, Mazzini G 2010 A Framework for Robust Audio Fingerprinting Journal of Communications **5**(5) 409-24
[9] Xiao Q, Saito N, Matsumoto K, Luo X, Yokota Y, Kita K 2013 Index Compression for Audio Fingerprinting Systems Based on Compressed Suffix Array International *Journal of Information and Education Technology* **3**(4) 455-60
[10] Logan B 2000 Mel Frequency Cepstral Coefficients for Music Modeling *ISMIR*
[11] Rubner Y, Tomasi C, Guibas L J 2000 The earth mover's distance as a metric for image retrieval *International Journal of Computer Vision* **40**(2) 99-121
[12] Pampalk E 2006 Computational Models of Music Similarity and their Application in Music Information Retrieval *Vienna University*
[13] Linde Y, Buzo A, Gray RM 1980 *IEEE Transactions on Communications* **28**(1) 84-95

Authors

Xin Luo, born in October, 1972, Shanghai, China

Current position, grades: assistant professor at the School of Computer science and technology at Donghua University (Shanghai).
University studies: PhD at the Faculty of Engineering at University of Tokushima, Japan (2007).
Publications: 1 book, 20 Papers.
Scientific interests: multimedia information retrieval and pattern recognition.

Xuezheng Liu, born in November, 1989, Qingdao, Shandong, China

Current position, grades: postgraduate at Donghua University.
University studies: BS degree in information security at Donghua University.
Scientific interests: artificial intelligence and cloud computing, music information retrieval.

Youqun Shi, born in September, 1964, Xuzhou city, Jiangsu Provence, China

Current position, grades: professor at Donghua University.
University studies: PhD in Application of Computer Technology, China University of Miming & Technology.
Scientific interests: component oriented programming, soft as a service.
Publications: 1 book, 15 papers.

Ran Tao, born in October, 1975,Shanghai, China

Current position, grades: Master of Computer Technology, Senior Engineer and Master supervisor at Donghua University
University studies: Master degree at Donghua University in Shanghai in 2007
Scientific interests: wisdom education and E-commerce in cloud computing
Publications: 4 patents, 10 papers, 2 books

Inversion strategy research of travel time tomography with sparse rays

Yali Guo*, Yan Han, Linmao Liu

National key laboratory of Electronic test Technology in North University of China, Taiyuan, Shanxi, China

Received 1 October 2014, www.cmnt.lv

Abstract

The incomplete data travel time tomography with sparse rays can result in the ill-posed inverse problem in practical engineering, so the inversion strategy is very important in order to obtain reasonable inversion result. In this paper, the generalized inverse theory is taken and the influences are discussed which the system layout, initial model and prior information will impose on the inversion. The indexes of system optimal layout, the selection principle of initial model and regularization methods are presented in this paper. A velocity model of explosion is imitated and the inversion results are compared. A conclusion can be gained that system optimal layout, initial model rational selection and regularization methods utilization can help to improve inversion precision farthest in practical project.

Keywords: sparse rays, inversion strategy, regularization methods

1 Introduction

Computerized tomography is one of the non-destructive testing methods [1]. We can achieve the distribution of physics parameter existing in detected target by laying sensors outside the detected target and resolve the engineering and technology problems [2-4]. Computerized tomography technology has already used in medical and industrial fields widely.

Elastic wave travel time tomography is one of the important methods of engineering physics survey, it has played an important role in industry testing and resources prospecting [5]. It is necessary that there are enough detecting rays through the target in order to achieve high accuracy image information inside the target, however, which will increase the practical project difficulty and improve the cost. The data of elastic wave computerized tomography in practical project is usually incomplete owing to the insufficient driving sources and detectors. This engineering problem results in the incomplete tomography data inevitably.

Incomplete data tomography has the problems of sparse inverse data and low inversion accuracy and the inversion result is determined by a good many factors [6]. This paper aiming at the problems of travel time tomography with sparse rays, analyses the influences that the system layout, initial model and prior information imposed on the inversion by numerical experiment. In practical project, we can improve inversion accuracy by system optimal layout, initial model rational selection and utilization of regularization methods.

2 Inversion strategy analysis

For travel time tomography, there is the equation below:

$$DS = T, \tag{1}$$

where $T = (t_1, t_2 \cdots t_m)'$ is the m-dimension column vector of travel time; $S = (s_1, s_2 \cdots s_n)'$ is an unknown *n*-dimensional column vector, it expresses unknown discrete element slowness value in discrete cell; D is the distance matrix of $m \times n$ and its element is d_{ij}.

2.1 SYSTEM LAYOUT OPTIMIZATION

2.1.1 Mesh generation

The tested area is divided into numbers of regular meshes and each mesh has a uniform wave velocity. The more meshes are generated, the higher resolution of computed tomography will be achieved and the more uncertain solutions will be achieved too. Mesh generation should accord to the tested area size, the prior information (such as velocity distribution characteristic, abnormal body size, sampling position etc.), reconstruction accuracy, the number of driving sources and detectors.

2.1.2 Optimal distribution of sensors and judgment indexes

When designing sensors position, we should meet the following principles: extensive coverage and uniform distribution of rays, reduced number of zero elements in distance matrix. In order to make the tested area covered by rays as much as possible and achieve the effective detection, we should have a rational distribution of the sensors according to the follow factors: the ray density, orthogonality and the condition number of matrix D.

The ray density represents the number of rays passing through each mesh. The ray orthogonality is measured by maximal sine value of angle between rays [7]. The greater

***Corresponding author** e-mail:guoyali@nuc.edu.cn

the ray density is and the better the orthogonality is, the smaller inversion error will be achieved.

The tomography inversion is to solve the ill - posed Equation (1). The inversion stability is determined by the condition number of matrix *D*. The bigger condition number can result in poor inversion stability. Supposing that observed data *T* has a minor change, the variation of solution is δS Equation (1) has the relation:

$$D(S+\delta S)=T+\delta T\ . \tag{2}$$

Then:

$$\delta S=D^{-1}\delta T\ . \tag{3}$$

According to the property of subordinate norm, there is the relation: $\|\delta S\|\le\|D^{-1}\|\|\delta T\|$ and $\|T\|\le\|D\|\|S\|$, so $\frac{\|\delta S\|}{\|D\|\|S\|}\le\frac{\|D^{-1}\|}{\|T\|}\|\delta T\|$. That is:

$$\frac{\|\delta S\|}{\|S\|}\le cond(D)\frac{\|\delta T\|}{\|T\|}\ . \tag{4}$$

Then,

$$cond(D)=\|D\|\|D^{-1}\|, \tag{5}$$

where $cond(D)$ is the condition number of matrix *D*. Rational distribution of the sensors can reduce the condition number and we can receive the more stable solutions.

2.2 THE INITIAL MODEL SELECTION

When the rays through the target are sparse, the inverse results depend on the initial model severely. If the deviation of initial model and true model is small, we can obtain better results, otherwise, the results aren't true.

This paper calculates the correlation degree of initial model and true model by correlation coefficient; analyses the change trend of correlation coefficient with inversion error and obtains the dependence degree of inversion result on initial model.

For two groups data of x_i and y_i , $i=1,2,\cdots n$, the correlation coefficient is defined as[8]:

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}\ . \tag{6}$$

The absolute value of *r* is closer to 1, the two groups data have higher correlation degree. We can set reasonable initial model by prior information and experience. We should analyse the character of model according to prior information before inversion. Experience formula is also a good way to set initial model. In a word, we should improve the correlation degree of initial model and true model as far as possible, only thus can we improve the inversion accuracy.

2.3 APPLICATION OF REGULARIZATION METHODS

Incomplete data travel time tomography with sparse rays makes the $n>m$ in Equation (1) and represents underdetermined problem in mathematics. In this condition, the inversion results have multiple solutions and credibility is poor. An effective way to resolve the underdetermined inversion problem of multiple solutions is adopting regularization methods to constrain the inversion process and the inversion results by prior information. At the same time, regularization methods make an important role in perfecting ill-conditioned problem.

The application of regularization methods includes: damping the asymmetrical covering of rays, damping the inaccuracy of observed data, tight constraint for true data, setting the value range of some parameters in the iterative procedure and so on. The ways of regularization adding have addition and multiplication. This paper presents some regularization methods as described below.

2.3.1 Damping the asymmetrical covering of rays

The ray coverage will affect the inverse results. The damping regularization methods aiming at the asymmetrical covering of rays is model covariance matrix, which weights different value in different mesh according to ray coverage. Generally, rays is denser, the inversion results are more accurate. Therefore, the meshes through by more rays have more information and should weight a larger value.

2.3.2 Damping the inaccuracy of observed data

The observed data contains noises. The data in different acquisition channel has different noises and the noises will be amplified in inversion. So, the data should have different weighting according to its accuracy. The damping matrix is data covariance matrix. For travel time tomography, the waveform signal is weakening with the transmission distance. The ray path is shorter, the signal to noise ratio is higher, so the smaller travel time should weight a larger value.

2.3.3 Tight constraint of true data

In practical application, we can measure the velocity of this position by individual sampling point in interested area and treat it as the prior information. Equivalent to increase the constraint equations in the linear Equation (1) with:

$$WS=S'\ , \tag{7}$$

where *W* is a $h\times n$ matrix. *h* is the number of sampling point. And the $w_i=[000\cdots10]$, the position of constrained parameter is 1, others are 0. S' is the observed data of sampling points. So the Equation (1) is rewritten as:

$$\begin{bmatrix} D_0 \\ W \end{bmatrix} \cdot S = \begin{bmatrix} T_0 \\ S' \end{bmatrix}, \tag{8}$$

where D_0, T_0 are the distance matrix and travel time matrix without prior information. Rewrite above equation becoming $DS = T$ and make calculation.

2.3.4 Setting the value range of some parameters

We can obtain the value range of some parameters by the prior information sometimes. Constraining the results by the value range of some parameters in the iterative procedure helps to improve the inversion precision.

3 Weighted generalized inversion algorithm

Solving the discrete inversion Equation (1) by matrix is the same as solving inverse matrix of D. However, in practical project, the problem of incomplete data owing to the insufficient number of driving sources and detectors makes Equation (1) become into a sparse, morbid, underdetermined, incompatible linear equation set [9]. The coefficient matrix-D in general is a singular matrix and its inverse matrix does not exist obviously. So it is necessary to adopt the generalized inverse theory to solve matrix-D.

Thinking of the regularization methods, we adopt model covariance matrix and data covariance matrix to damp the asymmetrical covering of rays and the inaccuracy of observed data. The weighted generalized inversion method is the generalized inversion algorithm combining with the regularization methods.

Given $A \in C^{m \times n}$, P and Q are positive definite matrix of $m \times m$ and $n \times n$ respectively. If $X \in C^{n \times m}$, satisfying:

$$\begin{cases} AXA = A, XAX = X \\ (PAX)^T = PAX \\ (QXA)^T = QXA \end{cases}. \tag{9}$$

Then X is defined as weighted generalized inverse of A: A_{mn}^{+}, A_{mn}^{+} is expressed as [10]:

$$A_{mn}^{+} = Q^{-1}(PAQ^{-1})^{+}P. \tag{10}$$

The data covariance matrix and model covariance matrix are P and Q, the diagonal element of P and Q is defined as:

$$diag\{P\} = T^{-1}, \tag{11}$$

$$diag\{Q\} = K. \tag{12}$$

T is the M-dimension vector of travel time of current model, the element of which is $t_i = \sum_{j=1}^{n} d_{ij} \cdot s_j$, $i = 1,2,3,\cdots,m$. K is the N-dimension vector, the element of K is $k_j = \sum_{i=1}^{m} d_{ij} \cdot v_j$ ($j = 1,2,3,\cdots,n$). v_j is velocity in mesh of No. j.

Consequently, travel time tomography based on weighted generalized inverse is rewritten as:

$$S = Q^{-1}(PDQ^{-1})^{+}P \cdot T. \tag{13}$$

4 Numerical simulation experiments

4.1 TESTING MODEL

The model is a velocity model of explosion field as shown in Figure1. The tested area is divided into 10×10 meshes, and bomb is placed in the centre of test area as shown in Figure 2. With the principle of symmetry, we only need make velocity inversion in the 1/4 area. The sensor number is no more than 20. This is a typical model of travel time tomography with single driving sources and sparse rays.

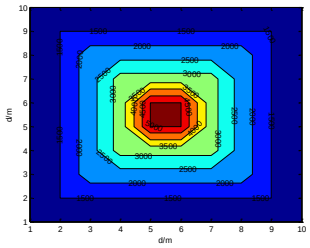


FIGURE 1 The velocity model of explosion field

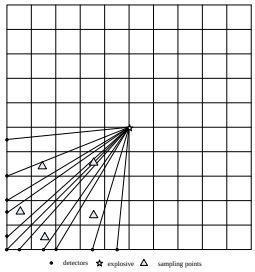


FIGURE 2 The tested area layout

Defining, the number of regular meshes is n, The relative velocity error in the No. j mesh is defined:

$$\delta_j = \frac{v_j - v_j'}{v_j'}, \tag{14}$$

where v_j' is the true velocity value in No. j grid , v_j is the inversion result.

The average relative error in all meshes is defined:

$$\bar{\delta} = \frac{1}{N}\sum_{j=1}^{N}\left|\delta_j\right| = \frac{1}{N}\sum_{j=1}^{N}\frac{\left|v_j - v_j'\right|}{v_j'}. \tag{15}$$

4.2 SIMULATION OF SENSORS DISTRIBUTION

We select the sensors number is 13; two different sensor layouts are presented in Figure 3. The first sensor layout is symmetrical as shown in Figure 3a, the second is optimized layout by the judgment indexes in this paper as shown in Figure 3b.

The ray density and orthogonality distribution in the whole tested area in the two layouts are in Figure 4 and Figure 5. The condition number of matrix D in the two layouts are $2.74 \cdot 10^{16}$ and 35.61 respectively. We can see that the second layout has more reasonable ray distribution, the better orthogonality and smaller condition number. The initial model is in Figure 6. We make velocity inversion and the relative errors in each mesh are in Figure 7, the average relative errors are 8.47% and 3.32%. The inversion results are in Figure 8. We can see that the inversion errors are smaller and inversion result is approaching the true model with the optimized sensors layout in this paper.

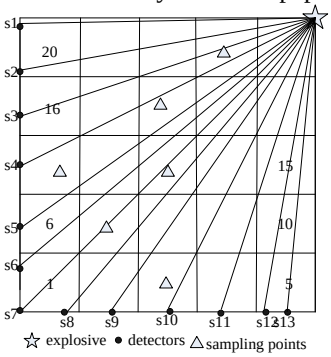


FIGURE 3a The first layout

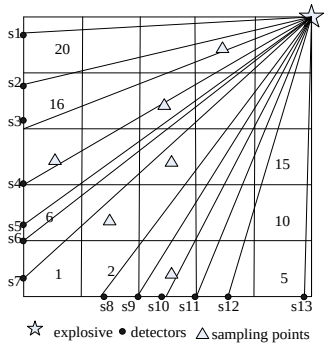


FIGURE 3b The second layout

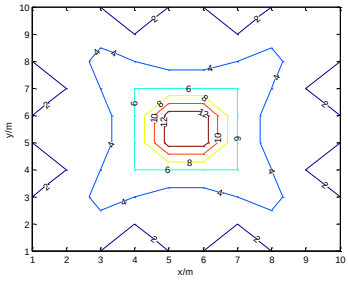


FIGURE 4a The first density distribution

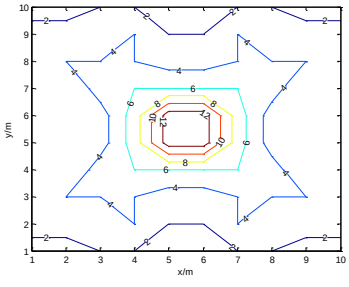


FIGURE 4b The second density distribution

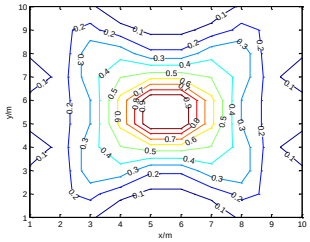


FIGURE 5a The first orthogonality distribution

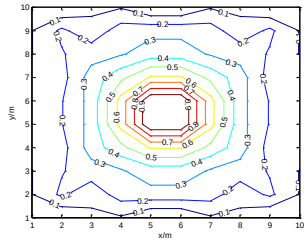


FIGURE 5b The second orthogonality distribution

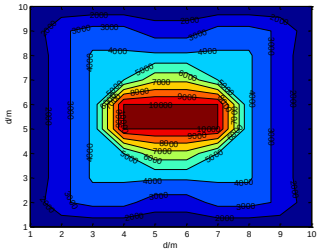


FIGURE 6 The initial model

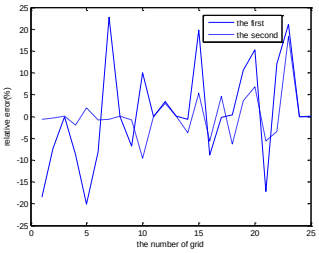


FIGURE 7 The relative error in each mesh

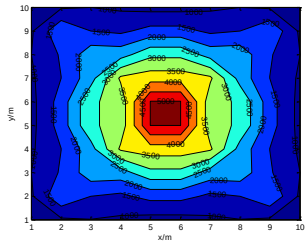


FIGURE 8a The first inversion results

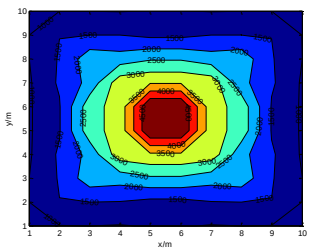


FIGURE 8b The second inversion results

4.3 INFLUENCE OF INITIAL MODEL

We generate the initial model by adding a certain proportion of random noise to true model and make the correlation coefficient of initial model and true model change from 0 to 1. We adopt the same weighted generalized inverse algorithm to simulate above model with different initial model. The relationship curve of correlation coefficient with inverse average relative error is in Figure 9. We can see that the correlation coefficient of initial model and true model is closer to 1, the inverse average relative error is smaller. Figure 10 is the relative error in each mesh when the correlation coefficient is 0.97, 0.75, 0.54 and the initial model is even velocity.

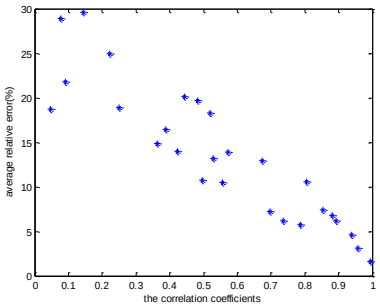


FIGURE 9The relationship curve of correlation coefficient with inverse average relative error

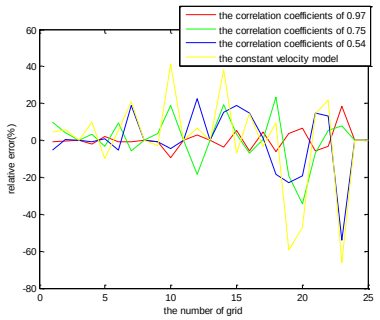


FIGURE 10 The relative error in each mesh with different correlation coefficient

From above analysis, we can conclude that inverse results depend on the initial model severely in the same condition. Selection reasonable initial model is very important for incomplete data travel time tomography with sparse rays.

4.4 INFLUENCE OF PRIOR INFORMATION

We make velocity inversion by generalized inversion algorithm with regularization methods. Figure 11 shows the different results of generalized inversion algorithm with regularization methods and without regularization methods. The average relative errors in all meshes are 4.2% and 19.81% respectively.

Figure 12 shows the influence of tight constraint of true data on inversion. When the number of sampling points is respectively 3, 5, 6, the relative errors in each meshes are shown in Figure 12. We can see that the relative error decreases with the number increase of sampling points.

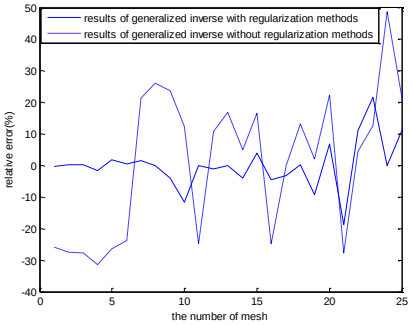


FIGURE 11 Inversion relative error in each mesh

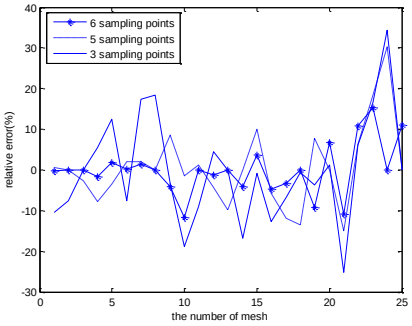


FIGURE 12The influence of tight constraint of true data on inversion

5 Conclusions

The particularity and complexity of the detected target result in the incomplete distribution of sensors. The travel time tomography owing to the insufficient driving sources and detectors is an incomplete data travel time tomography with sparse rays. This paper aiming at the problems of travel time tomography with sparse rays, analyses the influences that the system layout, initial model and prior information imposed on the inversion. We obtain the conclusions as follows:

In the condition of sparse rays, system optimal layout helps to improve the inversion accuracy. The indexes of system optimal layout are presents;

When the rays through the target are sparse, the correlation degree of initial model and true model is higher, the results are more accurate. We can set reasonable initial model by prior information and experience.

We can adopt regularization methods by prior information to overcome the underdetermined inversion problem of multiple solutions.

In practical project we can improve inversion accuracy by system optimal layout, initial model rational selection and utilization of regularization methods.

Acknowledgements

The authors would like to thank the reviewers and editors for their valuable comments and sincere helps, which contribute much to the improvement of this paper. This work was supported by the National Natural Science Foundation of China under Grant 61171179.

References

[1] Li G, Li T 2012 *Journal of Yan'an Vocational &Technical Institute* **26**(3) 90-3 *(in Chinese)*
[2] Liu H, Xue X 2003 Journal of Engineering Geology **11**(4) 435-9
[3] Watanabe T and Sassa K 1996 *International journal of rock mechanics and mining sciences & geomechanics abstracts* **33**(5) 467-77
[4] Wéber Z 2001 *Physics of the Earth and Planetary Interiors* **124** 33-43
[5] Liu G, Wang Z, Sun J 2003 *China Civil Engineering Journal* **36**(5) 76-81 *(in Chinese)*
[6] Cheng G, Ma Z, Zhang B 2003 *Progress in Geophysics* **18**(3) 512-8 *(in Chinese)*
[7] Pei Z, Yu Q, Di B 2006 *Geophysical and Geochemical Exploration* **26**(3) 218-24 *(in Chinese)*
[8] Chen J 1995 *Probability and statistics* Higher Education Press Beijing *(in Chinese)*
[9] Liu Y, Dong L, Xia J 2007 *Oil Geophysical Prospecting* **6**(42) 682-5 *(in Chinese)*
[10] Berryman J G 1989 *Physical Review Letters* **62**(25) 2953-6

Authors

Yali Guo, bbborn in June, 1980, Taiyuan, China

Current position, grades: instructor at North University of China.
University studies: master's degree in Signal Processing at North University of China.
Scientific interests: signal processing, reconstruction and inversion.
Publications: 2 patents, 10 papers.

Yan Han, born in June, 1957, Taiyuan, China

Current position, grades: professor and doctoral supervisor at North University of China.
University studies: doctor's degree in Signal Processing at Beijing Institute of Technology.
Scientific interests: signal processing, non-destructive testing, reconstruction and inversion.
Publications: 30 patents, 200 papers.

Linmao Liu, born in December, 1980, Taiyuan, China

Current position, grades: instructor in North University of China.
University studies: master's degree in Signal Processing at North University of China.
Scientific interests: signal processing, reconstruction and inversion.
Publications: 5 papers.

Diffusion limited aggregation of magnetic particles with exponential decreasing interactions in three-dimensional space

Wei Qiao[1, 2], Jie Sun[2], Qingfu Du[2]*

[1]*College of Control Science and Engineering, Shandong University, China*

[2]*School of Mechanical and Electrical Engineering, Shandong University at Weihai, China*

Received 1 March 2014, www.cmnt.lv

Abstract

Using the Monte Carlo simulation, we investigate the three-dimensional fractal growth of a magnetic diffusion-limited aggregation (MDLA), which consists of magnetic particles interacting with an exponential potential βCe-αr. The cluster morphology, fractal dimension and magnetic susceptibility of this MDLA are analysed with respect to the range facto rα and the coupling energy βC. In the case of long-range ferromagnetic interaction, our results show that the cluster morphology grows to be a hexagonal symmetry as the coupling energy increases, which is different from the two-dimensional simulation. For a proper coupling energy, the fractal dimension takes the maximal value and the cluster morphology becomes more compact. In the case of short-range interaction, the critical value of the cluster specific magnetic moment is much larger than the simulation result in the MDLA with the interacting potential of power law.

Keywords: diffusion limited aggregation, magnetic particle, exponential decreasing potential

1 Introduction

Since the "diffusion limited aggregation" (DLA) model was introduced by Witten and Sander [1] in 1981, it has been extensively studied for the cluster fractal growth simulation and applied in various fields such as the viscous fingering [2], the electrochemical deposition [3] and the thin film growth [4]. Among the simulation efforts are DLA models that have been proposed to investigate the influences by different physical conditions on the cluster growth [5-10]. Hassan et al. [11] investigated a simple model, which described the aggregation kinetics of two-component (two-different-size) particles with stochastic self-replication. The particle anisotropy and particle shape are also discussed by Liu et al. [12] and Li et al. [13]. Vandewalle and Ausloos [14] introduced the spin degree of freedom for the aggregating magnetic particles. They studied the dependence of the fractal growth on the nearest neighbor coupling among the magnetic dipoles as well as the external magnetic field. Indiveri et al. [15] studied the morphology and symmetry of a two-dimensional cluster growing in the long-range interaction scenario with the power law potential $r^{-\alpha}$ (r is the distance between any two-grid points in the system, α is a positive constant). Furthermore, Xu et al. [16] studied the two-dimensional cluster morphology when the power law function interaction exists among magnetic particles. Their results show that, for the long-range ferromagnetic interaction and when βC takes appropriate values, the cluster morphology appears four-degree symmetry.

In many physical systems, the interaction potential appears the exponential function with respect to the distance between two objects. In ion chemistry, some atomic and molecular repulsive forces and hydrophobic forces decay as $e^{-\alpha r}$ [17]. In the direct electrochemistry of proteins, the reaction rate of the electro active substance on the electrode decreases exponentially with respect to the distance between the electron donor and acceptor [18]. In condensed matter physics moreover, the interactions among some electric dipoles and magnetic dipoles decay exponentially with their distances. In some correlative many body systems, the exponential potential $e^{-\alpha r}$ is often used as an approximation of the practical interaction [19]. Tomohiko et al. [20] using the Yukawa potential $r^{-1}e^{-\alpha r}$ to study the phase diagram of the Pd-Mn alloy. By means of the Fourier transformation.

This paper discusses the magnetic diffusion limited aggregation (MDLA) with magnetic particles diffusing in the three-dimensional space. The dipole-dipole coupling between magnetic particles is assumed to decay exponentially with the distance $\beta Ce^{-\alpha r}$. Using the Monte Carlo method, we study the fractal growth dynamics. The cluster morphology, the fractal dimension, and they are analysed with respect to the decay rate α and the coupling strength βC.

2 Modeling and simulation methods

Based on the well-known DLA model, the MDLA model introduces the spin degree of freedom [21], where each particle is a two-level system with $\sigma = +1$ standing for spin-up and $\sigma = -1$ standing for spin-down. The cluster is assumed to grow on a three-dimensional cubic lattice. The simulation algorithm is as follows:

***Corresponding author** e-mail: shyshgl@163.com

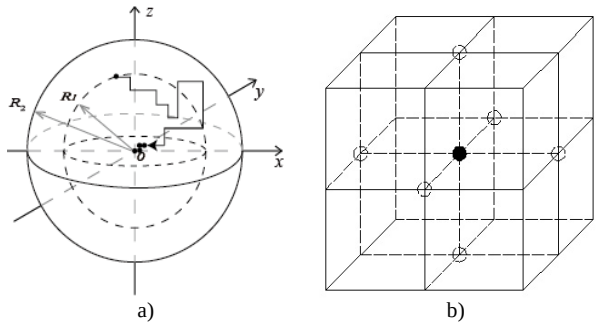


FIGURE 1 a) Illustration of the MDLA model and b) the six neighbor sites for a particle from the initial position diffusing

1) A seed particle with an initial spin state is placed at the centre of the cubic lattice (seed site).

2) Let the radius of the cluster aggregation is R_0, a diffusing particle with spin up or down is randomly generated on the spherical surface of radius $R_1 = R_0 + 20$, centred on the seed site.

3) The diffusing particle moves in the lattice according to the probability, which is determined by the change of the system energy with respect to the position and spin orienttation. Considering the Ising coupling among the spins and their potential energy in the outer magnetic field, the dimensionless system energy can be defined as follows:

$$\begin{aligned} &\beta E = -\sum_{\langle i,j \rangle} J r_{(i,j)} \sigma_i \sigma_j - \sum_i \beta H \sigma_i \\ &J\left(r_{i,j}\right) = \beta C e^{-\alpha r_{i,j}},\ \alpha \geq 0, \end{aligned} \quad , \qquad (1)$$

where $J(r_{i,j})$ is the coupling energy which is related to the distance $r_{i,j}$ between the particles *i* and *j*, βH is the outer magnetic field, and σ_i is the spin variable of the *i*-th particle. βC is the coupling strength related to temperature and α reflects the decay rate of the exponential function with the distance.

As illustrated in Figure 1, there are six neighbour sites where the diffusing particle from the initial site can move. The probability of moving to the *m*-th site is given by

$$P_m = \frac{\exp(-\Delta\beta E_m)}{\sum_{k=1}^{6} \exp(-\Delta\beta E_k)}, \quad m,k = 1,2,\cdots,6, \qquad (2)$$

where $\Delta\beta H$ is the energy difference induced by the change of the position and spin orientation. Using the roulette algorithm [22], we can select the location of the particle diffusion and spin orientation.

If the particle diffuses to the site connected to the condensed cluster, the particle will adhere to the cluster, and the spin orientation of the particle will be frozen. If the particle moves outside the spherical surface of radius $R_2 = R_0 + 30$, centred on the seed site, the particle disappears. In either case, one returns to the second step for another diffusing particle. The diffusing processes are repeated until a fractal cluster of *N* particles is formed.

3 Results and discussion

As an extension of the classical DLA model, the energy difference in the MDLA model is jointly determined by the diffusing particle's position and spin orientation. By tuning the parameters βH, βC and α, we can simulate the growth process of condensed clusters in the situation of different magnetic fields and temperatures. For simplicity, we consider the case of magnetic field strength $\beta H = 0$ in this paper. The spin of the seed particle is set to be spin-up. The two spin states, up and down, are illustrated by blue color and red color respectively. According to the energy Equation (1), we can know that, when βH and βC are close to zero, the energy difference before and after the particle diffusion is approximately zero. Therefore, the probabilities of the particle diffusing to different nearest neighbors are same. The motion is approximated to the uniform random walk process, and the formed cluster morphology is similar to the DLA model.

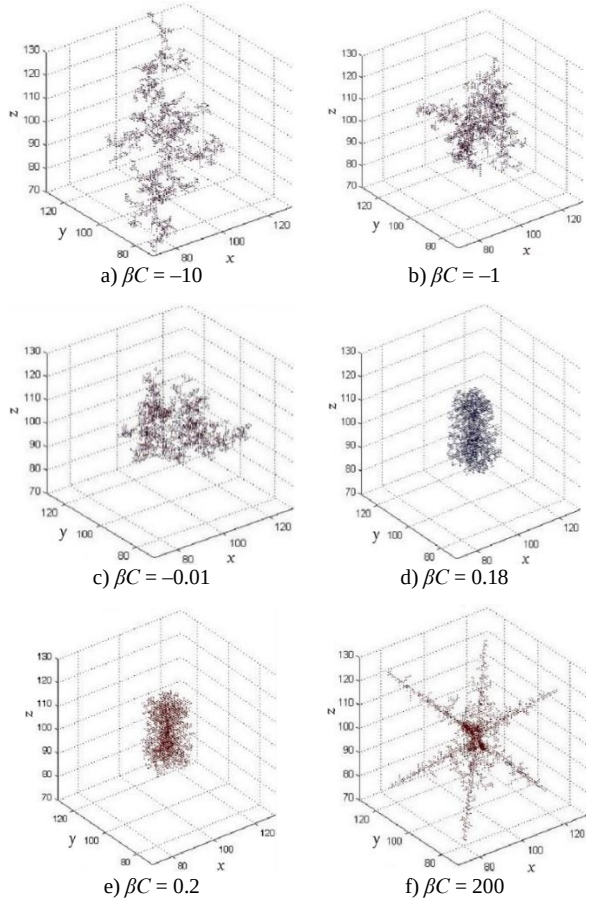
3.1 FRACTAL MORPHOLOGY OF THE LONG RANGE MAGNETIC INTERACTION SYSTEM

When the parameter α is small, the coupling energy between the magnetic particles damps slowly with their distance, and hence there is a long range interaction between the magnetic particles. Figure 2 shows the cluster morphologies with fixed $\alpha = 0.2$ and various $\beta C = -10, -1, 0.01, 0.18, 0.22, 200$.

Figures 2a and 2b show that when βC is negative, the cluster grows by the antiferromagnetic way. The cluster morphology becomes looser as $|\beta C|$ increases. The spin states of the magnetic particles present a regular crosswise distribution as shown in Figure 3a.

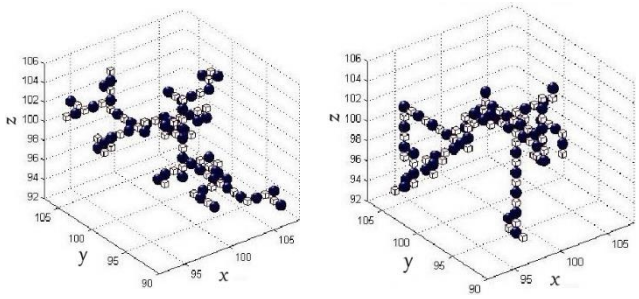
When βC is positive, the cluster grows by the ferromagnetic way. With the increasing βC, the cluster morphology becomes more and more regular as a six-degree symmetry structure distribution. As shown in Figure 1(f), the six-degree symmetrical structure corresponds to simple cubic lattice structure. In fact, the similar structure of six-degree symmetry widely exists in the systems (e.g., $\alpha = 0.1, 0.25, 0.3, 0.4$) of long range magnetic interactions.

When βC varies from 0.1 to 0.5, influenced by the random Brownian walking, the spin distributions of dominant up and dominant down appear alternately as shown in Figure 2c and 2d. However, when βC takes values greater than 0.8, the spins in the cluster are completely magnetized to the up state like the seed particle. This means that the force of the cluster imposing on the diffusing particle is increasing. The diffusing particles take the same spin states for the lower system energy. The certainty factor of the spin interaction plays a major role; while the effect the random factor is weaken.



a) $\beta C = -10$ b) $\beta C = -1$ c) $\beta C = -0.01$ d) $\beta C = 0.18$ e) $\beta C = 0.2$ f) $\beta C = 200$

FIGURE 2 The cluster morphologies for various βC but fixed $\alpha = 0.2$



a) $N = 200$, $\alpha = 0.2$, $\beta C = -50$ b) $N = 200$, $\alpha = 1$, $\beta C = -50$

FIGURE 3 The morphology of aggregates

3.2 FRACTAL DIMENSION OF THE LONG RANGE INTERACTIONS SYSTEM

In Figure 4, the fractal dimension of the cluster is plotted with respect to βC. In the case of antiferromagnetic growth, as $|\beta C|$ increases, the fractal dimension D_f gradually reduces to about 1.79. When βC reduces to - 100, the cluster no longer maintains the DLA fractal structure, but shows a broadband-like morphology.

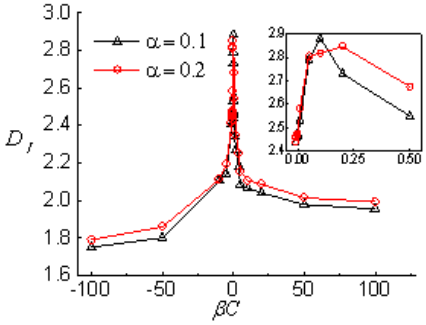
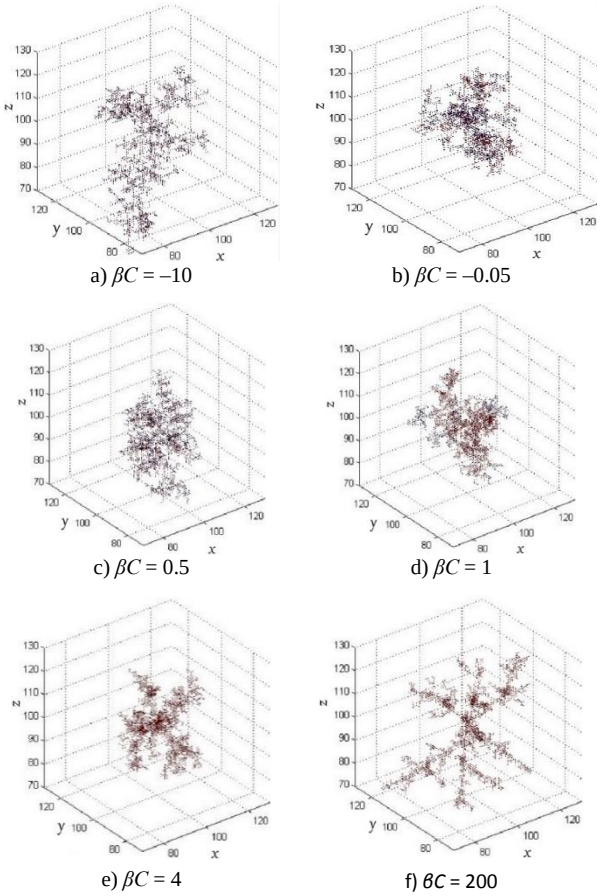


FIGURE 4 Aggregation clusters fractal dimension of different parameter βC when $\alpha = 0.1$, 0.2

In the case of $\beta C = 0$, clusters of fractal dimension is about 2.47. In the case of ferromagnetic growth, the value D_f of the cluster changes significantly with the increase of βC. Especially for the system of $\alpha = 0.1$, the value D_f has a maximum peak value of 2.88. In the system of $\alpha = 0.2$, D_f, also has a maximum peak close to 2.84. However, as βC continues to increase, the fractal dimension decreases sharply. As shown in Figures 2d and 2e, the morphologies of the clusters are relatively tight, and the system energies are the smallest, which means the systems are stable. With $|\beta C|$ increasing, the morphology varies similarly to the case of $\alpha = 0.1$, except that the value D_f drops relatively slow.



a) $\beta C = -10$ b) $\beta C = -0.05$ c) $\beta C = 0.5$ d) $\beta C = 1$ e) $\beta C = 4$ f) $\beta C = 200$

FIGURE 5 The the cluster morphology in different parameter βC but fixed $\alpha = 1$

3.3 FRACTAL MORPHOLOGY OF THE SHORT RANGE MAGNETIC INTERACTION SYSTEM

When the parameter α is large, the coupling energy between the magnetic particles damps quickly with their distance, and hence the magnetic particles have the short-range interactions with each other. Only when closing to the aggregation cluster, does the diffusing particles can be affected by the cluster heavily. Figure 5 shows the growth morphology of the clusters with $\alpha = 1$ and $\beta C = -10, -0.05, 0.5, 1, 4, 200$.

Figures 5a and 5b show that, when βC is negative, the clusters grow in the antiferromagnetic way. Similar to the long range interaction system, as $|\beta C|$ increases, the morphology of aggregation cluster is more and more loose, and the spin state presents a crosswise distribution as shown in Figure 3b. But under the long range interaction system, the morphology gets sparse relatively faster.

When βC is positive, aggregation clusters grow in the ferromagnetic way. Unlike the long rang interaction system, the cluster morphology does not vary obviously and won't appear the six degrees of symmetrical structure even though βC is very large. For the larger $\alpha(\alpha > 1)$, the branch tips of the cluster are forked. The cluster tends to be unstable and the morphology is not compact like that in Figure 2b any more. The reason is that, in the short range interaction system, when the diffusing particle is far away from the cluster, the influence made by the cluster to the diffusing particle is approximate to zero. Only when the diffusion particles move close to the aggregation clusters, does the magnetic interaction between the particles has a significant impact. However for the long range interaction system, from the beginning of its movement, the diffusing particle is affected by the cluster, and when βC takes appropriate values, the magnetic particles agglomerate more closely so that the cluster energy will be less and the system will be more stable.

When $\beta C < 1$ the spin-up particles and the spin-down particles appear in the aggregation cluster at the same time. But as βC increases, the number of spin-up particles increases. When $\beta C > 4$ all condensational particles are magnetized to spin-up state. In short-range interaction systems, the value of βC that magnetizes all aggregation clusters to spin-up state is much larger than that in the long range interaction systems.

3.4 THE FRACTAL DIMENSION IN THE SHORT RANGE INTERACTION SYSTEMS

Figure 6 shows the cluster fractal dimension curve with respect to βC when $\alpha = 1, 3$. We can see that when βC is negative, the cluster grows in the antiferromagnetic way. when $\alpha = 1$ with $|\beta C|$ increasing, the fractal dimension decreases gradually to about 2.13. When $\alpha = 3$, as $|\beta C|$ increases, the fractal dimension of clusters decreases gradually to about 2.4. When βC is positive, the cluster grows by ferromagnetic way. As βC increases, when $\alpha = 3$, $100 > \beta C > -100$, D_f, the fractal dimension distributes between 2.4 to 2.58 and does not appear the maximum peak that similar to the maximum peak in the long-range interaction systems. When βC is smaller, the value of D_f is about 2.48, and it is similar to the classic DLA model. With βC increasing, D_f, decreases gradually and tends to a stable value about 2.4. For short range interaction system, magnetic particles near the aggregation clusters determine direction of spin particles, therefore, regardless of ferromagnetic or antiferromagnetic growth way, it has a small influence to the cluster morphology, and the difference of its fractal dimension change is not obvious.

3.5 SPECIFIC MAGNETIC MOMENT

In order to study the evolution of the spin state under different forces, we analysed the dependence of the specific magnetic moment M of aggregation clusters on the parameter βC. The specific magnetic moment M of the cluster is the ratio between the total spin and the total number of the particles, where the total spin indicates the difference between the numbers of spin-up particles and the spin-down particles.

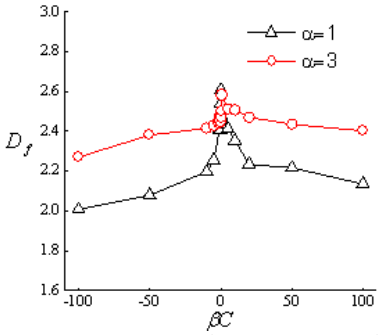


FIGURE 6 Aggregation clusters fractal dimension of different parameter βC when α = 1;3

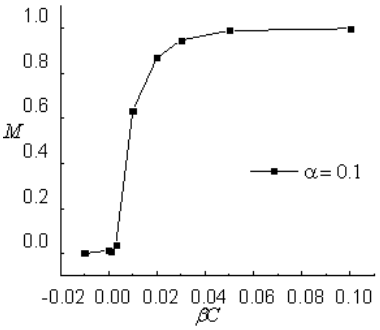


FIGURE 7 The curve of M changing with βC when α = 0.1

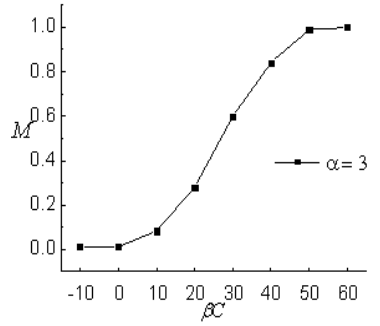


FIGURE 8 The curve of M changing with βC when α = 3

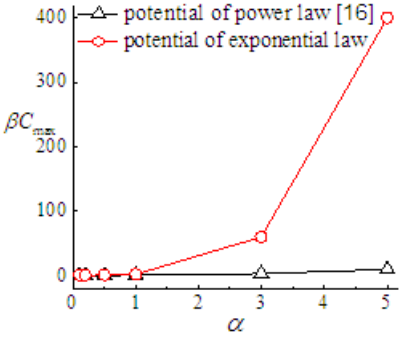


FIGURE 9 The curve of the critical value βC_{max} changing with different forcing process parameters α

As shown in Figures 7 and 8, in both cases of the long range interaction and the short range interaction, when βC is negative, the specific magnetic moment of aggregation clusters decreases and tends to 0 with $|\beta C|$ increasing. In the case of long range interaction $\alpha = 1$, the specific magnetic moment increases rapidly with βC increasing, and to the saturate value 1 as βC increases beyond 0.05. In the case of short range interaction $\alpha = 3$, the specific magnetic moment increases slowly with βC increasing, and when βC goes beyond 60. All of the aggregation particles are magnetized to spin-up state i.e., the specific magnetic moment $M = 1$. In the short-range interaction force, although the particle spin- flip probability is small, but once the spin flip, subsequent particle also will spin flip, resulting magnetization fluctuation phenomena.

For a given interaction range factor α, one can always find the threshold βC_{max}, which leads to the complete magnetization $M = 1$. Figure 9 shows the magnetization threshold βC_{max} as a function of the range factor α. In the case of long range interaction, the interaction between the particles did not change significantly with their distance, the magnetization threshold βC_{max} various slowly with the range factor [16]. As α increases to the short range interaction cases, the interaction strength between particles decays rapidly with their distance. The value of βC_{max} under the power law interaction is significantly greater than that under the exponential interaction. The influence that βC made to the specific magnetic moment in the power law interacting particles is far less than that made to the specific magnetic moment of exponential law interacting particles.

4 Conclusion

Using the MDLA model and assuming that the magnetic interaction force between the particles decays exponentially with their distance, we simulated the dynamic behavior of cluster fractal growth for different physical parameters. When the cluster grows in the antiferromagnetic way, the spin states of magnetic particles under different forces are in a crosswise distribution. Along with the decrease of the parameters of α and βC, the morphology of aggregation clusters gets more sparse, the fractal dimension decreases continuously, and the specific magnetic moment of aggregation clusters tends to be zero. When the cluster grows in the ferromagnetic way, for long range interaction systems, the cluster morphology and fractal dimension change significantly with the physical parameters. The cluster morphology evolves to regular six-degree symmetric structure from the ordinary DLA morphology, and there are also dense forms appear in the middle range. The specific magnetic moment M increases rapidly from 0 to 1 as the parameter βC increases. For short-range interaction systems, the cluster morphology and the fractal dimension change slowly with the physical parameters. Moreover, the compact structure and six-degree symmetric morphology does not appear. The specific magnetic moment M also increases slowly from 0 to 1 as the parameter βC increases. This conclusion provides a reference for explaining the related phenomenon of physics experiments and studying the fractal growth mechanism of magnetic particles, etc.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 61273088, 11271194 and 11201264).

References

[1] Witten T A, Sander L M 1981 *Physical Review Letters* **47**(19) 1400-2
[2] Herreman W, Molho P, Neveu S 2005 *Journal of Magnetism and Magnetic Materials* **289** 356-9
[3] Shaikh Y H, Khan A R, Pathan J M, Patil A, Behere S H 2009 *Chaos, Solitons & Fractals* **42**(5) 2796-803
[4] Hakan K, Mehmet B, Mursel A 2010 *Applied Surface Science* **256**(9) 2995-9
[5] Wang X Y, Meng Q Y 2004 *Acta Physica Sinica* **53**(2) 388-95 *(in Chinese)*
[6] Wang X Y, Liu W, Yu X J 2007 *Modern Physics Letters B* **21**(20) 1321-41
[7] Hou J G, Wu Z Q 1989 *Physical Review* B **40**(2) 1008-12
[8] Xiao R F, Alexander J I D, Franz R 1988 *Physical Review A* **38**(5) 2447-56
[9] Wu J, Liu B G, Zhang Z Y, Wang E G 2000 *Physical Review B* **61**(19) 13212-22
[10] Krapivsky P L, Ben N E 2000 *Journal of Physics A: Mathematical and General* **33**(31) 5465-77
[11] Hassan M K, Hassan M Z, Islam N 2013 *Physical Review E* **88**(4) 042137
[12] Liu X, Wang M, Li D, Strom C, Bennema P, Ming N 2000 *Journal of Crystal Growth* **208**(1-4) 687-95
[13] Li D, Wang Y T, Ou-Yang Z C 2012 *Communications in Theoretical Physics* **58**(6) 895-901
[14] Vandewalle N, Ausloos M 1995 *Physical Review E* 51(1) 597-603
[15] Indiveri G, Scalas E, Levi A C, Gliozzi A1999 *Physica A: Statistical Mechanics and its Applications* **273**(3) 217-30
[16] Xu X J, Wei G 2006 *China Phys Lett* **55**(8) 4039-45 *(in Chinese)*
[17] Kalinin A P, Yu D 2000 *Thermophysical properties of materials* **38**(6) 882-885
[18] Degani Y, Heller A 1987 *The Journal of Chemical Physics* **91**(6) 1285-9
[19] Puertas A M, Fernández-Barbero A, de las Nieves F J, Rull L F 2004 *Langmuir* **20**(22) 9861-7
[20] Tomohiko K, Takahiro S 2000 *Physica B: Condensed Matter* **347** 284-8
[21] Wu Y Q, Xu X J 2010 *Journal of Computational Physics* **27**(4) 608-12
[22] Wang L 2003 *Shop scheduling with genetic algorithms* Beijing: Tsinghua University Press *(in Chinese)*

Authors

Wei Qiao, born in October, 1978, Gaomi, China

Current position, grades: senior engineer at Shandong University.
University studies: master's degree in Control Theory and Control Engineering at Shandong University of China in 2007.
Scientific interests: application of fractal theory, fractal control and Chaos of nonlinear system, kinetic growth models, simulation and fractal growth control.
Publications: more than10 papers.

Jie Sun, born in October, 1978, Daqing, China

Current position, grades: lecturer at Shandong University.
University studies: PhD degree in Control Theory and Control Engineering at Shandong University of China in 2013.
Scientific interests: fractal theory, fractal control, parameter identification, chaos of nonlinear system and parameter identification.
Publications: more than 10 papers.

Qingfu Du, born in June, 1964, Junan, China

Current position, grades: associate professor at Shandong University.
University studies: master's degree in Software engineering at Harbin Institute of Technology of China in2007.
Scientific interests: fractal control, fractal theory, chaos, automatic detection technology.
Publications: more than10 papers.

Numerical simulation of multiphase flow inside hydrocyclone based on CFD

Yuekan Zhang, Peikun Liu*, Linjing Xiao, Xinghua Yang, Junru Yang

College of Mechanical & Electronic Engineering Shandong University of Science and Technology, 266590, Qingdao China

Received 1 October 2014, www.cmnt.lv

Abstract

This paper applied computational fluid dynamics (CFD) method to investigate the internal multiphase flow filed in a 75 mm hydrocyclone. The Reynolds stress model (RSM) and VOF model were employed in the numerical simulation. This study discussed the velocity and pressure distribution in the hydrocyclone, and analysed the formation and development mechanism of air core. The numerical simulation results showed that the flow field was very unstable in the region of the air core. The axial velocity gradient reached its maxima, and the turbulent fluctuation was strongest in the simulation region. This study provided theoretical basis on further research of the air core effect on separation efficiency and pressure drop.

Keywords: numerical simulation, multiphase flow, hydrocyclone, CFD

1 Introduction

Hydrocyclone is a typical device that is used to separate non-uniform multiphase fluids. Owing to its advantages of small space requirement, simple structure, high separation efficiency, large separation capacity, low maintenance costs, and no moving parts, it has been widely used in various industries such as chemical engineering, oil/gas, and ore dressing. Although the structure of hydrocyclone is simple, the inner flow field is very complex and the flow characteristics are complicated. High speed rotating turbulent flow, interaction between multiphase flows and interruption of the air core increase the challenge to the study of internal flow field in hydrocyclone [1-3]. For many years researches on hydrocyclone were typically carried out in experimental ways. For example, Kesall [4] applied light-speed microscopic measurement system, Knowles et al. [5] used imaging velocimetry system, and Hsieh et al. [6] employed a LDV to study the internal velocity distribution of hydrocyclone, in which small amount of tracing particles were added to the water. However, these methods require delicate experimental conditions, high cost, long working period, and have disadvantage of difficult measurement and non-universality. With the development of computation fluid dynamics (CFD) and continuous improvement of turbulent models, numerical simulation becomes research hot points in hydrocyclone studies. Through simulation, the internal flow field and velocity distribution can be clearly detailed, which provides important guide to structural design and optimization of hydrocyclones. This paper applied a CFD software, FLUENT, to simulate the internal two-phase flow field in a hydrocyclone. This study correctly reflected the internal flow field in the hydrocyclone, obtained the velocity and pressure distribution, and especially focused on the formation and development mechanism of the air core.

2 Construction of the model

Structural model This paper aimed at the standard hydrocyclone that was proposed by Hsieh, [6] where the structural dimension is shown in Figure 1.

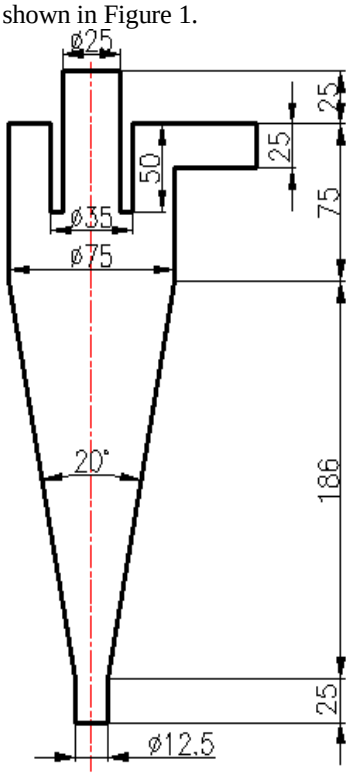


FIGURE 1 Geometry of the hydrocyclone

The diameter of cylindrical section Φ is 75mm, and the length is 75mm. The cone section is 186mm long, and the length of the underflow section is 25mm. The inner depth of the overflow section is 50mm. The overflow tube thickness

**Corresponding author* e-mail: lpk710128@163.com

is 5mm and the diameter of the overflow and underflow tubes is 25mm and 12.5mm, respectively. The inlet section area is 25×12.5mm, and the cone angle is 20°. The mesh was generated through ICEM with structural hexahedral mesh. The mesh in the central axis and wall area was refined, which is shown in Figure 2.

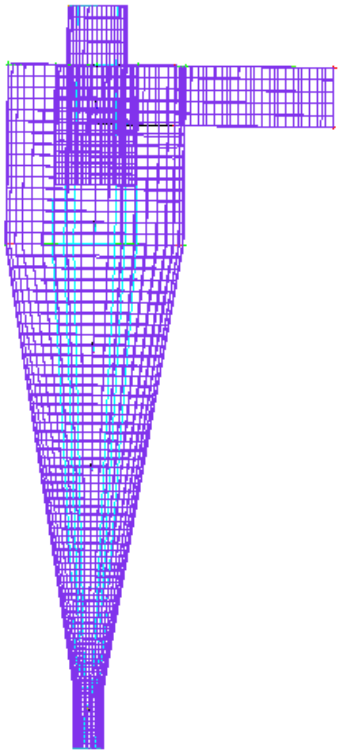


FIGURE 2 The mesh of the hydrocyclone

Determination of the turbulent and multiphase flow models Due to the complexity of turbulent flow, its generation mechanism has not been fully understood. Bhaskar et al. [7] used experimental studies to compare k-ε turbulent model, k-εRNG turbulent model and RSM model on the numerical simulation of hydrocyclone flow. The results indicated that the RSM model could provide the smallest error compared to the experimental data, because RSM model completely considers the rapid change of streamline curve, rotation and tension, which can well predict anisotropic turbulence. Therefore, it is very suitable for the study in hydrocyclone flow where the flow is highly rotational. Considering the flow in the hydrocyclone is two phase or multiphase flow, this paper applied volume of fluid (VOF) model to study liquid-air two phase flow. This study solved a set of momentum equation and analysed the VOF of two phase flow through the computational domain, which can well simulate the un-dissolved fluid distribution of air and water in the hydrocyclone.

Boundary condition and controlling parameters Boundary condition is the key to determine the exact solution of the governing equation. The governing equation and boundary condition constitute the complete mathematical description of the flow process of a physical variable. FLUENT provides various boundary conditions. In this study, the boundary condition was set as follows:

Inlet condition: the primary phase was set to be water. The temperature is constant, density of water is 998.2kg/m^3, viscosity is 0.001Pa·s, volume fraction is 1. The secondary phase was set to be air, the temperature is also constant, density is 1.225kg/m3, viscosity is 0.00001 Pa·s and volume fraction is 0. The velocity of water at inlet was 2.28m/s and the turbulent characteristics were determined based on the combination effect of hydraulic diameter and turbulent level.

Pressure outlet: Both overflow and the underflow exits were set exposed in the air, where the relative pressure was zero. The backflow coefficient of air was set to be 1. In this regard, the backflow of fluid would be air, if negative pressure presented at the flow outlet.

Wall condition: The wall condition was treated as non-slip boundary condition, where the boundary layer turbulence was determined according to the standard wall function.

3 Simulation results and discussion

The residual was set to be 10^{-5} and the time step was 1×10^{-4}s. The velocity field, pressure field, and the formation and development of the air core were determined through iteration. The results and discussion are presented as follows.

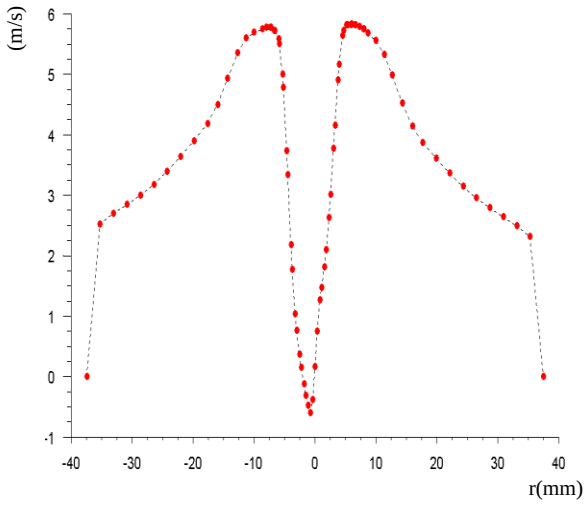


FIGURE 3 The tangential velocity distribution

Tangential velocity distribution in the hydrocyclone, tangential velocity has great influence on the separation efficiency. Large tangential velocity has great centrifugal force, and therefore high separation efficiency. From the tangential velocity distribution (Figure 3), tangential velocity varied clearly along the axis, where the minimum tangential velocity was at the wall. With the gradual decrease of radius, the value of tangential velocity increased continuously. After reaching the maximum value at the air core edge, the tangential velocity tended to decrease gradually. The distribution of the tangential velocity had the characteristics of combined vortex. An obvious transition area existed between forced vortex and free vortex, which agreed well with the research results of Kelsall [4] in 1952.

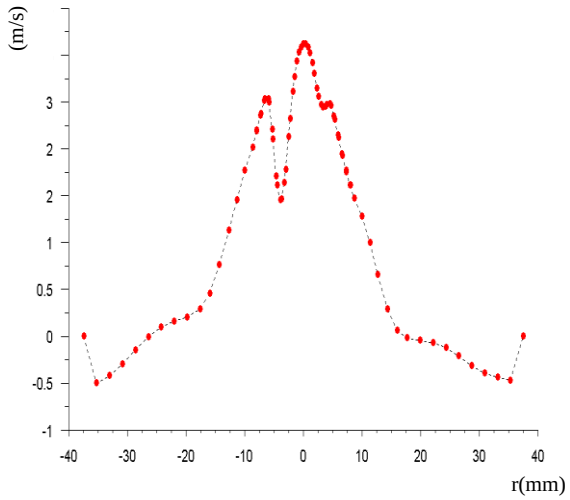


FIGURE 4 The distribution of axial velocity

Axial velocity distribution: during the separation process of the hydrocyclone, axial velocity directly affects the retention time of the fluid in the flow field, and determines the distribution relationship of fluid between overflow and underflow. Hsieh [6], Kelsall [4], and Bradley et al [2] conducted detailed experimental studies of axial velocity in hydrocyclone. Using experimental data, Xu [8] applied mathematical regression method to fit the mathematical equations for axial velocity distribution of fluids in the area under the overflow tube of hydrocyclone. The distribution of axial velocity in this study is shown in Figure 4. The direction of flow in the tube was downward close to the wall, towards the underflow exit. Towards the axial direction in the tube, the axial velocity decreased with the decrease of radius. The axial velocity was zero at the location of approximately half radius. After this zero point, the axial velocity changed to upward direction, i.e., the flow was towards the overflow exit, and increased with the decrease of radius. There was always a point with zero axial velocity at different cross-sections. Therefore, a contour surface shaped like a cone can be named locus of zero vertical velocity (LZVV), shown in Figure 5. This LZVV divided the axial velocity of fluids in the hydrocyclone into two parts: The flow direction of fluids outside of the zero axial velocity envelope plane was downward, forming outer helical flow; In contrast, the flow direction of fluids inside of the zero axial velocity envelope plane was upward, forming internal helical flow. Figure 6 gives axial velocity gradient of hydrocyclone flow. It can be seen that the gradient close to the wall was smallest. With the decrease of radius, in the area of air core, the gradient increased rapidly until reaching a certain value and then decreased fast, forming a hump-like distribution. It indicates that turbulence in the centre area of the hydrocyclone was very intense, and the flow field was extremely unstable. However, the turbulent level was weakest in the area of hydrocyclone wall, where the flow field was relatively mild.

FIGURE 5 The locus of zero vertical velocity

Pressure distribution: The pressure distribution in hydrocyclone is very important to separation efficiency and separation size, which is the primary basis for calculation of productivity and energy loss. Through the two-phase numerical simulation, the pressure distribution in the hydrocyclone is obtained and shown in Figure 7. It can be seen that the static pressure in the hydrocyclone gradually decreased from the wall towards the axis. The static pressure of the fluids was zero at the location of the air core interface. Afterwards, the negative pressure became stronger as closer to the axis, and decreased to the minimum value at the underflow and overflow exits.

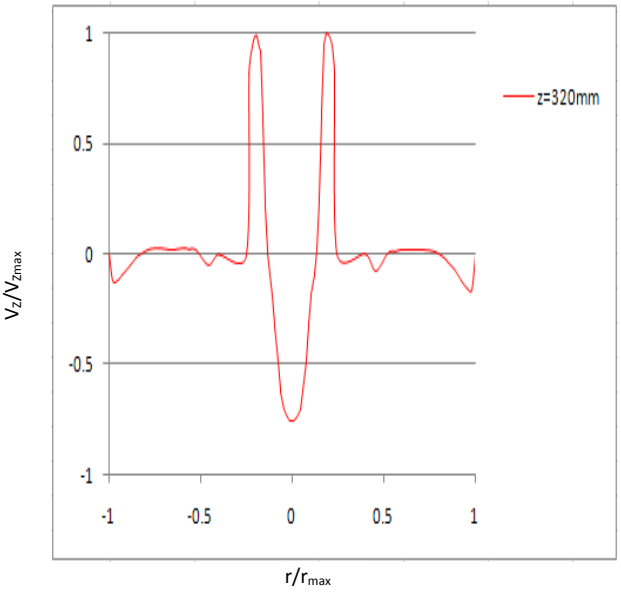


FIGURE 6 The axial velocity gradient of the hydrocyclone

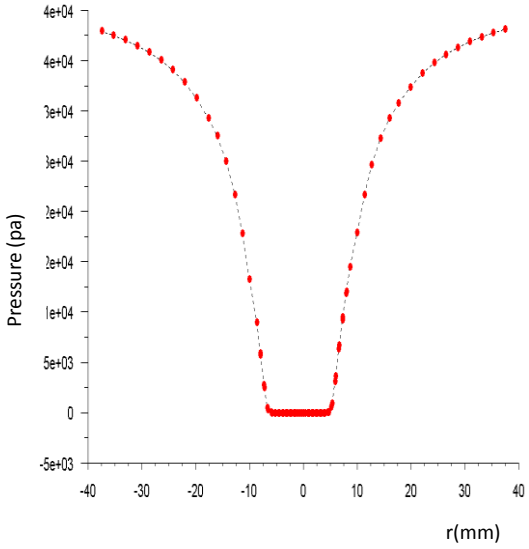


FIGURE 7 The pressure distribution of the hydrocyclone

Formation and development of the air core Figure 8 shows the formation and development of the air core in the hydrocyclone. In the figure, blue color represents air, whereas red part is water. At the initial time, the hydrocyclone is filled with air. With the evolution of the flow field, the volume of air in the hydrocyclone decreased gradually, and the ratio of water phase volume to total volume increased. When the time is about 1.1s, water occupied the whole internal area of the hydrocyclone. With the increase of time, air core initially formed under the overflow tube, and then in the area of underflow exit. The two air core grew and moved towards each other until they merged to a single air core. It is evident from this numerical simulation that, because the underflow and overflow tubes were directly connected to air, the fluids entering into the hydrocyclone rotated strongly as spiral vortices. When the tangential velocity reached a certain value, the negative axial pressure formed, and air entered into hydrocyclone. Under the effect of this negative pressure, the air core developed and reached the stable condition. The size and shape of the air core changed with the development of the flow field. Even when the flow condition reached stable, the sizes of the air core at different axial locations were not the same. The axial change of diameter of the air core is shown in Figure 9. The maximum diameter of the air core located at the bottom of the overflow tube, which is 15.57 mm. The diameter of the air core decreased to 9.91 mm with the decreasing height of the hydrocyclone, which was induced by structural change of bottom section of the overflow tube. In the cone section, the diameter of the air core varied irregularly at different locations in the hydrocyclone. The characteristics of the air core in the underflow tube were similar to that in overflow tube. The diameter at bottom of the underflow tube is maximum, i.e., 9.8 mm. The diameter changed suddenly to 8.91 mm at the interface between underflow tube and the cone section.

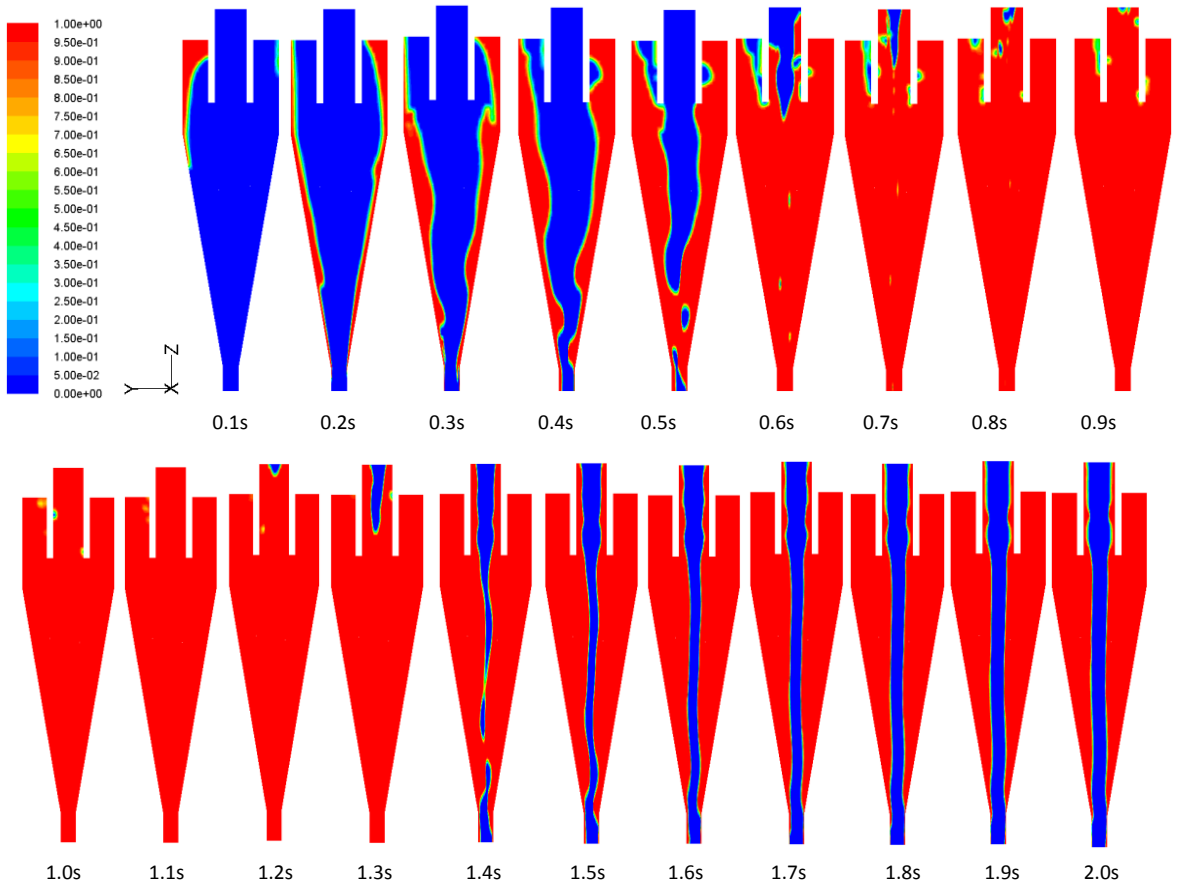


FIGURE 8 The development of the air core in the hydrocyclone obtained from CFD simulations

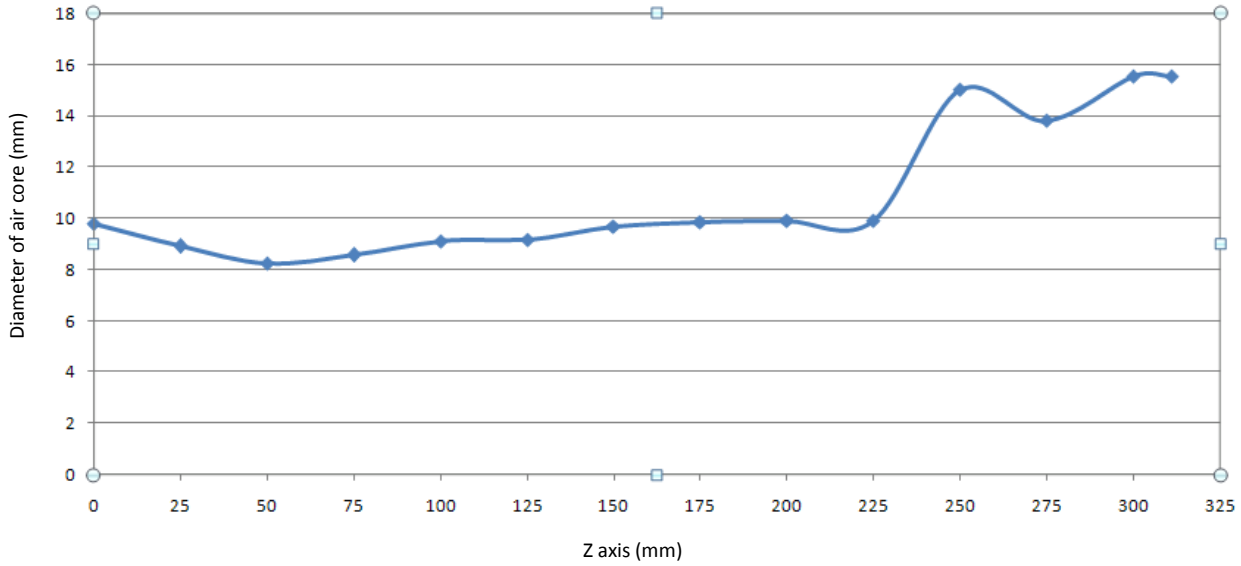


FIGURE 9 The axial change of diameter of the air core in the hydrocyclone

4 Conclusions

This paper used FLUENT and applied RSM and VOF models to simulate the two phase flow field in a hydrocyclone. The results of this study were consistent with the previous theoretical and experimental studies. The results explored the internal flow field in the hydrocyclone. This study emphasized the formation and development of the air core. The numerical results showed that the flow field was highly unstable in the area of the air core, with the largest axial velocity gradient and strongest turbulent fluctuation. This study also provides the theory basis of further investigation of air core effect on the separation efficiency and pressure drop.

Acknowledges

This work is supported by the National Natural Science Foundation of China (No.21276145) and guiding program for scientific and technological research of China National Coal Association (No. MTKJ2013-365).

References

[1] Wills B A, Napier-Munn T 2006 Will's Mineral Processing Technology *7th Oxford Butterworth-Heinemann*
[2] Bradley D 1965 The Hydrocyclone *London: Pergamon Press*
[3] Nowakowski A F, Cullivan J C, Williams R A, Dyakowski T 2004 Application of CFD to modelling of the flow in hydrocyclones. Is this a realizable option or still a research challenge? *Minerals Engineering* **17**(5) 661-9
[4] Kesall D F 1952 A study of the motion of solid particles in a hydraulic cyclone *Transactions of the Institute of Chemica Engineers* **30** 87-104
[5] Knowles S R, Woods D R, Feuerstein I A 1973 The velocity distribution within a hydrocyclone operating without an air core *The Canadian Journal of Chemical Engineering* 1973 **51**(3) 263-71
[6] Hsieh K T 1988 A phenomenological model of the hydrocyclone *Thesis Salt Lake City* University of Utah
[7] Bhaskar K U, Murthy Y R, Raju M R, Tiwari S, Srivastava J K, Ramakrishnan N 2007 CFD simulation and experimental validation studies on hydrocyclone *Minerals Engineering* **20**(1) 60-71
[8] Xu J, Luo Q 1988 Flow fields theory of the hydrocyclone *Beijing China Science Press (in Chinese)*

Authors

Yuekan Zhang, born in March, 1970, Xinji, Hebei Province, China

Current position, grades: instructor at the College of Mechanical & Electronic Engineering Shandong University of Science and Technology. PhD degree student in mineral process engineering at the Shandong University of Science and Technology, China.
University studies: MS degree in mechanical design and theory at Shandong University of Science and Technology, China in 2007.
Scientific interests: computational fluid dynamics, solid-liquid separation technology and equipment.
Publications: 25 papers.

Peikan Liu, born in January, 1971, Wenshang, Shandong Province, China

Current position, grades: professor at Shandong University of Science and Technology.
University studies: Doctor's degree in mineral processing at Wuhan University of Science and Technology China in 2008.
Scientific interests: computational fluid dynamics, solid-liquid separation technology and equipment.
Publications: 30 papers.

Linjing Xiao, born in January, 1966, Yishui, Shandong Province, China

Current position, grades: professor at the College of Mechanical & Electronic Engineering at Shandong University of Science and Technology.
University studies: PhD degree at Beijing University of Science and Technology, China, in 2001.
Scientific interests: virtual prototype and CAD/CAE/CAE.
Publications: 55 papers.

Xinghua Yang, born in October, 1978, Luoshan, Henan Province, China

Current position, grades: instructor at the College of Mechanical & Electronic Engineering Shandong University of Science and Technology.
University studies: MS degree at Shandong University, China, in 2001.
Scientific interests: computational fluid dynamics, solid-liquid separation technology and equipment.
Publications: 28 papers.

Junru Yang, born in October, 1969, Leting, Hebei Province, China

Current position, grades: associate professor at the College of Mechanical & Electronic Engineering Shandong University of Science and Technology.
University studies: PhD degree at Shandong University, China, in 2006.
Scientific interests: cladding part design and CAD/CAE
Publications: 40 papers.

The complexity analysis and performance comparison of MIMO systems based on antenna selection techniques

Yisong Lin[1], Mingjie Zhuang[2]*

[1]*School of Information Science and Engineering, Huaqiao University, Xiamen 361021, China*

[2]*College of engineering, MJZhuang176@163.com, Huaqiao University, Quanzhou, 362021, China*

Received 1 March 2014, www.cmnt.lv

Abstract

MIMO systems, also known as multiple-input multiple-output, can effectively improve the capacity and reliability of wireless communication. Antenna selection (AS) is a low-cost low-complexity attractive approach in MIMO systems that capture many advantages of these systems. In this paper, we comprehensively review and describe various kinds of AS schemes in MIMO systems. First, we discuss the techniques of antenna selection from the perspective of different channel environments. Analysis results show that exploiting the AS technique can avoid sending redundant information and improve the channel capacity in the low rank and correlated channel conditions. Then the applications of AS systems in spatial diversity and spatial multiplexing are considered. In order to reduce the complexity of AS algorithm, we propose a novel joint transmit and receive AS scheme in MIMO systems. The results of simulations demonstrate that the proposed AS method of performance outperforms other methods, and the proposed algorithm can significantly reduce computational complexity compared to the optimum algorithm. Finally, we summarize some conclusions about the antenna selection.

Keywords: MIMO, antenna selection, spatial diversity, spatial multiplexing

1 Introduction

Using multiple input, multiple output (MIMO) systems is one of the most significant technical breakthroughs in modern wireless digital communications. Compared with single inputs single output (SISO) systems, the capacity and reliability of a wireless communication system can be improved dramatically by employing multiple antennas at the transmitter and/or receiver without increasing bandwidth and transmit power [1]. MIMO technology has been drawn significant research interests recently due to its advantages. Most importantly, the standard for the third-generation cellular phones (3rd generation partnership project, 3GPP) has joined the MIMO technology. The MIMO technology has been widely applied to Beyond 3G and 4G systems. We can foresee that the mobile communication systems in the future, including the 5G system, will be implemented by massive MIMO technology [2].

It is shown that in an independent and identically distributed (i.i.d.) Rayleigh fading channel, the capacity of a MIMO system with N_T transmit antennas and N_R receive antennas scales almost linearly with the $min(N_T, N_R)$ in the high signal-to-noise ratio (SNR) regime [1]. The MIMO wireless communication systems have now demonstrated the potential for increased capacity and reliability in rich multipath environments, without any increase in bandwidth or transmit power. However, the application of multiple antennas has been restricted by the hardware cost and power consumption of the RF chains. Because it requires the same number of RF chains at the transmitter and receiver. The RF chains include amplifiers, up-down converters, as well as the analog-to-digital-to-analog (A/D/A) conversions. The cost and hardware complexity of the RF chains is often much higher than the antenna array. How to reduce the hardware complexity and at the same time maintain most of the advantages of MIMO systems is an important research topic. A MIMO system with antenna selection (AS) has been shown to significantly outperform a system exploiting the same number of RF chains without AS. Therefore, the AS approach that chooses a subset of the available transmit and/or receive antennas is an attractive low-cost and low-complexity technique. The MIMO technology improves the wireless communication system performance from two aspects, spatial multiplexing and spatial diversity. First, from the point of view of spatial diversity, MIMO systems can improve the signal-to-noise ratio (SNR) and the robustness of the communication links in terms of BER (Bit Error Rate) by collecting multipath signals between the transmitter and receiver. On the other hand, spatial multiplexing, which makes full use of independent space degrees of freedom between the transmitter and receiver can dramatically improve the capacity and the speed of data transmission of the communication system.

The remaining sections of this paper are organized as follows. We will introduce MIMO systems with AS techniques and discuss AS schemes in various kinds of operation environments of actual channel in Section II. AS systems based on both spatial diversity discussed and analysed In Section III. Section IV studies mainly the AS criterions based on spatial multiplexing, and analyses the complexity of various kinds algorithms. Section V describes the simulative results and discussion. Finally, the prospection and conclusions are included in Section VI.

***Corresponding author** e-mail: MJZhuang176@163.com

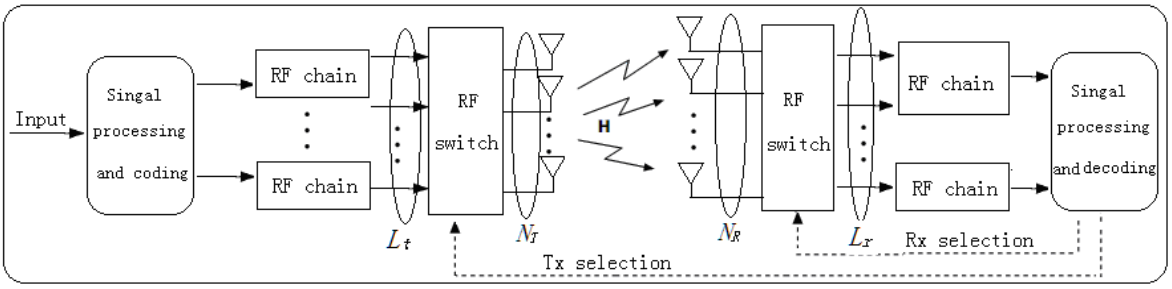


FIGURE 1 Antenna selection in MIMO system

2 Antennas selection of channel model

2.1 MIMO SYSTEM CHANNEL MODEL

Figure 1 shows the MIMO systems with N_T transmit antennas and N_R receives antennas. The transmitter includes an STBC encoder, RF chains, and a switch that connects the RF chains to select L_t out of N_T transmit antennas according to the message of *Tx* selection through feedback channel from the receiver. Likewise, at the receiver, a switch connects the RF chains to select L_r out of N_r receive antennas according to the message of Rx selection from the signal processing unit, in which during a period of the received signals process the channel state information (CSI) of MIMO systems is estimated. Assuming all N_T transmit antennas and N_R receive antennas are selected, the complex envelope of the received signal at the antenna array after matched filtering is given by:

$$y=Hx+n, \tag{1}$$

where x is a N_T dimensional column vector of the transmitted signals, and n is a N_R dimensional complex noise vector where each component is a complex white Gaussian noise sample with zero mean and variance σ_n^2. The channel matrix is defined in Equation (2)

$$H=\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1N_T} \\ h_{21} & h_{22} & \cdots & h_{2N_T} \\ \vdots & \vdots & \vdots & \vdots \\ h_{N_R1} & h_{N_R2} & \cdots & h_{N_RN_T} \end{pmatrix}, \tag{2}$$

where matrix H is a $N_R \times N_T$. It is assumed that the matrix H is independent complex Gaussian random variables with zero mean and unit variance, namely, all of the transmitted signals experience frequency selective flat Rayleigh fading channel. The element h_{ij}, i=1,2,…N_R; j=1,2,…N_T, of the matrix H represents the complex channel gain between the j-th transmit antenna and the i-th receive antenna.

2.2 THE EFFECT OF CHANNEL SITUATIONS ON AS

The capacity of wireless communication system has been severely influenced by the channel situations [3]. A lot of AS schemes on MIMO systems have been discussed in the different channel conditions [4-11]. Under the flat Rayleigh fading channels, assuming that MIMO wireless channels are mutually independent, the references [4-8] presented some AS schemes. These results indicated that AS technique can capture most of the advantages of MIMO systems and at the same time reduce the hardware complexity. If the wireless channels are independent, we can be sure that the performance of the system will be deteriorated in the correlated channels. The fact that the channel is correlated implies that some rows/columns in the channel matrix can be expressed as a linear combination of the others. This means that the information in these rows/columns is in some way redundant and does not contribute to capacity. It is thus clear that redundant information is easily produced in the correlated channel situation. To do this, a fast AS algorithm based on Frobenius norm and correlation coefficient was proposed in [7]. This algorithm chooses two rows of the channel matrix, which carry identical information, since they own similar signal components, one of them will be deleted in this scheme. In addition, the two rows have different powers, or rather squared Frobenius norm of the row, the corresponding row with higher power will be selected in the algorithm. When there are no two rows of identical information, the algorithm in [9] choose two rows of the channel matrix, which have higher correlation factor, and the corresponding row with lower power will be deleted. In a word, it allows us to obtain a sub-channel matrix whose rows are minimally correlated and have maximum powers. Further, in a low rank and correlation matrix channel, it is easy to send redundant information, which does not contribute to the channel capacity. It is necessary to select a set of antenna in a low rank channel conditions, so that the total transmission power may be allocated to a small amount of transmit antennas to improve the capacity of the system. An optimal AS algorithm in the low rank matrix channel was proposed in [10]. It has shown that AS in the low rank matrix channel can still improve the capacity of the channel. And in the seriously frequency selective fading channel, it has been shown that the effect on AS is very little. However, when the bandwidth of the system is greater than the correlated bandwidth of channel, an additional degree of diversity which is generated by AS is not prominent. Even so, the advantages of antenna selection are still reflected in a general frequency selective fading channels. According to reports in the [11], it shows that CDMA signals, in a general frequency selective channels, can be exploited to maximize the SNR at the receiver based on both generalized combined, in which combined multiple copy signals, and pre-coding technique as well, at the same time it can reduce the complexity of the transmitter and receiver.

3 AS systems based on spatial diversity

MIMO technique can greatly improve the performance of the wireless communication systems by exploiting the independence of the channel space. Namely, it takes advantage of the degrees of freedom provided by multiple antennas and the technique of time/spatial diversity to improve the performance and reliability of wireless communication. Diversity refers to two or more random fading signals are merged to reduce the fluctuation of signal level. As Figure 2 shows a block diagram of conventional receive diversity. All of the transmit signals experience a different fading channel with complex valued coefficient $h_i(t)$ and additional noise signal $n_i(t)$. Due to multiple fading paths are statistically independent, it is unlikely that they appear synchronous fading phenomenon. At any moment, the probability of all uncorrelated fading signals simultaneously in a deep fading is small. Therefore, diversity can significantly reduce fading degree of the combined signals. At present, diversity has been widely exploited in wireless communication systems to against the multipath fading.

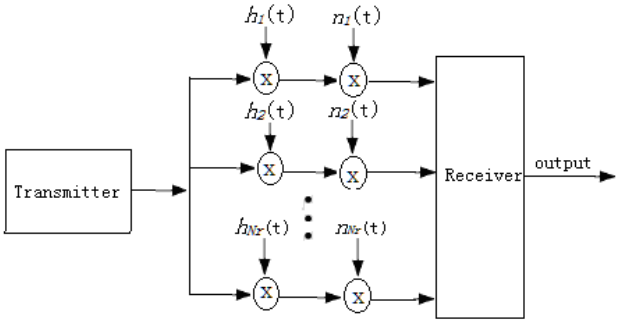


FIGURE 2 Block diagram of receive diversity

3.1 THE TRADITIONAL DIVERSITY

Figure 3 shows a generalized receive antenna selection diversity system, which is an extension of the conventional receive diversity. The independent signals after the antenna selection must be combined at the receiver to improve the signal-to-noise ratio (SNR) of the system. Diversity combining can be classified into three categories by the different ways of signals combination. They are maximal ratio combining (MRC), equal gain combining (EGC) and selection combining (SC). Although the implement of MRC would be relatively complicated, it is still an optimal technique of diversity combining which gives a better performance of the system. With the development of semiconductor and signal processing techniques, MRC technique now has been widely exploited in communication systems. A rapid incremental AS approach, in which at each step a new antenna is added to the selected antenna subset based on the maximum output SNR strategy, for MIMO diversity systems using maximum ratio transmit (MRT) at the transmitter and MRC at the receiver was proposed by Duan Hong, and Liu Feng in [12]. It shows that the MIMO AS systems can keep most of the diversity gain, and do not need too many RF chains. In addition, the performance analysis of a kind of reduced-complexity transmit/receive diversity systems and SNR and bit error rate (BER) of hybrid SC/MRC systems are discussed respectively in [13, 14], study results show that it shows that the system performance is obviously improved.

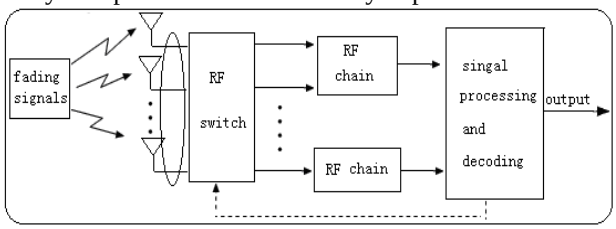


FIGURE 3 Block diagram of generalized receive AS diversity system

3.2 THE SPACE-TIME DIVERSITY

As mentioned above, through hybrid the AS technology and traditional MIMO diversity, we can greatly boost the diversity gain, thus output SNR and BER performance of the communication systems are obviously increased and improved. However, in the actual operating environments, it may bring the mutual interference of transmitting antennas by exploiting the multi-antenna. In order to reduce mutual interference of transmitting antennas, an optimal AS with space-time block coding was proposed in [15]. Research shows that MIMO AS systems with the space-time coding can obtain certain additional diversity advantages. The input serial data from signal constellation are performed series-to-parallel to multiplex parallel data streams, and then the processed data employing space-time coding are transmitted in sequence over the selected antennas. So transmit diversity advantage will be produced by space-time two dimensional coding in the transmitter. It shows that transmit antenna selection with space-time block coding (STBC) can greatly improve the performance of BER in communication systems [16]. At the same time, [17] shows that antenna selection combined with quasi-orthogonal STBC can improve the diversity gain and coding gain to reduce BER. The investigation results in [18] indicate that quasi-orthogonal STBC systems with AS can increase the capacity of the channel and improve the transmission rate of the system. [19] points out that the diversity gain and closed loop gain of the system can improved greatly by combining V-BLAST and antenna selection algorithm.

Finally, we can conclude that MIMO systems combining with space-time code will be a key technology in the future high-speed transmission communications. Employing MIMO AS systems not only can reduce tremendously the system implementation complexity and cost, but also provide a higher diversity gain and enhance the SNR magnitude. The following section, we will verify MIMO AS systems can increase effectively the channel capacity and improve the rate of transmission as well.

4 AS systems based on spatial multiplexing

As we known, MIMO can greatly improve the channel capacity and the rate of data transmission of the system by exploiting spatial multiplexing. With the increase of number of transmit and receive antennas, the channel capacity of

MIMO systems will increase linearly. As shown in the Figure 1, the MIMO system equips with N_T transmit antennas and N_R received antennas. Assume that the MIMO wireless channel is modelled as a Rayleigh fading distribution. Based on spatial multiplexing, the AS criteria of MIMO system is to select an antenna subset, in which subset can achieve the maximum channel capacity. On the basis of the channel matrix H, L_t out of N_T transmit antennas and L_r out of N_R receive antennas are selected from MIMO system, let the selected sub-matrix be denoted by $\tilde{H}$. Assuming that the system transmission power is evenly allocated to each of the transmitting antennas, thus the channel capacity of MIMO system can be expressed as [1]:

$$C_{\text{full}} = \log_2[\det(I_{N_T} + \frac{\rho}{N_T} H^H H)], \quad (3)$$

where ρ denotes the average signal-to-noise ratio (SNR), H is the $N_R \times N_T$ channel matrix. I_N denotes a $N \times N$ identity matrix. The superscript H denotes the conjugate transpose.

The channel capacity of sub-matrix $\tilde{H}$ corresponding to the optimal antenna subset can be written as:

$$C_{\text{sel}} = \underset{S(\tilde{H})}{max}\{\ \log_2[\det(I_{N_T} + \frac{\rho}{N_T} \tilde{H}^H \tilde{H})]\}. \quad (4)$$

However, the computational load required for an optimal selection through an exhaustive search over all possible antenna subsets grows exponentially with the total number of the antennas available. This is a computationally prohibitive problem. High-speed communications systems demand an efficient AS scheme with lower complexity, so the investigation of the sub-optimal AS algorithms is of great practical as well as theoretical interest.

4.1 RECEIVE ANTENNAS SELECTION

The Receive antennas selection (RAS) system based on spatial multiplexing is choosing the channel matrix to maximize the channel capacity in the receiver. The simplest RAS algorithm is norm-based selection (NBS), which was proposed in [4]. Selection criteria of NBS is that the row of channel matrix with the maximal Frobenius norm is chosen in each step selection algorithm. Due to lower computational complexity, NBS is the most suitable for low SNR or only a RF link at the receiver. However, NBS is not necessarily optimal in other scenarios. The lower complexity of NBS, the more losses of channel capacity, in order to reduce the losses of channel capacity, a fast antenna selection is proposed in [5]. The algorithm begins with the full channel matrix, and then removes one antenna per step. So the AS criteria is based on the antenna with the lowest contribution to the system capacity is removed per step. In contrast to [5], an incremental antenna selection algorithm is proposed [6,7]. The computational complexity is much lower than the algorithm of [5]. The algorithm starts with an empty antenna subset and adds one antenna per step to this subset. At each step, the objective is to select one antenna, which leads to the highest contribution to the channel capacity. Although the computational complexity of the algorithm [6, 7] will be lower than AS algorithm [5], there are still some losses of channel capacity. These algorithms [5-7] can substantially reduce the computational load in the AS systems, however, they are still restricted by so called local searching shortcomings. These local searching strategy, though reduce the computational load, give rise to the loss in optimality simultaneously. A kind of global and fast receiver antenna selection for MIMO systems was proposed [8], it takes into account the number of RF chains is the same both at the transmitter and receiver. It allows to finding the sub-matrix with the greatest determinant for a high SNR scenario. It shows that the fast AS algorithm can achieve almost the same capacity performance as the optimal selection method.

In order to illustrate the performance of the antenna selection, the channel capacities of different algorithms are simulated with 1000 random number seed by Monte Carlo method. Figure 4 shows that comparison of the channel capacity the incremental algorithm [6, 7], decreasing algorithm [5], global and fast antenna selection method [8] for $L_r = N_T = 3$, $N_R = 6$, and the value range of SNR is 0*dB* to 20*dB*. It can be seen that, obtain almost the same capacity performance as the optimal selection method. While the NBS [4] has a certain losses of channel capacity compared to the optimal algorithm. Table 1 shows the complexity of antenna selection algorithms. It shows that the complexity of NBS selection algorithm is smallest. While the complexity of global and fast antenna selection method [8], the incremental algorithm [6, 7] are smaller than decreasing algorithm [5].

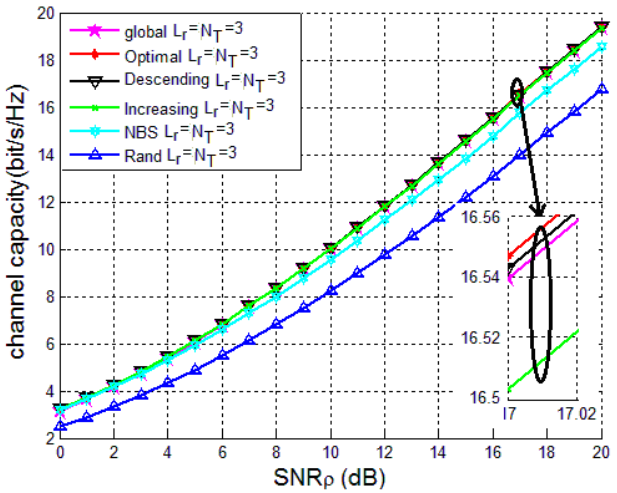


FIGURE 4 The channel capacity versus SNR

TABLE 1 Comparison of computational complexity for AS algorithms

Descending [5]	Increasing [6]	Global [8]	NBS [4]
$N_R^2 N_T^2$	$max\{N_T,N_R\}N_T L_r$	$N_R N_T^2$	$N_R N_T$

4.2 TRANSMIT ANTENNAS SELECTION

In the same way, transmit antenna selection (TAS) performs to select the optimal transmit antenna subset at the MIMO transmitter employing full or partial channel state information (CSI). However, TAS implementation is more difficult

than RAS. This is because TAS MIMO systems need a dedicated feedback channel to feed the CSI from the channel estimator unit at the MIMO receiver. In order to ensure performance of TAS system, known perfect CSI at the transmitter is very important. In practical scenario, the wireless channel is random variation, achieving perfect CSI is very complicated. Therefore, it is inevitable to receive a small fraction of the error CSI. Effect of imperfect CSI on the channel capacity of the MIMO systems with TAS is obvious. Like the RAS systems, compared to without TAS, the capacity performance of MIMO systems with TAS will be greatly improved [20].

4.3 JOINT TRANSMIT AND RECEIVE ANTENNA SELECTION

Figure 1 shows the MIMO system equips with N_T transmit antennas and N_R received antennas. There are L_t and L_r RF chains at the receiver and transmitter, respectively. Assuming all N_T transmit antennas and N_R receives antennas are selected. The object is to select the best L_r out of N_R antennas at the receive side and the best L_t out of N_T antennas at the transmit side so that the resulting system capacity is maximized. The channel matrix H is an $N_R \times N_T$ complex valued matrix. The ideal antenna selection technique chooses $\tilde{H}$ (is an $L_r \times L_t$) out of H. The optimal subset selection which can always be obtained by fully searching overall the possible row and column combinations of the channel matrix H. The performance of systems that using the optimal subset selection has much higher than that of the systems using the same number of antennas without any selection. However, the complexity of the optimal algorithm is $C_{N_R}^{Lr} \cdot C_{N_T}^{Lt}$, which may not be feasible. The joint transmit and receive antenna selection algorithm [21] was first proposed to decouple the joint antenna selection problem into separate transmit and receive antenna selection and carried out the exhaustive search. So the algorithm requires computing $C_{N_R}^{Lr} + C_{N_T}^{Lt}$ determinants of size $L_r \times L_t$ sub-matrices for the exhaustive search. In order to select the appropriate subset of antennas, antenna subset selection problems have been intensively studied in the [22-29], a similar decoupling strategy has also been used in the separable transmit/receive successive selection [22]. The selection algorithm adopted the fast incremental successive selection algorithm [7] for both transmit and receive side selection in the decoupled problem. The complexity of this algorithm is $max\{N_T, N_R\}^2 L^2$. It results in the huge complexity reduction compared to the algorithm [21]. But, it has a loss of performance compared to the optimal selection.

Although the decoupling-based strategies can reduce the computational burden of the problem of joint transmit and receive antenna selection, the complexity of this decoupling-based strategies is still large. To reduce the computational complexity, the successive joint transmit and receive antenna selection algorithm is proposed in [23-26]. The algorithms start with the empty set of selected antennas and add one pair of antenna per step to this set. At each step, the objective is to select one pair of antenna, which leads to the highest contribution to the system capacity. The joint transmit and receive selection strategies by exploiting stochastic optimization algorithms such as the genetic algorithm [27, 28] or particle swarm optimization [29] have also been investigated to find the subset of the antenna to achieve maximum instantaneous channel capacity. It is assumed that $N_T = N_R = N$. Then, the complexity of CJAS [25] is $N^2L^3/2$. It is large when the number of the selected antennas is more than $N/2$. A novel joint transmit and receive antenna selection algorithms is proposed. It is calculating the capacity reduce instead of the whole subsystem capacity when selecting a new candidate pair of antenna at each step. In each step, our objective is to select one pair of antenna, which leads to the lowest contribution of the capacity. The complexity of the proposed algorithm is $3\times N^4/4 - 3\times (N-L)^4/4$. Table 2 shows the complexity of three antenna selection algorithms, we can find that the computational complexity of the proposed antenna selection algorithm is far less than the optimal antenna selection algorithm. When the number of selected antennas is more than $N/2$, the complexity of the proposed algorithm is less than CJAS [25]. So it is suitable for more number of selected antennas.

TABLE 2 The complexity comparison for transmit and receive antenna selection algorithms

algorithm	**complexity**	***N*=8, *L*=2**	***N*=8, *L*=3**	***N*=8, *L*=4**	***N*=8, *L*=5**	***N*=8, *L*=6**
optimal algorithm	$C_N^L C_N^L L^3$	6272	84672	313600	392000	169344
proposed algorithm	$\frac{3}{4}N^4 - \frac{3}{4}(N-L)^4$	2100	2603	2880	3011	3060
CJAS	$\boldsymbol{N^2L^3/2}$	256	864	2048	4000	6912

5 Results and discussion

In this section, we consider the MIMO AS systems channel model as mentioned before, the elements of H are represented by independent complex Gaussian random variables with zero mean and unit variance. The simulation results will be vastly presented to compare the proposed algorithm with other joint transmit/receive AS algorithms, such as optimum selection (OS), NBS, and CJAS [25]. In order to illustrate more clearly the performance of the proposed algorithms and other AS approaches, we will plot a large number of graphs to describe the channel capacity and the cumulative distribution function (CDF), they are the curves of channel capacity as a function of SNR and different selected antenna numbers, and CDF as a function of channel capacity and SNR. These curves are plotted by computer Monte

Carlo simulation using a certain random number data (e.g., 1000, or 2000).

Figure 5 shows the comparison results of five algorithms with L_r=L_t =2, N_T=N_R=5, and the value range of SNR is 0*dB* to 20*dB*. In order to get these curves, in the two figures at each value of SNR, we average statistically to 1000 simulating values. It can be observed that the performance of our proposed algorithm is better than CJAS [25]. The figure also indicates that the proposed algorithm has almost the same performance with optimum AS. The reason is that our proposed algorithm removes the antenna serial number, which leads to the smallest contribution of channel capacity and yields a negligible lossed to channel capacity, compared to the optimal selection. Figures 5 and 6 show that the more number of selected antennas at the transmitter and receiver, the greater channel capacity of MIMO systems. This is because MIMO wireless systems have the potential for increased capacity in rich multipath environments.

Figures 7 and 8 show the CDF curves as a function of system channel capacity for the case of SNR = 8*dB* and 20*dB* with L_r=L_t=3, N_T=N_R=7. It can be observed that our algorithm also obtains closer curve to the optimal one with higher SNR. From the simulation results we can observe that our proposed algorithm perform better performance of channel capacity than CJAS [25], in addition, it can reduce computational cost significantly compared with OS scheme. So our proposed algorithm is an effective method, at the same time, it has multiple antenna diversity advantage.

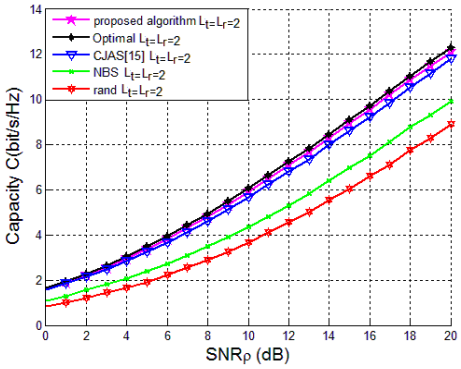


FIGURE 5 channel capacity versus *SNR* with L_r=L_t =2, N_T=N_R=5

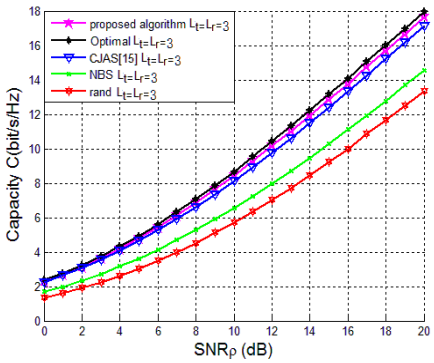


FIGURE 6 channel capacity versus *SNR* with L_r=L_t =3, N_T=N_R=7

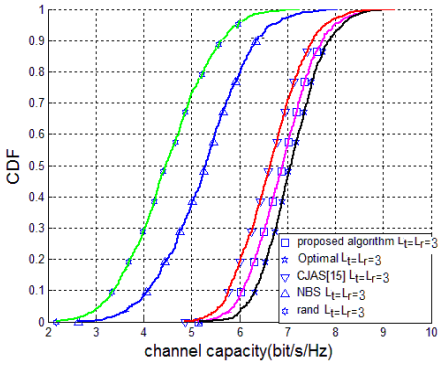


FIGURE 7 the CDF curves versus channel capacity with L_r=L_t=3, N_T=N_R=7 (*SNR*=8*dB*)

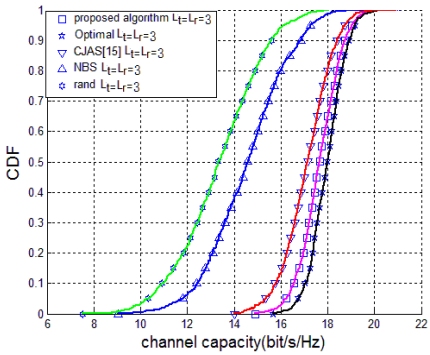


FIGURE 8 the CDF curves versus channel capacity with L_r=L_t=3, N_T=N_R=7 (*SNR*=8*dB*)

6 Conclusions

This paper presents an overview of antenna selection in MIMO systems. It focused on the spatial diversity and spatial multiplexing system. It shows that AS can reduce system implementation complexity and hardware cost, and achieve full diversity gain, and enhance the channel capacity and transfer rates. A novel joint transmit and receive antenna selection algorithm is presented in this paper. Our proposed algorithm is suitable for more number of selected antennas. Computer simulations show that the proposed algorithm can achieve near optimal channel capacity performance, while the overall complexity of the proposed algorithm is signifycantly lower than that of the optimal exhaustive search. However, there are still several problems in antenna selection, such as the problem of sub-optimal joint transmit and receive AS algorithm is still worth researching. How to evaluate performance of AS algorithms when there is a delay of feedback channel state information. At present, antenna subset selection problems have been intensively studied in Rayleigh fading channels, and Nakagami fading channels still needs more research.

Acknowledgments

The authors would like to thank the anonymous reviewers of this paper for their carefully reading of the manuscript as well as their many constructive comments. This paper is supported by the Technology Research and Development Projects of Quanzhou city (2012Z82), the Natural Science Foundation of Fujian Province of China (No. 2012J01270), and the Natural Science Foundation of Huaqiao University (Excellent Level Talents) (No.11BS430).

References

[1] Foschini G J, Gans M J 1998 On limits of wireless communications in a fading environment when using multiple antennas *Wireless Personal Communications* **6**(3) 311-35
[2] Yoshihisa Kishiyama 2013 Future radio access for 5G *NTT DOCOMO INC Copyright*
[3] Chiani M, Win M Z, Zanella A 2003 *IEEE Transactions on Information Theory* **49**(10) 2363-71
[4] Molisch A F, Win M Z, Winters J H 2001 Capacity of MIMO systems with antenna selection *Proc IEEE Int Contr Conf* **2** 570-4
[5] Gorokhov A 2002 Antenna selection algorithms for MEA transmission systems *Proc IEEE Int Contr Conf Acoustic, Speech, and Signal Processing* **3** 2857–60
[6] Gorokhov A, Gore D, Paulraj A, 2003 **49**(10) 2687-96
[7] Gharavi-Alkhansari M, Gershman A B 2004 *IEEE Transactions on Signal Processing* **52**(2) 339-47
[8] Bu H W, Hon T H, Mook-Seng L 2010 *IEEE Transactions on Communications* **58**(9) 2505-10
[9] Molisch A F, Win M Z, Winters J H 2005 *IEEE Transactions on Wireless Communications* **4**(4) 1759-72
[10] Gore D, Nabar R, Paulraj A 2000 Selecting an optimal set of transmit antennas for a low rank matrix channel *Proc IEEE Int Conf Conf* on *Acoustics, Speech, and Signal Processing* **5** 2785-8
[11] Irmer R, Fettweis G 2002 Combined transmitter and receiver optimization for multiple-antenna frequency-selective channels *In Proc. 5th Int. Symp on Wireless Personnal multimedia communications* **2** 412-6
[12] Duan H, Liu F 2007 A rapid antenna selection for MIMO diversity system *Journal of China Academy of Electronics and Information Technology* **2**(4) 605-10
[13] Molisch A F, Winm Z, Winters J H 2003 *IEEE Transactions on Signal Processing* **51**(11) 2729-38
[14] Winm Z, Winters J H 1999 *IEEE Transactions on Communications* **47**(12) 1773-6
[15] Gore D, Nabar R, Paulraj A 2001 Space-time block coding with optimal antenna selection *2001 IEEE Int Conf Acoustics, Speech, and Signal Processing* **4** 2441-4
[16] Zhuang M J, Lin B H, Peng G X 2005 Transmission performance study of multiple transmit antenna selection based on STBC *Navigation of China* **65**(4) 65-8
[17] Luo J Y, Wang J, Tang H C 2008 A sample algorithms based on antenna selection combined with quasi-orthogonal space-time codes *Communications Technology* **41**(1) 43-5
[18] Chen Q, Zhuang M J 2009 Capacity analysis of antenna selection based on quasi-orthogonal space-time codes *Journal of Dalian Maritime University* **35**(4) 71-4
[19] Chen C E, Chung W H 2013 Computationally efficient near-optimal combined antenna selection algorithms for V-BLAST systems *Digital Signal Processing* **23**(1) 375-81
[20] Zhuang M J, Chen R S, 2007 Effects of erroneous channel state information on performance of multiple transmit antenna selection *Chinese Journal of Radio Science* **22**(4) 652-8
[21] Gorokhov A, Collados M, Gore D, Paulraj A 2004 Transmit/receive MIMO antenna subset selection *Proc. Int. Conf. Acoustics, Speech, Signal Processing* **2** 13-6
[22] Sanayei S, Nosratinia A 2004 Capacity maximizing algorithms for joint transmit-receive antenna selection *Proc. 38th Asilomar Conf Signals, Systems, and Computers* **2** 1773-6
[23] Blum R S, Xu Z, Sfar S 2009 A near-optimal joint transmit and receive antenna selection algorithm for MIMO systems *Radio and Wireless Symposium RWS'09 IEEE* 554-7
[24] Chen C E 2010 *IEEE Communications Letters* **14**(5) 402-4
[25] Wang J L, Ana Isabel Pérez-Neira, Gao M G 2013 A concise joint transmit/receive antenna selection algorithm *Digital Communi-cations* **10**(3) 91-9
[26] Lan P, Liu J, Xu H J 2009 Joint transmit and receive antenna selection in MIMO systems for maximize channel capacity *Journal on Communications* **30**(7) 7-12
[27] Karamalis P D, Skentos N D, Kanatas A G 2004 *IEEE Transactions on Wireless Communications* **3**(6) 1994-8
[28] Hoang-Yang L, Wen-Hsien F 2007 Joint transmit/receive Antenna selection in MIMO systems based on the priority-based genetic algorithm *Antennas and Wireless Propagation Letters IEEE* **6** 588-91
[29] Naeem M, Lee D C 2011 Low-complexity joint transmit and receive antenna selection for MIMO systems *Engineering Applications of Artificial Intelligence* **24**(6) 1046-51

Authors

Yisong Lin, born in 1989, Anxi, Fujian province, China

Current positon, grades: Graduate student in the Hua Qiao University, China.
University study: Bachelor degree of Information Engineering Henna Institute of Science and Technology in 2012.
Research activities: wireless communications.

Mingjie Zhuang, born in 1964, Hui'an, Fujian province, China

Current positon, grades professor in Huaqiao University, China.
University study: Fudan University (Department of Electronic Engineering), Shanghai, 1982,Ph.D. degree in information and communications engineering, Xiamen University, Xiamen, Fujian, 2001.
Research activities: wireless communication technology, space-time processing, and stochastic processes.

Design and realization of atmospheric pressure altitude measuring system with temperature compensation based on FPGA

HongTao Zhang[1], Wei Zhao[2]*

[1]*ZhengZhou university software college, Zhengzhou, 450002, China*

[2]*ZhengZhou university software college, Zhengzhou, 450002, China*

Received 1 October 2014, www.cmnt.lv

Abstract

Starting from the analysis of the measuring principle of atmospheric pressure altitude and the necessity of the pressure sensor temperature compensation, this paper presents the design and realization of an atmospheric pressure altitude measuring system with high performance which use FPGA as the data processor and with the pressure sensor temperature compensation design. Article discusses in detail the hardware circuit design of the measuring system, and the internal structure of FPGA software design. At last, the results of system test verify the feasibility and effectiveness of the atmospheric pressure altitude measuring system. Because of the FPGA's characteristics of high reliability, strong data processing ability, high speed and so on, and the effective combination of temperature compensation design, the atmospheric pressure altitude measuring system has the advantages of high measuring precision, high reliability, good real-time and low power consumption.

Keywords: atmospheric pressure altitude, FPGA, temperature compensation algorithm, measuring system

1 Introduction

Atmospheric pressure altitude data is an important flight parameters of aircraft, it can guarantee the safety of the flight and ensure the ground command and control personnel to properly guide the flight to complete mission. The development of modern microelectronics technology and computer technology greatly promotes the development of aviation testing technology, and makes the development of aviation testing equipment to high precision, strong anti-interference, small size, and low power consumption [1].

The FPGA (Field Programmable Gate Array) is a programmable processor, it has strong data processing ability, high reliability, high speed, flexible design, intelligent development tools advantages etc, and it is more and more popular in the field of high-speed hardware electronic circuit design [2]. With the characteristics of FPGA, this paper introduces the design and implementation of a kind of high precision and real-time pressure measuring system with pressure sensor temperature compensation method.

2 Measuring principle of atmospheric pressure altitude

According to international standard ISO25332, which was developed by the international organization for standardization, we can get the relation formula between the geopotential altitude H (often referred to as the standard atmospheric pressure altitude) and the atmospheric pressure PH in 2000m to 80000m altitude range, and this relationship is shown in Equation (1):

$$H = \frac{T_b}{\beta}\left[\left(\frac{P_H}{P_b}\right)^{-\beta R/g_n} - 1\right] + H_b. \quad (1)$$

In Equation (1), *R* is the air special gas constant and the value is $287.05287\,m^2/\left(K\cdot s^2\right)$, g_n, is the free fall acceleration of gravity and the value is $9.80665\left(m/s^2\right)$, β is the rate of vertical temperature change and the value is dT/dH; T_b, H_b and P_b are the lower limiting value of atmospheric temperature, standard atmospheric pressure altitude and atmospheric pressure in the corresponding altitude range that are used by international standard atmosphere [1], and their values are shown in Table 1.

TABLE 1 Atmospheric temperature, temperature gradient and altitude stratified

standard atmospheric pressure altitude H_b(km)	temperature T_b(K)	vertical temperature change β(K·km^{-1})	atmospheric pressure P_b(Pa)
-2.00	301.15	-6.50	127 774
0.00	288.15	-6.50	101 325
11.00	216.65	0.00	22 632

Making these parameters what are in Table 1 into the Equation (1), we can obtain the single value functions of atmospheric pressure and atmospheric pressure altitude in different altitude stratified [3], and these functions are shown in Equation (2) and Equation (3).

In the range of -2000 to 0m altitude

$$H = -46330.8\left[\left(\frac{P_H}{127774}\right)^{0.190263102} - 1\right] - 2000. \quad (2)$$

***Corresponding author** e-mail: paul-2308@hotmail.com

In the range of 0 to 11000m altitude

$$H = -44330.8\left[\left(\frac{P_H}{101325}\right)^{0.190263102} - 1\right]. \quad (3)$$

3 Error analysis and compensation design

3.1 ERROR ANALYSIS

Piezoresistive pressure sensor is one of the most widely used sensors, but because the sensitivity of the semiconductor physics properties on temperature, the output of piezoresistive pressure sensor used in this measuring system will not only change with pressure but also change with the temperature. This phenomenon leads to the accuracy of the whole measuring system is reduced greatly, so a kind of software temperature compensation method for error correction used in FPGA processor is adopted to improve the accuracy of measuring system.

The idea of the temperature compensation method is to adjust the output values of pressure sensor at different environmental temperature to the output values at 25°C. The linear relationship between input p and output U of the pressure sensor at 25°C and t°C are shown in Equation (4) and Equation (5).

$$U_0 = a_0 + b_0 \times p, \quad (4)$$

$$U_t = a_t + b_t \times p, \quad (5)$$

a0, at, b0 and bt in Equation (4) and Equation (5) are zero position and sensitivity coefficient at normal temperature 25°C and t°C.

The output of pressure sensor with temperature compensation at t°C can be obtained by making the Equation (5) into the Equation (4) and it is shown in Equation (6).

$$U_t' = a_0 + \left(\frac{U_t - a_t}{b_t}\right) \times b_0. \quad (6)$$

System testing selects six groups of different atmospheric pressure value provided by the pressure pump for pressure sensor and six groups of different temperature value provided by high and low test temperature box for temperature sensor, the output values of pressure sensor without temperature compensation at different temperatures and pressures are shown in Table 2.

TABLE 2 Pressure sensor output values without compensation

Pressure/p_a	The output voltage/mV					
	-40°C	*-20°C*	*0°C*	*25°C*	*50°C*	*60°C*
49154.5	39.207	38.892	38.300	38.186	38.048	38.009
57728.3	46.023	45.661	44.975	44.849	44.688	44.645
67474.9	53.752	53.337	52.551	52.404	52.221	52.169
78513.1	62.526	62.049	61.146	60.981	60.771	60.711
90970.1	72.396	71.852	70.822	70.629	70.390	70.320
104981.0	83.529	82.908	81.727	81.512	81.238	81.158

It is obvious from Table 2 that the change magnitude of the pressure sensor output varies with temperature under the same pressure is large, and the maximum change amount under constant pressure output is 2.017 mV, to result in the decrease of sensor accuracy significantly [4].

3.2 COMPENSATION ALGORITHM DESIGN.

According to non-compensated pressure sensor data in Table 2, the linear relationship between the output *Ut* and the input pressure *p* of the sensor can be gotten by the linear interpolation algorithm using MATLAB software at different temperature, and these linear relationships are shown in Table 3.

TABLE 3 The relationships between input and output of sensor

Temperature	The linear relationship between sensor output and pressure
-40°C	$U_t = 0.19435 + 7.9379 \times 10^{-4} \times p$
-20°C	$U_t = 0.14724 + 7.8832 \times 10^{-4} \times p$
0°C	$U_t = 0.07012 + 7.7782 \times 10^{-4} \times p$
25°C	$U_0 = 0.04876 + 7.7596 \times 10^{-4} \times p$
50°C	$U_t = 0.02914 + 7.7354 \times 10^{-4} \times p$
60°C	$U_t = 0.02867 + 7.7278 \times 10^{-4} \times p$

According to the linear relationships between the output U_t and the input p of the pressure sensor at different temperature. The zero point data and sensitivity coefficient data of the pressure sensor at different temperature can be gotten, and then to use quadratic curve fitting formula with these data by MATLAB software to get zero point Equation (7) and sensitivity coefficient Equation (8).

$$a_t = 0.07143 - 2.17494 \cdot 10^{-3} \cdot t + 2.56347 \cdot 10^{-5} \cdot t^2 \quad (7)$$

$$b_t = 7.80122 - 2.6002 \cdot 710^{-3} \cdot t + 2.40446 \cdot 10^{-5} \cdot t^2 \quad (8)$$

Finally, the temperature compensation algorithm, which is shown in Equation (9), can be obtained by making the Equation (7) an Equation (8) into the Equation (6).

$$U_t' = 0.0487 + \frac{10000U_t - 0.256347t^2 + 21.7494t - 7143}{2.40446t^2 - 260.027t + 780122.0} \cdot 77.596 \quad (9)$$

4 Circuit design of measuring system

The circuit design of atmospheric pressure measuring system mainly consists of pressure sensor, temperature sensor, signal adjusting circuit, dual-channel A/D converter and FPGA processor, its working principle block diagram is shown in Figure 1.

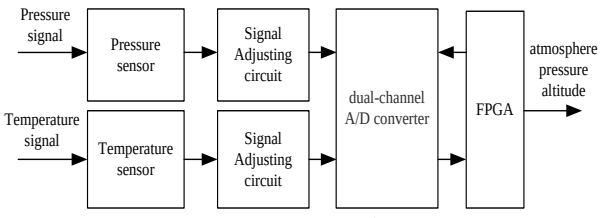


FIGURE 1 Principle diagram of system circuit

The output voltage signals of the pressure sensor and temperature sensor will be transmitted to the A/D converter through signal amplification, filtering and zero regulation of

the signal regulation circuit. Then the A/D converter will convert pressure and temperature analog signals into digital pressure value and temperature value, finally the FPGA processor will read the digital pressure and temperature value from the A/D converter, and output the atmospheric pressure altitude value by the calculation of the temperature compensation algorithm and atmospheric pressure altitude formula [5, 6].

4.1 SENSOR CIRCUIT

The pressure sensor used in this measuring system circuit is an absolute pressure sensors of NPC-1210-015A type made by GE NovaSensor company, it has the measuring range of 0Pa to 103425Pa, and the output voltage of the pressure sensor is in the range of 38.009 mV to 83.529 mV when the pressure sensor is tested in the pressure range of 49154.5Pa to 104981Pa.

The temperature sensor is AD590 made by AD Company, its measuring temperature range is the range of 55°C to150°C and the temperature coefficient is 1uA/K. The output current of the temperature sensor is in the range of 233.2uA to 333.2uA when the temperature sensor is tested in the temperature range of -40°C to 60°C If the output circuit of the temperature sensor connects an external 10k resistor, the sensor output voltage is the range of 2.332V to 3.332V.

4.2 SIGNAL ADJUSTING CIRCUIT

The signal adjusting circuit of the measuring system mainly completes the work of signal amplification, filtering and zero regulation of the sensor's output analog voltage signal.

The amplifier circuit is designed by AD620 made by AD Company which is low power consumption, low noise and high precision instrumentation, its use is very convenient that just connect a resistance R_G between pins 1 and pins 8 on a chip to adjust the amplification factor of amplifier, the Equation of amplification factor is shown in Equation (10).

$$G=49.4\ k\Omega/R_G+1. \quad (10)$$

Due to the reference voltage of the A/D converter is set to 1v, so the sensor output voltage range need to adjust the sensor voltage range of 0v to 1v, the value of resistance R_G can be calculated by Equation (10), the R_G resistance selection of 2.355kΩ resistance, magnification is about 21.968 times. In order to get more accurate output voltage range, one trimming resistor can be connected in parallel with R_G resistance.

Zeroing circuit in the system is implemented by anti-adder that is constituted by MAXIM OP07 operational amplifier. The role of zeroing circuit is to adjust the output voltage signal of pressure sensor and temperature sensor what are amplified to the 0 to 1V range in system testing scope [7, 8].

4.3 A/D CONVERTER

A/D converter used in the system circuit is AD7705 made by AD Company; it is a 16-bit dual-channel Sigma - Delta converter with 500Hz maximum sampling frequency and a three line serial communication port. AD7705 can work on two voltage range of 2.7V to 3.3V and 3.3V to 5.25V, in order to match with the working voltage of FPGA chip, the design choose 3V working voltage, and the reference voltage of AD7705 is VREF - 0V and VREF +1V.

The working status of AD7705 is set mainly by its 4 internal function registers which are communication registers, setting registers, clock registers and data registers. The communication register is used to select the register type will be working, the setting registers is used to set the AD7705 operating polarity and buffers, the clock register is used to select the AD7705 operating the master clock frequency and the sampling frequency, the data register is used to store finished converting 16 bit binary data, read these features and conversion data register settings are done through the FPGA control [3]. The connection diagram of AD7705 and FPGA chip is shown in Figure 2.

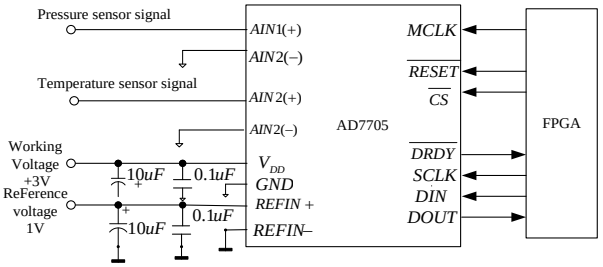


FIGURE 2 Connection diagram of AD7705

4.4 FPGA CIRCUIT

The type of the FPGA chip used in measuring system is Strax EP1S25F780C5 processor made by Altera Company, with an external crystal oscillator of 80MHz. FPGA is a core part of system control and data processing, to complete the work of reading the data from the A/D converter, implementing temperature compensation algorithm and calculating atmosphere pressure altitude.

5 Design of FPGA processor

The FPGA processor is designed by VHDL hardware description language using the QuartusII integrated development environment, the function structure of FPGA includes frequency dividing module, data reading and writing module, compensation calculating module and altitude calculating module.

5.1 FREQUENCY DIVIDING MODULE

The frequency dividing module plays a role of providing the clock signals that are required in FPGA control, A/D converter working and the serial port communication from main

clock signal, which is 80MHz. The internal structure of the frequency dividing module is shown in Figure 3.

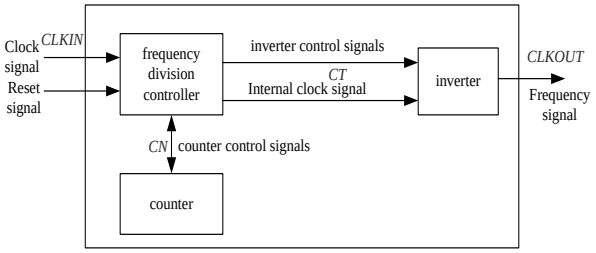


FIGURE 3 Inside block diagram of the frequency dividing module

As the system resets, the frequency division controller clears the counter through counter control signal *CN*, and then the counter starts counting according to the Rising edge of master clock *CLKIN* signal. When the counter reaches the frequency counts, a reverser control signal *CT* is sent to reverser by frequency division controller, and make the output clock signal *CLKOUT* reverse, at the same time counter will reset, so repeatedly.

5.2 DATA READING AND WRITING MODULE

Data reading and writing module mainly controls the working status of the A/D converter through the $\overline{CS}$ and $\overline{RESET}$ pins of AD7705, and sets the running status of A/D converter through the serial communication port, and ensures the FPGA reading the converted digital pressure value and temperature value fast and accurately. The internal structure of data reading and writing module is shown in Figure 4.

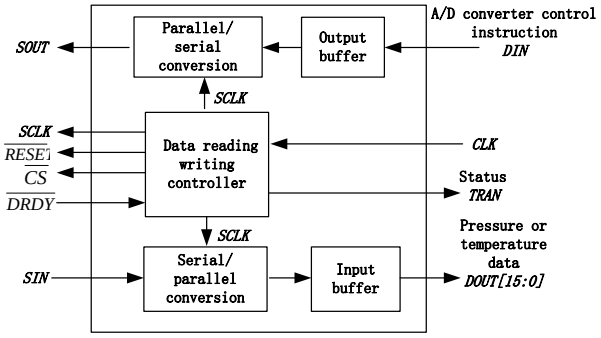


FIGURE 4 Inside block diagram of data reading and writing module

After the system resets, data reading and writing controller makes the A/D converter into working status through setting the $\overline{CS}$ and $\overline{RESET}$ signals, sends the operating status register command which has made parallel/serial conversion to A/D converter. Then monitors the status of pin *DRDY* in A/D converter, when *DRDY* status is valid, data can be read from the data register of A/D converter, and converts the serial data into parallel data [9].

5.3 COMPENSATION CALCULATING MODULE

Compensation calculating module is used to calculate the digital pressure value compensated according to the pressure sensor temperature compensation algorithm. The internal structure of compensation calculating module is shown in Figure 5.

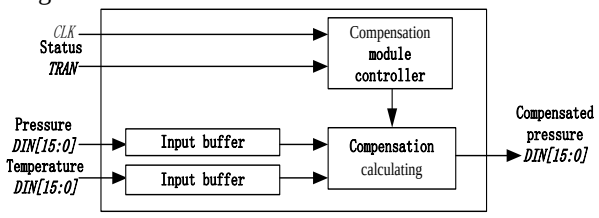


FIGURE 5 Inside block diagram of compensation calculating module

Calculating module controller monitors the value of *TRAN* at the trigger edge of the clock *CLK*, and reads pressure and temperature data when the value of *TRAN* is valid and calculates the pressure value with the temperature compensation algorithm.

5.4 ALTITUDE CALCULATING MODULE

The function of altitude calculating module is to calculate the atmospheric pressure altitude value according to the Equation (2) and Equation (3). In order to simplify the calculation for improving calculating efficiency, the linear interpolation method is adopted to make approximate treatment for atmospheric pressure altitude Equation. According to the error formula of the linear interpolation method, which is shown in Equation (11), and the precision requirements of the system design, we can calculate the maximum step length of the interpolation and the minimum interpolation section number in measuring range.

$$|R(x)| = \left|\frac{f''(\delta)}{(2)!}\omega(x)\right| \leq \delta. \tag{11}$$

At each interpolation interval, we can use Equation (12) to make approximate calculation, and the interpolation parameters y_i, k_i, x_i in Equation (12) can be calculated in advance and stored in memory.

$$y = y_i + k_i(x - x_i). \tag{12}$$

The internal structure of altitude calculating module is shown in Figure 6.

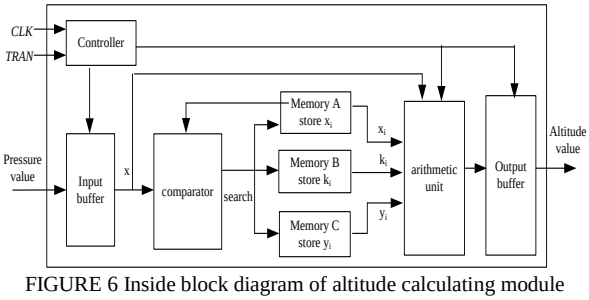


FIGURE 6 Inside block diagram of altitude calculating module

When TRAN value is valid, the controller will read the pressure value, which is compensated into input buffer. First of all, by comparing the pressure value decide the interpolation interval, and then the corresponding interpolation parameters can be got from the FPGA memory, finally to calculate the atmospheric pressure altitude value according

to the Equation (10). In order to improve the search efficiency and the computational efficiency, the binary search algorithm can be adopted when to search interpolation interval.

6 FPGA simulation and system testing

6.1 FPGA SIMULATION

After accomplishing program edit of FPGA using VHDL language, the functional simulation can be performed by Modelsim SE 6.0 software to verify the circuit function whether to meet the design requirements. The functional simulation picture of data reading and writing module is shown in Figure 7.

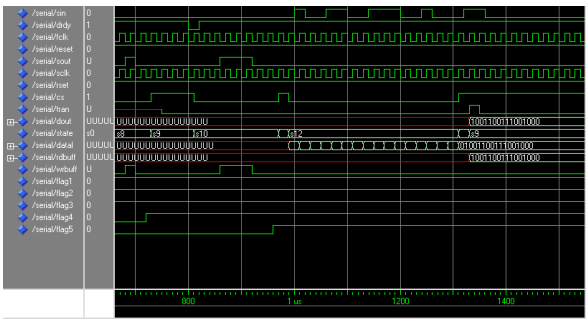


FIGURE 7 Functional simulation map of data read module

If every functional simulation of the FPGA's functional blocks is correct, the integrated design can be performed by QuartusII software to get the connection diagram of the basic logic units. The logic connection diagram of compensation algorithm is shown in Figure 8.

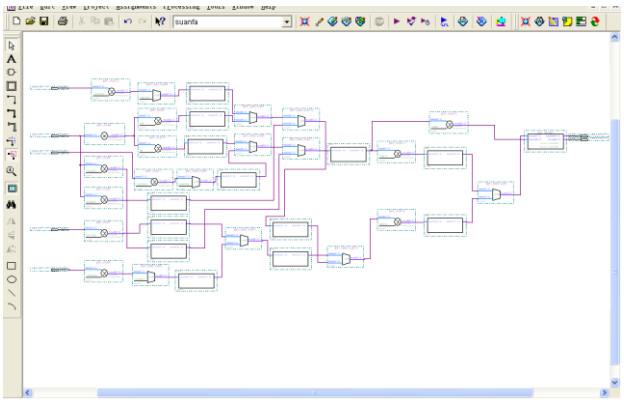


FIGURE 8 Logical connection diagram of compensation algorithm

FPGA's timing simulation is implemented after the completion of the placement and routing design in the FPGA, this kind of simulation contains the most comprehensive and accurate delay information, and reflect effectively the actual working state of the FPGA chip [10, 11]. The simulation diagram of compensation algorithm is shown in Figure 9.

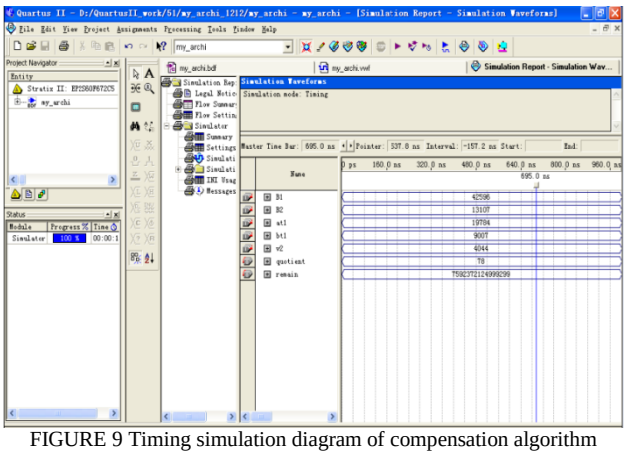


FIGURE 9 Timing simulation diagram of compensation algorithm

6.2 SYSTEM TESTING

After the FPGA simulation, the configuration file generated by source code program can be downloaded to the FPGA chip, and then connect FPGA chip to the whole measuring system. The measuring system is tested with the six groups of atmospheric pressure and three kinds of temperature, and the results of testing are shown in Table 4.

Table 4 The results of measuring system with compensation

atmospheric pressure (p_a)	theory altitude (m)	measuring altitude (m)		
		-20°C	*25°C*	*50°C*
49154.5	5700	5701.6	5700	5700.7
57728.3	4500	4497.9	4498.8	4501.1
67474.9	3300	3300.7	3300.5	3301.8
78513.1	2100	2097.7	2098.6	2102.6
90970.1	900	897.8	898.9	900.9
104981.0	-300	-297.5	-299.4	-298.6

As can be seen from Table 4, the maximum error of measuring system tested at different atmospheric pressure and temperature is 2.6m, and it is proven that this measuring system has the very high precision. Factors lead to error are the following:

1) Calculation error of FPGA data calculation.

2) The precision of reference voltage used in A/D converter.

3) Inherent quantization error of A/D converter.

4) External error of system test, such as the accuracy of the test equipment, human error, etc.

7 Conclusions

In this paper, a kind of atmospheric pressure measuring system is designed and implemented with temperature compensation design based on FPGA processor. Because FPGA has the advantageous characteristics of high reliability, high speed, strong processing capacity, small volume, low power consumption and so on, so this measuring system has a better performance in terms of stability, real-time, power consumption and so on, and the data processing time is only about 9us. The combination with the temperature compensation design in system also improves greatly the precision of the measuring system on the whole.

References

[1] Fan S C, Lv J F, Zhang Q R 2005 Aviation Test System *BeiHang University Press Beijing* 234-40 (*in Chinese*)
[2] Li H Z 2011 FPGA/ASIC High Performance Digital System Design *Electronic Industry Press Beijing* (*in Chinese*)
[3] Yan J M, Zhang H T 2007 Chinese Journal of Sensors and Actuators **20**(3) 707-10 (*in Chinese*)
[4] Zhang H T, Xue J X 2013 Applications of Electronic Technique **39**(4) 65-7 (*in Chinese*)
[5] Liang S Z, Guo X M, Yu X T 2013 *Chinese Journal of Sensors and Actuators* **26**(5) 654-9 (*in Chinese*)
[6] Wang J C, Wang F P, Hou R F 2011 *Transducer and Microsystem Technologies* **30**(12) 123-5
[7] He J T, Liu X Y 2009 Intelligent Sensor Principle, Design and Application *Electronic Industry Press Beijing* (*in Chinese*)
[8] Creed H 2009 Intelligent Sensor Design *People's Posts and Telecom Press Beijing*
[9] Cheng Z L 2011 *Applications of Electronic Technique* **37**(4) 7-9
[10] Zhou R J, Tu Y, Zhang L M 2007 FPGA/CPLD Digital System Design Example Based on Quartus II *Electronic Industry Press Beijing* (*in Chinese*)
[11] Xu X, Yu H Q 2005 Embedded System Design Based on FPGA *Mechanical Industry Press Beijing* (*in Chinese*)

Authors

HongTao Zhang, born in 1977, Xian City, Shanxi province, China

Current position, grades: lecturer in Zhengzhou University, China.
University studies: postgraduate degree in computer science and application from Northwestern Polytechnical University in 2007.
Scientific interest: design and development of embedded system, distributed data processing, web design and development.

Wei Zhao, born in 1976, Shangdan County, Gansu province, China

Current position, grades: lecturer in Zhengzhou University, China.
University studies: postgraduate degree in computer science and application from Zhengzhou University in 2007.
Scientific interest: distributed data processing, web design and development, data mining technology, mass data processing technology.

Design and implementation of a vibration cantilever energy harvester with suspended piezoelectric beam

Xiaojie Chen[1, 2], Chengjun Qiu[1,*], Hongmei Liu[1], Wei Qu[1], Yibo Liu[1]

[1]*Key Laboratory of Electronics Engineering of Heilongjiang Province, Heilongjiang University, Harbin, 150080, China*

[2]*The College of Electronic and Information Engineering, Heilongjiang University of Science and Technology, Harbin, 150027, China*

Received 1 March 2014, www.cmnt.lv

Abstract

The frequency response and energy conversion efficiency are the critical issue for vibration energy harvester. In this paper, a vibration energy harvester with suspended piezoelectric beam is proposed. This harvester structure is composed of a discrete support beam in the bottom and a suspended piezoelectric beam on top. This suspended beam structure is beneficial to the higher response frequency and the energy conversion efficiency by applying the tension on the shear mode piezoelectric layer. An analytical bending model of suspended piezoelectric beam is developed, based on the actual deformation: rotational and translational movement modes. The characteristic of suspended piezoelectric beam is also analyzed by Finite element method. Finally, a 1000:1 prototype is fabricated and measured. The optimums acquired experiment results show that it is matching well with the model. The maximum of output voltage is 6.5 V.

Keywords: energy harvester, vibration, suspended piezoelectric beam, piezoelectric layer

1 Introduction

Energy harvester is the device, which can convert environment energy, such as vibration, temperature, solar energy, etc., into useful electrical energy [1-6]. And with the development of NMES/MEMS technology, and wireless network sensors, the micro-scale energy harvester can generate electric energy at milli watt level. At the same time, the power consumption of wireless sensors drastically reduced to milliwatt and even nanowatt level. Thus, the micro-scale energy harvester is popular to power sources. Among the different energy forms, the vibration energy is a good choice due to its widespread existence. The vibration harvester can harvest vibration energy via different materials and structure [7-11]. And among different structure of vibration energy harvesters, the cantilever-mass structure is regarded as a promising structure due to the small size, little power consumption, low cost, and compatibility with integrated circuits [12-16].

However, the lower energy efficiency and lower response frequency limit the cantilever-mass energy harvester further application. To resolve these problems, several approaches were chose.

Firstly, selecting suitable material such as PZT, PMNT, and PDVF, etc. is a useful method to enhance efficiency [17].

Secondly, increasing the thickness of the piezoelectric layer is also a good choice since the normal strain experienced by the piezoelectric layer is proportional to its thickness [18, 19].

Thirdly, using the proper coupling mode of piezoelectric material is also worth considering, since the different coupling mode relate to the different electrode pattern [20]. Finally, choosing novel configuration such as tapered cantilever beam, L-shaped beam and so on can affect the frequency of harvester [21-24].

This paper reports an energy harvester with suspended piezoelectric beam for d31 mode. The cantilever structure is composed of a discrete support beam in the bottom and a piezoelectric beam on top. And a gap is forming between the top beam and the bottom beam. The electrode mode is selected as d31 mode based on the distribution of strain. To achieve more accurate analysis, the frequency formula of the suspended piezoelectric beam is derived, considering the two actual deformation modes. Furthermore, it also discussed how the geometrical parameters of suspended piezoelectric beam affect harvester's performance. This model is verified by a prototype and experiments afterwards.

2 Modeling analyses

The conventional mass-cantilever structure is shown in Figure 1, it consists of three parts, the piezoelectric layer, the support layer and the proof mass. And the piezoelectric layer is often attached on the support layer.

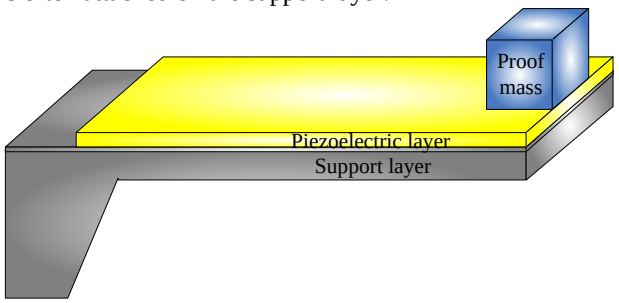


FIGURE 1 Conventional cantilever structure

[*]***Corresponding author*** e-mail: chiucj@hlju.edu.cn

Different from conventional mass-cantilever structure, the suspended piezoelectric beam energy harvester is constructed as Figure 2.

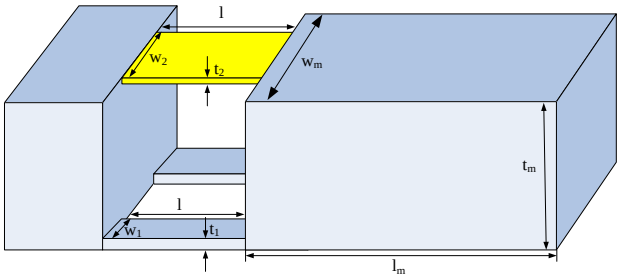


FIGURE 2 Structure of energy harvester with suspended piezoelectric beam

And the piezoelectric material operates in the d_{31} mode to enhance the output charge. The energy harvester is composed of a discrete support bottom beam and a piezoelectric top beam, so the gap structure exists between the top beam and bottom beam. And the discrete support bottom beam is composed of two small parallel beams, the geometric dimension of two parallel beams is same. The dimensions (width × thickness × length) of the bottom support beams, the piezoelectric top beam, and proof mass are $2\times w_1\times t_1\times l$, $w_2\times t_2\times l$, and $w_m\times t_m\times l_m$, respectively.

Cross-sectional view of suspended piezoelectric beam energy harvester is shown in Figure 3.

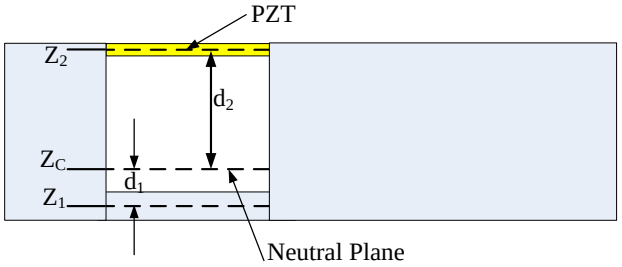


FIGURE 3 Cross-sectional view of energy harvester with suspended piezoelectric beam

Z_1 and Z_2 are the middle planes coordinates of the bottom and top beams respectively, and Z_C is the coordinate of neutral plane. In mechanics, the neutral plane is a conceptual plane where the stress is zero. When an inertial force is applied, the cantilever bends so that one surface of cantilever is in compression and the other surface of that is in tension. d_1 is the distance between the Z_1 and Z_C, d_2 is the distance between the Z_C and Z_2. Accordingly, the neutral plane coordinate Z_C of suspended piezoelectric beam can be expressed as:

$$Z_c = \frac{E_1A_1Z_1 + E_2A_2Z_2}{E_1A_1 + E_2A_2}. \tag{1}$$

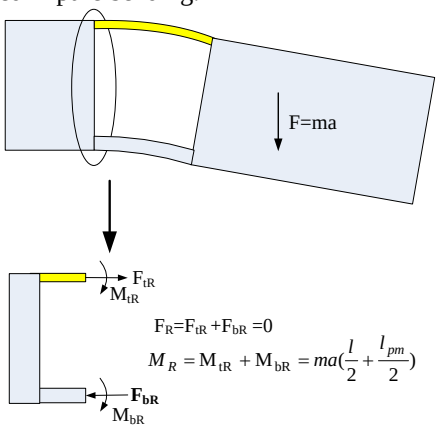
E_1, E_2, A_1 and A_2 are respectively defined as the Young modulus and cross-sectional area of bottom beam and top beam. A_1 is equal to the product of width $w1$ and thickness t_1 of bottom beam. A_2 is the product of width w_2 and thickness t_2 of top beam. Thus, the formulations of d_1 and d_2 can be expressed as:

$$\begin{cases} d_1 = \dfrac{E_2A_2(Z_2 - Z_1)}{E_1A_1 + E_2A_2} \\ d_2 = \dfrac{E_1A_1(Z_2 - Z_1)}{E_1A_1 + E_2A_2} \end{cases}. \tag{2}$$

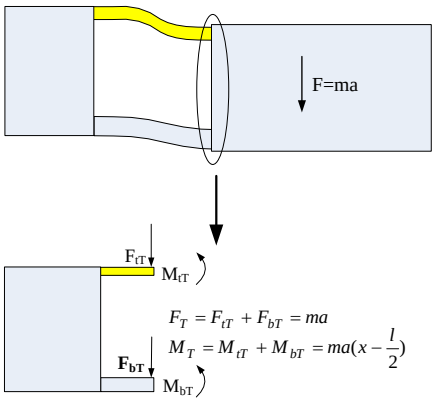
Since the thickness of PZT film prepared by Sol-Gel is generally less than 4 um, and then the bottom should be thick enough to support the big proof mass. As a result, the location of neutral plane will be shift down to close to the bottom beam. From the Equation (2) and Figure 3, the distance d_2 can approximately equal to the gap distance when considering the thickness of top beam and bottom beam.

Next the analysis model is discussed. The Euler-Bernoulli beam theory based on Kirchhoff's assumption is not accurate for the suspended beam structure. Therefore, it is necessary to establish a new model to describe this structure. In practice, we can observe two deformation modes: rotational movement and translational movement, shown in Figure 4. In our analysis, the beam's mass and the proof mass deformation are ignored.

For the rotational movement, the bending moments M of beam is constant and the shear force of beam is zero. Thus, the deformation of suspended piezoelectric beam can be considered in pure bending.



a) Rotational movement mode



b) Translational movement mode

FIGURE 4 Two movement modes

From the theorem of multilayer structure cantilever, rotary inertia of parallel axis is obtained by $I_R = I_i + A_i(d_i)^2$.

The bending rigidities of rotational movement can be expressed as:

$$(EI)_R = E_1 I_b + E_2 I_t + E_1 A_1 d_1^2 + E_2 A_2 d_2^2 . \quad (3)$$

For the translational movement, the mass remains in horizontal direction after deforming, this means the bending moment is a linear function of the horizontal position and the bending moment at the middle of the beam is zero, the shear force is constant. The bending rigidity of translational movement is $(EI)_T = E_1 I_b + E_2 I_t$.

Thus, the total bending rigidities of suspended piezoelectric beam is:

$$(EI)_Z = \frac{4(EI)_R}{[3\gamma^2 + \mu]}, \quad (4)$$

where $\gamma = \frac{l + l_{pm}}{l}$, $\mu = \frac{(EI)_R}{(EI)_T}$.

Finally, the resonant frequency of vibration energy harvester with suspended piezoelectric beam can be expressed as:

$$f = \frac{\pi}{11.24}\sqrt{\frac{(EI)_R}{ml^3(3\gamma^2 + \mu)}} . \quad (5)$$

Equation (5) describes the length of cantilever and the mass of proof mass as the function of frequency.

From the energy perspective, the piezoelectric layer was expected to store as much stretching energy as possible to transform into the electric energy from the total mechanical energy applied. Thus, the energy conversation efficiency should be discussed. As shown in Figure 4, the force F_{tR} of rotational mode stretch the piezoelectric layer, which means the strain is generated. But the force F_{tT} translational modes don't contribute the strain. As a result, the rotational mode is domain in contributing the strain of whole structure. Thus, the energy of rotational mode is composed of bending energy and normal stretching/compress energy. The value η defined as ratio of stretching energy stored in the top piezoelectric layer to the total mechanical energy, and η can be expressed as:

$$\eta = \frac{E_2 A_2 d_2^2}{(EI)_R} . \quad (6)$$

Finally, the output voltage and charge also can be considered. To generate as much output voltage as possible, the piezoelectric layer should be experience the maximized stretching stress accordingly. But the shear force of translational movement cannot contribute any stretching/compress strain for the top piezoelectric layer. Thus, the mainly electric energy is generated in the rotational movement and the contribution of translational movement can be negligent.

The output voltage and charge under d_{31} mode can be expressed as:

$$\begin{cases} V = \dfrac{F_{tR} \cdot d_{31}}{\varepsilon_{33} \cdot \varepsilon_0 \cdot w_2} \\ Q = -F_{tR} \cdot A_{elec} \cdot d_{31} \end{cases}, \quad (7)$$

where ε_{33} is the dielectric constant, A_{elec} the electrode area [25]. Substituting F_{tR} shown in Figure 4 into Equation (7), the output voltage and charge can be expressed as:

$$\begin{cases} V = \dfrac{ma \cdot (l + l_m) \cdot d_2 \cdot E_2 \cdot A_2 \cdot d_{31}}{2 \cdot (EI)_R \cdot \varepsilon_{33} \cdot \varepsilon_0 \cdot w_2} \\ Q = -\dfrac{ma \cdot (l + l_m) \cdot d_2 \cdot E_2 \cdot A_2^2 \cdot d_{31}}{2 \cdot (EI)_R \cdot \cdot w_2} \end{cases} . \quad (8)$$

3 Optimum design

To obtain the better performance, some parameters of suspended piezoelectric beam structure should be considered, especially in MEMS scale. The Figure 5 shows the relationship between frequency and the cantilever length and the proof mass length. The frequency is range from 1.5 KHz to 308 KHz. And the length of proof mass is from 10 um to 50 um and that of suspended piezoelectric beam of is 1 um to 50 um, respectively.

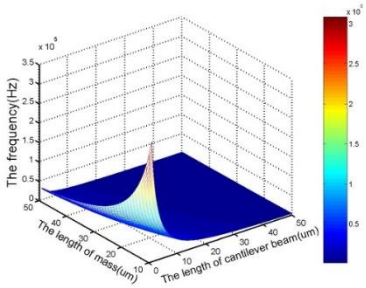
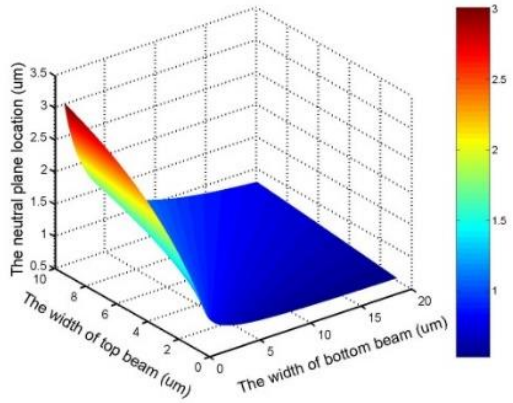


FIGURE 5 The frequency depending on the length of mass and that of cantilever

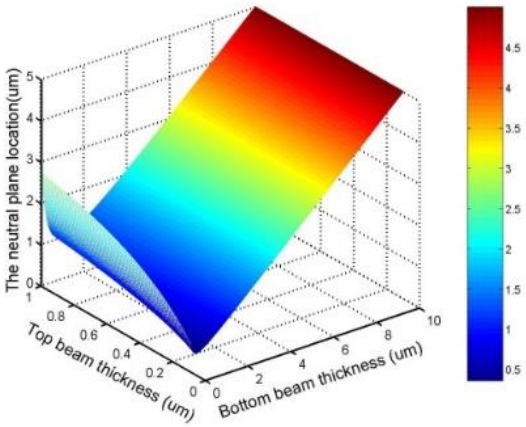
As shown in Figure 5, the smaller length of suspended beam and proof mass can lead to the higher frequency. Giving the width, the thickness, and the density, the length of proof mass can represent the mass. Thus, a conclusion can be derived that the frequency of suspended beam structure is affected the mass of proof mass and the length of suspended beam.

The neutral plane also should be discussed. From Figure 3 and Equation (1), the location of neutral plane is linked with the Young modulus, the width and the thickness of bottom beam and top beam. For conventional cantilever, the neutral plane should be far away the PZT layer, which can generate strong output voltage. This case is suitable for suspended beam structure. The relationship of neutral plane and width of top beam and bottom beams is shown in Figure 6a. The location of neutral plane is distinctly dropped with the decreasing of top width, but there is tiny change to the bottom beam. As shown in Figure 6b, the similar conclusion on the neutral plane and thickness of the top and bottom beams is derived. The thinner top beam generates the lower neutral plane location. But the relationship between the

thickness of bottom beam and neutral plane location isn't the linear function. And the neutral plane location is lowest when the bottom beam thickness is 1 um.



a) The relationship of neutral plane and thickness



b) The relationship of neutral plane and width

FIGURE 6 The parameter relationship: a); b)

Next, the optimum energy conversation efficiency should be considered. The relationship between the width ratio of two beams and energy efficiency is shown in Figure 7. It is obvious that the greater width of top beam can improve efficiency. When the width ratio is approximately equal to 0.5, the optimum efficiency is 87.5%. Similarly, the relationship between different thickness ratio of two beams and efficiency is shown in Figure 8, the optimum thickness ratio is 0.5.

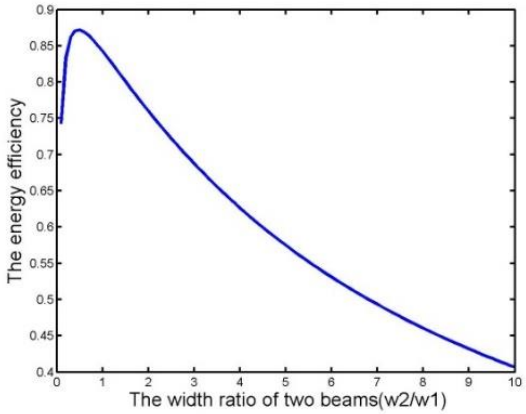


FIGURE 7 Relationship of width ratio and efficiency

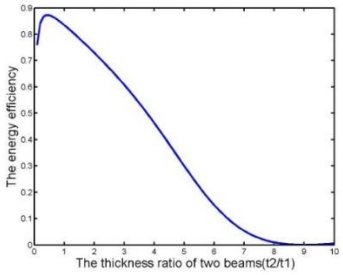


FIGURE 8 Relationship of thickness ratio and efficiency

Finally, the performance comparison of suspended piezoelectric beam structures and conventional structure is given in Table 1 based on Equation (5) and Equation (6). From the Table 1, we can observe that the suspended beam structure has the better performance than the conventional structure. Thus, a conclusion can be derived that the suspended beam helps to improve the resonant frequency and enhance the energy conversation efficiency.

TABLE 1 Comparison of suspended piezoelectric beam and conventional beam (l=8 um, w1=20 um, t1=1 um, w2=10 um, t2=0.5 um, L=40 um, wm=30 um, T=5 um)

	Suspended beam structure	**Conventional structure**
Frequency	372(K Hz)	102(K Hz)
Conversation efficiency	83.4	58.6

4 Measurement

In order to verify the analysis model and simulation results, the millimeters lever prototype is fabricated, and the dimension of prototype is a 1000:1 time than simulation parameter in Table 1.

To obtain the similarly performance compared with the silicon-based device, some candidate material should be considered. The bottom beam and proof mass is fabricated with 45# steel, since its mechanical property is very similar to that of silicon, and its Young modulus is 209 GPa.

The measurement system picture is shown in Figure 9. The device was fixed to a Vibrator (SINOCERA JZK-2), which is used to generate mechanical vibrations. Sinusoidal signals were applied to the vibrator at various frequencies. The voltages from the energy harvesting device were monitored by an Oscilloscope (YOKOGAWA DL850) at the same time, and the non-contact Doppler Vibrometer (Vibroducer V1002) is used to measure the contrastive voltage. The experiment frequency is 376.0 Hz, and the peak output voltage is 25.9 V, shown in Figure 10, which agrees well with the analysis frequency of 372 Hz.

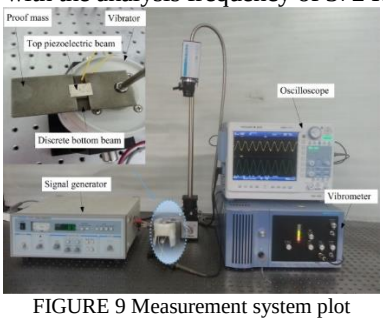


FIGURE 9 Measurement system plot

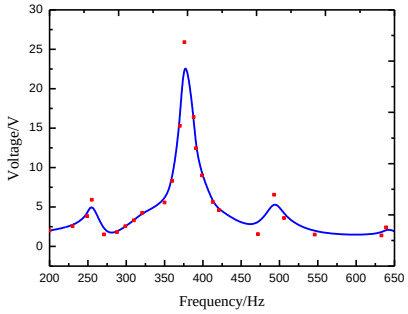


FIGURE 10 Frequency response of the suspended piezoelectric beam

Also, the plot of charging voltage/power is shown in Figure 11, the value of the capacitor was 100 μF. The capacitor voltage collected from the energy harvester is 6.5 V. Providing $E = \frac{1}{2}CV^2$ and $E = P \times t$, the 29.5 μW maximum power is derived.

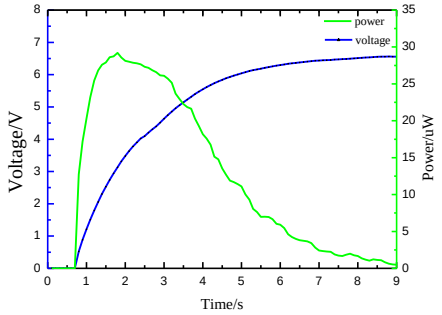


FIGURE 11 Plot charging voltage/power of the capacitor

5 Conclusions

The research on energy harvester with suspended piezoelectric beam, including structure design and optimum, modeling analysis, prototype fabrication and measurement are discussed in this paper. Considering the actual deformation in rotational and translational movement modes, the formulation of resonant frequency of suspended piezoelectric beam is derived. Furthermore, some parameters are considered to optimum the design. Accordingly, the acquired 1000:1 scale prototype is assembled. The prototype can generate the 6.5 V output voltage and 29.5 μW output power. The experiment frequency is 376.0 Hz that match well with 372.0 Hz of calculated frequency.

Compared to conventional cantilever, suspended beam structure can firstly enlarge the response frequency due to the existing gap which is affected the bending rigidities of system. The second advantage of suspended piezoelectric beam is the improvement of energy conversion efficiency by matching heavier mass, which improve output charge by stretching the piezoelectric layer. The finally advantage lies in the technology feasible of device fabrication [26].

Acknowledgments

This work was supported by Ph.D. Programs Foundation of Ministry of Education of China (20132301110005). This work is also supported by the Heilongjiang Province Graduate Student Innovate Scientific Research Foundation 2012 (YJSXC2012-317HLJ).

References

[1] Mateu L, Moll F 2005 Review of energy harvesting techniques and applications for microelectronics (Keynote Address) *VLSI Circuits and Systems II* **5737** 359-73
[2] Zhou L, Sun J, Zheng X J, Deng S F, Zhao J H, Peng ST, Zhang Y 2012 A model for the energy harvesting performance of shear mode piezoelectric cantilever *Sensors and Actuators A Physical* **179** 185-92
[3] Williams C B, Shearwood C, Harradine M A, Mellor P H, Birch T S, Yates R B 2001 *IEEE Proceedings on Circuits, Devices and Systems* **148**(6) 337-42
[4] Roundy S, Wright P K, Rabaey J M 2004 Energy Scavenging for Wireless Sensor Networks: With Special Focus on Vibrations *Springer*
[5] Roundy S, Wright P K, Rabaey J 2003 A study of low level vibrations as a powersource for wireless sensor nodes *Computer Communications* **26** 1131-44
[6] Noh J Y, Yoon G H 2012 Topology optimization of piezoelectric energy harvesting devices considering static and harmonic dynamic loads *Advances in Engineering Software* **53** 45-60
[7] Vullers R J M, Van Schaijk R, Visser H J, Van Hoof C 2010 *EEE Solid-State Circ Magazine* **2**(2) 29-38
[8] Singh K B, Taheri S, Priya S 2012 Piezoelectric vibration energy harvesting with an adaptive frequency tuning mechanism for intelligent tires *Mechatronics* **22** 970-88
[9] Van Schaik R, Elfrink R, Pop V, Vullers R 2012 Energy harvesting for self powered sensor systems-case study: vibration energy harvesting for 'intelligent tire' application *Proceedings of the 17th Workshop on Synthesis And System Integration of Mixed Information Technologies (SASIMI)* 141-6
[10] Lee J, Choi B 2014 Development of a piezoelectric energy harvesting system for implementing wireless sensors on the tires *Energy Conversion and Management* **78** 32-8
[11] Shu Y C, Lien IC 2006 Efficiency of energy conversion for a piezoelectric power harvesting systems *Micromech. Microeng* **16** 2429-38
[12] Kwon D, Rincón-Mora G A 2010 IC *IEEE Transactions on Biomedical Circuits and Systems* **4**(6) 400-9
[13] Shen Hui, Qiu JH, Ji HL, et al 2010 A lowpower circuit for piezoelectric bibration control by synchronized switching on voltage sources *Sensors and Actuators A* **161** 245-55
[14] Amarasinghe R, Dao D V, Toriyama T, Sugiyama S 2007 Development of miniaturized 6-axis accelerometer utilizing piezoresistive sensing elements *Actuators A Phys* **134** 310-20
[15] Dong P, Li X X, Yang H, Bao H, Zhou W, Li S Y, Feng S L 2008 High-performance monolithic triaxial piezoresistive shock accelerometers *Actuators A Phys* **141** 339-46
[16] Eklund E J, Shkel A M 2007 Single-mask fabrication of high-G piezoresistive accelerometers with extended temperature range *Micromech. Microeng* **17** 730-6
[17] Xue H, Hu Y T, Wang Q M 2008 *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control* **55**(9) 2104-8
[18] Kim S, Clark W W, Wang Q-M 2005 Piezoelectric energy harvesting with a clamped circular plate: Analysis *Journal of Intelligent Material Systems and Structures* **16** 847-54
[19] Kim S, Clark W W, Wang Q-M 2005 Piezoelectric energy harvesting with a clamped circular plate: Experimental study *Journal of Intelligent Material Systems and Structures* **16** 855-63

[20] Ramsay M J, Clark W W 2001 Piezoelectric energy harvesting for bio-MEMS applications SPIE's 8th Annual International Symposium on Smart Structures and Materials *International Society for Optics and Photonics* 429-38
[21] White N M, Glynne-Jones P, Beeby S P 2001 A novel thick-film piezoelectric micro-generator *Smart Mater. Struct.* 10 850-2
[22] Glynne-Jones P, Beeby S P, White N M 2001 *IEEE Proceedings-Science, Measurement and Technology* **148**(2) 68-72
[23] Glynne-Jones P, Beeby S P, White N M 2001 The modelling of a piezoelectric vibration powered generator for microsystems *Proceedings of the 11th International Conference on Solid-State Sensors and Actuators, Transducers* **1** 46-9
[24] Erturk A, Renno J M, Inman D J 2009 Piezoelectric Energy Harvesting from an L-Shaped beam-mass structure with an application to UAVs *Journal of Intelligent Material Systems and Structures* **20** 529-44
[25] Tang Gang 2013 Research on MEMS energy harvester based on piezoelectric thick film [PhD Dissertation] *Shanghai: Shanghai Jiao Tong University*
[26] Chen X J, Qiu C J, Liu H M 2014 Research of Complex PZT Film Base on Shear Mode Energy Harvesting *Advanced Materials Research* **884** 363-9

Authors

Xiaojie Chen, China

Current position, grades: graduate student.
University studies: Heilongjiang University, Harbin, 150080, China.
Scientific interest: functional materials and devices, inorganic non-metallic materials, MEMS and its application in engineering.

Chengjun Qiu, China

Current position, grades: PhD, professor.
University studies: PhD in College of Material Engineering from Harbin Engineering University in China.
Scientific interest: functional materials and devices, inorganic non-metallic materials, MEMS and its application in engineering.

Hongmei Liu, China

Current position, grades: PhD, associate professor.
University studies: PhD in College of Material Engineering from Harbin Engineering University in China.
Scientific interest: functional materials and devices, inorganic non-metallic materials, MEMS and its application in engineering.

Wei Qu, China

Current position, grades: PhD, associate professor.
University studies: PhD in College of Institute of electrical engineering from Hongjiang University in China.
Scientific interest: functional materials and devices, inorganic non-metallic materials, MEMS and its application in engineering.

Yibo Liu, China

Current position, grades: Graduate student.
University studies: Heilongjiang University, Harbin, 150080, China.
Scientific interest: functional materials and devices, inorganic non-metallic materials, MEMS and its application in engineering.

Prediction of capacitance of electrolytic capacitor with ripples

Pengchao Ye, Xiaochun Guan, Xiaojing Chen*

College of Physics and Electronic Engineering Information, Wenzhou University

Key Laboratory of Low-voltage Apparatus Intellectual Technology of Zhejiang, Wenzhou University, Chashan University Town, Wenzhou, Zhejiang Province, 325035, P.R. China

Received 1 March 2014, www.cmnt.lv

Abstract

The electrolytic capacitor is one of the most critical components in the switching power supply, in actual use of the power supply, the ripple is an important factor that leads to failure of the electrolytic capacitor. The research on the ripple's effect on the service life of the electrolytic capacitor is significant to the stability of the switching power supply. This paper designed a controllable circuit generating ripples of different phases, tested the electrolytic capacitors' changes in capacitance with ripples of different phases, analysed the reason of the changes in the capacitance, selected the fastest change, and used three prediction models of support vector regression, radial basis function and kernel-based partial least-squares algorithm to predict the change trends of capacitance of electrolytic capacitor with ripples of this phase. The results showed that compared with radial basis function and support vector regression model kernel-based partial least-squares algorithm can more accurately predict the service life of the electrolytic capacitor.

Keywords: electrolytic capacitor, ripples, different phases, prediction of capacitance, capacitor life

1 Introduction

The electrolytic capacitor is one of the most critical components in the switching power supply, and the service life of the electrolytic capacitor directly determines the service life of the power supply and other electronic devices [1]. In actual use of the power, the ripple is an important factor that leads to failure of the electrolytic capacitor [2]. Ripples of different initial phases reflect different circumstances of electrolytic capacitor failure. In the actual production process, the sampling method is often used to test quality of mass capacitors. In this paper, three prediction models of radial basis function, kernel-based partial least-squares algorithm and support vector regression can be used to test the failure time of mass capacitors to evaluate the applicability of the product. RBF network, that is, radial basis function neural network can approximate any nonlinear function, can handle the difficult regularity within the system, and has good generalization ability as well as quick learning and convergence rate, so it can be well used in the time series analysis in this experiment [3]. KPLS introduces RBF as a kernel function into partial least-squares regression, maps through the kernel function the input data into high-dimensional characteristic space, and then operates in the space, so that the nonlinear relationship between input and output variables can be established [4]. In SVR, RBF is used as the kernel function instead of the linear term in the linear equation, so that the non-linear regression can be achieved. At the same time, the kernel function is introduced to achieve the purposes of dimension rising, and increased adjustable parameters make the over-fitting still in control [5].

2 Experiment principle

2.1 EQUIVALENT FIGURE OF ELECTROLYTIC CAPACITOR

Figure 1 is the equivalent figure of electrolytic capacitor [6].

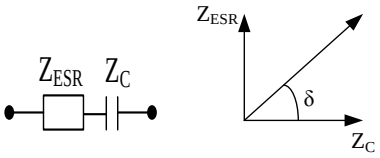


FIGURE 1 Series equivalent circuit and vector diagram

Where, Z_{ESR} is impedance of the equivalent series resistance ESR, Z_C is capacitance value of the electrolytic capacitor, and δ is the loss angle.

2.2 LOSS FACTOR OF ELECTROLYTIC CAPACITOR

Loss factor, i.e., the loss tangent, is defined as the ratio of ESR and capacitive impedance, as shown in the following Equation (1):

$$D = \tan\delta \times 100\% = \frac{ESR}{\frac{1}{\omega \times C}} = \frac{ESR}{\frac{1}{2\pi f \times C}} = 2\pi f \times ESR . \quad (1)$$

3 Experimental method and design

The experiment was carried out at room temperature. Due to room temperature for general projects, it was difficult to

* ***Corresponding author*** e-mail: chenxj@wzu.edu.cn

control the temperature. The experiment lasted one hour. The experimental subject was JAKEC electrolytic capacitor, +105°, 50V withstand voltage, 47uF. The experimental instruments used were YD2817 broadband LCR meter (to measure capacitance) and Tektronix TP2024B four-channel separated oscilloscope (to observe waveform). The frequency of ripple was 50Hz, more easily accessible, the peak value was 24V (effective value), and the ripple was not superposed on the DC component, with the effect of the DC component removed.

After each experiment, the measurement could not be done before the electrolytic capacitor was cooled to room temperature, about 2 hours. In the experiment A, five capacitors were measured for each phase, and two capacitance values were measured each time for the mean value to reduce the impact of errors. In experiment B, 20 capacitance values were measured for the mean value, and this is because as time progresses, the capacitance may vary to a small extent.

Figure 2 shows the experimental circuit. The adjusted ripple was actually the current obtained from the electric supply after transformed and half-wave rectified, whose initial phase was controlled by a single chip microcomputer. Figure 3 shows ripple voltages at different phases displayed by the oscilloscope, respectively 20°, 90° and 160°, with other circumstances omitted.

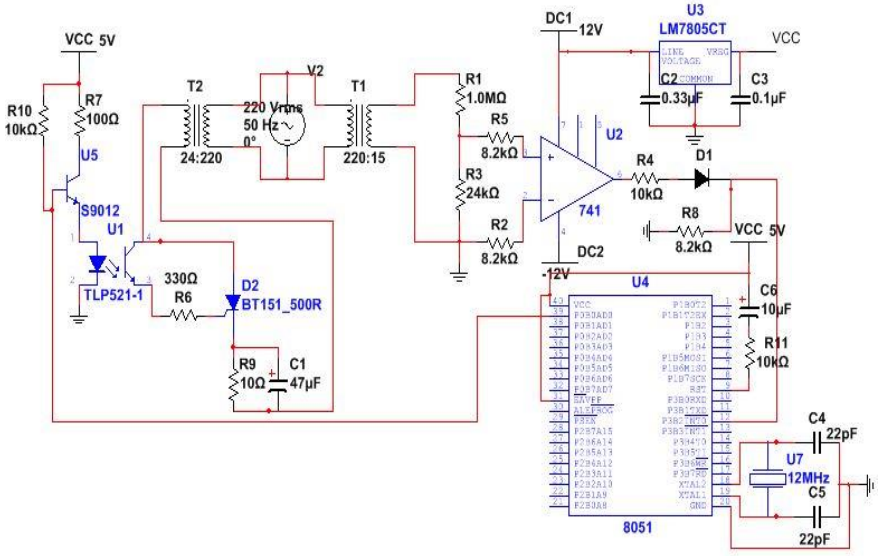


FIGURE 2 Experimental circuit diagram

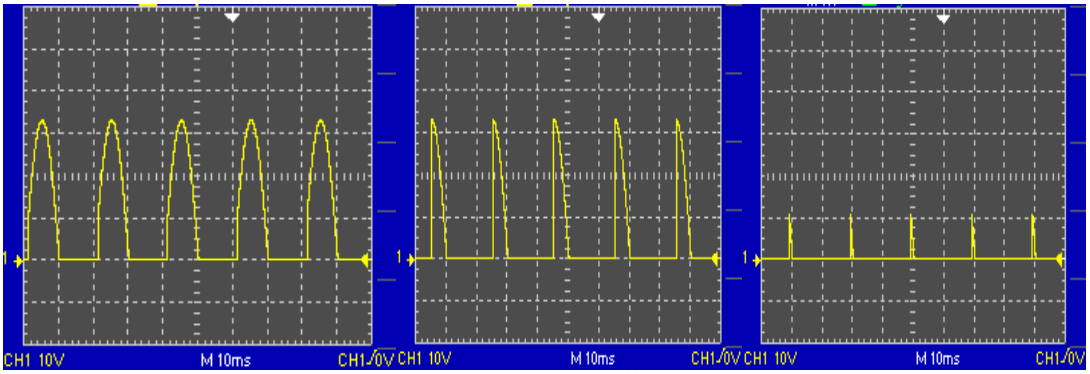


FIGURE 3 Ripple voltages at different phases

4 Results and discussion

4.1 VARIATION OF ELECTROLYTIC CAPACITOR CAPACITANCE WITH DIFFERENT PHASE RIPPLES

First, we will study the variation of electrolytic capacitor capacitance with different phase ripples, that is, loss of the electrolytic capacitor at different phases. Five capacitors were measured for each phase. The histogram (Figure 4) of the variation of electrolytic capacitor capacitance with different phase ripples can be obtained on the basis of results. Similarly, ESR of the capacitor in this case can be obtained, which is not stated here.

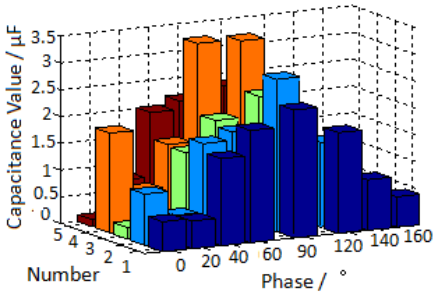


FIGURE 4 Histogram of variation of electrolytic capacitor capacitance with different phase ripples

Figure 4 shows that the ripple at 90° phase has the greatest "harm" to the electrolytic capacitor, embodied in maximum attenuation of the electrolytic capacitor of 90° in the same duration [7]. Compared to other special phases, such as 0° phase with the longest ripple duration and 160° phase with the shortest ripple duration, it is clear that at 90° phase with the highest ripple variation rate, that is, the maximum value of du/dt, rich harmonics components results in a rapid decrease of electrolytic capacitor capacitance [8, 9].

Then, we will study the ripple phase with most harms (i.e. 90° phase), and find out the variation of harm to the electrolytic capacitor with time. Figures 5 and 6 demonstrate the variation of the electrical capacitance with time.

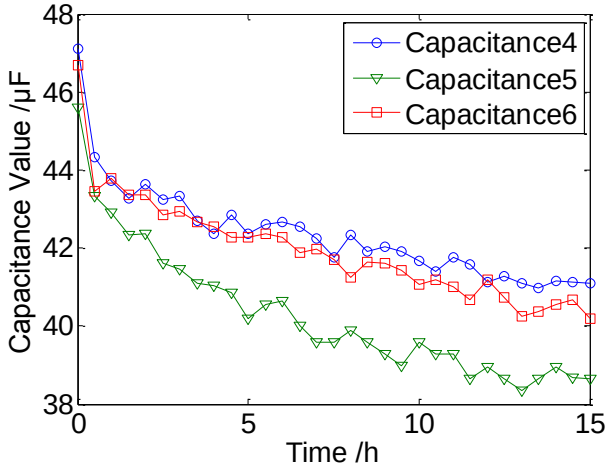


FIGURE 5 Variation of capacitance 4, 5, 6 with time (time interval of half an hour)

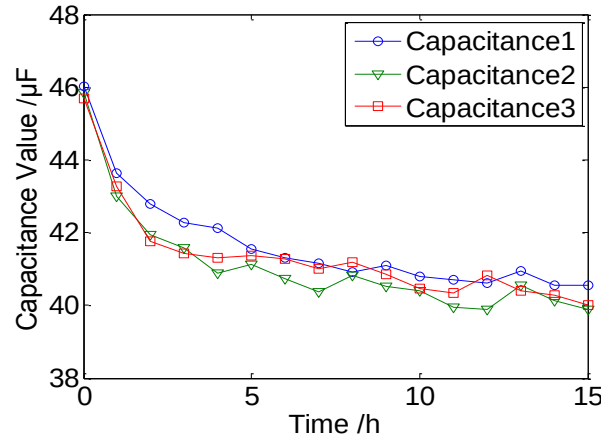


FIGURE 6 Variation of capacitance 1, 2, 3 with time (time interval of an hour)

Figures 5 and 6 show that with the ripple effect the overall trend of the electrolytic capacitor capacitance is gradual decrease, because the electrolyte repairs and thickens the oxidation film, the capacitance decreases, and ESR increases. Hydrogen generated at the cathode accelerates the volatilization of the electrolyte, resulting in degradation of the electrolytic capacitor [8]. In the early degradation, as time goes, the decrease of the capacitance becomes slow. But in the late degradation, the electrolyte has more loss and thickens, causing rise of resistivity and significantly increased loss of the capacitor [9]. But here we only study the impact of the capacitance failure in early degeneration on the circuit.

4.2 PREDICTION MODEL OF ELECTROLYTIC CAPACITOR CAPACITANCE

Figure 4 illustrates that the ripple at 90° phase has the greatest impact on the electrolytic capacitor capacitance. Figure 5 and 6 show the variation of the harm to electrolytic capacitor with time at 90° phase ripple with time, and then, we will study this trend and establish the prediction model. Capacitance 1, 2, 3 are taken as training data, Capacitance 4, 5, 6 are taken as predicted data.

First, data pre-treatment:

As in Equation (2), normalization is carried out.

$$S_{ij} = \frac{C_{ij}}{\max\limits_{j}\left(C_{ij}\right)}, \quad i = 1,2,3,4,5,6, \tag{2}$$

where, i and j are respectively the number and the moment of the capacitor sample, C_{ij} is the capacitance of the i-th capacitor at the j-th moment, $\max\limits_{j}\left(C_{ij}\right)$ is maximum value of C_{ij}, that is, the initial capacitance C_{i1} of the sample, and S_{ij} is the value of normalized C_{ij}.

As in Equation (3), equalization is carried out.

$$\bar{S}_j = \frac{\sum\limits_{i} S_{ij}}{n}, \tag{3}$$

where, n is the number of capacitor samples.

Second, as in Equation (4), by function $f(t)$ training, the data of one-hour time interval are used to train and fit the model $\bar{S}$:

$$\bar{S} = f(t), t > 0. \tag{4}$$

The training data used are $\bar{S}_j$, $j = 1, \cdots, 16$.

$\overline{s_j}$ is the value of $\bar{s}$ predicted by the fitting model at the moment j, as in Equation (5), the corresponding predicted value of the capacitor is:

$$C_{ij} = C_{i1} \times \hat{S}_j, \ i = 4,5,6, j \geq 0. \tag{5}$$

4.2.1 Neural network model

Capacitance 1, 2, and 3 are taken as modeling data, and the RBF network is applied. RBF network has two layers, the first is the implied radial basis layer, and the second is the linear output layer. Theoretically, RBF can approximate any nonlinear function.

4.2.2 KPLS model

KPLS introduces RBF as a kernel function into partial least-squares regression, maps through the kernel function the input data into high-dimensional characteristic space, and then operates in the space, so that the nonlinear relationship

before input and output variables can be established. The radial basis kernel function is selected as the kernel function.

4.2.3 SVR model

The support vector regression algorithm mainly after the dimension raising constructs in high-dimensional space a linear decision function to achieve linear regression. As e-insensitive function is used, it is mainly based on the e-insensitive function and kernel functions algorithm. If the fitting mathematical model is used to express a curve in multi-dimensional space, it, according to the results obtained by e-insensitive function, includes the curve and the "e-pipe" of the training point. In all sample points, only the sample points distributed on the "pipe wall" determine the position of the pipe. These training samples are called "support vector". To accommodate the nonlinearity of the training sample set, the traditional fitting method usually adds the higher-order term after the linear equation. Admittedly this method is effective, but the increased adjustable parameters accordingly increase risk of over-fitting. The support vector regression algorithm uses the kernel function to solve this contradiction. The kernel function is used instead of the linear term in the linear equation to make the original linear algorithm nonlinear, i.e., the nonlinear regression. At the same time, the introduction of the kernel function achieves the purpose of dimension raising, and the increase of adjustable parameters makes the over-fitting still in control. As in Equation (6), kernel functions in the above three methods are all radial basis functions and the kernel parameter is:

$$\sigma^2 = 1. \tag{6}$$

Figures 7, 8 and 9 respectively apply three prediction models of RBF, KPLS and SVR to predict the capacitance of the three capacitors of capacitor 4, 5, and 6.

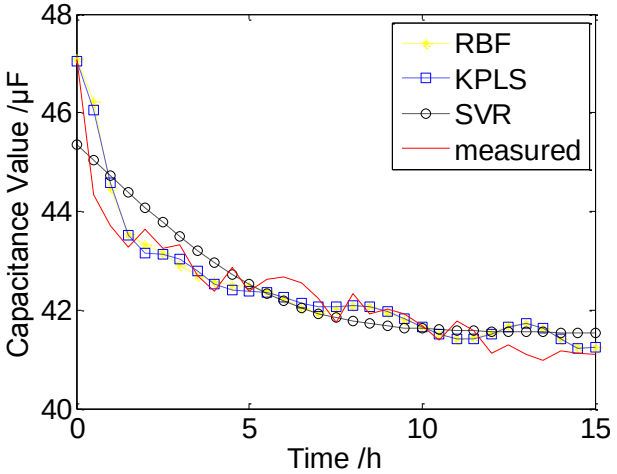


FIGURE 7 Capacitance of capacitor 4 predicted by models of RBF, KPLS and SVR

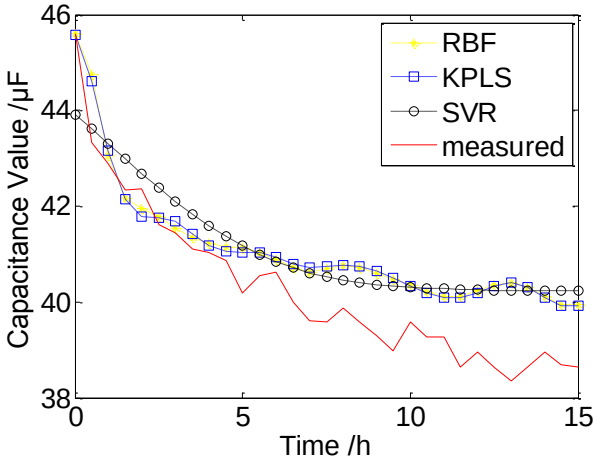


FIGURE 8 Capacitance of capacitor 5 predicted by models of RBF, KPLS and SVR

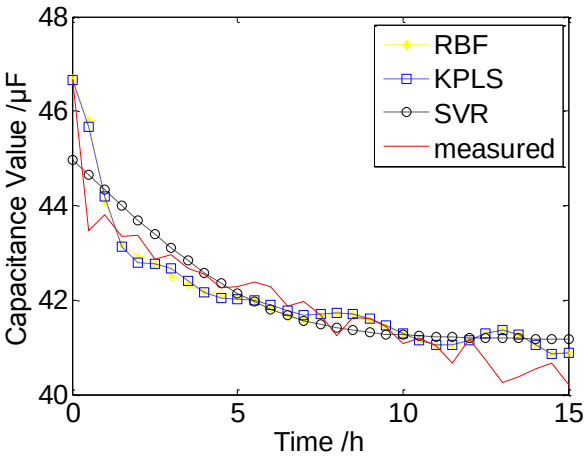


FIGURE 9 Capacitance of capacitor 6 predicted by models of RBF, KPLS and SVR

Table 1 shows mean absolute percentage error, root-mean-square error and correlation coefficient error between capacitance and measured value of the three capacitors in the three prediction models of RBF, KPLS and SVR. From Table 1, as can be seen, the root-mean-square error of KPLS is minimum.

TABLE 1 Errors of the three prediction models of RBF, KPLS and SVR

No. of capacitor	Mean absolute percentage error			Root-mean-square error			Correlation coefficient		
	RBF	KPLS	SVR	RBF	KPLS	SVR	RBF	KPLS	SVR
4	0.010	0.007	0.010	0.538	0.453	0.557	0.937	0.941	0.897
5	0.045	0.022	0.024	1.902	1.030	1.067	0.946	0.947	0.965
6	0.009	0.009	0.010	0.581	0.560	0.574	0.908	0.914	

5 Conclusion

The ripple at 90° phase has the greatest "harm" to the electrolytic capacitor, embodied in maximum attenuation of the electrolytic capacitor of 90° in the same duration. Compared to other special phases, such as 0° phase with the longest ripple duration and 160° phase with the shortest ripple duration, it is clear that at 90° phase with the highest ripple variation rate, that is, the maximum value of du/dt, rich harmonics components results in a rapid decrease of electrolytic capacitor capacitance. At the same time, we studied the curve of loss with time in case of the maximum loss. We then estimated the trend of the curve based on the above three algorithms models and found that these three models have high accuracy, KPLS is slightly better than SVR and RBF, which provides certain reference to the specific failure of the electrolytic capacitor.

Acknowledgements

The authors would like to acknowledge the financial support provided by the Program for Key Innovative Research Team of Zhejiang Province (2012R10006-0), and the Scientific Research Fund of Zhejiang rovincial Education Department (Y201223951).

Appendix A Experiment A

Table A1 Variation of electrolytic capacitor capacitance with different phase ripples

	Initial		**Final**		
Phase	*C*	*D*	*C*	*D*	*variation of C*
	46.01	0.0889	45.44	0.0961	0.56
	44.87	0.0771	43.91	0.0940	0.96
0°	45.40	0.0824	45.18	0.0863	0.22
	44.37	0.0868	42.50	0.1093	1.86
	44.35	0.0871	44.22	0.0938	0.13
	44.98	0.0751	44.46	0.0842	0.52
	45.89	0.0854	45.39	0.0907	0.51
20°	44.21	0.0849	43.86	0.0897	0.34
	45.08	0.0845	44.34	0.0898	0.73
	45.96	0.0858	45.14	0.0952	0.81
	45.63	0.0819	44.00	0.0911	1.63
	44.68	0.0860	42.89	0.0980	1.79
40°	44.78	0.0943	43.28	0.1009	1.50
	45.48	0.0830	43.94	0.0938	1.54
	45.23	0.0758	43.22	0.0960	2.02
	44.11	0.0893	42.01	0.1000	2.10
	45.40	0.0739	43.45	0.0809	1.94
60°	45.66	0.0692	43.61	0.0737	2.05
	45.33	0.0877	41.96	0.1066	3.37
	46.65	0.0731	44.48	0.0819	2.17
	46.03	0.0719	43.64	0.0747	2.39
	45.84	0.0898	42.99	0.0975	2.85
90°	45.69	0.0804	43.28	0.0825	2.41
	46.32	0.0789	42.99	0.0882	3.33
	45.28	0.0768	42.92	0.0833	2.37
	44.76	0.0934	42.88	0.0995	1.88
	46.07	0.0873	44.49	0.0951	1.59
120°	45.84	0.0882	44.49	0.0936	1.35
	45.45	0.0757	43.49	0.0882	1.96
	44.99	0.0847	43.54	0.0845	1.46
	44.99	0.0846	44.05	0.0917	0.94
	48.88	0.0624	48.05	0.0729	0.83
140°	44.98	0.0921	44.45	0.0965	0.53
	46.12	0.0890	45.47	0.0977	0.65
	45.62	0.0736	45.08	0.0786	0.53
	46.36	0.0806	45.79	0.0918	0.57
	44.99	0.1008	44.80	0.1033	0.19
160°	44.41	0.0824	44.22	0.0845	0.19
	45.68	0.0899	45.59	0.0961	0.09
	44.97	0.0757	44.77	0.0789	0.20

Appendix B Experiment B

Table B1 Variation of capacitance 1, 2, 3 with time (time interval of an hour)

	Number								
Time	*1*			*2*			*3*		
	C	*D*	*ESR*	*C*	*D*	*ESR*	*C*	*D*	*ESR*
Initial	46.03	0.0719	2.4856	45.84	0.0898	3.1197	45.69	0.0804	2.8003
1°	43.64	0.0747	2.7239	42.99	0.0975	3.6114	43.28	0.0825	3.0353
2°	42.78	0.0764	2.8438	41.96	0.1019	3.8656	41.76	0.0960	3.6606
3°	42.29	0.0742	2.7923	41.58	0.0989	3.7875	41.44	0.0929	3.5697
4°	42.12	0.0688	2.6010	40.89	0.1048	4.0817	41.30	0.0935	3.6050
5°	41.54	0.0766	2.9363	41.12	0.1015	3.9291	41.38	0.0893	3.4368
6°	41.32	0.0772	2.9751	40.73	0.1062	4.1500	41.27	0.0916	3.5328
7°	41.16	0.0804	3.1089	40.36	0.1116	4.4030	41.02	0.0965	3.7460
8°	40.92	0.0849	3.3042	40.83	0.1043	4.0677	41.20	0.0885	3.4185
9°	41.11	0.0788	3.0507	40.51	0.1088	4.2747	40.86	0.0936	3.6462
10°	40.79	0.0846	3.3026	40.42	0.1076	4.2375	40.48	0.0990	3.8948
11°	40.70	0.0856	3.3494	39.95	0.1132	4.5120	40.33	0.1009	3.9839
12°	40.60	0.0864	3.3887	39.90	0.1140	4.5482	40.83	0.0861	3.3579
13°	40.94	0.0755	2.9369	40.56	0.1034	4.0574	40.40	0.0960	3.7843
14°	40.55	0.0845	3.3186	40.14	0.1107	4.3900	40.29	0.0972	3.8401
15°	40.54	0.0843	3.3092	39.90	0.1135	4.5296	40.02	0.1023	4.0704

Table B2 Variation of capacitance 4, 5, 6 with time (time interval of half an hour)

	Number								
Time	*4*			*5*			*6*		
	C	*D*	*ESR*	*C*	*D*	*ESR*	*C*	*D*	*ESR*
Initial	47.10	0.0618	2.0893	45.61	0.0814	2.8419	46.69	0.0720	2.4556
30’	44.33	0.0849	3.0497	43.33	0.1008	3.7044	43.46	0.1038	3.8032
1°	43.71	0.0886	3.2277	42.90	0.1025	3.8046	43.80	0.0958	3.4828
1°30’	43.27	0.0904	3.3268	42.35	0.1068	4.0157	43.35	0.0968	3.5557
2°	43.64	0.0815	2.9738	42.37	0.1007	3.7845	43.37	0.0926	3.3999
2°30’	43.24	0.0843	3.1044	41.61	0.1077	4.1215	42.86	0.0979	3.6372
3°	43.32	0.0808	2.9700	41.45	0.1082	4.1566	42.95	0.0969	3.5925
3°30’	42.70	0.0892	3.3264	41.10	0.1106	4.2850	42.68	0.0982	3.6638
4°	42.37	0.0923	3.4688	41.03	0.1110	4.3079	42.55	0.0994	3.7199
4°30’	42.86	0.0842	3.1282	40.86	0.1108	4.3180	42.26	0.1009	3.8019
5°	42.38	0.0907	3.4079	40.18	0.1167	4.6249	42.28	0.1013	3.8152
5°30’	42.62	0.0849	3.1720	40.54	0.1144	4.4935	42.37	0.0959	3.6041
6°	42.67	0.0800	2.9854	40.63	0.1080	4.2327	42.28	0.0971	3.6570
6°30’	42.56	0.0799	2.9894	40.00	0.1141	4.5422	41.88	0.1022	3.8858
7°	42.24	0.0856	3.2269	39.60	0.1180	4.7449	41.96	0.1013	3.8443
7°30’	41.75	0.0923	3.5203	39.58	0.1189	4.7835	41.69	0.1044	3.9876
8°	42.34	0.0813	3.0576	39.88	0.1140	4.5519	41.24	0.1077	4.1585
8°30’	41.92	0.0878	3.3351	39.58	0.1179	4.7433	41.63	0.1047	4.0048
9°	42.03	0.0843	3.1938	39.29	0.1216	4.9282	41.61	0.1043	3.9914
9°30’	41.91	0.0852	3.2371	38.99	0.1212	4.9498	41.44	0.1053	4.0462
10°	41.68	0.0886	3.3849	39.59	0.1152	4.6335	41.08	0.1099	4.2600
10°30’	41.40	0.0928	3.5693	39.28	0.1199	4.8606	41.20	0.1100	4.2514
11°	41.77	0.0858	3.2709	39.28	0.1214	4.9214	41.02	0.1100	4.2701
11°30’	41.59	0.0878	3.3616	38.65	0.1261	5.1952	40.67	0.1143	4.4752
12°	41.12	0.0938	3.6324	38.96	0.1195	4.8842	41.18	0.1043	4.0331
12°30’	41.29	0.0907	3.4979	38.64	0.1262	5.2007	40.73	0.1132	4.4256
13°	41.11	0.0942	3.6487	38.35	0.1323	5.4933	40.26	0.1190	4.7067
13°30’	40.98	0.0956	3.7147	38.65	0.1301	5.3600	40.37	0.1183	4.6662
14°	41.17	0.0906	3.5042	38.96	0.1218	4.9782	40.54	0.1136	4.4621
14°30’	41.13	0.0901	3.4882	38.68	0.1213	4.9936	40.67	0.1108	4.3382
15°	41.10	0.0900	3.4869	38.65	0.1232	5.0758	40.20	0.1156	4.5790

References

[1] Hu X 2012 Fourier Analysis and Calculation of Electrolytic Capacitor Ripple *Skyworth Overseas Products Institute* (*in Chinese*)
[2] Xu L 2009 Modeling and Analysis of Single-phase Rectifier Capacitor Ripple Current *Nanjing University of Aeronautics and Astronautics* (*in Chinese*)
[3] Chen M 2013 MATLAB Neural Network Theory and Examples *Tsinghua University Press* 196-221 (*in Chinese*)
[4] Jia R 2013 KPLS Model based Product Quality Control for Batch Processes *School of Information Science and Engineering, Northeastern University* (*in Chinese*)
[5] Vapnik V 1995 The Nature of Statistical Learning Theory *Springer: Verlag New York Inc chapter* **6** 181-216
[6] Lin X, Hong X 2002 Aluminum Electrolytic Capacitor Engineering Technology *Xiamen University* 5-6 (*in Chinese*)
[7] Fang Y, et al 2012 Study on Ripple Rejection for Switching Power Supply *Shaanxi Province Key Laboratory of Thin Film Technology and Optical Test* (*in Chinese*)
[8] Chen Y 2003 Factors that Influence Life Test of Capacitor *Electronic Information Products Supervision Inspection Institute of Jilin Province* (*in Chinese*)
[9] Zhou H 2010 Research on Reliability of Aluminum Electrolytic Capacitor in SMPS *Harbin Institute of Technology* (*in Chinese*)

Authors

Pengchao Ye, China

Current position, grades: student of University of Wenzhou, China.
University study: Physics and Electronic Information Engineering, Wenzhou University (Communication Engineering), 2013.
Research field: information processing and testing technology.
Research direction: pattern recognition and digital signal processing.

Xiaochun Guan, China

Current position, grades: lecturer in the Wenzhou University, China.
University study: Master degree of electronic information engineering, University of Shanghai for Science and Technology in 2005.
Research activities: pattern recognition; digital signal processing; and electronic information engineering.

Xiaojing Chen, China

Current position, grades: associate professor in the Wenzhou University, China.
University study: Ph.D. degree of optical detection, Xiamen University in 2009.
Research activities: spectral analysis; pattern recognition; and digital signal processing.

Quadratic polynomial fitting of total energy of null subcarriers in underwater acoustic OFDM communication

Yang Chen[1*], Bin Zhou[2], Jingwei Yin[2]

[1]*School of Information Science and Engineering, Changzhou University, Gehu Road 1, Changzhou, China*

[2]*Science and Technology on Underwater Acoustic Laboratory, Nantong Str. 145, Harbin, China*

Received 1 March 2014, www.cmnt.lv

Abstract

Doppler scale estimation is one key module in underwater acoustic (UWA) orthogonal frequency division multiplexing (OFDM) communication. In this paper, null subcarrier based blind Doppler scale estimation is considered, which is an extremal problem. The cost function is constructed of the total energy of null subcarriers through DFT. The frequencies of null subcarriers are calculated according to non-uniform Doppler shifts at each tentative scaling factor. Then it is proofed that the cost function can be fitted as a quadratic polynomial near the global minimum. So the accurate location of the global minimum can be achieved through polynomial interpolation. This theory is a basement of developing new methods of Doppler scaling estimation and is verified through an experiment conducted in shallow water.

Keywords: orthogonal frequency division multiplexing (OFDM), underwater acoustic (UWA) communication, polynomial interpolation, Doppler scale estimation

1 Introduction

Underwater acoustic (UWA) orthogonal frequency division multiplexing (OFDM) communication has been under extensive investigation in recent years [1-5]. Unlike the radio channel, which has relative short delay spread and slow time variation, underwater acoustic channels suffer from exhibit long delay spread and fast time variation. The latter brings significant Doppler effects to UWA communication systems, so estimation of the Doppler scaling factor is of critical importance [6-9].

As an extension of the blind carrier frequency offset (CFO) estimation method [10], using the energy on the null subcarriers to find the best fit becomes a popular Doppler scaling factor estimation method in UWA OFDM communication [11-15]. The key is an extremal problem on the cost function constructed of energy of the null subcarriers. In [11] a two-step approach was used. First resample the received signal according to the Doppler scaling factor roughly measured by preamble and postamble, followed by resolution of residual Doppler, which is considered to be uniform. The rough measurement leads to a poor real-time performance and residual Doppler could not be considered to be uniform when the carrier frequency is low to achieve long range communication. In [12] the total energy of frequency measurements at null subcarriers of the block resampled with different tentative scaling factors and huge computational complexity is need. As an improvement, in [15] the cost function is sampled sporadically to find the rough place of the global minimum, and then get an accurate estimation by the method of steepest descent.

In this paper, the cost function is constructed of the total energy of null subcarriers through DFT. The frequencies of null subcarriers are calculated according to non-uniform Doppler shifts at each tentative scaling factor. DFT is more convenient and with much less computational complexity than resampling. Then the cost function is investigated and proofed that it can be fitted as a quadratic polynomials near the global minimum. So the accurate location of the global minimum can be achieved through polynomial interpolation, as a new method of Doppler scaling estimation. To verify the theory, an experiment was conducted in shallow water, whose results confirm its validity.

2 Cyclic prefix OFDM with non-uniform Doppler shifts

Let denote the OFDM symbol duration and T_g the cyclic prefix. The total OFDM block duration is $T + T_g$.

The frequency spacing is $\Delta f = 1/T$. The k -th subcarrier is at the frequency

$$f_k = f_c + k\Delta f \ , k = -\frac{K}{2}, \cdots, \frac{K}{2} - 1, \tag{1}$$

where f_c is the carrier frequency and K subcarriers are used so that the bandwidth is $B = K\Delta f$.

Consider one CP-OFDM block. Let $d(k)$ denote the information symbol to be transmitted on the k-th subcarrier. The non-overlapping sets of active subcarriers S_A and null subcarriers S_N satisfy $S_A \cup S_N = \{-K/2, \cdots, K/2 - 1\}$. The transmitted signal in passband is then given by

$$s(t) = \left[\sum_{k \in S_A} d(k) e^{jk2\pi\Delta ft} \right] e^{j2\pi f_c t} \ , t \in \left[-T_g, T \right]. \tag{2}$$

[*]***Corresponding author*** e-mail: chenyang.heu@163.com

Consider a multipath underwater channel that has the impulse response

$$c(\tau,t)=\sum_p A_p(t)\delta\left(\tau-\tau_p(t)\right), \quad (3)$$

where $A_p(t)$ is the path amplitude and $\tau_p(t)$ is the time-varying path delay. To develop our receiver algorithms, this paper adopt the following assumptions.

A1) All paths have a similar Doppler scaling factor a such that

$$\tau_p(t)\approx\tau_p-at. \quad (4)$$

In general, different paths could have different Doppler scaling factors. The method proposed in this paper is based on the assumption that all the paths have the same Doppler scaling factor. When this is not the case, part of useful signals is treated as additive noise, which could increase the overall noise variance considerably. However, we find that as long as the dominant Doppler shift is caused by the direct transmitter/receiver motion, as it is the case in our experiments, this assumption seems to be justified.

A2) The path delays $\tau_p(t)$, the gains $A_p(t)$, and the Doppler scaling factor a are constant over the block duration.

The received signal in passband is then

$$\tilde{y}(t)=\sum_p A_p\sum_{k\in S_A} d(k)e^{jk2\pi\Delta f\left(t+at-\tau_p\right)}e^{j2\pi f_c\left(t+at-\tau_p\right)}+\tilde{n}(t), \quad (5)$$

where $\tilde{n}(t)$ is the additive noise.

Base on the expression in (5), each subcarrier experiences a Doppler-induced frequency shifts $(f_c+k\Delta f)a$, which depends on the frequency of the subcarrier. Since the bandwidth of the OFDM signal is comparable to the center frequency, the Doppler-induced frequency shifts on different OFDM subcarriers differ considerably; i.e., the narrowband assumption does not hold.

3 Quadratic polynomial fitting of total energy of null subcarriers

The total energy of the null subcarriers is used as the cost function. Assume that coarse synchronization is available from the preamble. After truncating each CP-OFDM block from the received signal, CP is removed. The energy of null subcarriers whose frequency is measured according to tentative scaling factors are achieved by DFT as in (6).

$$Y\left(\gamma_\kappa\right)=\frac{1}{T}\int_0^T\tilde{y}(t)e^{-j2\pi\left(f_c+\kappa\Delta f\right)(1+a+\hat{a})t}dt,\kappa\in S_N \ , \quad (6)$$

where $\gamma_\kappa=(f_c+\kappa\Delta f)(1+a+\hat{a})$ is frequency of null subcarrier, $\tilde{a}=a+\hat{a}$ is a tentative scaling factor, a is the Doppler scale factor, $\hat{a}$ is the deviation of the Doppler scale factor and the tentative scaling factor.

The sum of the energy of null subcarriers is used as the cost function for the Doppler scale estimation.

$$J(\tilde{a})=\sum_{k\in S_N}\left|\frac{1}{T}\int_0^T e^{-j2\pi(f_c+k\Delta f)(1+\tilde{a})t}\tilde{y}(t)dt\right|^2, \quad (7)$$

$$\bar{a}=arg\ min J(\tilde{a}) \ , \quad (8)$$

where $\bar{a}$ is the estimation of a.

The Equation (6) can be transformed as follows:

$$Y\left(\gamma_\kappa\right)=\frac{1}{T}\int_0^T\sum_p A_p\sum_{k\in S_A}d(k)e^{j2\pi\left(f_c+k\Delta f\right)\left(t+at-\tau_p\right)}$$

$$e^{-j2\pi\left(f_c+\kappa\Delta f\right)(1+a+\hat{a})t}dt+\frac{1}{T}\underset{a\ll1,\left(f_c+\kappa\Delta f\right)\hat{a}T\ll1}{=}-$$

$$\frac{\left(f_c+\kappa\Delta f\right)\hat{a}T}{1+a}\sum_p A_p\sum_{k\in S_A,k\neq\kappa\pm1} \quad , \quad (9)$$

$$d(k)e^{-j2\pi\left(f_c+k\Delta f\right)\tau_p}+\frac{1}{T}\int_0^T\tilde{n}(t)e^{-j2\pi\gamma t}dt$$

$$=\alpha_\kappa\hat{a}+\beta_\kappa$$

where $\beta_\kappa=\frac{1}{T}\int_0^T\tilde{n}(t)e^{-j2\pi\gamma t}dt$,

$\alpha_\kappa=-\frac{(f_c+\kappa\Delta f)T}{1+a}\sum_p A_p\sum_k d(k)e^{-j2\pi(f_c+k\Delta f)\tau_p}$.

Then the cost function is:

$$J(\tilde{a})=\sum_{\kappa\in S_N}Y\left(\gamma_\kappa\right)Y^*\left(\gamma_\kappa\right)=$$

$$\sum_{\kappa\in S_N}\left(\alpha_\kappa\hat{a}+\beta_\kappa\right)\left(\alpha_\kappa\hat{a}+\beta_\kappa\right)^*= \quad . \quad (10)$$

$$\hat{a}^2\sum_{\kappa\in S_N}\left|\alpha_\kappa\right|^2+\hat{a}\sum_{\kappa\in S_N}\alpha_\kappa\beta_\kappa^*+\beta_\kappa\alpha_\kappa^*+\sum_{\kappa\in S_N}\left|\beta_\kappa\right|^2$$

Considering the irrelevance of signal and noise, $\sum_{\kappa\in S_N}\alpha_\kappa\beta_\kappa^*+\beta_\kappa\alpha_\kappa^*\to0$ when κ is large enough, thus $J(\tilde{a})$ becomes

$$J\left(\tilde{a}\right)=\hat{a}^2\sum_{\kappa\in S_N}\left|\alpha_\kappa\right|^2+\sum_{\kappa\in S_N}\left|\beta_\kappa\right|^2=$$

$$\left(\tilde{a}-a\right)^2\sum_{\kappa\in S_N}\left|\alpha_\kappa\right|^2+\sum_{\kappa\in S_N}\left|\beta_\kappa\right|^2 \quad . \quad (11)$$

From Equation (11) it can been seen that the cost function is a quadratic polynomial of the tentative scaling factor $\tilde{a}$, and minimized when $\tilde{a}=a$.

There are two conditions need in the derivation of (9):

$$a\ll1, \quad (12a)$$

$$(f_c+\kappa\Delta f)\hat{a}T\ll1\Rightarrow\hat{a}\ll\frac{1}{\left(\frac{f_c}{\Delta f}+\kappa\right)} \quad . \quad (12b)$$

In UWA communication, the Doppler scale factor is normally about the order of 10^{-3}, so (12a) is always meted. The condition (12b) indicates that the cost function can be fitted by a quadratic polynomials only when the tentative scaling factor being limited in a small range around the Doppler scale factor.

Based on (11), a Doppler scale factor estimating method can be proposed. First find the fitting function with several

sample of the cost function around its minimum, and then calculate the exact position of the minimum through the fitting quadratic polynomials function as the estimation of the Doppler scale factor.

4 Experimental results

An experiment was conducted in Lianhua lake of Heilongjiang province in September, 2010. The water depth was around 40 m, the transmitter was located at a depth of about 5 m and the receiver submerged at a depth of about 7 m. The receiver boat was anchored and the transmitter boat could move around. The range between the receiver boat and the transmitter boat was 2~3 km. OFDM signals were transmitted while transmitter boat was moving towards the receiver boat in the first voyage and away from in the second voyage.

The bandwidth of the OFDM signal is 4 kHz, and the carrier frequency is 6 kHz. The transmitted signal thus occupies the frequency band between 4 and 8 kHz. CP-OFDM with a CP of 85.3 ms per OFDM block. The number of subcarriers is 341. The subcarrier spacing is 11.72 Hz, and the OFDM block duration is 85.3 ms. QPSK modulation is used. Block-type pilot is adopted and a null subcarrier is inserted in every 4 subcarriers. So the number of active subcarriers is 256 and the number of null subcarriers is 85, as illustrated in Figure 1. Every frame transmitted contains 50 blocks. The transmitter is moving at a speed of up to 5 kn, at which the Doppler shifts of 8 kHz is 13.36 Hz, which is larger than the OFDM subcarrier spacing.

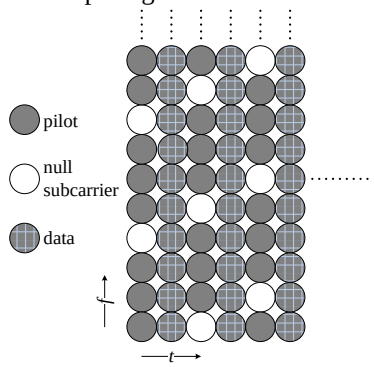


FIGURE 1 Illustration of pilot and null subcarrier pattern

Figure 2 depicts the cost function of one OFDM block. Three samples are made around the minimum, according to which a quadratic polynomials function is fitted. Zoom in Figure 2 to see the details as in Figure 3. The cost function $J(\tilde{a})$ has several minimums but just one unique global minimum, and can been fitted as a quadratic polynomial function around the global minimum. The position of the global minimum is the Doppler scale factor. Figure 4 depicts the errors between the cost functions and the corresponding fitting functions of different OFDM blocks. The error is expressed as $\left(J_p(\tilde{a}) - J_c(\tilde{a})\right)/J_c(\tilde{a})$, where $J_p(\tilde{a})$ is the quadratic polynomial function of (11) and $J_c(\tilde{a})$ is the cost function of (7). In this experiment, the area of accurate fitting is about $[-1.5 \times 10^{-4},\ 1.5 \times 10^{-4}]$, which is matched with (12*b*).

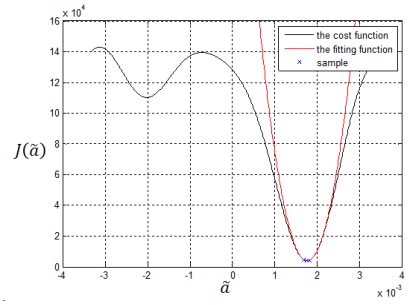


FIGURE 2 The fitting of the cost function of one OFDM block

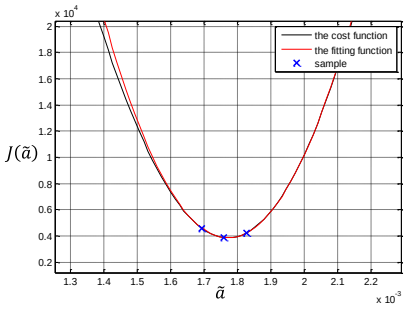


FIGURE 3 Zoom in around the global minimum

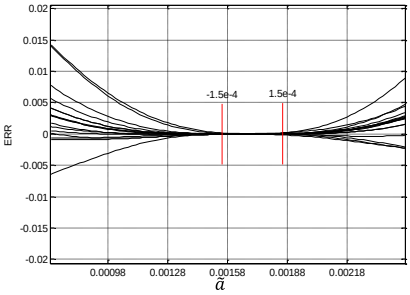


FIGURE 4 The errors between the cost functions and the corresponding fitting functions of different OFDM blocks

5 Conclusions

In this paper, a new cost function of total energy of null subcarriers measurement method is proposed in UWA channels with non-uniform Doppler shifts, which is of less computational complexity. The cost function has several minimums but just one unique global minimum whose position is the Doppler scale factor, and can been fitted as a quadratic polynomial function around the global minimum. To verify the theory, an experiment was conducted in shallow water, whose results confirm its validity.

Based on this theory, new methods of Doppler scaling estimation through polynomial interpolation will be investigated in future research, to develop simple and effective Doppler scaling estimation for UWA OFDM communication systems.

Acknowledgments

This work was supported by the State Key Laboratory of Ocean Engineering under Grant 1316, and the Natural Science Foundation of Changzhou University under Grant ZMF13020018.

References

[1] Berger CR, Chen W, Zhou S, Huang J 2010 A simple and effective noise whitening method for underwater acoustic orthogonal frequency division multiplexing ***Acoust Soc Am*** **127**(4) 358-67
[2] Wang Z-H, Zhou S, Giannakis G B, Berger C R, Huang J 2012 *IEEE J Oceanic Eng* **37**(1) 14-24
[3] Wang Z-H, Zhou S, Catipovic J, Willett P 2012 *IEEE Trans on SP* **60**(4) 1782-95
[4] Tu K 2011 Multi-carrier communications over underwater acoustic channels *PhD dissertation School of Electrical, Computer and Energy Engineering Arizona State University*
[5] Huang X 2011 Analysis and optimization of OFDM underwater acoustic communications *PhD dissertation Stevens Institute of Technology: Hoboken*
[6] Wang Z-H, Zhou S, Preisig J, Pattipati KR, Willett P 2012 *IEEE Tran on SP* **60**(6) 3079-91
[7] Mason S F, Berger C R, Zhou S, Willett P 2008 *IEEE J SEL AREA COMM* **26**(9) 1638-49
[8] Walree P A, Otnes R 2013 *IEEE J Oceanic Eng* **38**(4) 678-88
[9] Berger C R, Zhou S, Preisig J C, Willett P 2010 *IEEE Trans on SP* **58**(3) 1708-21
[10] Ma X, Tepedelenlioglu C, Giannakis G B, Barbarossa S 2001 *IEEE J Sel Area Comm* **19**(12) 2504-15
[11] Li B, Zhou S, Stojanovic M, Freitag L L, Willett P 2008 *IEEE J Oceanic Eng* **33**(2) 2078-98
[12] Wan L, Wang Z, Zhou S, Yang T C, Shi Z 2012 Performance comparison of doppler scale estimation methods for underwater acoustic OFDM *Elect. Comput. Eng.* 1-11
[13] Hwang S J, Schniter P 2008 *IEEE J. Sel. Areas Commun* **26**(9) 1674-83
[14] Mason S, Berger C R, Zhou S, Ball K, Freitag L, Willett P 2009 Receiver comparisons on an OFDM design for Doppler spread channels *Proceeding of MTS/IEEE OCEANS Conference of 2009 Bremen Germany* 1-7
[15] Yuen C H, Farhang-Boroujeny B 2013 Doppler scaling correction in OFDM *IEEE International Conference on Communications Budapest Hungary* 4713-7
[16] Berger CR, Chen W, Zhou S, Huang J 2010 A simple and effective noise whitening method for underwater acoustic orthogonal frequency division multiplexing ***Acoust Soc Am*** **127**(4) 358-67
[17] Wang Z-H, Zhou S, Giannakis G B, Berger C R, Huang J 2012 *IEEE J Oceanic Eng* **1**(37) 14-24
[18] Wang Z-H, Zhou S, Catipovic J, Willett P 2012 *IEEE Trans on SP* **60**(4) 1782-95
[19] Tu K 2011 Multi-carrier communications over underwater acoustic channels *Ph.D. dissertation, School of Electrical, Computer and Energy Engineering Arizona State University*
[20] Huang X 2011 Analysis and optimization of OFDM underwater acoustic communications *Ph.D. dissertation, Stevens Institute of Technology: Hoboken*
[21] Wang Z-H, Zhou S, Preisig J, Pattipati K R, Willett P 2012 *IEEE Trans on SP* **60**(6) 3079-91
[22] Mason S F, Berger C R, Zhou S, Willett P 2008 *IEEE J SEL AREA COMM* **26**(9) 1638-49
[23] Walree P A, Otnes R 2013 *IEEE J Oceanic Eng* **38**(4) 678-88
[24] Berger C R, Zhou S, Preisig J C, Willett P 2010 *IEEE Trans on SP* **58**(3) 1708-21
[25] Ma X, Tepedelenlioglu C, Giannakis G B, Barbarossa S 2001 *IEEE J Sel. Area Comm* **19**(12) 2504-15
[26] Li B, Zhou S, Stojanovic M, Freitag L L, Willett P 2008 *IEEE J Oceanic Eng* **33**(2) 2078-98
[27] Wan L, Wang Z, Zhou S, Yang T C, Shi Z 2012 *Elect Comput Eng* 1-11
[28] Hwang S J, Schniter P 2008 *IEEE J Sel Areas Commun* **26**(9) 1674-83
[29] Mason S, Berger C R, Zhou S, Ball K, Freitag L, Willett P 2009 *Proceeding of MTS/IEEE OCEANS Conference of 2009 Bremen Germany* 1-7
[30] Yuen C H, Farhang-Boroujeny B 2013 *IEEE International Conference on Communications Budapest Hungary* 4713-7
[31] Berger CR, Chen W, Zhou S, Huang J 2010 ***Acoust Soc Am*** **127**(4) 358-67
[32] Wang Z-H, Zhou S, Giannakis G B, Berger C R, Huang J 2012 *IEEE J Oceanic Eng* **1**(37) 14-24
[33] Wang Z-H, Zhou S, Catipovic J, Willett P 2012 *IEEE Trans on SP* **60**(4) 1782-95
[34] Tu K 2011 Multi-carrier communications over underwater acoustic channels *PhD dissertation, School of Electrical, Computer and Energy Engineering Arizona State University*
[35] Huang X 2011 Analysis and optimization of OFDM underwater acoustic communications *PhD dissertation, Stevens Institute of Technology Hoboken*
[36] Wang Z-H, Zhou S, Preisig J, Pattipati KR, Willett P 2012 *IEEE Trans on SP* **60**(6) 3079-91
[37] Mason S F, Berger C R, Zhou S, Willett P 2008 *IEEE J SEL AREA COMM* **26**(9) 1638-49
[38] Walree P A, Otnes R 2013 *IEEE J Oceanic Eng* **38**(4) 678-88
[39] Berger C R, Zhou S, Preisig J C, Willett P 2010 *IEEE Trans on SP* **58**(3) 1708-21
[40] Ma X, Tepedelenlioglu C, Giannakis G B, Barbarossa S 2001 *IEEE J Sel Area Comm* **19**(12) 2504-15
[41] Li B, Zhou S, Stojanovic M, Freitag L L, Willett P 2008 *IEEE J Oceanic Eng* **33**(2) 198-209
[42] Wan L, Wang Z, Zhou S, Yang T C, Shi Z 2012 Performance comparison of doppler scale estimation methods for underwater acoustic OFDM *Elect Comput Eng* 1-11
[43] Hwang S J, Schniter P 2008 *IEEE J Sel Areas Commun* 26(9) 1674-83
[44] Mason S, Berger C R, Zhou S, Ball K, Freitag L, Willett P 2009 Receiver comparisons on an OFDM design for Doppler spread channels *Proceeding of MTS/IEEE OCEANS Conference of 2009 Bremen Germany* 1-7
[45] Yuen C H, Farhang-Boroujeny B 2013 Doppler scaling correction in OFDM *IEEE International Conference on Communications Budapest Hungary* 4713-7

Authors

Yang Chen, China

Current position, grades: lecturer in Changzhou University, China.
University study: PhD degree in underwater acoustic engineering, Harbin Engineering University, Harbin, China, 2010.
Research activities: communication signal processing, array signal processing, underwater acoustic theory.
Professional Activities and Memberships: reviewer of journal of systems engineering and electronics.

Bin Zhou, China

Current position, grades: PhD student in Harbin Engineering University, China.
University study: PhD student in underwater acoustic engineering, Harbin Engineering University, Harbin, China, since 2011.09.
Research activities: communication signal processing, array signal processing, underwater acoustic system.

Jingwei Yin, China

Current position, grades: professor in Harbin Engineering University, China.
University study: PhD degree in underwater acoustic engineering, Harbin Engineering University, Harbin, China, 2007.
Research activities: underwater acoustic communication signal processing, underwater acoustic metering
Professional Activities and Memberships: reviewer of journal of systems engineering and electronics.

Research of the vibrating infrasonic sensor based on Fiber Bragg Grating

Haohua Liu, Zhou Wan*, Yiyang Li, Zhunen Chen

College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

Received 1October 2014, www.cmnt.lv

Abstract

Current research on infrasound more extensive and in-depth, the sensor is essential prerequisite for infrasonic detection, in order to overcome some lack of current infrasound sensor for detecting the infrasound, paper innovative design a vibrating wire sensor structure, and make it capable of receiving a full rang infrasound, range adjustable, easy installation, etc. and use fiber Bragg grating which has obvious advantages in terms of low-frequency detection as the sensing element. Test the performance of the sensor and the results show that the average of the sensor is 6.11%, within 1 Hz-20 Hz range, the sensor has good linearity and small error, has practical value.

Keywords: Fiber Bragg grating; vibration wire; infrasonic sensor

1 Introduction

In the acoustic frequency, frequency of less than 20Hz of the band called infrasonic wave. Infrasound has the particularity of low frequency, long wavelength, and has spread far, strong penetrating power, small interference characteristics. Infrasound exists widely, for example: the storm at sea, earthquake, volcano eruptions and other natural disasters are likely to produce infrasonic wave [1]. Our life is often accompanied by infrasonic wave, for example: Ohashi Ko, car racing, even audio and mixer at hone will produce [2] infrasound wave. At present there are many countries are devoted to the study of infrasonic wave, infrasound weapons, infrasonic wave exploration, infrasonic wave forecasting and warning are the recent hot, there must be foresee that future of infrasonic wave will have broader applications. Therefore, there will be great practical value and scientific significance to detect infrasound, existing for infrasound detection of infrasonic sensor have some shortcomings such as low sensitivity, frequency range is small, large volume, inconvenient installation and high demands on the environment, the fiber grating sensor has high sensitivity, wide dynamic range, without electromagnetic interference, corrosion resistance, small volume, light weight and other advantages [3]. Based on this, this paper designs a vibrating sound sensor based on fiber Bragg grating.

2 The measuring principle

2.1 THE SENSING PRINCIPLE OF FIBER BRAGG GRATING

Fiber grating wavelength of reflection or transmission peak concern with the refractive index modulation of the grating period and the core refractive index, and when a broadband light source is incident to the Bragg grating, refraction, transmission or reflection due to refractive index change occurs, the reflection need to meet Bragg condition, namely the reflection wavelengths of light to meet the optical [4] equations:

$$\lambda_B = 2\Lambda n_{neff} . \tag{1}$$

Type (1): λ_B as the center wavelength of fiber Bragg grating; n_{eff} as the core region of the fiber refractive index; Λ as the Bragg grating period.

By the Equation (1) shows, the Bragg center wavelength of fiber Bragg grating λ_B changes with the change of the n_{eff} and Λ, or change any physical quantity of n_{eff} or Λ will lead to reflection or transmission peak wavelength of fiber Bragg grating drift. Through the detection of Bragg wavelength reflection or transmission spectrum peak wavelength, also is the detection of the central fiber Bragg grating wavelength, can detect the corresponding variables.

When the fiber grating under strain and strain is uniform, on one hand, external strain induced grating period changes [5]:

$$\Delta\Lambda / \Lambda = \varepsilon . \tag{2}$$

On the other hand, photoelastic effect caused by the change of effective refractive index:

$$\Delta n_{eff} = \frac{n^3_{eff}[P_{12} - \nu(P_{11} + P_{12})]}{2}\varepsilon . \tag{3}$$

Type (3), P_{11} and P_{12} are the fiber optic strain tensor components, ν is the Poisson's ratio.

Changes of these two aspects led to the change of the wavelength of the fiber grating, by Equation (1) can be obtained:

* ***Corresponding author*** e-mail: liuhaohua@126.com

$$\Delta\lambda_B = 2n_{eff}\Delta\Lambda + 2\Delta n_{eff}\Lambda . \quad (4)$$

The types (2) and (3) into type (4) have

$$\Delta\lambda_B = 2n_{eff}\Lambda\{-\frac{1}{2}n_{eff}{}^2[P_{12} - \nu(P_{11} + P_{12})]\}\varepsilon + 2n_{eff}\Lambda\varepsilon . \quad (5)$$

Valid elastic optic constants of optical fiber P_e is defined:

$$P_e = \frac{n^2{}_{eff}}{2}[P_{12} - \nu(P_{11} + P_{12})] . \quad (6)$$

Then type (5) can be simplified as:

$$\Delta\lambda_B = (1 - P_e)\lambda_B\varepsilon = K_\varepsilon\varepsilon = K_\varepsilon\frac{\Delta\Lambda}{\Lambda} . \quad (7)$$

Equation (7) expresses the relationship between the wavelength shift and the external when the fiber Bragg grating strains uniformly. In Equation (7), K_ε is the strain sensitivity coefficient of fiber grating sensor. Its value is closely related to the size of the optical fiber material, by the decision of the effective refractive index of optical fiber, elastic-optic coefficient and Poisson's ratio. Therefore, When the materials of the fiber is determined, K_ε is a constant so the change of the fiber grating wavelength is determined by the change of the fiber Bragg grating period, in the other word, which is determined by fiber Bragg grating lattice spacing changes. When the lattice spacing changes at a high frequency, the change frequency of the center wavelength is also high. But the rapid changes of the center wavelength are not conducive to the detection. On the contrary, for small frequency strain, the change of the fiber Bragg grating lattice spacing frequency is small; its centre wavelength change frequency is also small, which is easier to be tested. Therefore, optical fiber Bragg grating has obvious advantages in detection of low frequency and ultra-low frequency.

2.2 WORKING PRINCIPLE OF THE SENSOR

As designed in this paper, the Vibration Wire infrasound sensor based on the fiber Bragg grating, using vibrating string as a receiving component, can receive a full range of infrasound and vibration. The vibration is transmitted to the elastic diaphragm through the string bridge. Elastic diaphragm produces strain due to the vibration, driven the fiber Bragg grating wavelength attached to the elastic diaphragm to shift.

At this point, the vibration is converted into the center wavelength shifting of the fiber Bragg grating. The center wavelength shift values of the fiber Bragg grating are obtained by the use of demodulation instrument, thus the corresponding relation of the vibration frequency and wavelength. In addition, the effective length of the vibrating wire is determined by two string bridges, one of the two can move within a certain range, so that the effective length of the vibrating string is adjustable, so the range of the designed sensor in this paper is adjustable.

3 Structure and design of the sensor

The structure of the sensor is shown in Figure 1, In order to be able to receive the full range of infrasound, this design uses a using vibrating string as a receiving component. Vibrating strings are fixed by the clamp device on both ends of the bracket. In order to make the range of the designed sensor adjustable, the effective length of the vibrating string must be adjustable. The vibrating wire length between the two string bridges is the effective length of the vibrating wire. So install a guide rail on the base plate of the right said of the fixed bracket to make the string bridge on the right side move on the guide rail to transform the effective length of the vibrating string. The string bridge on the left side can transmit the vibration of the vibrating wire. So, install the left side string bridge on center of the elastic diaphragm and paste the fiber Bragg grating under the elastic diaphragm then design a circular section on the left side of the fixed bracket to fix the elastic diaphragm.

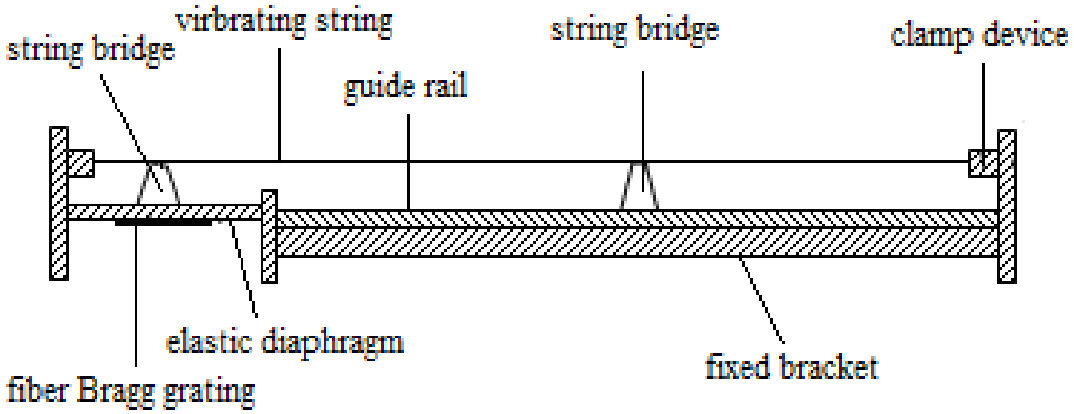


FIGURE 1 The structure of the vibrating wire infrasound sensor based on fiber Bragg grating

For the fixed strings at each end, the correspondent nth vibration displacement generated by receiving the infrasound of a series of frequencies is [6]:

$$y_n = B_n \sin\frac{n\pi}{l} x\cos(2\pi f_n t - \varphi_n) . \quad (8)$$

In Equation (8), B_n and φ_n are the correspondent undetermined coefficients of the n-th vibration. f_n is the frequency of the n-th vibration, and l is the length of the vibrating wire.

The fundamental frequency vibration is the most stable form of vibration. That is, in Equation (1), when n=1, the frequency is:

$$f = \frac{1}{2l}\sqrt{\frac{T}{\delta}} = \frac{1}{2l}\sqrt{\frac{E\varepsilon}{\rho}} \ . \tag{9}$$

In Equation (9), T - vibrating string tension, δ - vibrating wire density, $\delta = \rho s$, ρ means the vibrating wire material density, s means the cross-sectional area of the vibrating string, t - time, l - vibrating string length, E - vibrating wire elastic modulus; ε - the internal stress of vibrating. By increasing the length of the vibrating string, increasing the radius of vibrating strings can reduce the frequency of vibrating wire, thereby increasing sensor's sensitivity to low-frequency. At the same time, we can also know that the bigger elastic modulus of the material is helpful to improve the sensitivity of the sensor, Therefore, we chose the experimental calculated Young's modulus of E=200Gpa, a density of ρ=OCr$_{18}$Ni$_9$ 7.93g/cm3, the chord diameter of 1.20mm, the chord length is 1000mm stainless steel vibrating string.

By the Equation (9) shows that the longer vibrating wire means the better characteristics of low-frequency, but it is represented by the Equation (8) shows that the longer vibration string will get very small strain when receiving the low frequency infrasound. So it is not suitable for the fiber Bragg grating pasted directly on the vibrating string, In this paper an innovative structural design makes vibrating accept vibration, the vibration through the bridge to the elastic diaphragm. This put the vibration focus on the elastic diaphragm, the diaphragm can produce relatively large strain, to facilitate detection.

While the elastic diaphragm not only played the role of isolation and protection, but also increase the damping effect of sensor system. An ideal sensor should be an ideal linear time invariant system, namely the sensor output can be linear to truly reflect the input signal. Specifically, should make the output amplitude is to maintain a constant for all the frequency of the measured signal, to ensure that the harmonic components are the same amplification, and the output phase is linear to the all measured frequency, to ensure the phase shift that each harmonic component is directly proportional to the frequency. This requires sensors with larger bandwidth, and the bandwidth depends largely on the system's natural frequency and damping ratio. When the damping ratio is too large, with the increase of frequency amplitude frequency characteristic curve decreased rapidly, while the damping ratio is too small, with the increase of frequency amplitude frequency characteristic curve will rise, in the above two cases, the amplitude frequency characteristics flat areas are narrow, therefore, appropriate damping ratio can make the sensor has a wide operating frequency range. The sensor is designed in this paper, and the effective length of string vibration adjustable makes the sensor range adjustable, thus the need for a larger operating frequency range. After testing, this design uses the polyester film as diaphragm can enable the sensor has the appropriate damping ratio and wide operating frequency. The elastic diaphragm is round, the radius of R=120mm, thickness of h=0.3mm. The bridge is made of wooden materials, the bottom surface of radius r0=30mm.

According to the principle of the elastic mechanics, when the uniform pressure of perpendicular to the circular diaphragm clamped films, film will produce bending deformation. The center (R=0) of the maximum deformation, the deformation of the diaphragm can be expressed as a Equation:

$$Y_{\max} = \frac{3(1-\mu^2)\mathrm{P}}{16Eh^3}R^4. \tag{10}$$

In the Type (10), E, μ, h and R respectively refers to Young's modulus of the polyester film, Poisson's ratio and thickness and the effective radius.

Analysis of elastic diaphragm were carried out by the finite element software ANSYS [7, 8], circular flat diaphragm radial strain distribution in the diameter direction of the Figure 2:

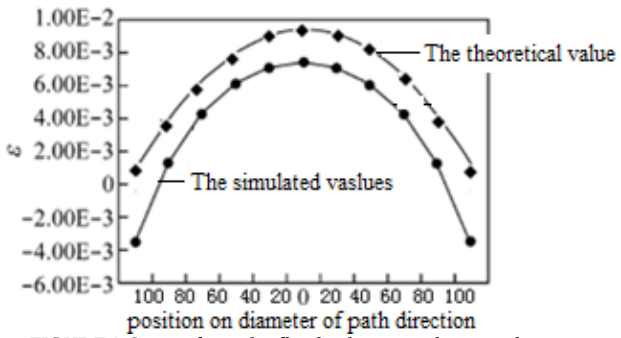


FIGURE 2 Strain of circular flat diaphragm in diameter direction

We can see from Figure 2 strain at the center of the diaphragm is maximum, and at about 1/2 in radius radial strain have another turning point, combine type (10), so as to determine the paste position of fiber Bragg grating on the membrane should be at 1/2 in the diaphragm to the radius of the center. Figure 3 is the schematic diagram of circular flat diaphragm sticking position of fiber Bragg grating.

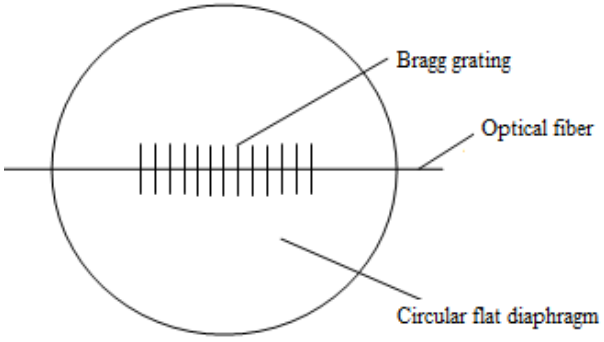


FIGURE 3 Paste position of fiber Bragg grating in the round flat diaphragm

4 Test data and analysis

Produce 0-20Hz sine excitation signal by signal generator, amplified by the power amplifier and speaker produce infrasound wave, the cut-off frequency to the 21Hz test will be low pass filter, 10 tests were carried out for each test frequency, test results, by the Equations (11) and (12) to calculate the average value, the result in Table 1.

$$\bar{f} = \frac{1}{n}\sum_{1}^{n} f_i , \tag{11}$$

$$\bar{\lambda} = \frac{1}{n}\sum_{1}^{n} \lambda_i . \tag{12}$$

TABLE 1 Output frequency of the sensor and fiber Bragg grating center wavelength with Different input frequency

The given input frequency M(Hz)	The average value of measured output f(Hz)	The center wavelength of fiber Bragg $\lambda(\mu m)$	The given input frequency M(Hz)	The average value of measured output f(Hz)	The center wavelength of fiber Bragg $\lambda(\mu m)$
0	0	1.137236	5.00	5.42	1.137236
0.01	0	1.137205	6.00	6.44	1.137241
0.05	0.04	1.137206	7.00	7.56	1.137244
0.10	0.09	1.137207	8.00	8.36	1.137247
0.20	0.19	1.137208	9.00	9.47	1.137251
0.30	0.25	1.137208	10.00	10.21	1.137255
0.40	0.37	1.137209	11.00	11.17	1.137258
0.50	0.52	1.137211	12.00	11.56	1.137261
0.60	0.63	1.137214	13.00	12.53	1.137265
0.70	0.74	1.137216	14.00	13.80	1.137270
0.80	0.83	1.137218	15.00	14.55	1.137274
0.90	0.94	1.137220	16.00	15.69	1.137279
1.00	1.12	1.137223	17.00	16.42	1.137283
2.00	2.32	1.137226	18.00	17.79	1.137289
3.00	3.35	1.137230	19.00	18.80	1.137294
4.00	4.16	1.137233	20.00	19.74	1.137298

From Table 1 can draw the diagram of the relationship between each given frequency and the correspond center of fiber Bragg grating wavelength, as shown in Figure 4:

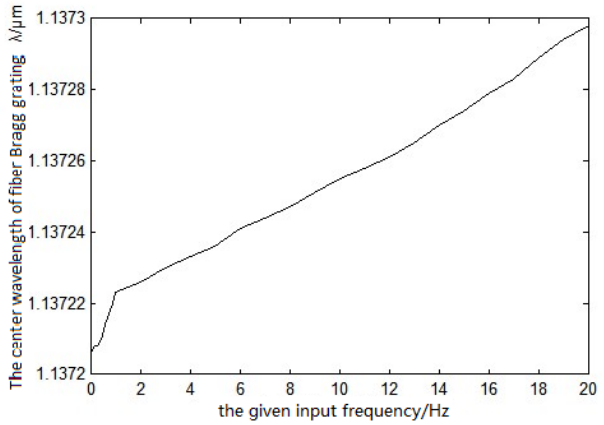


FIGURE 4 Corresponding relations between the given input frequency and the center wavelength about fiber Bragg grating

We can see from Figure 4, in the range of 0 Hz-1 Hz relationship of wavelength and frequency is relatively large, and in the 1 Hz-20 Hz range, relationship between wavelength and frequency are g in a straight line that in this range the sensor has good linearity.

According to Table 1 can also draw the input frequency and the measured output frequency contrast diagram, as shown in Figure 5:

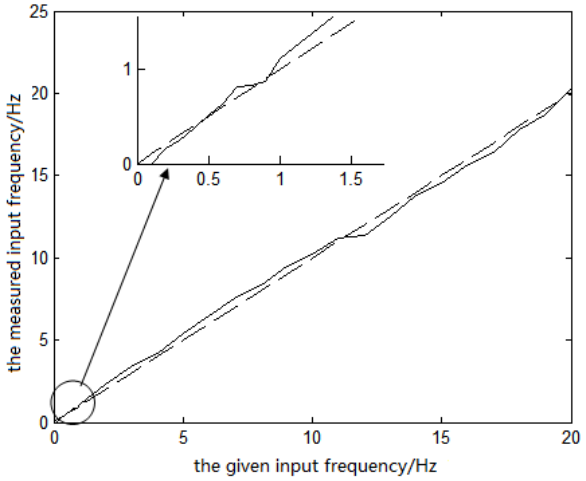


FIGURE 5 The comparison chart between the input frequency (dashed line) and the actual output frequency (solid line)

The Equation of relative error of E:

$$e_i = \frac{f_i - M_i}{M_i} \cdot 100 . \tag{13}$$

Relative errors of given input frequency and the measured frequency is shown in the table below:

TABLE 2 The relative error of the given input frequency and the measured output frequency

The given input frequency M(Hz)	The average value of measured output f(Hz)	Relative error (%)	The given input frequency M(Hz)	The average value of measured output f(Hz)	Relative error (%)
0	0	--	5.00	5.42	0.084
0.01	0	--	6.00	6.44	0.073
0.05	0.04	-0.200	7.00	7.56	0.080
0.10	0.09	-0.100	8.00	8.36	0.070
0.20	0.19	-0.050	9.00	9.47	0.045
0.30	0.25	-0.167	10.00	10.21	0.021
0.40	0.37	-0.075	11.00	11.17	0.015
0.50	0.52	0.040	12.00	11.56	-0.037
0.60	0.63	0.050	13.00	12.53	-0.036
0.70	0.74	0.057	14.00	13.80	-0.014
0.80	0.83	0.038	15.00	14.55	-0.030
0.90	0.94	0.044	16.00	15.69	-0.019
1.00	1.12	0.120	17.00	16.42	-0.031
2.00	2.32	0.160	18.00	17.79	-0.012
3.00	3.35	0.117	19.00	18.80	-0.011
4.00	4.16	0.040	20.00	19.74	-0.013

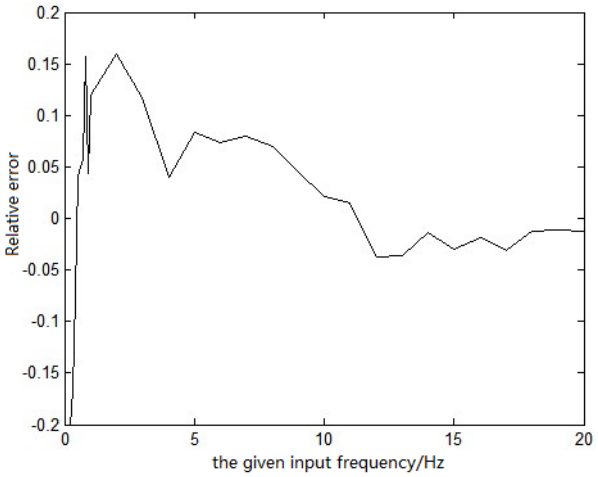


FIGURE 6 Curves of relative error

According to the Equation (13), using the data in Table 2 dawn relative error curves plotted as in Figure 6. We can see from Figure 6, with the increase of the test frequency, the error will be gradually reduced, the average error is E=6.11%. According to Figure 4 and Figure 6 shows, the sensor in the range of 0 Hz-1 Hz designed in this paper, the measurement results is not very stable, the error is large, but in the 1 Hz-20 Hz range, the sensor has good linearity and small error, the result show that the design of the sensor is of practical value.

5 Conclusions

This paper design the vibrating string type infrasonic sensor based on fiber Bragg grating, the sensor using optical fiber Bragg grating as the sensitive element and structure design innovation make the sensor has the advantages of convenient installation, electromagnetic interference resistance, suitable for high temperature, corrosive environment, also can receive infrasound in full range, the measuring range can be adjusted according to the actual situation, so as to improve the sensitivity of the sensor, avoid reducing the sensitivity by using the same range to measuring different frequencies , analysis the test data, and the results show that the sensor error is small, high precision, and has great practical value.

Reference

[1] Chen X 2011 Talk to you Infrasound *The middle School Math* **8** 54-5
[2] Han B, Wang Y, Meng F 2012 Design of pressure sensor based on PVDF piezoelectric materials *Journal of Jilin University (Natural Science Edition)* **3** 333-6
[3] Wang Y, Liang D, Zhou B 2008 Monitor of Corrosion about concrete structure analysis based on fiber Bragg grating spectrum *Spectroscopy and Spectral Analysis* **11** 2660-3
[4] Li T, Li C, Liu H 2012 Method of test Cable optical and the end of the cable anchorage stress measurement based on fiber Bragg grating *Chinese Journal of Highway* **5** 134-8
[5] Zhou J 2010 Study on transmission characteristics of fiber grating sensors *Wuhan University of Science and Technology*
[6] Yang B, Zhang Y 2003 Engineering elastic-plastic Mechanics *Beijing: Machinery Industry Press*
[7] Shi F 2012 Two Development and Application Examples about ANSYS *Beijing: Chinese water conservancy and Hydropower Press*
[8] Wu Z, Cui T 2006 Analysis of clamped circular diaphragm elastic characteristics based on ANSYS technology *Research and Application of Mechanical* **2** 47-54

Authors

Haohua Liu, China

Current position, grades: Graduate student of Kunming University of Science and Technology, China.
University study: Bachelor degree in Measurement and control technology and instrument, Zhongyuan University of Technology, 2008; Master's degree in Instrument and Meter Engineering, Kunming University of Science and Technology jointly, 2012.
Research activities: biomechanical study of sensor; research on human-computer interaction technology based on force sense; new sensor technology

Zhou Wan, China

Current position, grades: associate professor of Kunming University of Science and Technology, China.
University study: Bachelor degree in automatic control and systems engineering, Kunming University of Science and Technology,1984; Master's degree in Automation of Tsinghua University and Kunming University of Science and Technology jointly, 1997
Research activities: biomechanical study of sensor; research on human-computer interaction technology based on force sense; new sensor technology
Professional Activities and Memberships: Yunnan Province Association of automation (promote the development of automation in Yunnan province)

Yiyang Li, China

Current position, grades: Graduate student of Kunming University of Science and Technology, China.
University study: Bachelor degree in Measurement and control technology and instrument, Yunnan University, 2008; Master's degree in Instrument and Meter Engineering, Kunming University of Science and Technology jointly, 2012.
Research activities: biomechanical study of sensor; research on human-computer interaction technology based on force sense; new sensor technology

Zhunen Chen, China

Current position, grades: Graduate student of Kunming University of Science and Technology, China.
University study: Bachelor degree in Measurement and control technology and instrument, Nanchang University of Aeronautics and Astronautics, 2007; Master's degree in Instrument and Meter Engineering, Kunming University of Science and Technology jointly, 2012.
Research activities: biomechanical study of sensor; research on human-computer interaction technology based on force sense; new sensor technology

The error analysis of TDOA based ultrasonic position

Ang Gao[1], Yansu Hu[2*], Weijun Duan[1]

[1]*School of Electronics and Information, Northwestern Polytechnical University, No.127, West Youyi Road, Xi'an, China*

[2]*School of Electronics and Control Engineering, Chang'an University, Southern Middle Section of the Second Circular Road, Xi'an, China*

Received 1 March 2014, www.cmnt.lv

Abstract

How to enhance the accuracy and range is a typical issue in the ultrasonic position technology. The position accuracy depends on the hardware architecture and the detecting methods. This paper designs a hardware platform called Hexagon-Buck with six transceivers around. For the inherent property of TDOA (Time of Difference of Arrival) ranging, the RF synchronous time and embedded software processing time are two inevitable factors that leads to the distance error. At the same time, the ultrasonic velocity is sensitive to the environment temperature. The paper considers both the temperature compensation and Least Squares Linear Regression to enhance the ranging accuracy. By taking the physical test data to identify the linear regression model, we can estimate the real distance. The validity of the method has been proved by the final experiment.

Keywords: ultrasonic ranging, TDOA, least squares linear regression, hexagon-buck

1 Introduction

Each node how to locate themselves in the large sensor network systems has become a challenging subject. According to the measuring distance and the angle information, the wireless location technology can be classified into two types: the range-based technology and the range-free technology. Although range-free technology has advantages in the power consumption and hardware cost, its measured accuracy can not meet the requirement of WSN (Wireless Sensor Networks), such as DV-Hop [1] and APIT [2]. On the contrary, range-based technology like TOA (Time of Arrival) [3], TDOA (Time of Difference of Arrival) [4, 5], AOA(Angle of Arrival) [6, 7], RSSI(Received signal strength indication) [8, 9] and TOF(Time of Fly) [10] is more commonly used in engineering.

TDOA technology is used to range the distance between the sender and receiver by the time difference, which has been widely proposed as a necessary ingredient in WSN self localization systems. Compared with other range technologies, TDOA works without the network time synchronization, which can significantly reduce the resource calculation cost and hardware expenditures. In order to ensure the ranging accuracy, the paper takes TDOA to calculate the distance between two nodes and adopts the Least Squares Linear Regression model to revise the measured error. The hardware called Hexagon-Buck with six transceivers around the board is also implemented to test the ranging performance.

The ultrasonic based TDOA leads to the nodes distance but not the nodes position. This paper also discusses the multi-dimensional scaling positioning algorithm based on the nodes' distance matrix [11]. Theoretically, if the nodes ranging is zero-defection, then the position is zero-error. So we are pursuing an error eliminating method, and concerns how to enhance the ranging accuracy by means of software algorithm optimizing as well as the hardware improvement.

2 System architecture

In this paper, the Hexagon-Buck which is a novel ultrasonic platform featured with six transceivers is presented. The hardware architecture is shown as Figure 1.

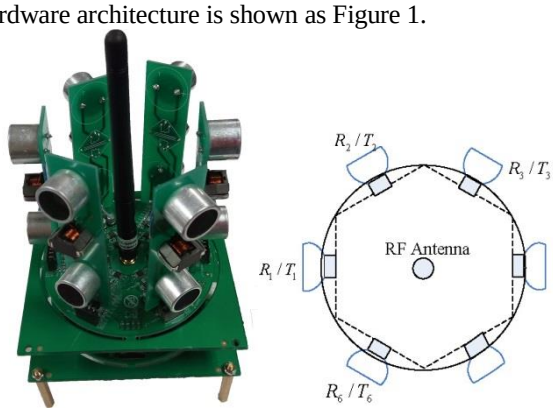


FIGURE 1 Hexagon-Buck platform with six transceivers around (the board has a diameter of 8 centimeters)

Since the radio transmitting time is measured by the local clock of the host node, and the arrival time is calculated on the ultrasound board, so it is important that the clock drift between the two microcontrollers keeps minimal. Thus the time synchronization among the sending nodes and the receiver nodes is necessary. For our hardware platform, the ad-hoc network is constructed by two layers: the physical layer with IEEE 802.15.4 and the network layer with Zigbee. The core of Hexagon-Buck is a dedicated micro-controller to control the ultrasonic transmission, receive signals, as well as manage the Zigbee stack and network application. The MCU of Hexagon-Buck is JN5148 produced by NXP. It is combined with an IEEE802.15.4 compliant transceiver on chip. JN5148 also integrates a temperature sensor. By providing the absolute measure of

[*]***Corresponding author*** e-mail: huyansu@163.com

the device temperature, we can easily amend the sound velocity according to Equation (3).

Typical ultrasonic transceiver is directional. In order to realize the omni-directional detection, we equip six transceivers around to solve the limited beam angle problem, just like the previous platforms (e.g. the SpiderBat nodes [7] and the AHLoS nodes [12]). 40C16T/R-1 Ultrasonic Sensor produced by OSENON is designed as the transceiver. The centre frequency is 40kHz and the bandwidth is 2kHz. The ultrasonic transmission is cone shaped with the angle of $60\pm15^{\circ}$. So six 40C16T/R-1 transceivers can cover 360° omni-direction. Further more, the separate ultrasound transmitters and receivers can reduce the complexity without switching the circuits between receive and transmit mode.

1) The transmission circuit is driven by the square wave with 50% duty ratio and 40kHz frequency. The square wave is operated by the Pulse Width Modulation (PWM) of the micro controller. DC-DC module as power amplifier provides $5V$ and $\pm10V$ voltage to 40C16T/R-1.

2) Since the receiver signal is only a few millivolt, it needs to be amplified for reliable detection. As shown in Figure 2, the first amplification stage provides 40dB amplification. The second amplification stage equipped with a digital potentiometer is to adjust the detection threshold and prevent the saturation of the sampled signal. This detection signal is then to compare with a reference voltage, which is also can be adjusted by a digital potentiometer. If the signal exceeds the reference voltage, the comparator will trigger an electrical level change captured by the micro controller, which leads to hardware interrupt.

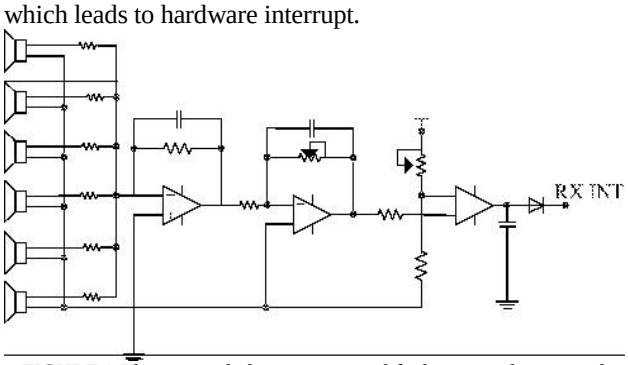


FIGURE 2 The received ultrasonic is amplified twice and compared with a reference voltage

3 Temperature and frequency feature of ultrasonic

3.1 TEMPERATURE FEATURE

Like other acoustic wave, the ultrasonic velocity is about $340.2\,m/s$ in the air at the condition of normal pressure and temperature (NPT). But actually it is affected by the status of transmission medium such as the moisture, pressure and density. Compared with the moisture and pressure, sound velocity is more sensitive to the temperature. The propagation velocity in normal atmosphere can expressed as bellow:

$$v(t)=\sqrt{\gamma RT/\mu},T=273+t, \tag{1}$$

where $\gamma=1.4$ is the air specific heat ratio. R is the perfect gas constant, and the value in the air is $R=8.134kg\cdot mol^{-1}\cdot K^{-1}$. $\mu=0.00283kg\cdot mol^{-1}$ is the air molar mass and t is the centigrade temperature. The Taylor Series Expansion of Equation (2) is as bellow:

$$v(t)=\sum_{n=0}^{\infty}\frac{v^{(n)}(0)}{n!}t^n=v(0)+v'(0)t+\frac{1}{2!}v''(0)t^2+\cdots+\frac{1}{n!}v^{(n)}(0)t^n+\frac{1}{(n+1)!}v^{(n+1)}(\xi)t^{n+1}, \tag{2}$$

where $\xi\in(0,t)$. Omit the higher order term:

$$v(t)\approx331.5+0.607t(m/s). \tag{3}$$

According to Equation (3), the velocity increases about $0.607\,m/s$ along with every additional centigrade. Later we will take Equation (3) to correct the ultrasonic velocity.

3.2 FREQUENCY FEATURE

On one hand, the sound intensity is damping along with the transmission for the energy keeps decreasing during the wave diffusion, reflection and scattering. When the frequency is higher, the intensity declines faster.

On the other hand, the directivity of ultrasonic also changes with the frequency, which is described by the index of half-power beam width (the angle of two half-power point with -3dB damping). The sound energy dumps along with the axial direction and fades out on both sides. If the frequency is higher, the half-power beam width is smaller and the directivity is better.

So the balance between a better directivity and a lower sound intensity decline must be considered when choose the ultrasonic frequency during the hardware design. This paper selects 40 kHz ultrasonic for its better robust and anti-interference.

4 Accuracy deterioration factors

The ultrasound ranging is unidirectional. The sender node initiates a single measurement by broadcasting a radio packet and indicates a time stamp. While the nodes nearby listen for the incoming ultrasound waves. During the procedure, there are several factors which may cause the accuracy deterioration.

4.1 SYNCHRONOUS TIME ERROR

In Figure 3, the coordinator and router are taken as the sender and receiver respectively to show the error schematic. t_{rf} is the RF transmission time and δt_{rp} is the synchronization signal processing time at the receiver. After receiving a synchronization signal, the receiver sends an ACK signal back to the sender and starts a hardware timer to count the ultrasonic fly time. As soon as the sender node gets this ACK, it starts a PWM wave to drive 40C16T/R to generate

an ultrasonic. The processing time at the sender is marked as δt_{sp}. The measured fly time of ultrasonic counted by receiver is marked as t_{uf}, and the theoretical fly time is t'_{uf}.

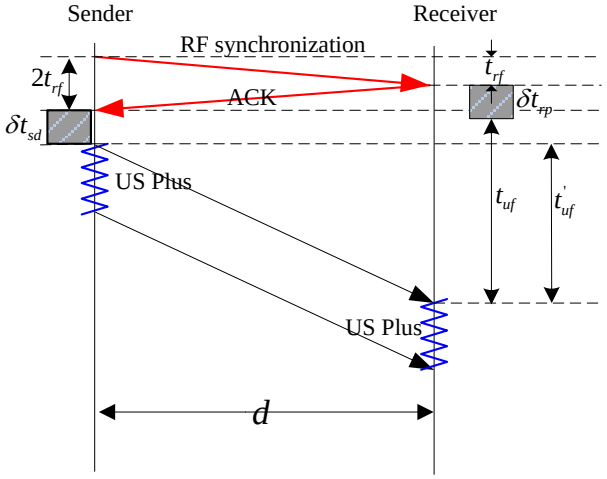


FIGURE 3 Schematic of time synchronization error

As shown in Figure 3, there is the equation: $2t_{rf} + \delta t_{sp} + t'_{uf} = t_{rf} + \delta t_{rp} + t_{uf}$.

Define $\Delta t = t'_{uf} - t_{uf}$, then the Equation above can be rewritten as:

$$\Delta t = \delta t_{rp} - \delta t_{sp} - t_{rf} . \quad (4)$$

Because of the tremendous transmission velocity of RF signals and the span limitation of $20m$, its maximum transmission time is only $0.07us$, during which the ultrasonic only travels $0.0024cm$. So t_{rf} can be omitted during the calculation. But no matter the sender node or the receiver node, the software procession time and hardware start-up time are inevitable and immeasurable. So t'_{uf} must be estimated by t_{uf}:

$$t'_{uf} = t_{uf} + \Delta t \approx t_{us} + (\delta t_{rp} - \delta t_{sp}) . \quad (5)$$

4.2 ULTRASONIC DETECTION ERROR

At the receiver node, the ultrasonic should firstly be transmitted into the voltage signal and then be amplified and filtrated. Next the voltage signal will be compared with a specified threshold to identify if the ultrasonic has reached the receiver node. When the voltage amplitude is greater than the threshold, the comparator triggers a hardware interrupt to tell the MCU that an ultrasonic pulse has been detected and ready to enter the interrupt service routine.

The ultrasonic detection error is because there is time delay t_{sw} before the ultrasonic amplitude reaches the threshold. This delay changes along with the ranging distance. For example, this time is nearly $112.3us$ in our platform and the corresponding distance is about $3.8cm$.

4.3 THE SHAPE ERROR OF HEXAGON-BUCK

As previous mentioned Hexagon-Buck has six parts of transceivers around. In Figure 4, d' is the theoretically transmitting distance. But we need the central to central distance d in practice. So the shape caused error is: $d_{hw} = d - d'$.

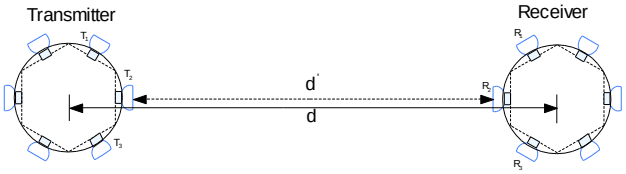


FIGURE 4 Schematic of shape caused error

4.4 ENVIRONMENT EFFECT

If the environment moisture is random distribution in 10~90%RH, the maximum normalized velocity error is only 1.5×10^{-3}, which means that there is only $3cm$ bias at the maximum measure range (about $20m$). This accuracy is enough for the implementation, so need no compensate.

As we discussed in section 3.1, the ultrasonic velocity is more sensitive to the temperature. Define Δv_{temp} be ultrasonic velocity error caused by the temperature. In the following section, the velocity compensation is only based on temperature without considering the environment moisture.

5 Ranging errors correction

5.1 TEMPERATURE COMPENSATION

Use the following equation to calculate the ultrasonic velocity compensation for the distance measurement:

$$d^* = \left(v + \Delta v_{temp}\right) \cdot \left(t_{us} + (\delta t_{rp} - \delta t_{sp})\right) + d_{hw} . \quad (6)$$

Combined the Equation (5), there is: $d^* = v_{temp}\Delta t + d_{hw}$.

The experiments are designed to test the ranging performance of Hexagon-Buck with only temperature compensation. The distance moves every $1m$ and is corrected according to Equation (6). Take the average value of 50 data at each distance. Figure 5 shows the results of ranging error.

The maximum range is about $15.8m$ and the ranging data is stable at $13m$. The average error is about $28.6cm$ and the extreme is $33.6cm$. The phenomenon is because there is no consideration on the RF synchronization error t'_{uf}, the ultrasonic detection error t_{sw}, and the shape error d_{hw}. In order to enhance the accuracy, use Least Square method to correct to error affected by other factors.

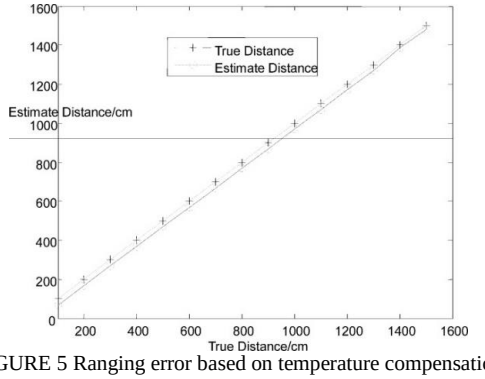


FIGURE 5 Ranging error based on temperature compensation

5.2 LSF CORRECTION

The ranging data correction is a typical liner regression problem: $D_n(x) = \mathbf{AX} = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$.

Define $\{(x_i, D_n(x_i)) \mid i = 1, 2 \ldots m\}$ to approximate the actual ranging data $\{(x_i, y_i) \mid i = 1, 2 \ldots m\}$, where x_i is the measured ranging data corrected by the temperature compensation, and y_i is the actual distance between the sender and receiver. Considering the CPU processing ability, take first-order least square regression to estimate the value of constant a_0 and a_1.

$$E = \sum_{i=1}^{m} (y_i - a_1 x_i - a_0)^2 . \tag{7}$$

In order to minimize the variance E, calculating the partial differential of Equation (7):

$$\begin{cases} \dfrac{\partial E}{\partial a_1} = -2\sum_{i=1}^{m} x_i (y_i - a_1 x_i - a_0) = 0 \\ \dfrac{\partial E}{\partial a_0} = -2\sum_{i=1}^{m} (y_i - a_1 x_i - a_0) = 0 \end{cases} .$$

Solve the equation, there is:

$$a_0 = \frac{\sum_{i=1}^{m} x_i \sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i}{\left(\sum_{i=1}^{m} x_i\right)^2 - m\sum_{i=1}^{m} x_i^2}, \tag{8}$$

$$a_1 = \frac{\sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i - m\sum_{i=1}^{m} x_i y_i}{\left(\sum_{i=1}^{m} x_i\right)^2 - m\sum_{i=1}^{m} x_i^2}, \tag{9}$$

where x_i is the average value of 50 measured data at specific distance, and the distance is increased every 1*m*. Because the maximum range is about 15.8*m*, so the regression only effective in 15*m*, etc. *m*=15. The regression model expression is:

$$D(x) = 0.9922 d^* + 0.347 . \tag{10}$$

Figure 6 is the comparison of the measured distance calculated by regression model and the actual distance. The average bias is only about 3*cm*.

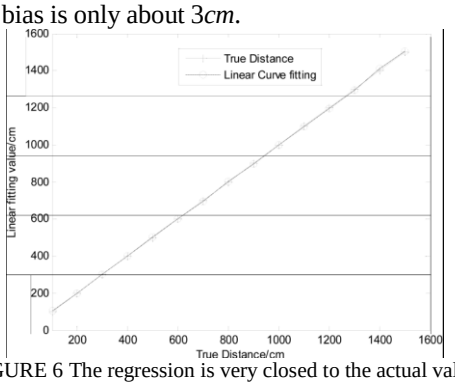


FIGURE 6 The regression is very closed to the actual value

5.3 EXPERIMENT VALIDATION

A further experiment is implemented to evaluate the validity of the linear regression model with the ultrasonic velocity compensation. The experiments are operated twice: first only with the temperature compensation (UN-Calibration) and second considering both the temperature compensation and LSF correction (Calibration). The ranging is also in the range of 15*m* and changes every 1*m*. Figure 7 shows the results. In the condition of only velocity calibration, the average bias is 29.8*cm* and the extreme is 35.3*cm*. However, it is noteworthy that the LSF correction with velocity calibration can greatly reduce the ranging error. The average bias is only 3.5*cm* and the extreme is 8.4*cm* (at the distance of 9*m*).

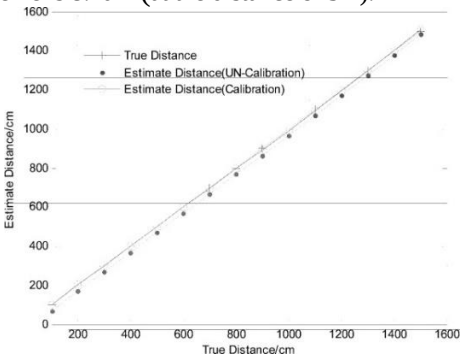


FIGURE 7 The ranging error correction with two methods

6 Conclusions

The paper designs and implements the Hexagon-Buck to enforce the node ranging and position. Considering the ultrasonic velocity error caused by the temperature and the RF synchronous time error caused by TDOA, the paper adjusts temperature compensation and least squares linear regression model to enhance the ranging accuracy. First design the physical experiment and liner regression model to match the unknown parameter, and then calculate the real distance. The experiments show the validity of this method.

Beyond these, the further study will continue consider the impact of ranging error distribution as well as the WSN nodes distribution to the position estimating.

Acknowledgments

This work was supported by Chinese Universities Scientific Foud (GEKY8004, GEKY9004), the Young Scientists Fund of the National Natural Science Foundation of China (61203233) and the China Postdoctoral Science Foundation (2013M540772).

References

[1] Bulusu N, Heidemann J, Estrin D 2000 *Personal Communications IEEE* **7**(5) 28-34
[2] Wang Z J, Jin H 2009 Improvement on apit localization algorithms for wireless sensor networks *International Conference on Networks Security Wireless Communications and Trusted Computing NSWCTC'09* **1** 719-23
[3] Mogi T, Ohtsuki T 2008 Toa localization using rss weight with path loss exponents estimation in nlos environments *14th Asia-Pacific Conference on Communications APCC* 1-5
[4] Girod L, Estrin D 2001 Robust range estimation using acoustic and multimodal sensing *Proceedings of 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems* **3** 1312-20
[5] Yang L, Ho K 2009 *IEEE Transactions on Signal Processing* **57**(12) 4598-615
[6] Steinbauer M, Molisch A F, Bonek E 2001 The doubledirectional radio channel *Antennas and Propagation Magazine IEEE* **43**(4) 51-63
[7] Oberholzer G, Sommer P, Wattenhofer R 2011 Spiderbat: Augmenting wireless sensor networks with distance and angle information *10th International Conference on Information Processing in Sensor Networks (IPSN)* 211-22
[8] Alippi C, Vanini G 2006 A rssi-based and calibrated centralized localization technique for wireless sensor networks *Proc IEEE Int Conference on Pervasive Computing and Communications Workshops (PERCOMW)* 301-6
[9] Sugano M, Kawazoe T, Ohta Y, Murata M 2006 Indoor localization system using rssi measurement of wireless sensor network based on zigbee standard *Target* **538** 50
[10] Pettinato P, Wirstrom N, Eriksson J, Voigt T 2012 Multichannel two-way time of flight sensor network ranging *Wireless Sensor Networks Springer* 163-78
[11] Liang H, Fubao W, Weijun D, Chao M, Guoqiang Y 2013 The multidimensional scaling positioning algorithm for wireless sensor networks based on distance reconstruction *Chinese Journal of Sensor and Actuators* **26**(9) 1284-7
[12] Savvides A, Srivastava M B 2002 A distributed computation platform for wireless embedded sensing *Proceedings 2002 IEEE International Conference on Computer Design VLSI in Computers and Processors* 220-2

Authors

Ang Gao, China

Current position: lecturer of Northwestern Polytechnical University, China.
University study: PhD in control science and engineering, Northwestern Polytechnical University, Xi'an, China, 2010.
Research activities: WSN, location, web cluster, network control.
Professional Activities and Memberships: Chinese society of Astronautics from 2011.

Yansu Hu, China

Current position: lecturer Chang'an University, China.
University study: PhD in control science and engineering, Northwestern Polytechnical University, Xi'an, China, 2012.
Research activities: WSN, cloud computing, network control.

Weijun Duan, China

Current position: professor of Northwestern Polytechnical University, China.
University study: MD. degree in Signals and Systems, Northwestern Polytechnical University, Xi'an, China, 1989.
Research activities: WSN, signal processing.

The study and design of inter-harmonic detection device based on quasi-synchronous technique in power system

Zheng Fan[1] Xianfeng Zheng[2*]

[1]*Faculty of Electrical Engineering Henan Mechanical and Electrical Engineering College, Xinxiang, Henan, China*

[2]*Faculty of Automatic Control Engineering Henan Mechanical and Electrical Engineering College, Xinxiang, Henan, China*

Received 6 February 2014, www.cmnt.lv

Abstract

On the basis of analyzing the characteristics of harmonic signals, the inter-harmonic detection algorithm of quasi-synchronous sampling is introduced in this paper. This algorithm pre-processes the original sampled signal and separates all the harmonics and inter-harmonics of the reconstructed signal with designed comb FIR filters, to restrain the mutual interference between them, and finally achieve accurate measurement of harmonic and inter-harmonic in power system. At last, the validity of this algorithm is verified with harmonic and inter-harmonic signal generated from standard signal source.

Keywords: quality of electric energy, inter-harmonics, quasi-synchronization, comb FIR filter

1 Introduction

With higher demand on power quality, the inter-harmonics in power quality monitoring is more and more concerned. Nevertheless, due to the uncertainty of inter-harmonic frequency and much smaller amplitude with respect to fundamental and harmonic waves, the spectrum leakage in the process of harmonic signal sampling may cause serious impact on inter-harmonic measurement, or even annihilating of inter-harmonic signal in serious case.

Thus, the current sophisticated algorithms for harmonic detection (such as FFT, etc.) can hardly locate inter-harmonics in the spectrum accurately, and even mistake spectrum leakage or noise interference as inter-harmonics [1, 2].

It is difficult to achieve the desired complete synchronous sampling in practice considering the fluctuation of fundamental wave and non-integer ratio between harmonic frequency and fundamental frequency in actual electric power signal. However, in actual asynchronous sampling process, the reciprocal serious interference among sub-harmonics and neighbor inter-harmonics on the spectrum can influence final measure results. Therefore, the harmonics and inter-harmonics of signal must be first separated, then, their respective parameters are calculated and the parameters of harmonics and inter-harmonics are obtained accurately.

In order to reduce mutual interference between harmonics and inter-harmonics, the quadratic measurement can be adopted to increase accuracy, i.e. the entire measurement process is divided into two stages: first, the harmonic component is extracted from original signal and parameters are calculated; then, the various parameters of inter-harmonics are calculated. In this method, because the harmonics interference has been wiped out before calculating parameters of inter-harmonic signal, so the accuracy of inter-harmonic measurements can be effectively increased [3].

The method of reducing mutual interference through separating harmonics and inter-harmonics of signal, can increase the accuracy to some extent, but the ignorance of inter-harmonics' influence on harmonics when harmonic signals are extracted directly and subtracting them from original signal makes this method incomplete. Due to the mutual interference between harmonics and inter-harmonics, a better way to calculate their parameters should be to separate them precisely from the original signal before calculating their respective parameters [4, 5].

Undoubtedly use comb filter can be ideal separation between the harmonic and harmonic signal, but the comb filter is suitable for the synchronous sampling signal processing, in order to enable it to application in asynchronous sampling condition, must be the actual asynchronous sampling signal accurate synchronization processing to meet the needs of a comb filter for synchronous sampling.

2 Detection model of quasi-synchronous sampling and inter-harmonic

Synchronous sampling requires that the number of signal cycles T_0 contained N sample sequences must be integer, namely: $N \cdot T_S = P \cdot T_0$.

Wherein, T_S is the sampling period; for synchronous sampling, P is an integer; in the asynchronous sampling, P is not an integer. And comb filter requires p to meet integer condition, T_S can be adjusted to make the equation take holds, set the adjusted quasi-ampling period λ_s satisfies [6]:

$$\lambda_s = \frac{P \cdot T_0}{\mathrm{K}}, \tag{1}$$

wherein, K is the sampling points in P sampling periods, usually an integer. According to Equation (1), quasi-sampling period λ_s can be calculated. Then the interpolation

[*] ***Corresponding author*** e-mail: hnxxzxf@126.com

calculation is carried out among K sampling points with equal interval λ_s, a group of new signal sample sequence $K\lambda_s$ with a total length of K can be obtained. P sampling periods are contained in this sequence, so it is the desired quasi-synchronous sampling sequence, and DFT operation or some other operations applied will not result in spectral leakage.

In this case, by simply setting the space width of comb filter according to different situations (usually the fundamental frequency), the harmonic and inter-harmonic signal components can be separated from signal.

3 Measurement of fundamental frequency from inter-harmonic signal

In accordance with IEC61000-4-7, due to the uncertainty of inter-harmonic frequency, at least 10 cycles of fundamental signal sequence are analysed in order to obtain a more accurate calculation value of fundamental cycle. Meanwhile, in order to ensure that the sample sequence after the reconstruction is no shorter than the sum of the length of both the comb filter and the analysis window, 20 cycles of sample sequence will be taken and it's mean value will be operated to obtain a more accurate value of the fundamental cycle [6-8].

$$\mathrm{T}^*_{\mathrm{MEAN}} = \frac{1}{20}[(k_{n+20} - k_n)T_s + (I_{n+20} - I_n)]. \tag{2}$$

Quasi-sampling period is: $\lambda_s = \dfrac{20T^*_{mean}}{L^*}$.

The reconstitution of sampling results after Newton interpolation is:

$$\mathrm{P}(m) = x(g_u T_s) + \tau_m \Delta f_0 + \frac{\tau_m(\tau_m - 1)}{2!}\Delta^2 f_0 + \cdots + \frac{\tau_m(\tau_m - 1)\cdots(\tau_m - 4 + 1)}{4!}\Delta^4 f_0 . \tag{3}$$

4 The design of equiripple comb FIR filter

Comb filters are generally divided into FIR and IIR filters. In the same filter performance parameters requirement, although IIR filter has lower order, its nonlinear and stability of phase characteristics does not meet the requirements of the inter-harmonic measurement; therefore, comb FIR filter will be chosen [9].

Supposing FIR filter with order Mi, if Mi is odd, then the transfer function can be expressed as:

$$H(z) = \sum_{m=0}^{2K} h(n)z^{-m} = z^{-K}[h(K) + 2\sum_{K=1}^{2K} h(K-m)\frac{1}{2}(z^m + z^{-m})] . \tag{4}$$

According to the first species of Chebyshev polynomials $T_n(\omega) = \cos(n \cdot \arccos(\omega))$, $\omega \in [-1,1]$ $\omega = \left(z + z^{-1}\right)/2$, then:

$$\mathrm{T}_{\mathrm{n}}(\omega) = \frac{z^n + z^{-n}}{2}, \tag{5}$$

$z = e^{j\omega}$.

The transfer function can be reduced as

$$H(z) = z^{-k}\sum_{m=0}^{k} a(m)T_m(\omega) . \tag{6}$$

Considering the impact of the filtering effect and filter order on the computational efficiency, design parameters of comb FIR filter with inter-harmonic detection algorithm will be selected. As mentioned earlier, first, the sample sequence with length of 20 fundamental cycles is selected; after reconstructing the sampling sequence, fundamental cycle is determined as 50.1 Hz, and sampling points 102. Then, the filter order of Chebyshev polynomial series is obtained through experiment. The amplitude-frequency response of designed comb FIR filter is shown as Figure 1 [10-12].

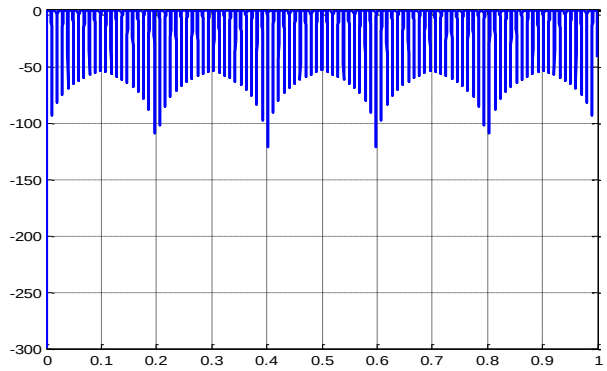


FIGURE 1 Amplitude-frequency Response of Comb FIR

In the harmonic measurement, impulse response effect needs to be considered, the impact of the filter impulse response is shown in Figure 2 [13, 14].

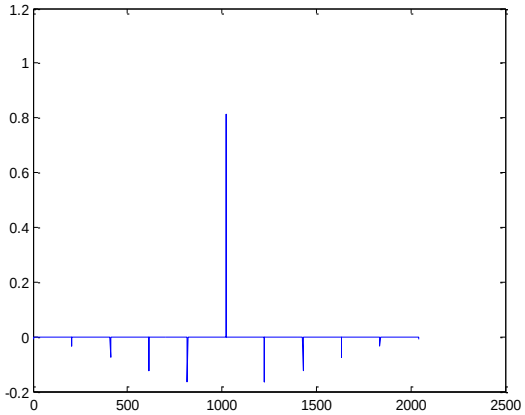


FIGURE 2 Impulse Response of Comb FIR Filter

5 Hardware structure diagram of harmonic and inter-harmonic analyzer

Hardware Structure Diagram of Harmonic and Inter-harmonic Analyser are shown in Figure 3. Three parts are

mainly included: data sampling module, the data processing module data management module [15] based on DSP and ARM.

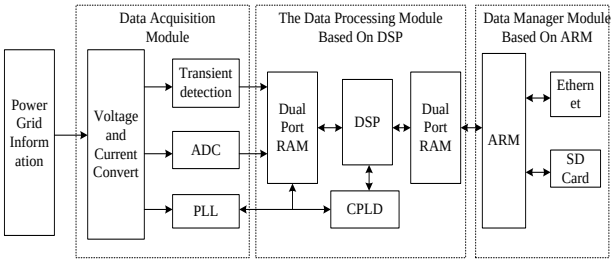


FIGURE 3 Hardware Structure Diagram of Harmonic and Inter-harmonic Analyzer

Among them, the data sampling module includes voltage and current transmitter, PLL, analog to digital converter (ADC) and transient detection and other components, which is mainly responsible for collecting original signal; data processing module is in charge of processing the collected power system voltage, current data, and transient information, meanwhile calculating power quality parameters, which is the core part of analyzer; data management module implements the function of data storage, management, human-computer interaction and supporting network communication.

For inter-harmonic detection, it needed to separate harmonics from inter-harmonics with comb FIR filter, which is a relatively large amount of data processing, and due to the power quality analyser's high requirements of accuracy, the high-performance TMS320C6713 DSP chip produced by TI will be chosen as the arithmetic unit to improve system performance.

In order to achieve inter-signal synchronization between phase position and frequency, a phase locked loop (PLL) circuit is used to automatically adjust the sampling frequency so, as to correspond to the synchronous measurement.

6 Experimental results

In order to carries on the analysis comparison, using the device on the analog signal test and analysis. The experiment uses FLUKE 6100A standard electrical power to produce a distorted signal as a test signal. The actual output signal is as shown in Equation (7):

$$x(t) = 10sin(49.9\omega t) + 0.6sin(99.8\omega t + 0.3) + 0.05sin(149.7\omega t - 0.15) + 0.2sin(75.5\omega t) + 0.05sin(154.5\omega t) \quad (7)$$

In the above Equation, the first three are the fundamental wave and second and third harmonics, the last two are inter-harmonics. The fifth inter-harmonic and third harmonic frequency are closer, and are prone to mutual interference, which can cause some difficulties to accurate measurement of inter-harmonics. The specific waveform is shown in Figure 4:

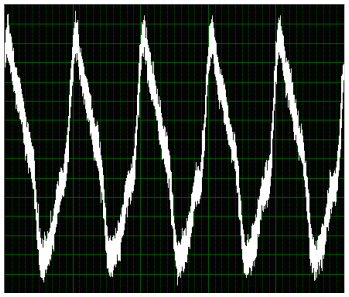


FIGURE 4 The measured signal waveform

Under the proposed IEC standard, the analysis window length should be 200 ms (approximately 10 cycles). The sampling frequency of inter-harmonic and harmonic analyser is the fixed value of 10240 Hz, which is 200 times of fundamental wave (about 50 Hz) in order to avoid aliasing, while ensuring high harmonic detection accuracy.

In addition, the analyser can upload the collected original signal to a computer to facilitate the comparison with a variety of other algorithms and calculation results. In this experiment the original signal is compared with the FFT and second-level WIFFTA methods. The results are in Tables 1-3.

TABLE 1 Results of different algorithm - FTP

Actual value		**FFT**			
		Frequency		*Amplitude*	
Frequency	*Amplitude*	*Measured value*	*Deflection*	*Measured value*	*Deflection*
49.9	10	50	0.1	9.99	0.01
99.8	0.7	100	0.2	0.75	0.05
149.7	0.05	150	0.3	0.07	0.02
75.5	0.1	75	0.5	0.15	0.05
154.5	0.05	-	-	-	-

TABLE 2 Results of different algorithm - Two-level WIFFTA Method

Actual value		**Two-level WIFFTA Method**			
		Frequency		*Amplitude*	
Frequency	*Amplitude*	*Measured value*	*Deflection*	*Measured value*	*Deflection*
49.9	10	49.9	0	9.99	0.01
99.8	0.7	99.78	0.02	0.7	0
149.7	0.05	150.9	1.2	0.07	0.02
75.5	0.1	75.48	0.02	0.11	0.01
154.5	0.05	156.9	2.4	0.02	0.03

TABLE 3 Results of different algorithm - Quasi-synchronous sampling algorithm

Actual value		**Quasi-synchronous sampling algorithm**			
		Frequency		*Amplitude*	
Frequency	*Amplitude*	*Measured value*	*Deflection*	*Measured value*	*Deflection*
49.9	10	49.9	0	10.0	0
99.8	0.7	99.8	0	0.7	0
149.7	0.05	149.68	0.02	0.05	0
75.5	0.1	75.5	0	0.1	0
154.5	0.05	154.65	0.15	0.04	0.01

7 Conclusions

Judging from the analysis, traditional FFT method is affected by spectral leakage, and accurate parameters of harmonics and inter-harmonics cannot be obtained; though the secondary WIFFTA method is with higher accuracy in the measurement, it cannot effectively suppress mutual interference

between Inter-harmonics and harmonics. Therefore, the accuracy of measurement will be affected, in case of serious mutual interference. In the three times measuring results for harmonics and 154.5 Hz inter-harmonics, both frequency and amplitude errors are relatively large. Yet, the present method, which provides better stability and accuracy, even in the case of serious mutual interference, and in which the accuracy of sub-harmonic measurement is able to meet the requirements, is significantly superior to the other two methods.

References

[1] Vatansever F, Ozdemir A 2008 A new approach for measuring RMS value and phase angle of fundamental harmonic based on wavelet packet transform *Electric Power Systems Research* **78**(1) 74-9

[2] Kusljevic M D 2010 *IEEE Transactions on Instrumentation and Measurement* **59**(4) 954-62

[3] Zhang Q M, Liu H J 2008 *IEEE Transactions on Power Delivery* **23**(4) 1728-35

[4] Diego R I, Barros J 2009 Global method for time-frequency analysis of harmonic distortion in power systems using the wavelet packet transform *Electric Power Systems Research* **79**(8) 1226-39

[5] Chang G W, Chen C I, Wu M C 2008 Measuring power system harmonics and interharmonics by an improved fast Fourier transform-based algorithm *IET Generation, Transmission & Distribution* **2**(2) 193-201

[6] Liu Z D, Himmel J, Bonfig KW 2005 *IEEE Trans on Power Delivery* **20**(4) 2370-80

[7] Chang G W, Cheng I C 2010 *IEEE Transactions on Power Delivery* **25**(3) 1787-95

[8] Salor O 2009 Spectral correction-based method for interharmonics analysis of power signals with fundamental frequency deviation *Electric Power Systems Research* **79**(7) 1025-31

[9] Min W, Rowe G I 2004 *IEEE Transactions on Power Delivery* **19**(3) 1496-503

[10] Zahradnik P, Vlcek M 2009 *IEEE Transactions Circuits and Systems II-Express Briefs* **56**(12) 941-5

[11] Zhou F, Zhao C 2011 Based on the time domain linear interpolation method and error analysis of signal cycle *Chinese Journal of Scientific Instrument* **32**(8) 1724-30 (*in Chinese*)

[12] Zhou B, Zhao C 2008 Design of synchronous sampling clock generator in power quality monitor system *Electronic Measurement Technology* **31**(6) 65-8

[13] Zhou F, Huang Z, Zhao C, Chen D 2011 *IEEE Transactions on Instrumentation and Measurement* **60**(8) 2804-12

[14] Zahradnik P, Vlcek M 2006 *IEEE Transactions Circuits and Systems II-Express Briefs* **53**(1) 72-6

[15] Wang L, Zhao C 2008 Power quality monitoring system based on DSP and ARM *Electronic Measurement Technology* **31**(10) 64-7

Authors

Zheng Fan, China

Current position, grades: the associate professor of Henan Mechanical and Electrical Engineering College, China.
University studies: BSc. in Electrical Engineering and Automation from Yanshan University, MSc. from Xi'an Jiaotong University in China.
Scientific interest: computer measurement and control technology, electrical testing and analysis of power quality.

Xianfeng Zheng, China

Current position, grades: the associate professor of Henan Mechanical and Electrical Engineering College, China.
University studies: BSc. in Industrial Electrical Automation from Shanxi Institute of Mining Technology, M.Sc. from Xi'an Jiaotong University in China.
Scientific interest: measurement and control technology and electrical insulation testing.

An optimization method of signal de-noising in discrete wavelet transform based on generalized cross-validation

Xiaojing Chen, Ke Liu, Peng Ye*

College of Physics and Electronic Engineering Information, Key Laboratory of Low-voltage Apparatus Intellectual Technology of Zhejiang, Wenzhou University, Chashan University Town, Wenzhou, Zhejiang Province, 325035, P.R. China

Received 1 July 2014, www.cmnt.lv

Abstract

A method for automatically selecting the asymptotical optimal parameters is presented for signal de-noising in discrete wavelet transform. The parameters of wavelet de-noising were first encoded. A generalized cross-validation algorithm was then used to select these parameters automatically. The parameters that obtained the smallest generalized cross-validation were asymptotically optimal. Simulation signals with different features and range signal-to-noise ratios were used to demonstrate the optimality of the proposed method. In addition, the Raman spectrum of edible oil and nuclear magnetic resonance spectrum of quinine and Boc-protected proline were employed as real-world data to validate the proposed method. The proposed method achieved superior performances in both real-world data and in artificial simulation.

Keywords: signal de-noising, wavelet transform, generalized cross-validation, parameter optimization

1 Introduction

Obtaining accurate characteristics of instrumental signal is important, however, the required signal is often polluted by noise. To extract reliable information from instrumental signals such as infrared spectra (IR), Raman spectral, unwanted noise should be removed [1, 2]. The effects of noise in spectra can be reduced in several ways. One of the most recent methods is based on wavelet transform (WT). Because of the advantages of the localization of time-frequency characteristics and the sparse representation of signal in the wavelet domain, discrete wavelet transform (DWT) is becoming an increasingly important tool in and compression [3-7]. The most popular tool in wavelet de-noising is the wavelet coefficient shrinkage method known as wavelet threshold de-noising [8-10]. The quality of wavelet threshold de-noising is strongly affected by several essential parameters, including wavelet function, decomposition levels, threshold estimate, and thresholding policy [2]. How to obtain these optimal parameters is an important issue for wavelet de-noising. To overcome this problem, cross-validation (CV) and generalized cross-validation (GCV) have been proposed to optimize the threshold estimate [2, 11-16]. Pasti et al. proposed a method to optimize individually the parameters, including the optimal decomposition level, wavelet function, and threshold estimate [2], Cai et al. employed minimum description length (MDL) algorithm to select the threshold and wavelet functions [17]. However, individual optimal parameters do not guarantee a global optimal when all parameters are considered simultaneously. To the best of our knowledge, no single method can optimize these parameters automatically and simultaneously in DWT threshold de-noising.

The present study obtains the optimal parameters by maximizing DWT threshold de-noising. The DWT threshold de-noising parameters are first encoded using Arabic numerals. The GCV algorithm, which was originally developed to optimize the soft threshold value, is then applied to determine automatically the optimal parameter combinations so that the optimal performance of DWT threshold de-noising can be reached.

2 Theory

2.1 DISCRETE WAVELET TRANSFORM (DWT)

Discrete wavelet transform is a linear transform, and the wavelet transform of a discrete signal f can be described as the follows:

$$w = Wf \,, \tag{1}$$

where W is an orthonormal matrix represented as wavelet basis or wavelet filter coefficients, and w I is the wavelet transform coefficients. f is decomposed by the filterbank with a lowpass filter and highpass filter of W into set of wavelet coefficients w:

$$w = [cD_1, cD_2, ..., cD_J, cA_J] \,, \tag{2}$$

where J is decomposition level, and $cD_1, cD_2, ..., cD_J$ represent the detail information of signal f and cA_J represents the approximation information. This transform localizes the most important spatial and frequent characteristics of f in a limited number of wavelet coefficients since the wavelet basis can be derived from a common function called

* ***Corresponding author*** e-mail: yep@wzu.edu.cn

mother wavelet, which has two operations: translation and dilation.

Signal reconstruction can be processed by the same highpass filter and lowpass filter in the W can be described by the simple equation:

$$f = W^t w, \tag{3}$$

where W^t is inverse wavelet transform. In practice, a fast algorithm developed by Mallat [18] is commonly used for performing the transformation.

2.2 DWT THRESHOLD DENOISING

An input instrumental signal can be represented as the sum of two components:

$$y = f + \varepsilon, \tag{4}$$

where f represents the ideal signal and ε represents the noise. A DWT yields the same situation in terms of wavelet coefficients:

$$w = v + \omega. \tag{5}$$

The vector $v = Wf$ contains the wavelet coefficients of the original instrumental signal. Here, $\omega = W\varepsilon$ is the noise coefficient and $w = Wy$ is the noisy signal coefficient. To obtain the ideal signal f, v and ω are separated from one another using a comparison of their wavelet coefficients. The simplified procedure for DWT threshold de-noising can be described as follows:

1) A wavelet transform is applied to signal y to obtain the wavelet coefficient w.

2) The wavelet coefficient ω is removed to obtain the estimated wavelet coefficient f_d.

3) The inverse discrete wavelet transform is applied to f_d to obtain the de-noised signal y_d.

One of the key issues of DWT threshold de-noising is how to estimate ω (i.e., how to distinguish ω from v). Many possible approaches can be used to estimate the noise level, a systematic analysis of the performance of these approaches can be found in [1, 19]. The universal threshold has the following format:

$$\mathrm{TUV} = \sigma\sqrt{2\ln N}, \tag{6}$$

where N is the length of signal y, σ is the standard deviation of the noise, and estimated from the median of the detail coefficients at the first level of signal decomposition.

$$\sigma = \mathrm{median}\left|(\mathrm{de\,tai1})\right| / 0.674. \tag{7}$$

Once the threshold value has been calculated, the key question is how to obtain f_d. In general, DWT threshold de-noising methods use two different policies: hard and soft thresholding. The hard thresholding policy simply sets all the wavelet coefficients below a certain threshold th to zero:

$$\begin{cases} w_{ij}^t = w_{ij} & if\ \left|w_{ij}\right| > th \\ w_{ij}^t = 0 & if\ \left|w_{ij}\right| < th \end{cases}. \tag{8}$$

In soft thresholding, the values of the wavelet coefficients are shrunk by a certain threshold if they are above a certain threshold th:

$$\begin{cases} w_{ij}^t = \mathrm{sgn}(w_{ij})(\left|w_{ij}\right| - th) & if\ \left|w_{ij}\right| \geq th \\ w_{ij}^t = 0 & if\ \left|w_{ij}\right| < th \end{cases}, \tag{9}$$

where sgn is the sign function that returns the sign of the wavelet coefficient w_{ij}.

2.3 GENERALIZED CROSS-VALIDATION (GVC)

The goal of signal de-noising is to minimize the difference between the de-noised and the ideal (i.e., noiseless) signal, thereby minimizing the mean square error (MSE), as shown below:

$$\mathrm{MSE} = \frac{1}{N}\left\| f_d - f \right\| = \frac{1}{N}\sum_{i=1}^{N}\left(f_d(i) - f(i) \right)^2, \tag{10}$$

where f_d is the de-noised signal, f is the ideal signal. However, normally, the ideal signal is not known. Thus, MSE cannot be obtained.

The GVC theory was developed to estimate the best threshold value to optimize DWT threshold de-noising using the soft thresholding policy. GVC defines the risk estimate function as follows:

$$GCV(\xi) = N\left\| w - w_\xi \right\|^2 / N_0^2, \tag{11}$$

where N_0 is the number of the coefficients replaced by zero, N is the total number of the wavelet coefficients, w_ξ is the modified wavelet coefficient after applying a threshold ξ, and w is the wavelet coefficient of the original noisy signal. [12] proved that the threshold ξ that results in the smallest GCV is asymptotically optimal under certain conditions. Moreover, [20] improved the GCV algorithm for the hard thresholding policy. The detailed description of the GVC theory can be found in the document [20]. In general, when GCV is smallest, MSE is asymptotically at minimum, which means searched parameter for de-noise might be optimal, and might also rank among the top 10%, and implies that the de-noising result is asymptotically optimal. Moreover, GVC only depends on input and output data, which are of vital importance in practice. Therefore, GCV is able to measure MSE when the shape of the signal is unknown or the noise energy is difficult to estimate.

3 Experimental and calculations

3.1 ENCODING OF DWT THRESHOLD DENIOSING PARAMERERS

Many types of wavelet functions exist, and some wavelet families contain wavelets of different orders. The Db, Bior, Sym, and Rbio wavelets which are the different filters have been proven effective in signal processing [21]; however, the present study is limited only to the daubechies, symmlet, bior, and rbio families of filters. A total of forty-eight filters were investigated in the present study. To obtain the minimum GCV, these wavelet functions need to be encoded as variables denoted as Arabic numerals. For example, "1" denotes the wavelet function "sym1," "11" denotes the wavelet function "db1," and "48" denotes the wavelet function "dmey." At the same time, other parameters, such as decomposition levels, threshold policy, and threshold estimation, are also encoded as Arabic numerals. The definitions and the encoding of these parameters are shown in Table 1.

TABLE 1 Definition and Encoding of wavelet de-noising parameters

Parameters description	Definition domain	Remark
Wavelet function	[1, $N_{wavelet\ name}$][1]	$N_{wavelet\ name}$ is the number of wavelet function in definition domain. Here, $N_{wavelet\ name}$ is 48.
Decomposition scales	[1, N_{max}]	N_{max} is the maximum of decomposition scales. Here, N_{max} is 10.
Threshold estimation	[5Tuv, 8Tuv]	Tuv is universal threshold.
Thresholding policy	[1, 2]	Hard or soft thresholding.

[1] Wavelet filters from 1 to 48: sym1, sym2, sym3, sym4, sym5, sym6, sym7, sym8, sym9, sym10, db1, db2, db3, db4, db5, db6, db7, db8, db9, db10, db12, db15, db18, db20, db30, bior1.1, bior1.3, bior2.2, bior2.4, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8, rbio1.1, rbio1.5, rbio2.2, rbio3.3, rbio3.5, rbio3.7, rbio3.9, rbio4.4, rbio5.5, rbio6.8, Dmey;

3.2 SIMULATION SIGNALS AND CALCULATIONS

Different simulation signals were considered in order to evaluate the performance of the proposed method. The simulation signals of Blocks, Bumps, Heavysine and Doppler signal in the wavelet tool are employed in present study. The four signals contain abundant frequency components, many peak information and lots of catastrophe parts, so they are always used as the typical signals for comparing the quality of de-noising algorithms. The four original signals and noisy signals are showed in Figure 1.

In this study, the Wavelet Toolbox 3.0 (The Math Works, Natick, USA) was used, and the Pieflab Toolbox for GCV algorithm can be obtained from http://homepages.ulb.ac.be/~majansen/software/ thresh -lab.html, and all the calculations were performed on the platform of MATLAB 7.6 (The Math Works, Natick, USA).

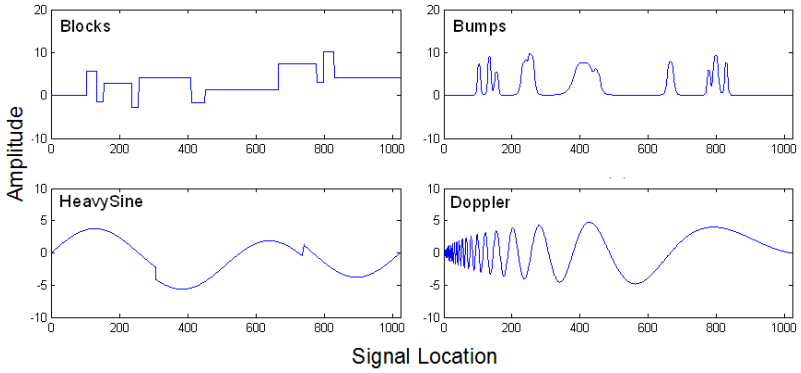


FIGURE 1 Four simulation signals

4 Results and discussions

4.1 SIMULATION DATA

In this study, the definition domain of the threshold was from 5Tuv to 8Tuv with the interval of 1 unit. As a result a total of 40 definition points are available.

As mentioned above, when the minimum GCV value is achieved, MSE is also near to the optimal value under certain conditions. This means that the parameter combinations of DWT threshold de-noising corresponding to the minimum GCV are also near to optimal. For real-world signals, MSE cannot be estimated, hence, rather than MSE, the proposed GCV approach was adopted to optimize the parameter combinations.

Random noise was added 100 times to each type of simulation signal with the signal-to-noise ratio (SNR) ranging from 16.34 to 21.18 (Level 1) to validate whether the parameter combinations corresponding to the minimum GCV (or optimal GCV) are optimal. Each type of simulation signal can obtain 100 noisy signals. Here, only Doppler noisy signals are listed in Figure 2a (Level1). The MSE value corresponding to the minimum GCV (GCV-MSE) can be obtained from 38,400 (48 * 10 * 40 * 2 parameter combinations) different MSE values. Observing the arrangement locations of GCV-MSE in all 38,400 MSE

values is a simple way to determine whether GCV-MSE is minimum. When GCV-MSE is minimum (or ranks first), and imply the obtained parameter combination is optimal. Similarly, if GCV-MSE is near to minimum (or ranks in the top ten or top one hundred), the searched parameter combination is asymptotical optimal. The GCV-MSE concrete arrangement distribution of 100 noisy signals of each type of simulation signals is shown in Figure 3a. In Figure 3a, the abscissa values for Blocks, Bumps, Heavysine, and Doppler signals are 1–100, 100–200, 201–300, and 301–400, respectively. The small circles in Figure 3a represent the arrangement position of each noisy signal. Each discrete point is connected with solid lines for easy comparison of several different signals.

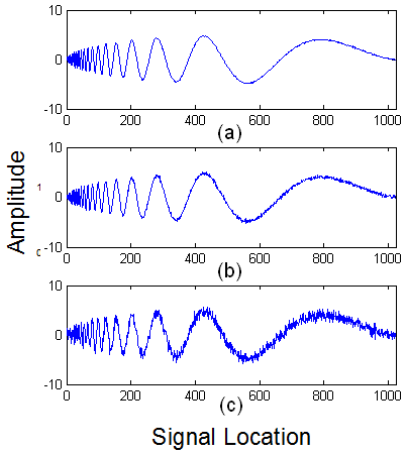


FIGURE 2 Noisy Doppler signals with different SNR level: (a) Level 1 (b) Level 2 (c) Level 3

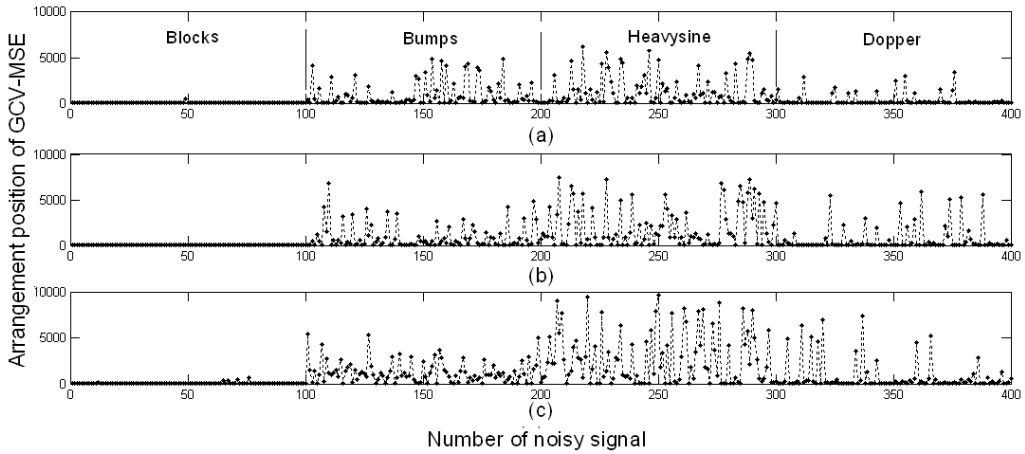


FIGURE 3 GCV-MSE Arrangement locations of four simulation signals with different SNR, 1-100 is the arrangement locations of GCV-MSE for the Blocks signals, 100-200 are for Bumps signals, 201-300 are for Heavysine signals, 301-400 are for Dopper signals: (a) level 1 SNR (b) level 2 SNR (c) level 3 SNR

As seen in Figure 3a, the arrangement locations of GCV-MSE is relatively small for signals with the SNR level 1, and most of them are lower than 5,000, which indicates that the identified parameter combination is working well. A comparison of the optimization results of four signals shows the results of the Block signal sequence 1–100, as shown in the Figure 3a to be the best. This result can be explained by the GCV algorithm principle. Because the GCV algorithm is more suitable for regular signals, it is more appropriately applied in signals with less frequency information. The mean and standard deviation of GCV-MSE arrangement positions listed in Table 2 further validate the optimization effect of the proposed method. As shown in Table 2, the mean of the arrangement locations of the GCV-MSEs is top-ranked (corresponding to 38,400 combinations), indicating that optimization results are very satisfactory. Compared with these results, the mean of Heavysine is the largest, which indicates that the optimization result is the worst among the four simulation signals. Moreover, the standard deviations of the Blocks signals are smaller than that of the other three signals, which indicate that the proposed method is relatively stable for Blocks signals.

TABLE 2 Mean and standard deviations of GCV-MSE arrangement locations of four simulation signals

Signals SNR level	**Blocks**		**Bumps**		**Heavysine**		**Doppler**	
	Mean	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*
Level 1	21	48	876	1327	1249	1651	285	665
Level 2	17	16	847	1297	1947	2215	580	1339
Level 3	38	92	1239	1176	2729	2840	690	1586

TABLE 3 Mean and standard deviations of MDL-MSE arrangement locations of four simulation signals

Signals SNR level	Blocks		Bumps		Heavysine		Doppler	
	Mean	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*
Level 1	71	74	2007	981	3897	1895	1131	692

The MDL algorithm is free from any parameter setting or any subjective judgment when used for data compression and de-noising [17]. In the present study, the MDL algorithm can also be used in the parameter optimization of the DWT algorithm, with the coding and setting of the parameters the same as before. The optimization result shown in Table 3 also shows a comparison of the results of GCV algorithm of the noise level 1. Based on the Table 3, the MDL algorithm is inferior to the GCV algorithm, which is consistent with the result proposed by the document [2]. Moreover, MDL can only be used for optimization of hard threshold values. Hence, the following studies are all based on GCV algorithm.

To investigate the influence of adding signals with different noise levels on the proposed method, 100 noisy signals with SNRs ranging from 11.17 to 16.22 (Level 2) and from 8.82 to 11.37 (Level 3) are employed. The noisy signals with different SNRs are shown in Figure 2b and 2c, along with a list of the Doppler noisy signals. Similarly, the GCV-MSE specific arrangement distributions of the SNR levels 2 and 3 are shown in Figures 3b and 3c. Clearly, the performance of the proposed method is influenced by different levels of noises. Based on Figure 3, the Heavysine and Doppler signals are influenced the most, followed by the Bumps signal, whereas little influence is generated on the Blocks signal. These results indicate that the proposed method is suitable for several regular signals, further verifying that the GCV algorithm is more suitable for regular signals. In Figure 3b, most of the GCV-MSE arrangement locations are within 5,000, and only a few Bumps and Heavysine signals exceed 5,000, and the GCV-MSE arrangement locations greater than 5,000 are slightly increased for the SNR level 3, especially for Heavysine signal, as seen from Figure 3c. However, compared with the 38,400 MSE values, the arrangement locations are still placed at the front location. The mean and standard deviations of the arrangement positions of 100 GCV-MSE of the four types of signal with SNRs at Levels 2 and 3 are listed.

As seen from Table 2, the means of all the arrangement locations are small, which indicates that the parameter optimization results are satisfactory. The Heavysine and Doppler signal are greatly affected by noise on the aspect of standard deviation values, whereas noise has little influence on Blocks signals.

At the same time, the specific numbers of different arrangement scopes have been summarized out, as shown in the Table 4. The optimization results of the parameters of three types of noises which is shown in the Table 4 indicate that the numbers of those whose arrangement locations are below 384 are 292, 269 and 227 respectively, which means that there are probability values of 73.00%, 67.25 and 56.75% to select the optimal combinations of top 1% correspondingly. Meanwhile, this also means that there are probability values of 94.50%, 91.75% and 89.25% for the optimal combinations of top 10% to be selected respectively. This result is very satisfactory for such a huge combination of 38400.

In order to investigate the relation between the GCV value and the corresponding MSE values, each GCV value of the 38,400 different parameters and the corresponding MSE values should be listed. However, illustrating such massive data would be impossible because of page limitations. Hence, only 200 data were selected randomly from Block and Heavysine signals, their GCV values and the corresponding MSE values are reported in Figure 4. As one can see from Figure 4a, small GCV is not always corresponds to the small MSE, however, in the region of GCV smaller than 0.5, the trends of GCV and MSE are similar, based on Figure 4b, it was found that the smaller GCV and MSE correspond to same data point (parameter combination). Similar situation is also happened in the Heavysine signal; the detailed results are showed in Figures 4c and 4d.

TABLE 4 Numbers of arrangement location

SNR level	Range		
	<384	*<1920*	*<3840*
Level 1	292	356	378
Level 2	269	337	367
Level 4	227	321	357

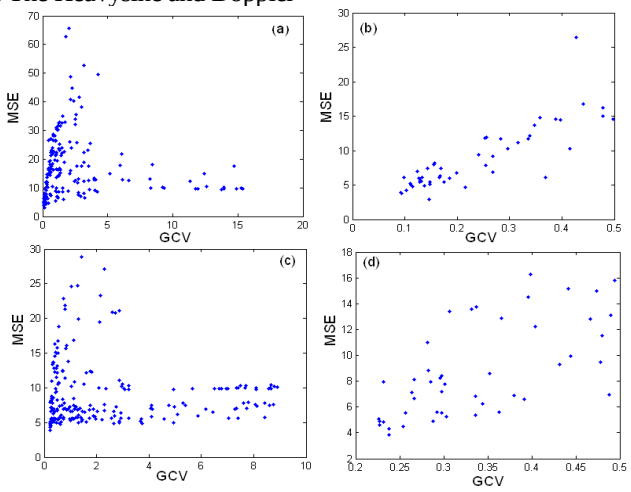


FIGURE 4 Relations between GCV and MSE of block and heavysine signals

4.2 REAL-WORLD DATA

The real-world data used in the present study is the Raman spectrum of edible oil (mainly composed of protein and fat). The T64000 Raman spectrometer used was produced by the French HORIBA Jobin Yvon Company. All Raman tests were conducted at room temperature. The excitation light source used was an argon ion laser with a wavelength of 514.5 nm, laser power of 120 mw, and scanning wavenumber range of 500–3,500 cm^{-1}. The experimental light path was backscattered. Laser-formed light spots were shown to have a diameter of approximately 1 um on the sample surface after being focused 100 times by the object lens. The original Raman spectrum in Figure 5a shows four peaks with different widths; the figure also shows the height and the area of the spectra peaks. In general, the shapes of the peaks, including width, height, and area should be maintained in the de-noising process to avoid influence on the analysis results. This makes signal de-noising very challenging.

Using the proposed GCV methodology, the optimal parameter combinations were obtained as 8, 5, 4, and 1. Correspondingly, wavelet function of sym8, decomposition level of 5, a 4/5 Tuv threshold and hard threshold were employed. De-noising result in Figure 5b shows that the shapes of several peak values to be well kept, whereas most of the noises were eliminated. The attributes of the peak, including position, height, width, and area, are the main indices that define the components in the Raman spectrum analysis. The changes of the attributes of the peak before and after de-noising are shown in Table 5. The position, width, height, and area of the peak showed no prominent changes after the noise of the original spectrum was eliminated.

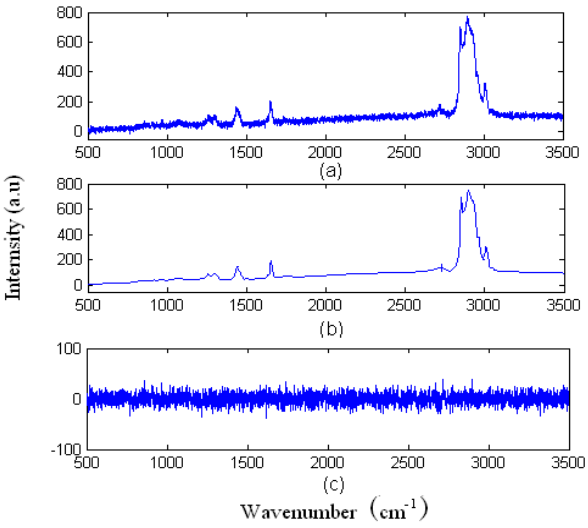


FIGURE 5 Raman spectrum of edible oil: (a) noisy spectrum (b) denoised spectrum (c) noisy signals

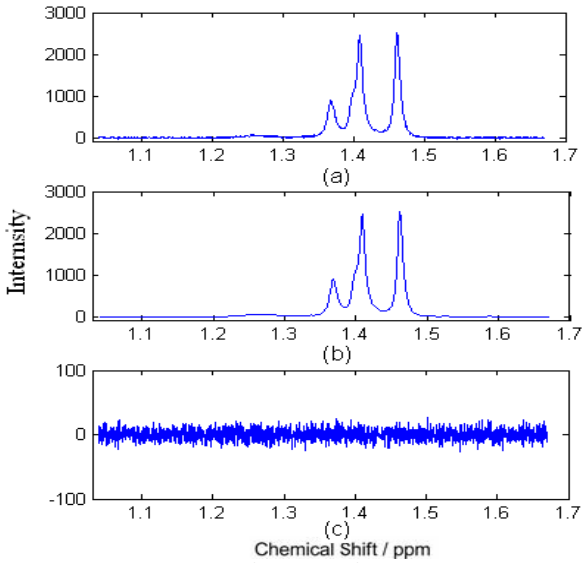


FIGURE 6 NMR spectrum of mixtures of quinine and Boc-protected proline: (a) noisy spectrum (b) denoised spectrum (c) noisy signals

Another real-world data is nuclear magnetic resonance (NMR) spectrum in Figure 6a, which was obtained 16:84 mol ratio mixtures of quinine and Boc-protected proline with enantiomeric excess of 10% in CDCl3. The NMR spectra were recorded after several minutes of thermal equilibration time. NMR spectrum was recorded on a Bruker Avance 500 MHz spectrometer. Spectrum was recorded using 16 scans at 298K. A full spectrum of sample was recorded referenced to TMS. An exponential window function with a line-broadening factor of 1Hz was applied to the FID before Fourier transformation. The 1H-NMR spectrum was phased and baseline-corrected using Topspin 2.1(Bruker) and was automatically reduced by using the AMIX (Bruker GmbH, Germany) software package to continuous integral segments of equal width of 0.004ppm corresponding to the chemical shift range 1H, δ 6.9–8.8 after removing the solvent resonance region (δ 7.2-7.3). Similar to the situation of Raman spectrum, to de-noise NMR spectrum is a very challenging thing because it requires that the shape of the peak should remain unchanged as much as possible. Since several peaks of this NMR spectrum are connected together, there is impossible to calculate the area and width of each peak. Therefore, we give the height of the peak and total area and width of three peaks prior to de-noising here. Using the proposed GCV methodology, the obtained optimal parameter combination is 4, 5, 4 and 1, correspondingly, parameters of wavelet function is sym8, decomposition level is 5, and 4/ 5Tuv threshold and hard threshold are employed. De-noising results are shown in the Figure 6b. The changes of the attributes of the peak before and after de-noising are shown in Table 6, as one can see from the Table 6, the spectral profile after de-noising, the height of the peak changes relatively less, and the shape of peaks almost remain unchanged, indicating that the de-noising effect is relatively satisfactory. Hence, it can be concluded that the parameter combination thus generated is relatively ideal.

TABLE 5 Comparison of attribute of peak before and after de-noising for Raman spectrum

Peak	Original				Denoised			
	Position	*Width*	*Height*	*Area*	*Position*	*Width*	*Height*	*Area*
1	1148	102	73	7079	1147	104	73	7076
2	1346	46	124	7790.	1345	51	121	7788
3	1652	23	174	6456.	1652	23	173	6451
4	3421	153	668	120529	3422	151	668	120525

TABLE 6 Comparison of attribute of peak before and after de-noising for NMR spectrum

	Original Peak			Denoised Peak		
	1	*2*	*3*	*2*	*2*	*3*
Position	1367	1409	1461	1367	1409	1461
Height	903.5	2446.3	2516.6.	900.5	2445.8	2512.6.
Width		403			404	
Area		310320			310350	

5 Conclusions

The present paper proposed an algorithm that automatically selects DWT threshold de-noising parameters based on the GCV algorithm. Parameter optimization includes wavelet function, decomposition level, threshold estimation, and threshold policy. Four simulation datasets and real-world data of Raman spectral and NMR signals were used for validating the proposed algorithm. The results show that the identified minimum GCV value produces a better MSE value. Therefore, DWT threshold de-noising can be optimized automatically through reasonable parameter encoding based on GCV value. At the same time, the optimization ability of the proposed method slightly degrades with the increase of the magnitude of added noise.

Acknowledgements

The authors would like to acknowledge the financial support provided by the Program for Key Innovative Team of Research of Zhejiang Province (2012R10006-03).

References

[1] Alsberg BK, Woodward AM, Winson MK, Rowland J, Kell D B 1997 *Analyst* **122**(7) 645-52
[2] Pasti L, Walczak B, Massart DL, Reschiglian P 1999 *Chemometrics and Intelligent Laboratory Systems* **48**(1) 21-34
[3] Alsberg BK, Woodward AM, Kell DB 1997 *Chemometrics and Intelligent Laboratory Systems* **37**(2) 215-39
[4] Komsta L 2009 *Journal of Chromatography A* **1216**(12) 2548-53
[5] Chen X, Wu D, He Y, Liu S 2009 *Analytica Chimica Acta* **638**(1) 16-22
[6] Zhu D, Ji B, Meng C, Shi B, Tu Z, Qing Z 2007 *Journal of Agricultural and Food Chemistry* **55**(14) 5423-8
[7] Mittermayr CR, Nikolov SG, Hutter H, Grasserbauer M 1996 *Chemometrics and Intelligent Laboratory Systems* **34**(2) 187-202
[8] Donoho DL 1995 *IEEE Transactions on Information Theory* **41**(3) 613-27
[9] Donoho DL, Johnstone IM 1994 *Biometrika* **81**(3) 425-55
[10] Donoho DL, Johnstone IM 1995 *Journal of the American Statistical Association* **90**(432) 1200-24
[11] Nason GP 1996 *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(2) 463-79
[12] Jansen M, Malfait M, Bulthell A 1997 *Signal Processing* **56**(1) 33-44
[13] Jansen M, Bulthell A 1999 *IEEE Transactions on Image Processing* **8**(7) 947-53
[14] Jansen M, Uytterhoeven G, Bulthell A 1999 *Medical Physics* **26**(4) 622-30
[15] Weyrich N, Warhola GT 1998 *IEEE Transactions on Image Processing* **7**(7) 82-90
[16] Tiwari AK, Shukla KK 2004 *Digital Signal Processing* **14**(2) 38-157
[17] Cai C, Harrington PB 1998 *Journal of Chemical Information and Computer Sciences* **38**(6) 1161-70
[18] Mallat SG 1989 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7) 674-93
[19] Barclay VJ, Bonner RF, Hamilton IP 1997 *Analytical Chemistry* **69**(1) 78–90
[20] Jansen M 2001 *Noise Reduction by Wavelet Thresholding Springer Press New York*
[21] Singh BN, Tiwari AK 2006 *Digital Signal Processing* **16**(3) 275-87

Authors

Xiaojing Chen, born on January 3, 1978, Wenzhou, Zhejiang Province, P.R. China

Current position, grades: associate professor in Wenzhou University, China.
University study: PhD degree was earned in major of optical detection, Xiamen University in 2009.
Research activities: spectra analysis; pattern recognition; digital signal processing.

Ke Liu, born on July 15, 1989, Changsha, Hunan Province, P.R. China

Current position, grades: Graduate student in Wenzhou University,China.
University study: BsC in electronic and information engineering, Hunan University of Science and Engineering in 2012.
Research activities: pattern recognition, digital signal processing.

Peng Ye, born on September 19, 1970, Wenzhou, Zhejiang Province, P.R. China

Current position, grades: lecturer in Wenzhou University, China.
University study: BsC in applied physics, Fudan University, in 1994.
Research activities: optical analysis; electronic engineering.

Research on the machinability when dry turning hardened steel with ceramic tool

Chunjuan Tu[1], Xuhong Guo[2*]

[1]*Suzhou Institute of Industrial Technology, Suzhou- Jiangsu, China*

[2]*School of Mechanical and Electric Engineering, Soochow University, Suzhou- Jiangsu, China*

Received 24 April 2014, www.cmnt.lv

Abstract

The experiments of dry turning hardened steel with ceramic tool (CC6050) were carried out. The cutting forces of ceramic tool with different cutting parameters were measured by Kistler cutting force acquisition system, the prediction model of the surface roughness was built by Particle Swarm Optimization (PSO), the wear morphology of the tool and chip characteristics were observed by scanning electron microscope (SEM), analyzed the wear zone of the tool by the EDAXPV990(EDS). Results indicated that the cutting depth was the dominant factor affecting the cutting force, next was feed speed, the minimum was cutting speed. The feed speed was the dominant factor affecting the surface roughness, next was cutting depth, the minimum was cutting speed. The crater wear was shown on the rake face and the evenly distributed strip shaped wear was shown on the flank of the CC6050 tool, the main wear mechanism was abrasive wear, diffusion wear, adhesive wear and oxidation wear. The shape of chip was saw-tooth, the rough slice layers with periodic flow were shown on the upside of the chip and the streaks were evenly distributed on the underside of the chip. The cutting speed and feed speed were larger, the serrated level of the chip was higher and the angle of shear was larger.

Keywords: Ceramic Tool, Dry Turning, Hardened Steel, Machinability

1 Introduction

Hardened steel has been used in manufacturing all kinds of mechanical parts with high hardness and high wear resistance attribute to its excellent mechanical properties. Grinding was the most traditional processing technology on finishing hardened steel, however, it would cause lower production efficiency, high cost and heavy environmental pollution [1-3]. With the development of the super-hard tool materials, dry turning instead of grinding on finishing hardened steel is expected to become the new technology. However, hardened steel has high hardness and poor machinability, in order to achieve the technology by using dry turning instead of grinding, the tool material should have good resistance of thermic wear, tool edge and tool nose should have enough strength and low breakage rate of wear, all of above can ensure that the machining error is within the scope of the working accuracy.

Ceramic tool has high micro hardness and heat stability. It has good resistance of thermic wear when turning the hardened steel and the materials of ceramic tool are abundant, its price is not half of the CBN tool. The strength and fracture toughness would be increased greatly by advanced milling, sintering process and toughening technology. It has good comprehensive mechinablity and high cost performance [4]. So it is the ideal tool in dry turning hardened steel.

Xiaobin cui etc researched the evolution of the tool failure in intermittent turning hardened steel with ceramic tool, built the formulas of the tool life in the different failure stage of evolution under the condition of intermittent turning [5]. Professor Gabriel C. Benga in Romania did the research on the influence of tool life and surface roughness by cutting depth when turning 100Cr6 bearing steel with ceramic and PCBN cutting tools [6]. Xiaoguang Wang etc did the experiments on dry turning hardened steel by ceramic tool, researched the endurance, wear morphology and wear mechanism of the ceramic tool [7]. Xuhong Guo etc researched the wear morphology, wear properties and wear mechanism when turning GCr15 hardened steel with ceramic tool and PCBN tool [8]. Many scholars did the studies of dry turning hardened steel by ceramic tool, but most studies focused on the tool endurance and tool wear, it was lacked for intensive study on cutting force, surface quality, tool wear and chip characteristics in dry turning hardened steel, reports on these aspects at home and abroad were few.

The experiments of dry turning hardened steel by ceramic tool (CC6050) were carried out in this paper. Measured the values of cutting force in different cutting parameters, built the prediction model between cutting parameters and surface roughness by PSO, researched the impact of cutting parameters on the cutting force and surface roughness, micro analyzed the wear morphology of the CC6050 and the chip characteristics, researched their affects on dry turning hardened steel. We hope we can provide some experimental basis for popularization of the high hardened tool material.

2 Materials and Methods

The work material was GCr15 bearing steel，its chemical composition was shown in Table 1. It was provided by Suzhou Jingji Mech & Elec CO., LTD, which was φ40×300mm cylindrical specimen. The hardness of the

* ***Corresponding author*** e-mail: tucj@siit.edu.cn

workpiece quenched by machine oil at 840℃was 60HRC (58HRC~61HRC).

TABLE 1 Nominal chemical composition of GCr15 bearing steel (wt%)

C	Si	Mn	Cr	Mo	P	S
0.95~ 1.05	0.15~ 0.35	0.25~ 0.45	1.40~ 1.65	≤0.1	＜0.027	＜ 0.020

Dry turning experiments were carried out on the CA6140A lathe. The lathe was the product of Shenyang No.1 Machine Tool Plant. The ceramic cutting tool was the product of SANDVIK Company in Sweden. The type was SNGA120408S01525 6050. After clamping, the geometric parameters of the tool were shown in Table 2.

TABLE 2 Geometric parameters of the tool after clamping

γ_o	α_o	λ_s	κ_r	r_ε	b_{rn}	γ_n
-6°	0°	-6°	75°	0.8 mm	0.15 mm	25°

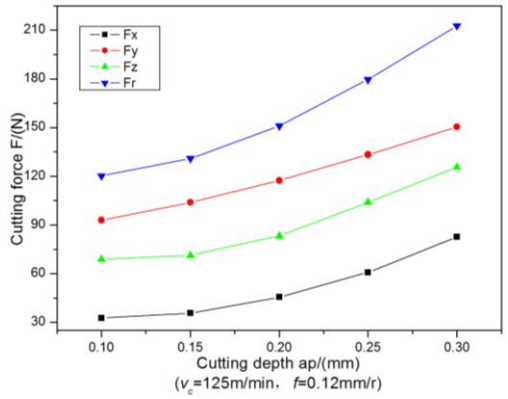
The values of the cutting force in different cutting parameters were measured by the cutting force data acquisition system, which was made up of Kistler9257B piezoelectric crystal sensor dynamometer, Kistler 507A10100 charge amplifier, Kistler9403 steel tool rack and computer. The surface roughness experiment based on $L_9(3^3)$ orthogonal, the factors of experiment were cutting speed, feed speed and cutting depth. The values of the surface roughness in different cutting parameters were measured by 2025 surface roughness tester. Enlarged photomicrograph of the wear morphology on the rake face and flank of CC6050 and chip characteristics were acquired by S-4700 scanning electron microscope (SEM), analysed the wear zone of CC6050 by EDAXPV9900 (EDS).

3 Results and Discussion

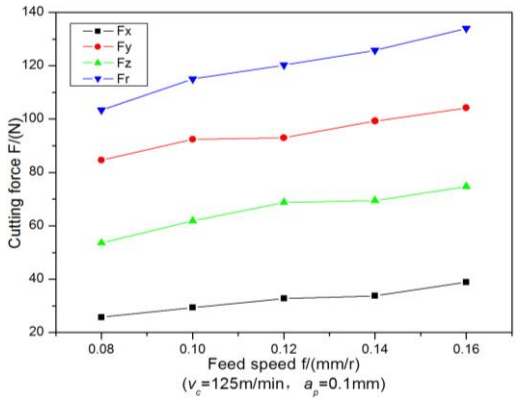
3.1 IMPACT OF CUTTING PARAMETERS ON CUTTING FORCE

The influence curves of cutting force in different cutting depth were shown in Figure 1(a) when cutting speed was 125m/min and feed speed was 0.12mm/r. The influence curves of cutting force in different feed speed were shown in Figure 1(b) when cutting speed was 125m/min and cutting depth was 0.1mm. The influence curves of cutting force in different cutting speed were shown in Figure 1(c) when feed speed was 0.12mm/r and cutting depth was 0.1mm. F_x was axial cutting force, F_y was radial cutting force, F_z was main cutting force and F_r was resultant tool force in three directions.

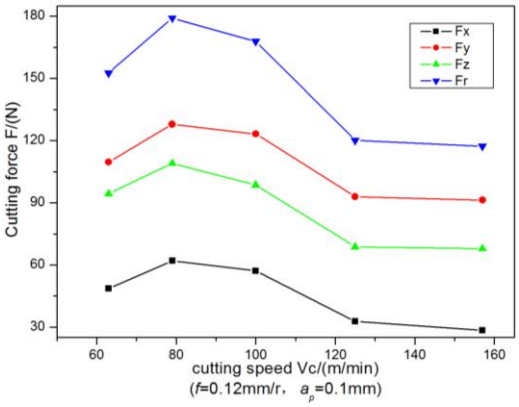
Figure 1(a) showed that the cutting forces in three directions of CC6050 increased gradually with cutting depth increasing when feed speed and cutting speed were certain and the relation was approximately linearity. When the cutting depth was less than 0.15mm, the trend of cutting force increasing was slow, when the cutting depth was larger than 0.15mm, the trend of cutting force increasing was fast. Figure 1(b) showed that the cutting forces in three directions of CC6050 also increased with feed speed increasing when cutting speed and cutting depth were certain, but the trend of cutting force increasing was not obviously, especially the cutting force in three directions keep basically stable when feed speed was within the scope of 0.12mm/r and 0.14mm/r. Figure1(c) showed that the cutting forces in three directions increased with cutting speed increasing when cutting speed was less than 80m/min and decreased when the cutting speed was larger than 80m/min. Especially the trend of cutting force decreasing was obviously when cutting speed was within the scope of 80m/min and 125m/min, the trend was slow when the cutting speed was larger than 125m/min. Compared with Figure 1(a), (b) and (c), the radial force was maximum, next was the main cutting force, the minimum was axial force when dry turning hardened steel by CC6050.



(a) Impact of cutting depth



(b) Impact of feed speed



(c) Impact of cutting speed

FIGURE 1 Influence curves of cutting force in different cutting parameters

The reason was: cutting width a_w increased in direct proportion to cutting depth ($a_w=a_p/\sin K_r$), it caused cutting layer area, the deformation resistance and friction force increased, so cutting force increased. Cutting force increased with cutting power increasing when feed speed increased and cutting thickness a_c increased in direct proportion to feed speed ($a_c=f\sin K_r$), the deform coefficient ξ and friction coefficient μ decreased, so cutting force decreased. Therefore the trend of cutting force increasing was not obviously when feed speed increasing.

When the cutting temperature raised with cutting speed increasing, the metal softening effect was produced in the workpiece [9, 10], the friction coefficient μ decreased with cutting speed increasing, the angle of shear φ increased, it caused the deform coefficient ξ decreased, so the cutting force decreased.

3.2 IMPACT OF CUTTING PARAMETERS ON SURFACE ROUGHNESS

The values of the surface roughness after dry turning hardened steel by ceramic tool were shown in Table 3 based on L_9 (3^3) orthogonal experiment.

The theoretical prediction model of surface roughness was shown in Eq.1. The objective function built by Particle Swarm Optimization (PSO) was shown in Eq.2. Optimized the objective function with 1stOpt mathematical analysis software, the constraint expression was $x_{min} \leqslant x_i \leqslant x_{max}$ (x_{min} was the lower limit value of the variable and the x_{max} was the upper limit value of the variable) [11].

The each parameter in Particle Swarm Optimization was as follows: the populations size was 100, adjacent population was 2, the value of weighting factor was 0.1, learning factor was 2.05, the max speed was 2, the iterations was 5000. The model simulated with experiment results was shown in Eq.3.

TABLE 3 The measure results of surface roughness experiment

NO.	v_c(m/min)	f (mm/r)	a_p (mm)	R_a (um)
1	100	0.08	0.1	0.51
2	100	0.12	0.2	0.67
3	100	0.16	0.3	1.01
4	125	0.08	0.2	0.48
5	125	0.12	0.3	0.76
6	125	0.16	0.1	0.78
7	157	0.08	0.3	0.49
8	157	0.12	0.1	0.57
9	157	0.16	0.2	0.93

$$R_a = ka_p^x f^y v_c^z = e^{x_1} a_p^{x_2} f^{x_3} v_c^{x_4}, \quad (1)$$

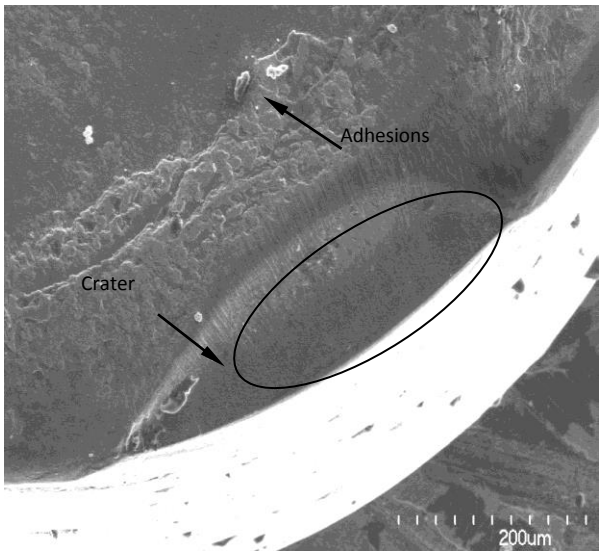
$$\min(f(x_1,x_2,x_3,x_4)) = \sum_{i=1}^{9} \left| e^{x_1} v_{ci}^{x_2} f_i^{x_3} a_{pi}^{x_4} - R_{ai} \right|, \quad (2)$$

$$R_a = e^{2.1599} v^{-0.0189} f^{0.9739} a_p^{0.2314}. \quad (3)$$

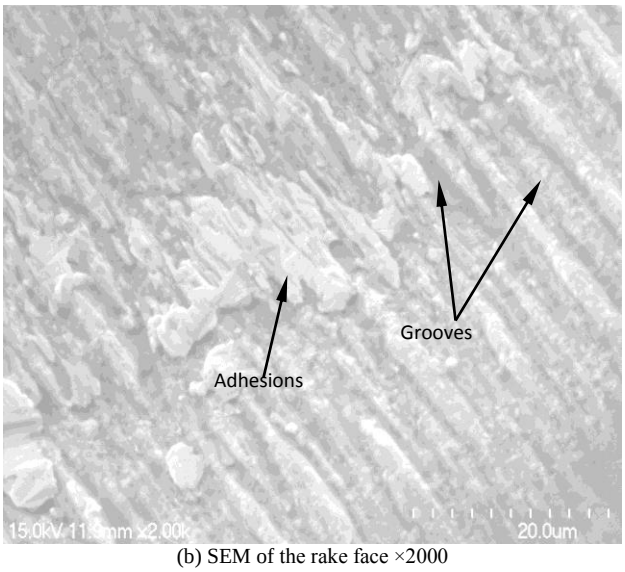
Eq.3 showed that the index of feed speed and cutting depth were positive value, so the surface roughness increased when feed speed and cutting depth uninterrupted increasing, the index of cutting speed was negative value, the surface roughness decreased when cutting speed uninterrupted increasing. It was apparent from the value of every index that the dominant influence factor of cutting parameters was feed speed, next was cutting depth, the minimum was cutting speed.

3.3 WEAR MORPHOLOGY ANALYSIS

The wear morphology of the rake face and flank were shown in Figure 2 and Figure 3 when dry turning hardened steel by ceramic tool (CC6050). The cutting parameters were v_c=125m/min, a_p=0.2mm, f=0.12mm/r, cutting time was 25min.



(a)SEM of the rake face ×200

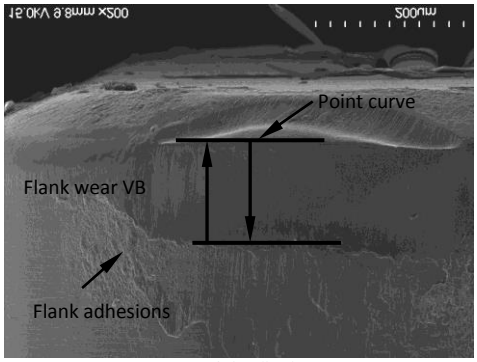


(b) SEM of the rake face ×2000

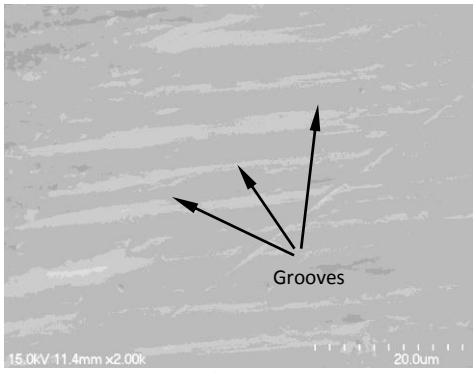
FIGURE 2 The Wear morphology of the rake face of CC6050 in later period of wear

Figure 2(a) showed that there was obviously crater wear on rake face of CC6050, wear zone was basically parallel to the rake face, it was wide at center and narrow at both sides. The wear area grouped mainly on the negative chamfering of the tool nose, it was directly connected with the tool edge,

the depth of the crater was steady. Figure 2(b) showed that the groove like plowed covered on the wear area of the rake face, accompanied with more adhesions. The features of abrasive wear and adhesive wear were obviously showed in Figure 2.



(a)SEM of the flank ×200



(b) SEM of the flank ×2000

Figure 3 The wear morphology of the flank of CC6050 in later period of wear

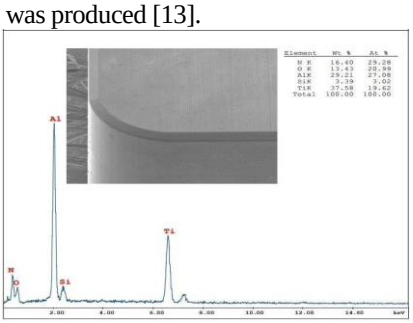
Figure 3(a) showed that there was strip shaped wear on the flank of CC6050 and its width was steady, the abrasion was relatively large on the arc chamfering of tool nose, accompanied with more adhesions near by the flank wear area, the wear marks was not obviously on the border of wear area, large slough was not exist. Figure 3(b) showed that the groove like plowed on the wear area of flank was steady. The steady feature of abrasive wear was showed in Figure 3.

The main reason to produce the wear morphology of the above was: in the cutting progress, the acutely friction was produced by high stress contact between the rake face of tool and chips, the average temperature measured in these cutting areas was higher than 1000℃ according to work-tool thermocouple principle, adhesive, diffusion and oxidation wear were intensified, the high temperature zone of the tool was close to the tool edge and the cutting force was also focused on the near to the tool edge, the softened tool edge with heat deformed and collapsed [12]. Elastic and plastic deformation were produced because the contact force between the flank and surface of the workpiece was very large, the contact area between flank and workpiece was small, the strength of tool nose was lower and the cooling condition was poor, so the wear was serious.

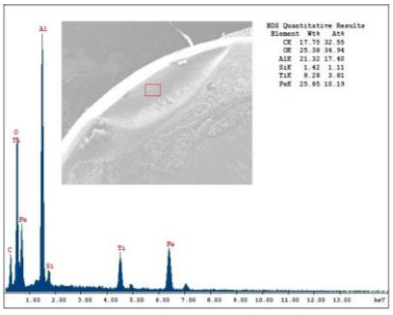
3.4 WEAR MECHANISM

The EDS spectra of the tool face on CC6050 before and after dry turning were shown in Figure 4.

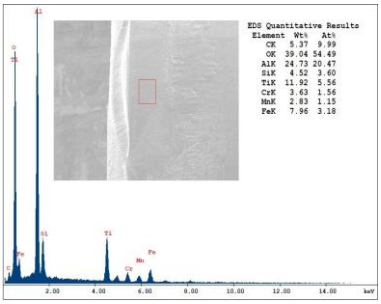
Compared with Figure 4(a), (b) and (c), the mass fraction of oxygen in the rake face and frank of CC6050 was improved from 13.43% to 25.38% (in the rake face) and 39.04% (in the flank), it proved that the oxidation wear was produced. The mass fraction of Ti was decreased from 37.58% to 8.28% (in the rake face) and 11.92% (in the flank), it proved that Ti in the tool had been diffused to the material and chip, the diffusing phenomenon in the rake face was more obviously. The mass fraction of Si and Al had little change, it proved the chemical affinity between Si, Al and material was small, it was hard to adhere in turning. Otherwise, the Fe, C and Cr was added in the flank of the tool, the mass fraction was 7.96%, 5.37% and 3.63% respectively, the three elements of the material were separated out because of the high temperature in turning and diffused to the flank of the tool. There had no Cr element in the rake face, but the content of the Fe and C in the rake face was higher than flank, its mass fraction was 25.85% and 17.75% respectively, the chemical affinity between the tool and material was improved with the content of the Fe and C improved, it caused the adhesive wear of the tool was produced [13].



(a) The EDS spectra of the tool before dry turning



(b) The EDS spectra of the rake face after dry turning



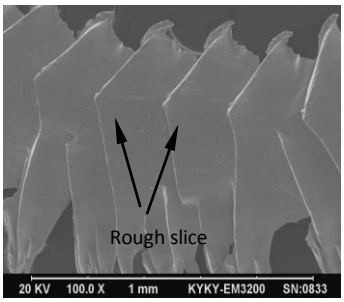
(c) The EDS spectra of the flank after dry turning

FIGURE 4 The EDS spectra of CC6050 before and after dry turning

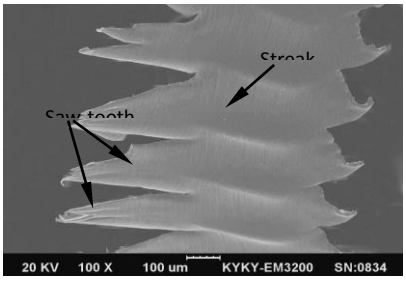
3.5 CHIP CHARACTERISTICS ANALYSIS

The shape of chips was shown in Figure 5 when dry tuning hardened steel by CC6050. Figure 5(a), (b), (c) showed the shape of chips when the cutting parameters were v_c=157m/min, f=0.2mm/r, a_p=0.3mm. The values of three cutting forces measured in these cutting parameters by Kistler cutting force acquisition system were F_z=262.95N, F_x=142.44N, F_y=371.04N. Figure 5(d),(e),(f) showed the shape of chips when the cutting parameters were v_c=125m/min, f=0.12mm/r, a_p=0.4mm. The values of three cutting forces were F_z=213.06N, F_x=145.67N, F_y=259.61N.

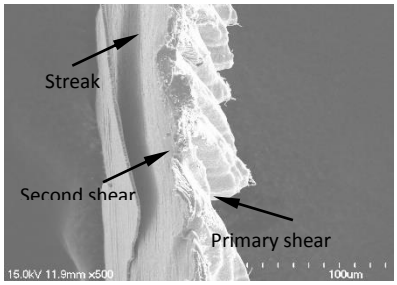
Figure 5 showed that the chip characteristic was obviously saw-tooth when dry turning hardened steel by CC6050. The extrusion plasticity shear of the chip was produced in the primary and second shear zone [14]. The severe plastic deformation was produced under extreme stress in the primary shear zone, so the rough slice layers with periodic flow was shown on the upside of the chip and the underside of the chip was smooth, the streak was evident as a result of the extrusion and friction of the chip against the rake face of tool.



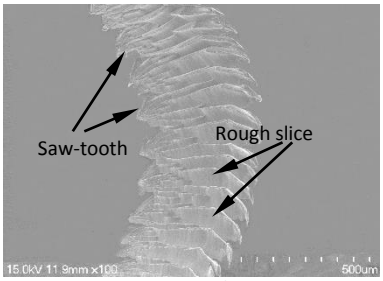
(a)Upside SEM of chip×100
(vc=157m/min, f=0.20mm/r, ap=0.3mm)



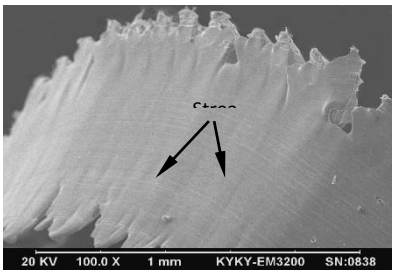
(b) Underside SEM of chip×100
(v_c=157m/min, f=0.20mm/r, a_p=0.3mm)



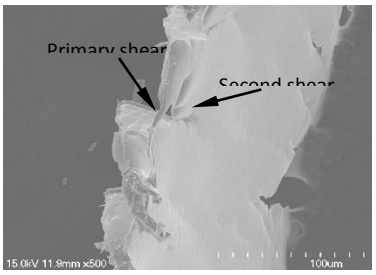
(c)SEM of the chip in profile×500
(v_c=157m/min, f=0.20mm/r, a_p=0.3mm)



(d)Upside SEM of chip×100
(v_c=125m/min, f=0.12mm/r, a_p=0.4mm)



(e) Underside SEM of chip×100
(v_c=125m/min, f=0.12mm/r, a_p=0.4mm)



(f)SEM of the chip in profile×500
(v_c=125m/min, f=0.12mm/r, a_p=0.4mm)

FIGURE 5 The shape of chip

Compared with the Figure 5(a),(b),(c) and (d),(e),(f), when the values of the cutting speed and feed speed were larger, the serrate level was higher, the spacing of rough slice layers was wider and the angle of shear was larger[15]. The reason was: the strain rate was increased when the cutting speed increased, moreover the metal removal in unit time was increased when the feed speed increased, the power consumption and the cutting heat was increased, the deformation time of the chip was shorten, it caused that the cutting heat was produced more fast and diffused more harder. These changes all of the above caused the adiabatic shear was aggravated, improved the deformational degree of the chip [16]. The saw-tooth chip caused the cutting force fluctuated and then influenced the machining precision, surface roughness and the tool wear [17].

4 Conclusion

1. The dominant factor affecting the cutting force is cutting depth, next is feed speed and the minimum is cutting speed when dry turning hardened steel by CC6050, the radial cutting force is maximum, next is main cutting force and the minimum is the axial cutting force.

2. The dominant factor affecting surface roughness is feed speed, next is cutting depth and the minimum is cutting speed.

3. There is obviously crater wear on the rake face and steady strip shaped wear on the flank of CC6050. The mainly wear mechanism is abrasive, diffusion, adhesive and oxidation wear.

4. The chip characteristics is saw-tooth when dry turning hardened steel with ceramic tool, the cutting speed and feed speed are larger, the serrated level of the chip is higher, the spacing of the rough slice layers on the upside of the chip is wider and the angle of shear is larger.

Acknowledgements

This research was funded by the industrial support projects of JiangSu Province in china (BE2011076) and the technology special project of Suzhou city in china (ZXG201131). Thanks go to the Soochow University, for its support in these experiments.

References

[1] Elbestawi M A, Chen L, Becze C E, El-Wardany 1997 *High-Speed Milling of Dies and Molds in their Hardened State*, CIRP Annals - Manufacturing Technology **46** 57-62

[2] Sokvic M, Mijanovic K 2001 Ecological Aspects of the Cutting Fluids and its Influence on Quantifiable Parameters of the Cutting Processes *Journal of Materials Processing Technology* **109** 181-9

[3] Yu S, Ning H, Liang L 2010 Effect of Refrigerated Air Cutting Tool Wear in High-speed Cutting of Difficult-to-cut Materials *Tribology* **30**(5) 485-90

[4] Bairu L, Wanhua H 2005 Application of Ceramic Cutting Tool in Machining of Hardness Material *Coal Mine Machinery* **2** 65-7

[5] Xiaobin C, Jun Z 2011 Research on Failure Evolution of Ceramic Cutting Tool in Intermittent Turning of Hardened Carbon Steel *Machine tool and hydraulics* **5** 1-3

[6] Benga G C, Abrao A M 2003 Turning of Hardened 100Cr6 Bearing Steel with Ceramic and PCBN Cutting Tools *Journal of Materials Processing Technology* **143** 237-41

[7] Xiaoguang W, Dongyang L 2006 Experimental Study on Dry Turning of Hardened Steel with Ceramic Cutters *Manufacturing Technology and Machine Tool* **2** 90-2

[8] Xuhong G, Shenlin Z, etc. Research on Wear Mechanism of Ceramic and PCBN Cutting Tools *Mechanical Engineer Material* **11** 10-2

[9] Liguo C 2000 Cutting Performance of PCBN Tools on Dry Turning *Paper of master's degree* **3** 22-4

[10] Baodai W, Huangbing G, etc. Analysis of Cutting Progress on Hardened Steel *Journal of Guangxi University (Science & Philosophy)* **3** 39-43

[11] Chunjuan T, Xuhong G, Xiao G 2012 Experimentation Research on Surface Roughness in Turning Hardened Steel with High Hard Cutting Tool *Materials for Mechanical Engineering* **3** 89-92

[12] Barry J, Byrne G 2001 Cutting Tool Wear in the Machining of Hardened Steels Part I: Alumina / TiC Cutting Tool Wear *Wear* **247** 139-51

[13] Chunjuan T, Xuhong G, Shaohua W 2013 Wear Property of Cutting Tools with High Hardness in Dry Turning Hardened Steel *Materials for Mechanical Engineering* **10** 55-9

[14] Katuku K, Koursaris A, Sigalas I 2009 Wear, Cutting Forces and Chip Characteristics When Dry Turning ASTM Grade 2 Austempered Ductile iron with PCBN Cutting Tools under Finishing Conditions *Journal of Materials Processing Technology* **209** 2412-20

[15] Dong L, Honghai X 2009 Research on the Shear Localization Plasticity in Serrated Chip Formation *Journal of Plasticity Engineering* **4** 203-7

[16] Chunzheng D, Minjie W, Haiyang Y 2012 Microstructure and Formation Mechanism of Deterioration Layer in Hardened Steel Chips During High Speed Machining *Materials science & Technology* **6** 127-31

[17] Guosheng S, Zhanqiang L, Jing D, Qibiao Y 2010 Description of Serrated Chip Deformation and Its Morphology Evolution *Transactions of the Chinese Society for Agricultural Machinery* **11** 223-7

Authors

Tu Chunjuan, born on January 29, 1982, Jiangsu, China

Current position, grades: Lecturer in Suzhou Institute of Industrial Technology
University studies: Department of Mechanical Engineering in Jiangsu University of Technology
Scientific interest: Cutting mechanism, Precision finishing, Process parameters Optimization.
Publications: 1 Patent, 10 Papers
Experience: 2004-present, working in Suzhou Institute of Industrial Technology. 2008-2010, studying for a master's degree in mechanical engineering in Soochow University. 2000-2004, studying for a bachelor's degree, majoring in mechanical and electrical integration in Jiangsu University of Technology.

Guo Xuhong, born on June 22, 1963, Gansu, China

Current position, grades:: Professor in Soochow University
University studies: Department of Mechanical Engineering in Lanzhou University of Technology
Scientific interest: Cutting mechanism, Precision finishing, Process parameters Optimization.
Publications: 2 Patents, 40 Papers
Experience: 1998-present, working in school of mechanical and electric engineering, Soochow University. 1997- 1998, further studying in Sunny electronic(Nagano, Japan) Co. 1996-1997, teaching in Department of Mechanical Engineering of Soochow silk engineering college. 1990- 1995, teaching in process department of Hubei University of Automotive technology. 1987- 1990, Studying for a master's degree in Metal processing laboratory of South China University of Technology. 1983-1987, working in Hubei University of Automotive technology. 1979-1983, studying for a bachelor's degree, majoring in manufacturing technology and equipment in Lanzhou university of technology.

Design of reflective optical fiber displacement sensor using double optical paths

Wenlong Du*

Computer and Communication Engineering, Huaian College of Information Technology, China

Received 1 March 2014, www.cmnt.lv

Abstract

A reflective optical fiber displacement sensor using double optical paths is designed based on the principle of reflective optical fiber displacement sensor. The sensor has double optical paths composed of random fiber and semicircle fiber. The output signal of double optical paths is disposed by using the preamplifier, intermediate amplifier, band-pass filter and AC to RMS circuit. The ratio and the displacement of the output signals of the two optical paths have linear relationship, effectively enlarging the measuring range on basis of the accuracy of measurement.

Keywords: Displacement Measurement, Reflection, Optical Fiber

1 Introduction

Optical fiber displacement sensor makes up of LIOFDS (Laser Interferometric Optical Fiber Displacement Sensor) based on phase modulation and ROFDS (Reflective Optical Fiber Displacement Sensor) based on intensity modulation. The former is used for detection by interference principle of coherent light, so it has a good performance in accuracy, but it requires the optical source stable and has a high price [1]. However, the later has lots of advantages – simple structure, stable performance, low price, flexible design and working in harsh environments, thus being widely used and occupying a very important position in optical fiber sensors. The output characteristics of ROFDS present a parabolic approximation, so the linearity is low within the measurable range. Only the front slope with a high linearity is used for measurement in actual application, thus limiting the measuring range and application [2]. Because the semicircle type fiber displacement sensor and the random type fiber displacement sensor have different peak positions, the ratios are calculated as the outputs of the former divided by those of the latter. So the obtained ratios and the displacements have a good linear relation. The function is realized by circuit design, making good compensation and enlarging the range of fiber displacement sensor.

2 Principles for temperature measurement with fluorescent lifetime

Figure 1 shows the principle diagram of ROFDS. The fiber, divided into two parts – the transmitting fiber and the receiving fiber, is used for signal transmission. The surface of the reflector corresponding to the fiber probe is a reflecting surface. The optical emitted by the optical source goes through the transmitting fiber and irradiates displacement reflection. Then the reflected optical goes through the receiving fiber and is received by the photo sensor. The output intensity is determined by the distance between the reflector and the fiber probe, that is, if the displacement changes, the output intensity will change as well. The displacement is got by the detection of the output intensity.

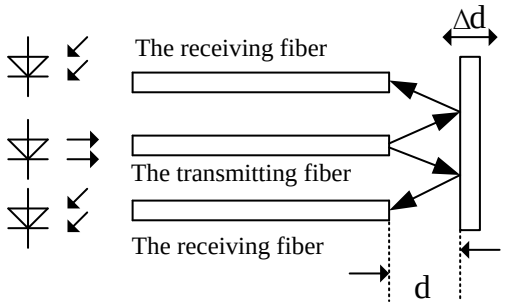


FIGURE 1 Principle diagram of reflective displacement fiber sensor

Figure 2 shows the method of equivalent coordinates analysis. Firstly, the mirror of the receiving fiber about the reflection is drawn; secondly, the optical intensity received from the optical field of the transmitting fiber end; at last, the equivalent results of the system is calculated as the optical intensity multiplied by the reflectivity of the reflection.

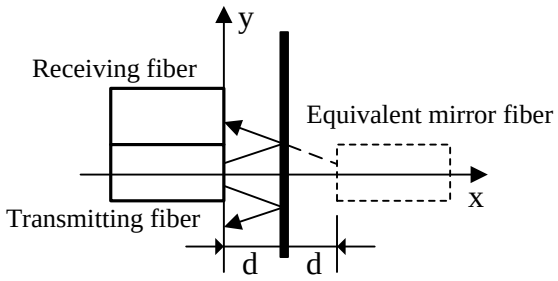


FIGURE 2 Equivalent coordinates analysis

For the multimode fiber, the output optical field flux distribution function of the fiber end is calculated as:

**Corresponding author* e-mail: duwenlongha@163.com

$$\Phi(d,r) = \frac{k_0 I_0}{\pi(a_0 + K\tan\theta_c d^{3/2})^2}\exp(-\frac{r}{a_0 + K\tan\theta_c d^{3/2}}), \quad (1)$$

where I_0 is the optical intensity of the transmitting fiber coupled with the optical source, $\Phi(d,r)$ is the optical flux density in the position (d,r) of output optical field, k_0 is the optical power loss coefficient of the transmitting fiber, a_0 is the fiber radius, K is the coupling parameter, θ_c is the fiber maximum incidence. If the fibers of the same kind are put in the output optical field as the detection receivers, the received optical intensity can be expressed as:

$$I(d,r) = \iint_s \frac{kk_0 I_0}{\pi R^2(d)}\cdot\exp(-\sum_i \eta_i r_i)\cdot\exp(-\frac{r^2}{R^2(d)})ds . \quad (2)$$

In equation (2) $R(d) = (a_0 + K\tan\theta_c d^{3/2})$, where k is the optical power loss coefficient of the receiving fiber, $\exp(-\sum_i \eta_i r_i)$ stands for the added loss because of the winding of the receiving fiber, S is the core area. For convenience, in the output optical field, the optical intensity in the center point of the receiving fiber end can be as the average optical intensity of the whole core surface. Approximately, the detected optical intensity of the receiving fiber end can be calculated by:

$$I(d,r) = \frac{Skk_0 I_0}{\pi R^2(d)}\cdot\exp(-\sum_i \eta_i r_i)\cdot\exp(-\frac{r^2}{R^2(d)}) . \quad (3)$$

Figure 2 shows that the optical intensity received by the fiber in the reflected optical field is equivalent to the receiving optical intensity of the mirror fiber multiplied by the reflection coefficient R of the reflector. The coordinate of the equivalent receiving fiber center (shown as the dashed line) is $(2d,r)$, where r is the distance between the center of the transmitting fiber and that of the receiving fiber. The coordinate is substituted in equation (3):

$$I(2d,r) = \frac{S_1 k_0 k_1 I_0}{\pi R^2(2d)}\cdot R\cdot\exp(-\sum_i \eta_i r_i)\cdot\exp(-\frac{r^2}{R^2(2d)}) . \quad (4)$$

Equation (4) shows that when the reflector, that is the detected object, is determined and the incident optical source remain stable, the modulation function of the fiber probe will be related to the displacement d. Namely the change of optical intensity reflects that of the displacement. So, the measurement of the displacement is transferred into that of the optical intensity [3].

3 Design of reflective fiber using double optical paths

At the end of the fiber probe, the transmitting fiber and the receiving fiber have four kinds of distribution: the random distribution, the semicircle distribution, the coaxial internal emission distribution and the coaxial external emission distribution. In the work, the random type and the semicircle type fiber are used. The transmitting and receiving fiber at the end of the random type are randomly distributed; the transmitting and receiving fiber at the end of the semicircle type occupy the semicircle region. Figure 3, the output curve of random type fiber and the semicircle type fiber, shows that the ROFDS has a poor linearity within the measuring range. Both the semicircle type and the random type have a peak value in the output curve, and the displacement related to the peak value of the random type is smaller than that of the semicircle type. Meanwhile, the curves have a bad linearity in the top region. In the actual measurement, only the front slope with a good linearity is applied, limiting the range of measurement and application. According to reference [4], the ratio of the output signals of the semicircle type and the random type probe (B/A) has a good linear relation with the displacement. Figure 4 shows that if the B/A output characteristic curve is fitted by the least square method, the nonlinear error will be less than 2.5%. So the double optical paths are used in the design. Figure 1 shows that one receiving fiber uses the random type, the other uses the semicircle type. The nonlinear error of the sensor is decreased by the method and the using range of the fiber displacement sensor is enlarged to about 4 mm.

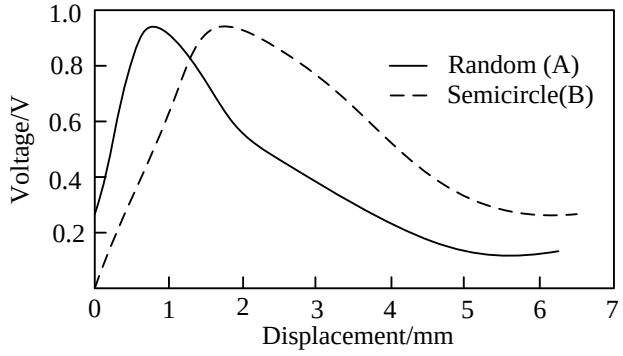


FIGURE 3 The output curve of random type fiber and semicircle type fiber

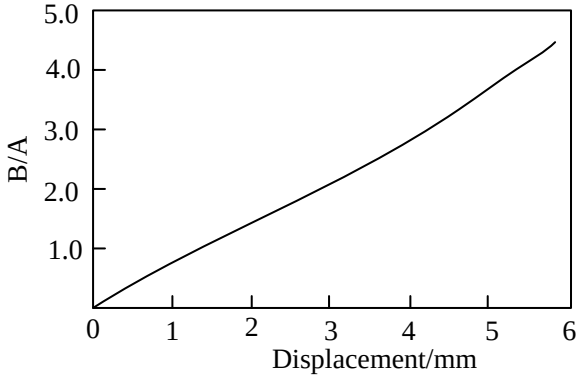


FIGURE 4 B/A output characteristic curve

Figure 5 shows the reflective fiber displacement measurement system using double optical paths. It is made up of the light source driver, the photoelectric detector, the preamplifier, the filter and the AC to RMS conversion circuit. The

working process is as follows: firstly, The LED, driven by the square wave generator made up of 555, emits the optical pulse waves with a wavelength of 940 nm suitable to the low loss window of the quartz fiber. The optical pulse waves are focused by the lens, poured into the transmitting fiber, received by the semicircle and random type fiber after being reflected and transformed to the electric signals through the photoelectric devices. The electric signals are transformed to the DC signals through the preamplifiers, the filters and AC to RMS conversion. The DC signals are sent to the divider to get the ratio of the optical intensities of the two receiving fibers in the form of DC. After A/D module transformation of the SCM MSP430F449, the ratio-displacement function is obtained by linearization of the measuring data, using polynomial interpolation. In actual measurement, the detected displacement can be obtained by substituting the ratio of a position back into the function.

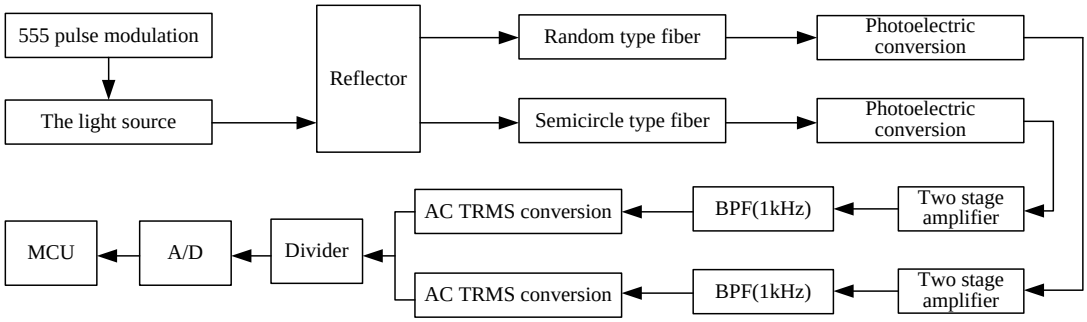


FIGURE 5 Diagram of the displacement measuring system

4 Design of light source driving and signal processing circuit

4.1 LIGHT SOURCE DRIVING CIRCUIT

The colorless and transparent LED L51R4 is selected as the light source in the circuit. It is considered that the error signals caused by external disturbing light are usually slowly changing DC or random signals, and those caused by the dark current of the photoelectric detector and the drifting of the amplifier circuit are also DC error signals. In order to eliminate these error signals, the LED is driven by the square wave driving circuit. The pulse is generated from the 1 KHz square wave generator, made up of the timer 555 and the peripheral resistance capacitance element. By using this method, the LED is in the intermittent working state with low power consumption, making the fiber sensor system work steadily for a long time. When the displacement changed, the amplitude of the pulse light will be modulated by the detected displacement signals. So, the signals arriving at the photoelectric detector are the wave signal after amplitude modulation with a message of displacement.

4.2 PREAMPLIFIER AND THE INTERMEDIATE AMPLIFIER

To get stable outputs of the preamplifier, the integral type light I/V conversion circuit is used in the system. The integrated circuit LF351 with low power consumption and noise is used for the preamplifier. Figure 6 shows the practical preamplifier circuit with low noise, where C_1 is the anti-hunt capacity used to make the performance of the preamplifier more stable. C_1, usually valuing from 30 to 100 pF, is 68 pF here. And the output voltage of the preamplifier is expressed as:

$$U_o(t) = -I_{ph}R_f \,, \tag{1}$$

where I_{ph} is the output light current of the detector; R_f is the feedback resistance. In order to make the preamplifier have enough linear dynamic range and provide enough output voltage signals for the lower amplifier; R_f values 2.2 $M\Omega$. The dynamic range of the output signals is about 0.7 V, being in the linear part of the preamplifier [5].

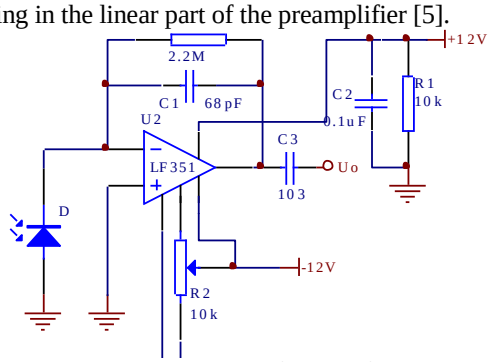


FIGURE.6 Circuit of preamplifier

In order to make the preamplifier have stable performance and good linearity within the dynamic range of the signals, the I/V conversion coefficient should not be too large. Meanwhile, considering the working performance of the narrow band filter, an intermediate amplifier is set between the preamplifier and the filter circuit. The circuit LF353A, integrated by double operation amplifiers with low power consumption and noise, is used to make up of intermediate amplifier with double paths, and the gain G, determined based on the whole condition of the system, values 5 more or less.

4.3 BAND-PASS FILTER

In this system, the static displacement is measured. If the light source is driven by using the pulse driving mode, then

only the 1 KHz pulse optical intensity modulation signals will be related to the displacement variable. Besides, all the high and low frequency signals caused by the stray light and the electromagnetic field, except the above signals, are disturbing signals. The voltage controlled voltage source filter circuit is used to filter the noises and unnecessary frequency components added in the detected signals. The 1 KHz signals can pass through the filter, and the signals with different frequencies decay fast, increasing the signal-to-noise ratio of the system. Figure 7 shows the circuit and Fig. 8 shows the filter amplitude-frequency characteristics. The band-pass width is a narrow frequency range making the modulation frequency as the center frequency.

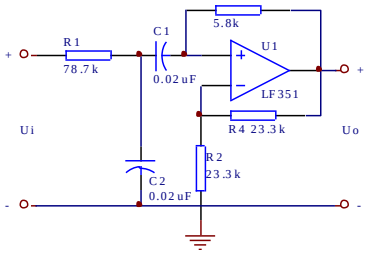


FIGURE 7 Circuit of BPF at 1.0 kHz

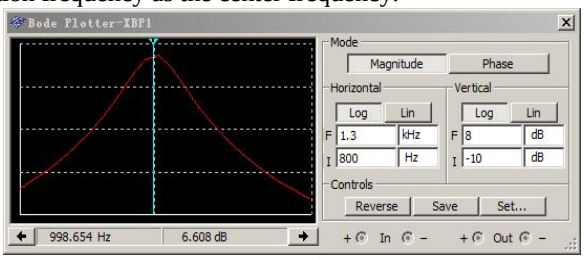


FIGURE 8 Filter amplitude-frequency characteristics

4.4 AC TO RMS CIRCUIT AND DIVIDER

Compared with the average conversion circuit, the RMS to DC integrated chip has many advantages, e.g. high integrated level, perfect function, less peripheral component, simple circuit connection and good performance in electric. So, the chip AD536 is selected to measure the RMS of all kinds of voltage wave form accurately and in real time. The high precision conversion within the measuring range can be achieved by only a capacitance, without considering the wave form parameter and distortion. AD536, supplied by double sources voltage or single source voltage ranging from +5V to +36V, compute the RMS of the input signals of AC or AC plus DC and output a equivalent DC signal. The maximum error within the wider dynamic range is 0.5%.

The divider is used to eliminate the wave of the light source, the difference of the reflectivity and the influence of the fiber loss [6]. The preset adjustment single chip multiplier or divider AD532 can control the maximum multiplication error within ±1.0% and the range of the output voltage within ±10V. The output data are converted to digital signals after entering the A/D converter. Then the digital signals are entered to the MCU and the results are achieved and shown on the MCU.

5 Data analysis and results

The high precision dial gauge calibrator with a resolution of 0.5 μm and a range of 0.5 μm is used as the calibration device to make a calibration to the system within effective measuring range. The reflector is controlled to move along the radial of the fiber, making the received light quantity decrease, and the output voltage is measured at certain intervals of displacement. Then the linearization revise is performed to the measuring results by using the least square method. The MCU system reads all the A/D conversion results, calibrates them, computes the gradient k and the intercept b of the fitting lines and stores them in ROM. In actual measurement, the fitting coefficients are called to compute the displacement according the measuring results. Table 1 shows the results of the experiment.

TABLE. 1 Experiment data of the displacement sensor

Displacement (μm)	Actual Output	Fitting Value	Displacement (μm)	Actual Output	Fitting Value	Displacement (μm)	Actual Output	Fitting Value
0	3.573	3.7690	400	2.749	2.6954	750	1.776	1.7560
50	3.474	3.6348	450	2.627	2.5612	800	1.649	1.6218
150	3.349	3.3664	500	2.478	2.4270	850	1.498	1.4876
200	3.251	3.2322	550	2.326	2.2928	900	1.379	1.3534
250	3.148	3.0980	600	2.179	2.1586	950	1.124	1.2192
300	3.026	2.9638	650	2.024	2.0244	1000	1.001	1.0850
350	2.873	2.8296	700	1.901	1.8902	1050	0.903	0.9508

The fitting linear equation of the working curve by using the least square method can be calculated as:

$$y = -0.002684 * x + 3.769$$

The degree of linearity of the sensor is 2.3%, the resolution of the sensor reaches 0.5 μm within the range of 0~1.2 mm and the repeatability error is 0.6%. Fig. 9 shows the experimental results and the fitting line. In application, the reflector cannot be replaced because the surface roughness of the reflector has influence on the change of the reflective optical.

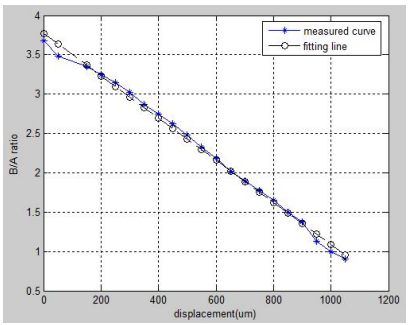


FIGURE 9 Measured result of displacement sensor

6 Conclusions

The output characteristic curve of ROFDS has a good degree of linearity in the fore slope. The traditional fiber displacement sensor only uses the fore slope, ensuring good degree of linearity, but greatly reducing the measuring range of the fiber sensor. In the work, ROFDS is designed to do division to the output signals of double optical paths using the different output characteristics of different types of fiber probes to get the ratio that has good linear relationship with the displacement, thus ensuring the degree of the linearity and enlarging the measuring range. In addition, the fiber is sensitive to temperature and vibration, and easily affected by environment. These problems are to be improved in the process of application for fiber sensors.

References

[1] Grattan K T V, Sun T 2000 Fiber Optic Sensor Technology: An Overview *Sensors and Actuators* **82**(l-3) 4-61
[2] Culshaw B 2000 Fiber optics in sensing and measurement (Selected Topics) *IEEE Journal of Quantum Electronics* **6**(6) 1014-21
[3] Arun Kumar, Varshney R K, Rakesh Kumar 2004 SMS fiber optic microbend sensor structure: effect of the modal interference *Optics Communications* **232**(2004) 239-44
[4] Faria J B 1998 A theoretical analysis of the bifurcated fiber bundle displacement sensor *IEEE Transactions on Instrumentation and Measurement* **47**(3) 742-7
[5] Fairchild Semiconductor Corporation. LF351 Single Operational Amplifier (JFET) 2001
[6] Analog Devices Inc. Internally Trimmed Integrated Circuit Multiplier AD532 2001

Authors

Wenlong Du, born in November, 1973, Tongshan, Jiangsu Province, China

Current position, grades: Lecturer, Master, Engineer
University studies: Communication Technology
Scientific interest: Communication Technology
Publications: 15 papers published in the international or national journals
Experience: He is a lecturer of Huaian College of Information Technology, China. His research interests include communication technology and computer engineering.

Optimization of immune particle swarm algorithm and application on wireless sensor networks

Jiang Fei*

Laboratory of Intelligent Information Processing, Suzhou University, Suzhou 234000, China

School of Information Engineering, Suzhou University, Suzhou 234000, China

Received 1 June 2014, www.cmnt.lv

Abstract

As a new network technology, wireless sensor network (WSN) have been applied to military, intelligent transportation, environmental monitoring and other fields. Localization is one of the important support technologies of Wireless Sensor Networks. Location information is important to network monitoring. It is meaningless, if there is no location information. We need to adopt a certain mechanism and algorithm to implement the localization of Wireless Sensor Networks. Based on the analysis of the features of Wireless Sensor Networks on range-free positioning algorithm and DV-Hop positioning algorithm error sources, this paper focuses on the improvement research on DV-Hop positioning algorithm. Inspired by biological immune system and mechanism, this paper introduces the immune information processing mechanisms in the immune system to the particle swarm optimization algorithm and thus gets an immune particle swarm optimization (IPSO) algorithm. By applying into the running of DV-Hop positioning algorithm, the paper proposes a DV-Hop improved algorithm, which is, the WSN positioning algorithm based on IPSO algorithm. Simulation experiments show that the improved algorithm can significantly reduce positioning errors to improve positioning accuracy.

Keywords: wireless sensor networks, immune particle swarm optimization, DV-Hop positioning algorithm

1 Introduction

Wireless Sensor Networks positioning control is a basic problem in WSN applications. It is a study on how to maximize the network coverage to provide reliable monitoring and tracking services in the guaranteed quality of service conditions. Effective strategies of the coverage control and algorithms can be used to optimize the allocation of resources of WSN, increase the efficiency of the energy usage of network nodes, and improve the perceived quality of service and the overall survival time. How to combine different environmental demands and design a practical strategy for positioning is a significant research field.

Particle Swarm Optimization (PSO) have the advantages of natural paralleled, strong robustness, and global optimization, and have obvious advantages in dealing with the complex problems. Therefore, the PSO will be very suitable to solve the nodes positioning problem in the WSN. This paper makes the investigations on the PSO in solving the nodes positioning problem in the WSN, and proposes coverage mechanism based on an immune particle swarm optimization algorithm to optimize the problem. The paper gives a review on the WSN and its nodes optimal coverage problem. Moreover, the implementation details for the IPSO solving the nodes optimal positioning are described, including the particle code representation, and the complete algorithm flowchart. Simulations have been conducted and the results are compared with the ones obtained by other algorithms. The experimental results and comparisons demonstrate the effectiveness and efficiency of our proposed algorithm.

At last, based on simulation platform, the paper makes a simulation experiment on the existing DV-Hop improve algorithms and the new DV-Hop positioning algorithm based on IPSO algorithm. Through the simulation experiment results, this paper analyses the merits of the IPSO algorithm. This paper is highlighted with the contributions to the theoretical researches and practical applications of the PSO and the WSN as follows: Firstly, the IPSO is successfully applied to provide a new solution to the nodes optimal positioning in the WSN. Secondly, the proposed algorithm is simpler than the traditional algorithms, and with a higher computational effectiveness. Thirdly, the proposed algorithm has a better performance that can obtain higher positioning accuracy.

2 Basic particle swarm optimization

The basic PSO could be described as follow: supposing in a n-dimension research space, the population is $X=\{x_1,...,x_i,...,x_m\}$, among which the position of I particle is $X_i=\{x_{i1},x_{i2},\cdots,x_{in}\}^T$, the speed is $v_i=\{v_{i1},v_{i2},\cdots,v_{in}\}^T$, the individual extremum is $p_i=\{p_{i1},p_{i2},\cdots,p_{in}\}^T$, the global extremum is $p_g=\{p_{g1},p_{g2},\cdots,p_{gn}\}^T$. Based on the principle of following the optimized particle, the particle X_i will change

* ***Corresponding author's*** e-mail: 2954043892@qq.com.

the speed and position in accordance with the following Equation, as is shown in the Figure 1:

$$v_{id}^{(k+1)} = v_{id}^{(k)} + c_1 r_1 (p_{id}^{(k)} - x_{id}^{(k)}) + c_2 r_2 (p_{gd}^{(k)} - x_{id}^{(k)}) . \quad (1)$$

$$x_{id}^{(k+1)} = x_{id}^{(k)} + v_{id}^{(k+1)} . \quad (2)$$

In the Figure 1, m is the sum of the particles in the population. d=1,2,…,n, n is the dimensions of particle search space. $v_{id}^{(k)}$ is the iterative particle i's Flight velocity vector's d dimension component at the K time, which is between $-v_{d\max}$ and $v_{d\max}$. $v_{d\max}$ could be set according to different controlling variables by the users. $x_{id}^{(k)}$ is the iterative particle i's d dimension component at the K time. $p_{id}^{(k)}$ is the iterative particle i's best position at the K time p_{best} 's d dimension component. $p_{gd}^{(k)}$ is the iterative group's best position at the K time g_{best} 's d dimension component. c_1, c_2 is the learning factor or speeding up factor. r_1, r_2 is the random figure among [0,1] produced by random function [1, 2].

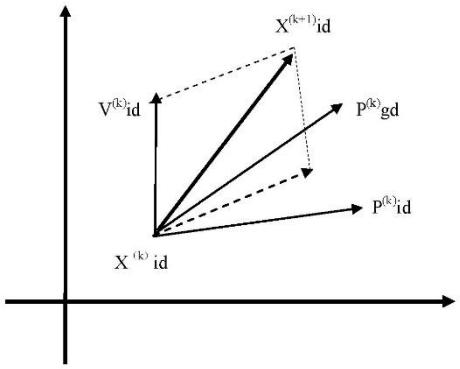


FIGURE 1 The d dimension variable diagram of basic particle

In Equation (1), a constant inertia weighting factor ω could be added before $v_{id}^{(k)}$. Normally, the ω's Figure 1 and Figure 2 is the formula of basic particle swarm algorithm formula.

In Equation (1), three are three parts to calculate the new speed of particle i: the first part is the $\omega v_{id}^{(k)}$, the speed at the former time, indicating the status of the particle at that moment. The second part is the $c_1 r_1 (p_{id}^{(k)} - x_{id}^{(k)})$, the distance of the current position and the best position of particle i, which is the part for self-recognition, indicating the thoughts for itself. The third part is the $c_2 r_2 (p_{gd}^{(k)} - x_{id}^{(k)})$, distance of the current position of particle i and the best position of the group, which the social part and the best vector from current point to the group, indicating the cooperation and knowledge share between particles.

All the three parts determine the spatial search capabilities of the particles. The first part has played a role in the balance of global and local search capability; the second part of the particles has a sufficiently strong global search capability; third part reflects the sharing of information between the particles. Under the combined effect of these three parts, the particles thus could effectively reach the best position. Particle i calculates the coordinates of the new location by the Equation (2), and determines the position and velocity of movement next by Equations (1) and (2).

There are few parameters need to be adjusted in PS algorithm. But the setting on the parameters makes a great impact on the performance of the algorithm. The parameters as well as experience setting are listed below: inertia weight ω has a better performance in [0.9, 1.2], and performs better from 1.4 to 0 than the fixed. The largest speed $v_{d\max}$ also influences the ω, generally speaking, when the $v_{d\max}$ is small, ω is about 1 and when $v_{d\max}$ is big, the ω is about 0.8 with the better performance of PSO algorithm.

Learning factor c_1 and c_2 are used to control the particle's own memory and the memory of companions. Selecting the appropriate algorithm can improve the speed and avoid local minima. Experiments showed that $c_1 = c_2 = 2$ or $c_1 = c_2 = 0.5$ is a good choice, but $c_1 + c_2 \le 4$ is better.

When the particles constantly adjust their positions according to the speed, they are restricted by the maximum speed $v_{\max}$. When $v_{id}^{(k)}$ exceeds $v_{d\max}$, the particles would be defined as $v_{d\max}$. Maximum speed $v_{d\max}$ determines the maximum travel distance of the particles in a loop, which typically should not exceed the width of the particles. If $v_{d\max}$ is too large, the position of the particle may fly optimal solution; If $v_{d\max}$ is too small, it may reduce the global search ability of particles [3].

Particles flying speed cannot exceed the maximum speed $v_{d\max}$ of the algorithm set. In general, in

$$v_{\max} = (v_{1\max}, ..., v_{d\max}, ..., v_{n\max})^T =$$
$$a * (v_{1\max} - v_{1\min}, ..., v_{d\max} - v_{d\min}, ...v_{n\max} - v_{n\min})^{T_1}$$

$x_{d\max}$ and $x_{d\min}$ represent the upper and lower limit values of the particles of the d-dimensional variables. a is a control factor, 0<a<1. Setting a large $v_{\max}$ could promote the global search capability of particle populations, while when $v_{\max}$ is small, the local search capability could be strengthened. The numbers of particles (population size or population size) generally could use 20~60. For general optimization problems, 10 particles are enough. For some special problems, the number of particles can take up to 100~200. Particle dimension is determined by the dimension of the optimization problem (solution space dimension) [4].

3 Immune particle swarm optimization process

The design flow chart of IPS algorithm used in this paper is as shown in Figure 2:

1) Initialization parameters. The algorithm is given an initial value of the inertia factor $\omega(0)$, the acceleration factor c_1, the number of immune population of particles N, the maximum number of iterations (the number of generations) K_{max} and other parameters.

2) Initial antibody (population) yields. N immune particles are randomly generated and the "flying" speed is V_i, so that to form the initial immune particle population P_0. Set the value of each individual particle initial immunization optimal solution P_{best} and the global optimal solution G_{best}.

3) Antibodies fitness evaluation. In the current population, to calculate all antibodies (immune particle) fitness. In the immune particle swarm algorithm, the fitness function usually takes the transformation of some function's objective function which needs to be optimized [5]. In the design of the main fitness function, it should satisfy the following conditions:

a) Reasonable and conformed requirements. The fitness should reflect the degree of solution of the pros and cons of the corresponding value solution; small;

b) small task of calculation. The amount of the fitness function should be designed as simple as possible;

c) strong versatility. The fitness for certain specific issues should be commonly used as possible as it can.

d) Calculation of antigen and antibody binding strength and concentration of antibody. Corresponding antigen immune system optimization objective function corresponding antibody solution optimization problems. To decide antigen and antibody binding force based on the value of the objective function; to decide the binding force of the antibody based on the degree of similarity solution and use the combination of these two forces to make solution evaluation and selection.

The binding force between antibody and antigen is defined as A_v, and could be changed by the target function into:

$$A_v = \mu[f(v)]. \tag{3}$$

In Equation (3): $f(v)$ is the target function, $\mu(x)$ is the monotone function of x, which reflects the evaluation to the target function [6].

This article taken sorting selection method, the size of the binding force A_v between the antibody and antigen has no absolute meaning. A_v value does not need to guarantee positive, so the reactive power optimization problem studied in this paper can be the binding force A_v between the antibody and antigen and directly take reactive power optimization objective function model:

$$A_v = f(v). \tag{4}$$

Process of solving the objective function is as follows: the control variable is calculated based on the decision of the antibody, and then derived value of the state variables into the objective function optimization problems. Then put the objective function into the Equation (4), which can obtain antigen and antibody binding force. Since the target function of this paper is to get the minimum value of the solution function, the smaller the antigen and antibody binding force, the closer the optimal solution antibody to the optimization.

The binding force between the antibodies reflects the degree of similarity. When two antibodies are similar, the binding force is larger, on the contrary, the smaller, as is shown in the following equation:

$$B_{v,w} = \frac{1}{1+H_{v,w}}. \tag{5}$$

$H_{v,w}$ is the vector from the definition of the two antibodies type. In this paper, Euclidean space 2-norm can be used to define the sense of distance:

$$H_{v,w} = \left\{\sum[(V_{Giv}-V_{Giw})^2+(D_{civ}-D_{ciw})^2+(T_{iv}-T_{iw})^2]\right\}^{\frac{1}{2}}. \tag{6}$$

Thus the calculation of antibody density c_v could use the following Equation:

$$c_v = \frac{\sum_{w=1}^{N} B_{v,w}}{N}. \tag{7}$$

N is the number of antibody populations. The larger the concentration of the antibody is, the greater the similarity of such antibodies and other antibodies is.

5) Memory structure and unit prohibition. The largest concentration of antibody population antibody showed that it accounted for in the antibody group in an absolute advantage, namely to achieve a relatively optimal solution, which reflects the evolution of the results of this generation of population. The largest concentration of antibody should be retained by the principle to retain it which is the role of memory structure. The high concentration of antibodies and other antibodies showed large similarities between large groups, in order to maintain the diversity of the antibody, it needs higher concentrations of antibody suppression. Setting the antibody according to the size of its concentration, the author sort through the set-out rate (This article set it as 10%), finding that the higher concentrations of antibodies were eliminated and new randomly generated antibodies and antibody replaced the eliminated. Inhibiting the high concentrations of antibodies can well prevent the development of antibodies from being a single group so as to maintain the antibody population diversity and avoid local optimal solution. That is the role of inhibition of cell.

Fitness value of antibody v showed its performance. The fitness value calculation function could be formed by the binding force A_v and the concentration c_v:

$$p = \alpha A_v + (1-\alpha)c_v. \tag{8}$$

α is a scaling factor ($0<\alpha<1$) antibody V's adaptation has the following characteristics: the smaller the antibody and the antigen binding capacity A_v, the smaller

the value of the corresponding adaptation; the smaller the antibody concentration, the smaller the value of the corresponding adaptation; the smaller the adaptation value, the closer the near-optimal solution. Setting the size of α could adjust the weight between antibody-antigen binding force and concentration, which can also retain the antigen binding force small (i.e., the objective function value is small) antibodies, but to ensure the diversity of individuals through inter-antibody concentration based on mutual promotion and inhibition mechanisms, so that the optimal solution for improving the convergence of favorable vicinity [7].

6) Group updating and the generation of new immune particle. To update particle speed and produce new immune particle *N* in accordance with the following equation:

$$v_{idid}^{(k+1)} = \omega^{(k)} v_{id}^{(k)} + c_1 r_1 (p_{id}^{(k)} - x_{id}^{(k)}) + c_2 r_2 (p_{gd}^{(k)} - x_{id}^{(k)}) \quad (9)$$

$$x_{id}^{(k+1)} = x_{id}^{(k)} + v_{id}^{(k+1)} \quad (10)$$

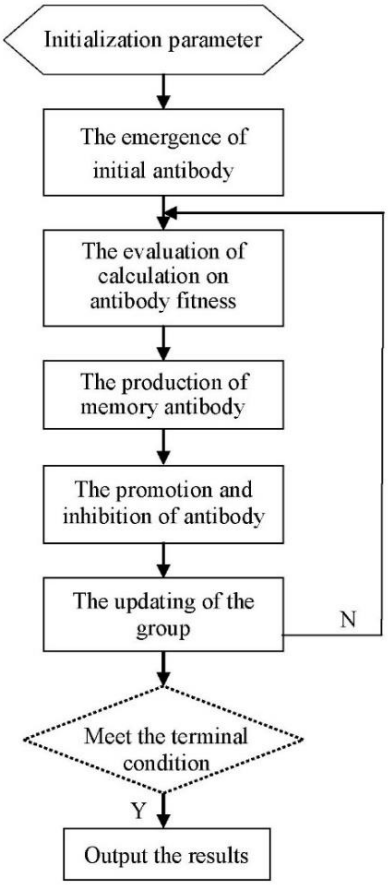


FIGURE 2 The flow chart of immune particle swarm optimization algorithm

PSO in the early age has a fast convergence but has a not high accuracy in the late convergence stage, which is easy to fall into local optimum. Therefore, the author designed the inertia factor ω, which will have a great impact to optimize the performance of the algorithm. The large ω values could help to improve the convergence speed while the small ω could help to improve the convergence of the algorithm accuracy. According to this feature, this paper used the Equation (11) to improve ω, so that it can adaptively adjust to achieve the purpose of decreasing ω with the increasing of the value of the iterations

$$\omega^{(k)} = (2/(1+e^{\sigma k / K\max}))\omega(0) . \quad (11)$$

σ is the positive coefficient; K_{max} is the set maximum number of iterations (the number of generations); $\omega(0)$ is the initial value; k is the current iteration [8].

7) To judge if it meets the termination condition. If being satisfied, to stop the running and put out the results, otherwise, go to step 3.

4 Simulation of improved algorithm

In order to evaluate the performance of immune particle swarm optimization algorithm, this paper makes a comparison on the DV-Hop algorithm based on IPSO and several other DV-Hop algorithms through simulation experiment. In the simulation, all the sensor nodes are arranged in the 100M, 100M square area. According to the relevant literature, anchor nodes arranged in a border area is relatively good accuracy. For this case, the paper put 100 unknown nodes and anchor nodes evenly on the boundary of the square area. The total proportion of the anchor node of the node changes as the experimental conditions.

While maintaining the other parameters constant, the author changed the proportion of anchor nodes in the total nodes, the communication range of the unknown nods and the three parameters respectively to make simulation experiments. Meanwhile, in order to make the results more convincing, this paper made a simulation experiment on the DV-Hop positioning algorithm based on IPSO and the improved several DV-Hop algorithms under the same experimental conditions and then made an analysis on the positioning results. By making a plurality of simulations on the same algorithm, the author took the average value of the final simulation results in order to reduce the error caused by the experiment itself and make the results more accurate. In the immune particle swarm optimization algorithm, the number of particles is 30, the largest and the smallest particle fitness values were max, min.

4.1 THE RELATIONSHIP DIAGRAM OF THE ANCHOR NODE PROPORTION AND THE AVERAGE POSITIONING ERROR OF THE NODE

Figure 3 shows the relationship diagram of the anchor node proportion and the average positioning error of the node. From this, it can be seen as follows:

(a) The positioning error ratio of several positioning algorithms decreased with of increasing of the anchor nodes. That is because if the anchor node ratio increases, it will provide more information and the corresponding positioning accuracy would be promoted. There is an extreme case, when all nodes are anchor nodes, the node can position itself, which is also conformed with the achievement situations.

(b) When the proportion of the anchor nodes is the same, the improved DV-Hop algorithm has a smaller positioning error compared to the original location DV-Hop algorithm, among which the DV-Hop algorithm based on immune particle swarm optimization has the minimal positioning error, which also shows a good performance of IPSO.

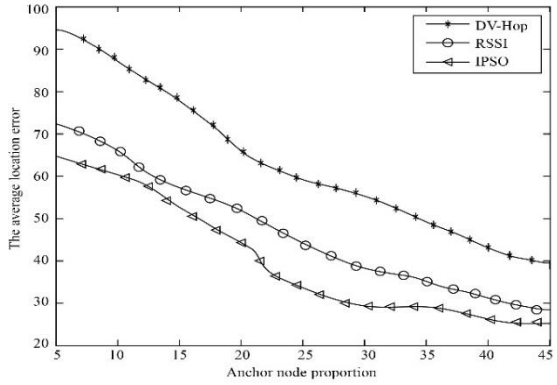


FIGURE 3 The relationship diagram of the anchor node proportion and the average positioning error of the node

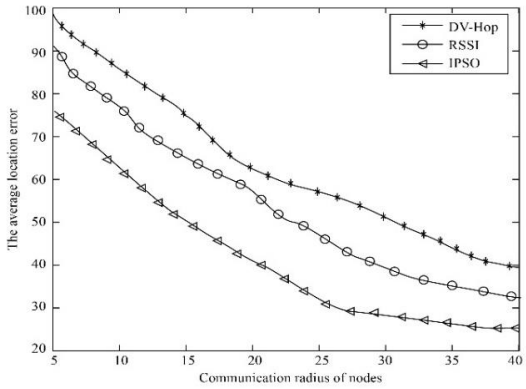


FIGURE 4 The relationship diagram of the communication radius and the average location error of the node

4.2 THE RELATIONSHIP DIAGRAM OF THE COMMUNICATION RADIUS AND THE AVERAGE POSITIONING ERROR OF THE NODE

Figure 4 shows the relationship diagram of the communication radius and the average positioning error of the node. From this, it can be seen as follows:

(a) The positioning errors of the shown positioning algorithm declines with the expansion of the unknown node communication radius for the expansion of the unknown node communication radius strengthens the corresponding connectivity of the unknown node with other nodes, thus the node can get a larger range of information exchange and to improve the positioning accuracy of the algorithm.

(b) Although the positioning errors of several positioning algorithms were decreased, it is relatively small, because even though the communication range of the node were increased to some extent, the exchange information of the number of the anchor node exchanged with the unknown node may be increased. And not every anchor node can provide useful information for locating the unknown nodes.

4.3 EFFECT OF TOTAL NODES ON AVERAGE POSITIONING ERROR OF THE NODE

Figure 5 shows the relationship diagram of total number of nodes and average positioning error. From this, it can be seen as follows:

(a) When the total number of nodes increases from 5 to 30, the positioning of several errors in the positioning algorithm has a certain impact. The average error of nodes declines with the increasing of the total number of nodes.

(b) With the change of the total number of nodes, the magnitude of change in average bit error localization algorithm is relatively large, which also reflects the scalability of the algorithm needs to be further improved.

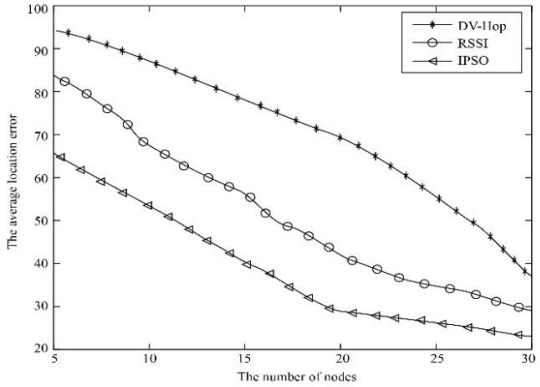


FIGURE 5 The relationship diagram of total number of nodes and average location error

5 Conclusion

This paper introduced the research background and research status at home and abroad for the WSN, and then introduced the Wireless Sensor Networks characteristics, as well as its structure and key technologies. While making a classification on positioning algorithms, this paper illustrated the basic principles of the typical positioning algorithm. On the basis of the analysis of the basic ideas of immune particle swarm optimization, this paper analysed the positioning error sources of DV-Hop positioning algorithm. By combining with the characteristics of immune particle swarm optimization algorithm, this paper proposed a new DV-Hop algorithm based on immune particle swarm optimization algorithm. Finally, the simulation experiments showed that the improved algorithm can significantly reduce positioning errors to improve coverage accuracy.

Acknowledgments

This work was supported by the Anhui Province Comprehensive of Course Specializations Pilot Projects NO. 2012zy075, the Suzhou University research platform Open Projects NO. 2012YKF37 and the Suzhou University Master Scientific Research Foundation Project NO. 2010yss02.

References

[1] Pratyay Kuila, Prasanta K Jana 2014 Energy efficient clustering and routing algorithms for wireless sensor networks: Particle swarm optimization approach *Engineering Applications of Artificial Intelligence* **33**(8) 127-40

[2] Mahmoodabadi M J, Salahshoor M Z, Bagheri A 2014 HEPSO: High exploration particle swarm optimization *Information Sciences* **273**(20) 101-11

[3] Panagiotis N, Kechagiopoulos G, Beligiannis N 2014 Solving the Urban Transit Routing Problem using a particle swarm optimization based algorithm *Applied Soft Computing* **21**(8) 654-76

[4] Ramazan Coban 2014 Power level control of the TRIGA Mark-II research reactor using the multifeedback layer neural network and the particle swarm optimization *Annals of Nuclear Energy* **69**(7) 260-6

[5] Zhao F, Li G 2014 Human–computer cooperative particle swarm optimization based immune algorithm for layout design *Neuro computing* **132**(20) 68-78

[6] El-Sherbiny M M, Alhamali R M 2013 A hybrid particle swarm algorithm with artificial immune learning for solving the fixed charge transportation problem *Computers & Industrial Engineering*, **64**(2) 610-20

[7] Zhai T, He Z 2013 Instance selection for time series classification based on immune binary particle swarm optimization *Knowledge-Based Systems* **49**(9) 106-15

[8] Kuo R J, Hung S Y, Cheng W C 2014 Application of an optimization artificial immune network and particle swarm optimization-based fuzzy neural network to an RFID-based positioning system *Information Sciences* **262**(20) 78-98

Author

Fei Jiang, born on October 186 1980, China

Current position, grades: lecturer at School of Information Engineering Suzhou University, China.
University studies: Master degree in Computer Application Technology, Hefei University, China in 2011.
Scientific interest: image processing and wireless sensor networks.

Authors' index

Cumulative Index

Mathematical and Computer Modelling

Jing Zeng An image threshold segmentation method based on multi-behaviour global artificial fish swarm algorithm

*Computer Modelling & New Technologies 2014 **18**(11) 38-42*

Firstly, this paper describes how the histogram analysis method pre-processes the images to be segmented. Then is makes a detailed analysis of the working principles and behaviour pattern of basic artificial fish swarm algorithm (AFSA); dissects the defects of AFSA in principle and proposes an improved AFSA with global convergence. Finally, it presents the main steps of image threshold segmentation method based on AFSA; compares the performances of AFSA with those of other intelligent algorithms and proves that this improved AFSA makes all-around improvements on image segmentation.

Keywords: Image Threshold Segmentation, Artificial Fish Swarm Algorithm

Ganzhou Wu, Haiyan Liang A modified BFGS method and its convergence

*Computer Modelling & New Technologies 2014 **18**(11) 43-47*

In this paper, a new modified BFGS method for unconstrained optimization problems is presented. The algorithm preserves the convergence properties of the famous BFGS algorithm. The relation between the new algorithm and a self-scaling quasi-Newton algorithm is revealed. If we assume the objective function is twice continuously differentiable and uniformly convex, we prove the iteration converge globally to the solution. And under some additional conditions, the superlinear convergence is given. Finally, the experimental results show that the proposed algorithm performs very well, which indicate that the numerical performance of the new algorithm is somewhat like the self-scaling quasi-Newton algorithm.

Keywords: modified BFGS method, convergence, unconstrained problems

Qian Chen, Junhao Liu, Chaowei Fu, Haitao Wang Performance analysis and side lobe suppression in radon-Fourier transform based on random pulse repetition interval

*Computer Modelling & New Technologies 2014 **18**(11) 48-54*

In order to solve the Blind Speed Side Lobe (BSSL) appeared in Radon-Fourier Transform (RFT) method used for dim target detection, a novel method of BSSL suppression is proposed in this paper. It is based on Random Pulse Repetition Interval (RPRI). The process of RPRI-RFT and the BSSL properties are described, the performance of coherent integration and the modulation noise of RFT algorithm based on RPRI are analysed in detail. Both the theoretical analysis and the numerical experimental results show that RPRI-RFT can be used to improve signal-to-noise ratio (SNR) and suppress BSSL effectively, and the influence of modulation noise of RPRI can be suppressed by the long-time integration characteristic, thus significantly improve the ability of low pulse repetition frequency radar to detect and measure long-range weak high-speed multi-targets.

Keywords: Blind Velocity Side Lobe, Radon-Fourier Transform, Random Pulse Repetition Interval

Kuangfeng Ning, Xiliang Zeng Improvements of ant colony algorithm and its applications in artificial neural network

*Computer Modelling & New Technologies 2014 **18**(11) 55-59*

Ant colony algorithm (ACA) is a bionic intelligent optimization algorithm with positive feedback, distributed computing and heuristic search. As an important branch of computational intelligence and swarm intelligence, ACA has been successfully applied in solving many combinatorial optimization problems. Artificial neural network is a large-scale distributed parallel processing system with the characteristics of self-organization, self-study, self-adaptation and non-linear dynamic processing and it has a broad prospect in settling the complicated non-linear problems. This paper has proposed an algorithm used to solve multi-objective optimization problems and the applications of ACA.

Keywords: Ant Colony Algorithm; Neural Network

Chang Liu, Zhenyu Na, Xin Zhang An improved algorithm for multi-factor fuzzy correlation

*Computer Modelling & New Technologies 2014 **18**(11) 60-65*

As a means for Vessel Traffic Service (VTS) to oversee the vessels, the traditional radar and the new navigation method of Automatic Identification System (AIS) are the two sources of getting the vessels' information. Tracks Fusion of the data received from these two sensors becomes the fundamental problem to be resolved in VTS. The tracks correlation is the premise and basis of the tracks fusion. This paper proposed an improved algorithm of multi-factor integrated fuzzy correlation based on the least square-time interpolation. We make generous correlation

decision of distance and achieve the targets set in a fixed range, and then after time correction based on the least square-time interpolation we get the correlated tracks set and make fuzzy correlation used the membership function of normal distribution. The simulation experiment shows the proposed fuzzy correlation algorithm is more precise and the data are more close to the actual data of the vessel. The result of this effort can become an efficient method that impacts greatly on the vessel traffic management.

Keywords: Fuzzy Correlation, Tracks Fusion, Automatic Identification System (AIS)

Shenghai Hu, Xiulian Liu, Manhui Zhang Kinematics analysis and simulation of a novel metamorphic parallel cutting mechanism

Computer Modelling & New Technologies 2014 ***18****(11) 66-75*

Based on the metamorphic mechanisms, we proposed a new type of metamorphic kinematic pair, with which a novel metamorphic parallel cutting mechanism was designed to achieve changing pose and trajectory movement. Using the screw theory and the product of exponentials formula, we constructed a nonlinear system of equations for each configuration framework of the metamorphic parallel cutting mechanism, and the Sylvester resultant elimination method was employed to simplify equations of motion for the mechanism, thereby completing the research about the forward and inverse kinematics solution of the original and sub-configuration framework and verifying the feasibility of this novel mechanism via MATLAB and ADAMS simulation.

Keywords: Metamorphic Mechanism; Screw Theory; Product of Exponentials Formula; Forward and Inverse Kinematics Solution

Wenfa Wang, Shiyao Wang Test paper composition based on fish swarm algorithm

Computer Modelling & New Technologies 2014 ***18****(11) 76-80*

Fish swarm algorithm was applied to seek optimization solution of test paper composition. Imitation results showed that fish swarm algorithm had better performance than random drawing algorithm in both composition accuracy and running time. The validity and superiority of the algorithm in this work, therefore, were verified.

Keywords: fish swarm algorithm, test paper composition, mathematical model, running time

Chang Zinan, Shao Fei Epidemic spreading in weighted homogeneous networks with community structure

Computer Modelling & New Technologies 2014 ***18****(11) 81-85*

Community structure has been proven to have great impact on epidemic spread in weighted networks. To understand the epidemic propagation in weighted homogeneous networks with community structure, a model of pseudo-random network is presented with adjustable community structure. By changing the number of edges connecting to the nodes in the same community and the weight of edges connecting to the nodes in the same community, we investigate the epidemic spreading in weighted homogeneous networks with different community structure. Simulations show that both the number of within-community edge and the weight of within-community edge have great impact on epidemic spreading behaviour.

Keywords: epidemic spread, weighted network, homogeneous network, community structure

Shengneng Hu, Xiaoming Lu, Juan Han Design and realization of road geometry alignment simulation

Computer Modelling & New Technologies 2014 ***18****(11) 86-91*

Embarking from road's characteristic and road design's demand to the simulation system, road simulation system's characteristics have been discussed, and then road simulation system frame based on multi-agent has been constructed. The simulation agents have been divided into response agent, corporation agent and interface agent, and their structures have been analysed too. The car dynamics model, road ground view model and driver model have been built, the simulation results are validated by basic experiments. The simulation system is an efficient tool for road design evaluation.

Keywords: road simulation, road engineering, multi-agent, simulation model

Hao Yan, Xiaoyong Yan, Rulin Dou A multivariate analysis-based for range-free localization algorithm

Computer Modelling & New Technologies 2014 ***18****(11) 92-98*

Proposed an improved DV-Hop localization algorithm (PLS-DVHop) based on partial least squares, which uses the partial least squares to model of hop-count and the Euclidean distances, along with the maximum covariance of input matrix and output matrix to estimate the location of unknown nodes. PLS-DVHop has strong adaptability for different deployment network, and overcomes the shortage of only suitable for isotropic networks in the original

algorithm. Simulation results show that PLS-DVHop algorithm has high estimate precision and stable performance, can adapt to different network topologies, and is very suitable for large scale deployment network.

Keywords: range-free localization, wireless sensor network, partial least squares

Hualong Xie, Xiaopeng Mei, Yongxian Liu The dynamic characteristic analysis of spindle based on ANSYS
Computer Modelling & New Technologies 2014 ***18****(11) 99-107*

The basic concept of dynamic characteristics of spindle is introduced. The first ten order vibration mode and natural frequency of shaft are obtained by modal analysis and harmonic response analysis using software ANSYS. The deformation trend of spindle under resonance state is analysed and the contrast between theory and simulation critical speed is done. The frequency response curve of spindle is given. The first order, second order dynamic stiffness and dynamic stiffness under general conditions are calculated. The influence of bearing damping to spindle dynamics characteristics is discussed. In the end, the measures of improving spindle performance are summarized.

Keywords: Shaft, Dynamic characteristic, ANSYS, modal analysis, harmonic response analysis

Ran Ding, Caoyuan Ma, Yongyi Zhao, Yanfang Luo, Jianhua Liu Anti-synchronization of a class of fractional-order chaotic system with uncertain parameters
Computer Modelling & New Technologies 2014 ***18****(11) 108-112*

In order to pull the fractional-order theory to better application, the detailed of computer numerical simulation of the Adams-Bashforth-Moulton Algorithm is proposed in this paper. Anti-synchronization of a class of fractional-order chaotic system with uncertain parameters is realized on this basis and the stability theorem of the system is presented at the same time. And thus it indicates that this method can be adapted to chaotic system with certain parameters and a class of chaotic system with not equal fractional-order. And corresponding implementation conditions is given as well. Besides, it is pointed out that the method, which unites the synchronization and anti-synchronization is also suitable for synchronization issues of the system. Finally, take classic Lorenz system for instance, track time domain and error map about drive system and response system of anti-synchronization are given. The results prove the effectiveness of the control method in the realization of anti-synchronization of a class of fractional-order chaotic system with uncertain parameters and the feasibility of fractional order computer numerical simulation

Keywords: fractional-order, chaotic system, anti-synchronization, chaos control

Liang Zhao, Zhanping Yang Validation assessment with uncertain model inputs
Computer Modelling & New Technologies 2014 ***18****(11) 113-118*

This paper presents a validation assessment method to measure the discrepancies between the model predictions and experimental observations under both aleatory and epistemic uncertainty. The model inputs considered in the paper are sparse point data or interval data, which leads to uncertain parameters for the distribution of the model inputs. A likelihood based method is used to represent the stochastic model inputs and it yields a single probability distribution which integrates the aleatory and epistemic uncertainty of model inputs. This representation of model inputs provides an advantage in computation efficiency for the conventional double loop sampling requirement in uncertainty propagation is collapsed into a single loop sampling. An area based validation metric is extended to compare the probabilistic model predictions obtained from uncertainty propagation with the empirical distribution function of the experimental observations, it reflects an objective quantification of the entire discrepancies between predictions and observations. The confidence interval for the validation metric, which just depends on the amount of experimental observations and confidence level is also developed. A numerical example is used to illustrate the proposed method.

Keywords: validation assessment; likelihood based method; validation metric; confidence interval

Yun Peng Fuzzy set and rough set based evaluation algorithm of web customers
Computer Modelling & New Technologies 2014 ***18****(11) 119-122*

A fuzzy algorithm of web customers evaluation based on rough set is presented. Key attributes can be gotten through rough set. The evaluation from the data objects based on key attributes can reduce the data size and algorithm complexity. After Clustering analysis of customers, then the evaluation analysis will process to the clustering data. There are a lot of uncertain data in customer clusters, so the traditional method of classification and evaluation to the incomplete data is very difficult. Superposition evaluation algorithm based on fuzzy set can improve the reliability and accuracy of web customer evaluation. Evaluation of the web customer also can improve efficiency, service quality and profitability of web businesses.

Keywords: Rough Set, Membership Function, Cluster Analysis, Fuzzy Set

Cunchang Gu, Xiaoyan Xu Coordinated scheduling on single serial batching machine with transportation considerations

Computer Modelling & New Technologies 2014 ***18****(11) 123-128*

The coordination of production scheduling and transportation has recently received a lot of attention in logistics and supply chain management. We study a coordinated scheduling problem, in which each job is transported to a single serial batching machine for further processing, each batch to be processed occurs a processing cost, and the objective is minimizing the sum of the makespan and the total processing cost. Under the condition of the jobs' processing times are equal, if the job assignment to the vehicles is predetermined, we provide a polynomial time dynamic programming algorithm, for the general problem, we prove it is NP-hard. When the returning time of vehicle is zero, we present the approximation algorithm and prove that the worst case ratio of the algorithm is not greater than $2-\frac{1}{m}$, and the bound is tight.

Keywords: supply chain scheduling, dynamic programming algorithm, complexity, worst case analysis

Yaolin Huang, Shuangli Wang, Chengjun Xie, Zhuo Jiang, Jinyan Zhao Chaos control of unified chaotic system base on tridiagonal matrix stability theory and adaptive hybrid synchronization

Computer Modelling & New Technologies 2014 ***18****(11) 129-133*

This paper presents a new chaos control method to control the unified chaotic system to zero. In order to control a unified chaotic system to zero, the first step is to design a stable system, which stable to zero base on tridiagonal matrix stability theory, the stable system as the master system. The second step is to make the unified chaotic system with controlled as slave system, the third step to make the master system and the slave system synchronization. Different system state apply different adaptive synchronize method to realize synchronization. The adaptive control law and parameter update law are obtained base on Lyapunov stability theory. Numerical simulations are presented to demonstrate the effectiveness of theoretical analysis.

Keywords: chaos control, adaptive hybrid synchronization, tridiagonal structure matrix stability theory

Lirong Qiu The impact on collaborate level of cluster

Computer Modelling & New Technologies 2014 ***18****(11) 134-138*

Due to the impact of the characters of nodes in complex network on collaboration level, we put forward a new iterated game model based on conformist mechanism. In this model, nodes can update tactic not only according to their payoffs but also to their species, which they belong to. The new model can assure that nodes in the same species adopt the same tactic. Simulation results show that the collaboration level of the networks that adopt conformist mechanism is higher than the networks that adopt normal mechanism. In the other words, the collaboration level is in inverse proportion to the species number. On the other hand, we find that the average payoffs increased with the penalty gene instead of increasing alternately. So the new model can promote the collaboration level and the average payoffs of the nodes in network at the same time.

Keywords: Conformist Mechanism, Iterated Game, Collaborate Level, Complex Network

Tengda Shi, Kaifeng Su, Linqian He, Lei Yan A gait recognition system based on BP neural network and plantar pressure

Computer Modelling & New Technologies 2014 ***18****(11) 139-142*

In order to get a faster, more effective and stable control over lower extremity exoskeleton of power assist robot, precise examination on gait information is necessary, thus it is so important to design and establish a gait recognition system with accurate detection. In this paper, a wireless in-shoe wearable plantar pressure acquisition system based on ATmega16 and 8 FSR sensors will be applied to data acquisition for the gaits which consist of standing, walking, jumping and going upstairs. And four volunteers (2 males and 2 females) will be invited in this research to collect the pressure information. The NNT of MATLAB will be applied to establish an 8-12-4 BP neural net model. The input factors come from the eight sensors of plantar pressure system, the output is gait category. Proved by a great deal of experiments, the gait recognition method proposed in this research is quite feasible.

Keywords: gait recognition, BP neural networks, plantar pressure, lower extremity exoskeleton of power assist robot

Jinjiang Liu, Jingjing Liang Automatic verification of embedded system based on EFSM

Computer Modelling & New Technologies 2014 ***18****(11) 143-147*

To ensure the correctness of embedded system, automation of test case generation is necessary in industrial. This

paper present a technique for specifying coverage criteria and a method for generating test suites for embedded systems whose behaviours is depend on its interactive environment. The embedded system under test can be described as extended finite state machines (EFSM) and the coverage criteria can be specified as monitor automata with parameters, which monitor and accept traces that cover a given test criterion of an EFSM. The flexibility of the technique is demonstrated by specifying a number of well-known coverage criteria based on control- and data-flow information using observer automata with parameters. We also develop a method for generating test cases from coverage criteria specified as observers. It is based on transforming a given observer automata into a bitvector analysis problem that can be efficiently implemented as an extension to an existing state-space exploration such as, e.g. SPIN or Uppaal.

Keywords: EFSM (Extended Finite State Machine), embedded system, test case generation, model-based testing

Baowen Hu, Changhong Li Damage characteristic of rock sample with circular defect based on the distinct element and moment tensor methods
Computer Modelling & New Technologies 2014 ***18****(11) 148-153*

Based on the distinct element and moment tensor theory, the damage process and mechanism of rock sample with circular hole were researched. The crack evolution patterns corresponding to the laboratory test showed that there were mainly four stages. Firstly, the microcracks were randomly distributed in the rock sample. Secondly, the larger primary cracks are formed around top of circular hole with the increasing of microcracks. Then, more microcracks were localized near the boundary of circular hole, which formed the competition of several possible sets of fractures. Finally, the rupture zone was formed along one of crack zones. The size effect of circular hole showed that there would take on different rupture forms with increasing of diameter. Meanwhile, for better probing into damage mechanism, the acoustic emission (AE) algorithm based on moment tensor theory was implanted into the whole loading process. The AE magnitudes of all parts of rock sample were shown in AE contour maps, and these contour maps showed that the formations of rupture zone were contributed by different stress or energy levels.

*Keywords: c*ircular hole, damage mechanism, acoustic emission, distinct element, moment tensor

Hongwei Ye, Lianjiao Zhang, Xiaozhang Liu Network intrusion clustering based on fuzzy C-Means and modified Kohonen neural network
Computer Modelling & New Technologies 2014 ***18****(11) 154-158*

Kohonen neural network recognizes and clarifies substantive network data, but with a long running time and a slow convergence process. To solve this problem, a network intrusion clustering method is presented in this paper. Specifically, the training data is pretreated using Fuzzy C-Means (FCM). Then some selected data will be trained with using Kohonen neural network. Meanwhile, to speed up the convergence process of Kohonen neural network and to form a better optimized network topology, a neighbourhood function is established for the competing neuron. Each neuron has neighbourhood topology collections. The data simulation results demonstrate the efficiency and effectiveness of the proposed algorithm.

Keywords: Kohonen neural network; neuron, FCM, network intrusion clustering

Jiansheng Xia, Shasha Dou Numerical analysis of sheet metal U-Bending process
Computer Modelling & New Technologies 2014 ***18****(11) 159-165*

The sheet metal U-bending forming is a complex process, based on the assumption of Prandtl-Reuss flow rule and Von Mises yield criterion, the incremental elastoplastic large deformation finite element model was established based on the Updated Lagrangian Formulation (ULF). The elastoplastic conversions of boundary and deformation are reduced with rmin rule. The friction phenomenon of slippage and viscosity at the boundary interface is revised with increment of revision Coulomb rule. The increment rules are led into the whole stiffness matrix, and derived out the stiffness equation. The studies show that the influence on U-Bending deformation of sheet metal is influenced by die structure and parameter. The results show there is a good consistency between the finite element simulation and experimental result.

Keywords: elastoplastic, FEM simulation, sheet metal, U-Bending, warpage

Ming Tao Frequency analysis of the area error in the triangular partition of a discrete global grid
Computer Modelling & New Technologies 2014 ***18****(11) 166-173*

The discrete global grid is a method of hierarchical space expression with global scale, multi-resolution, and multi-scale transform features. It is a basic theory related to global geographic information systems and spatial grid computation. The main challenge in the practical application of discrete global grids is the generation of errors. Thus,

four typical kinds of global discrete grid models were investigated in this study of partition errors. Area was considered an indicator of partition errors. Furthermore, this study computed the area errors in levels 3 to 10 partitions of each kind of global discrete grid model. The frequency of each indicator was also analysed. Analysis results show that the Synder models have the smallest area deformation.

Keywords: discrete global grid, triangular partition, frequency analysis

Chen Ye Research on the application of clustering algorithm based on minimum spanning tree
Computer Modelling & New Technologies 2014 ***18****(11) 174-177*

Cluster analysis is one of the most important technologies in data mining. Minimum spanning tree (MST) is an advanced algorithm in cluster analysis. Studying MST has important practical significances. Firstly, this paper analysed partitioning, hierarchical, density and grid clustering algorithms based on MST thoroughly. Secondly, implementation principles and shortcomings of these four algorithms were discussed. Finally, practical applications of clustering algorithm based on MST were introduced, aiming to solve some practical problems.

Keywords: data mining, clustering algorithm, MST model

Jing Yang Routing method of quantum genetic algorithm
Computer Modelling & New Technologies 2014 ***18****(11) 178-181*

Quantum genetic algorithm was applied in the work to solve multiple-QoS routing problems of bandwidth-delay constraint, thus meeting current requirements for multimedia messaging. Mathematical model and constraint condition of QoS routing were established, analysing algorithms principles and steps that how quantum genetic algorithm optimized QoS routing. By comparing average fitness and maximum polymerization fitness under different running times, quantum genetic algorithm can effectively find optimal solutions to multiple-QoS routing problem.

Keywords: genetic algorithm, quantum genetic algorithm, mathematical model, route choice, fitness

Jinguo Zhao A novel method for K-Means clustering algorithm
Computer Modelling & New Technologies 2014 ***18****(11) 182-188*

This paper investigated K-means algorithm, a well-known clustering algorithm. K-means clustering algorithms have some shortfalls and defects, and one defect is reviewed in this study. One of the disadvantages of K-means clustering algorithms is that they can produce clusters that do not always include all the correct components. It is due to the presence of the error rate during the clustering process. The purpose of this research was to decrease error rates in the K-means clustering algorithm and to reduce iteration of running this algorithm. A novel method is proposed to calculate the distance between cluster members and cluster centre. To evaluate the algorithm proposed in this study, seven well-known data sets consisting of Balance, Blood, Breast, Glass, Iris, Pima and Wine data sets were used. This investigation revealed that the performance of K-means algorithms was increased and resulted in valid clusters and that it reduced error rates, run time and iteration.

Keywords: K-means, clustering algorithm, error rate, iteration, reduction, stable

Hong Wang, Tao Zhang Research on transition priorities in group based on CPN
Computer Modelling & New Technologies 2014 ***18****(11) 189-196*

Transition priorities might be a useful mechanism when modelling using Petri nets. Accordingly, the newest CPN Tools, widely used for modelling and simulating the Coloured Petri net, implements transition priorities. Whereas, the algorithms compute enabling for all transitions in a highest-priority-first order. In the real system, it is usually that there are priorities relationships not for all transitions but only some ones. Based on the above analysis, this paper put all the transitions, having priority relations, into one group and advances relative theoretical definitions of transition priorities in group, such as absolute preset of transition, key place set, key colour set, etc. Furthermore, it proposes new algorithms when the systems have different key place set and key colour set, and construct the model of the interrupt priorities to solve the problem of software model checking for interrupt system.

Keywords: Petri net, CPN, transitions, transition priorities in group, counter place

Yuejun Liu, Aiping Tang, Ketong Liu, Jiewen Tu Mechanical characteristics and form-finding analysis of iced transmission lines
Computer Modelling & New Technologies 2014 ***18****(11) 197-203*

Form-finding is the important problem to be solved in the cable structure analysis, to the different forms of loads, the direct iterative method is used to determine the initial configuration of the cable structures. Horizontal tension or

cable tension is used as iterative convergence condition, the form-finding of the cable is researched under its own gravity, uniform ice and non-uniform ice load. As for the multi-span transmission lines, two conditions of uniform ice and non-uniform ice loads on the whole span were analysed. The results of initial configuration are consistent with the analytical method, which verified the correctness of the direct iterative method, under the condition of non-uniform ice load, the stress of conductor is larger than the maximum stress, which is very dangerous in the actual operation.

Keywords: mechanical characteristics, form-finding, uniform ice, non-uniform ice, finite element method

Ying Sun, Meikui Deng, Shenghua Ye Mobile ECG QRS detection algorithm and implementation
Computer Modelling & New Technologies 2014 ***18****(11) 204-211*

The ECG signal contains a lot of interference in the mobile ECG monitoring system. Reasonable selection of signal filtering and QRS wave detection method are the key to ECG signal analysis. According to the characteristics of ECG signals, the design of low-pass filtering and the improved median filter that can filter the interference has been conducted in the paper. Meanwhile, addressing to the limitation generated by selecting and fixing the median threshold using the traditional differential slope method, the paper has proposed the adaptive dynamic threshold and used quadratic difference algorithm to process signal in order to obtain R wave, and then locate Q and S wave based on R wave location. In addition, combining the characteristics of the QRS wave group, judgments on the missing and over detection are conducted, so that the algorithm is robust and has good fault-tolerant ability. The experimental results show that, this algorithm can not only satisfies the need of the real-time QRS wave detection, but is also more suitable for the transmission and processing of ECG signals in mobile ECG monitoring.

Keywords: mobile ECG signal, QRS detection, algorithm

Xie Wu, Ouyang Shan, Hailin Xiao Concurrent quantum key distribution using quantum teleportation and time division multiplexing
Computer Modelling & New Technologies 2014 ***18****(11) 212-218*

In view of the problem that it is difficult to use only one typical channel to deal with many quantum channels during multi-channel QKD (Quantum Key Distribution), the method of TDM (Time Division Multiplexing) in typical channel is introduced to construct a new CQKD (concurrent QKD) system. Using multi-channel quantum teleportation, every concurrent quantum channel is mapped to a time slot of TDM. The results of case study with EPR (Einstein-Podolsky-Rosen) pairs show that this problem can be solved with the CQKD methods. Moreover, this CQKD scheme also has the advantage of unconditional security, while the QKD bit rates can be improved. It opens a new way to develop large-capacity long-distance quantum secure communications.

Keywords: quantum information security, concurrent quantum key distribution, quantum teleportation, time division multiplexing

Kaishi Sun, Fasheng Liu A few expanding integrable models of WKI hierarchy and their Hamiltonian structures
Computer Modelling & New Technologies 2014 ***18****(11) 219-226*

The integrable coupling of the WKI hierarchy is obtained by the perturbation approach. With the help of a higher dimensional loop algebra, the coupling integrable couplings of the WKI hierarchy are obtained, respectively. Their Hamiltonian structures are worked out by employing the component-trace identities and variational identity.

Keywords: coupling integrable couplings, component-trace identities, perturbation equation

Lijuan Hong, Ju Qian, Jifeng Cui Automated unit-level testing of java memory leaks
Computer Modelling & New Technologies 2014 ***18****(11) 227-233*

Java programs may suffer from serious memory leak bugs. To resolve these bugs, various leak diagnosing and even fixing techniques have been proposed. However, in literature, there is very few work, which focuses on memory leak testing. Without revealing leak phenomenon by testing in advance, even excellent leak diagnosing and fixing techniques can be hard to work. In software testing, unit testing is a technique to avoid faults at early stage of software development. This paper proposes an automated unit-level memory leak testing approach to find potential leak bugs in Java methods. It firstly identifies the methods with high leaking risks. Then, strengthened unit tests are generated accordingly to check whether those risky modules can really cause leaks. Cases studies show that our method could be valuable for real programs.

Keywords: Java, memory leak, unit testing, test generation

Minsheng Tan, Huan Zhou, Yangwei Li, Jianxue Liu Improved PSO clustering routing algorithm for WSN
*Computer Modelling & New Technologies 2014 **18**(11) 234-241*

For cluster head selection randomness of clustering algorithm, and PSOC algorithm were not considered the distance from cluster head to base station, an improved particle swarm optimization (I-PSOC) routing algorithm was brought out. The improved algorithm particle swarm fitness function was improved by considering the node residual energy, nodes' distance and the distance between nodes and base station. At the same time, the optimal nodes were selected as the cluster head and the cluster head has transmitted the data to base station in a single or more jumps through searching right path in the improved algorithm. The simulation results show that the nodes energy consumption of network was reduced significantly and the network life cycle was extended

Keywords: PSO, Clustering Routing Algorithm, cluster head, Wireless Sensor Network

Wei Zhong, Ming Lei, Zhou Tian Theoretical analyses and numerical simulation of the interaction time and the separation time of two elastic bars after the loading of a triangular wave
*Computer Modelling & New Technologies 2014 **18**(11) 242-248*

With theoretical analyses, the stress-wave propagation and reflection between two elastic bars, whose left side achieved a triangular wave load, is mainly studied. Considering the effects of pulse width of the load, length and wave impedance of the bars and the relationship between the length of the load wave and the two bars, the regularities of the stress-wave propagation and reflection between the two bars is analysed, and the formulas for calculating the interaction time and the separation time of the two bars are derived, in three different conditions. The same problems in three different conditions are simulated by using AUTODYN, and curves of displacements of points at contact surface of the two bars varying the time are given. According these curves, the simulating results of the separation time can be obtained. By the comparison of the simulating results and the theoretical calculated results using formulas derived in this paper, the correctness of the theoretical analyses and formulas here is demonstrated.

Keywords: triangular wave load, elastic bar, separation time, theoretical analyses, AUTODYN

Zhiyan Li, Baoxia Jin Research of numerical solutions of differential equations model based on the finite element method
*Computer Modelling & New Technologies 2014 **18**(11) 249-252*

Using the finite element method solving a class of second order ordinary differential equations, analyses the two-point boundary value problem of a class of second order ordinary differential equations, through numerical examples to validate its effectiveness.

Keywords: ordinary differential equations, finite element method, two-point boundary value problem

Jing Cao, Xuefeng Xing, Shan Liu Design and application of iterative Monte Carlo localization for mobile wireless sensor networks based on MCL
*Computer Modelling & New Technologies 2014 **18**(11) 253-258*

In recent years, wireless sensor network had been more and more widely used in our daily life, and with the propose of Monte Carlo localization (MCL) algorithm, node localization of the mobile wireless sensor network had been solved effectively. But it needed to have a large number of samples if it used the Monte Carlo localization algorithm to obtain a high positioning accuracy. This paper proposed a new improved algorithm (iterative Monte Carlo localization algorithm) based on the Monte Carlo localization algorithm. In iterative Monte Carlo localization (IMCL) algorithm, each anchor node location information was forwarded by its neighbour nodes only once and preserved by the receiving node in each period. Then the next period, merge it and the sent/forwarding information into a packet and forward. Make sure that points have more observations for estimating previous location sets. IMCL, meanwhile, also can make full use of observation to filter out some samples that were far from the real position of node, so as to improve the accuracy of node localization. We finally confirmed by experiment: IMCL algorithm had higher positioning accuracy compared with other algorithm.

Keywords: IMCL, immune genetic algorithm, wireless sensor network, improve the accuracy of localization

Jian-Wei Li, Xiao-Wen Li Algorithm design and the application for cluster validity based on geometric probability
*Computer Modelling & New Technologies 2014 **18**(11) 259-263*

Determining optimum cluster number is a key research topic included in cluster validity, a fundamental problem unsolved in cluster analysis. In order to determine the optimum cluster number, this article proposes a new cluster

validity function for two-dimension datasets theoretically based on geometric probability. The function makes use of the corresponding relationship between a two-dimension dataset and the related two-dimension discrete point set to measure the cluster structure of the dataset according to the distributive feature of the point set in the characteristic space. It is designed from the perspective of intuition and thus easily understood. Through TM remote sensing image classification examples, compare with the supervision and unsupervised classification in ERDAS and the cluster analysis method based on geometric probability in two-dimensional square, which is proposed in literature 2. Results show that the proposed method can significantly improve the classification accuracy.

Keywords: Cluster validity, Geometric probability, optimum cluster number

Information and Computer Technologies

Si Tian, Shui Wang, Yang Liu, Le Wang An algorithm for mining frequent itemsets on uncertain dataset
Computer Modelling & New Technologies 2014 ***18****(11) 264-272*

Mining frequent itemsets from uncertain transaction dataset is a research topic in data mining. Some algorithms are based on FP-Growth, but they construct the tree structure in a manner that cannot be as compact as the original FP-Tree, so the tree is easily developed to huge size and this hinders their performance. In this paper, we propose a new tree structure called IT-Tree (Itemset Tail-node Tree) to efficiently maintain probability information of itemsets in tail-nodes; we also propose a corresponding algorithm IT-Mine to mine frequent itemsets from IT-Tree without additional dataset scans. We evaluate our approach on real sparse and dense datasets with different minimum support numbers that can produce non-null frequent k-itemsets ($k \geq 2$); the results show that IT-Mine outperforms other algorithms in terms of execution time, especially for large dataset or small minimum expected support number.

Keywords: frequent itemset, frequent pattern, uncertain transaction dataset, data mining

Changsu Liao, Xiaojing Niu, Meili Wang, Dongjian He Simplification of 3D point cloud data based on ray theory
Computer Modelling & New Technologies 2014 ***18****(11) 273-277*

To effectively reduce the amount of 3D point cloud data, whose shape is symmetrical or spherical, this paper proposes an efficient simplification algorithm based on ray theory. Meanwhile, a boundary retention method based on the distribution uniformity of neighbouring data points is used to keep the model complete. Avoiding time-consuming recursion and curvature estimation, the proposed method is much efficient and achieves good simplification results.

Keywords: 3D Point Cloud, Data Simplification, Ray Generation, Boundary Retention

Lei Zheng Virtual machine resource allocation algorithm in cloud environment
Computer Modelling & New Technologies 2014 ***18****(11) 279-284*

To resolve the problem that virtual machine deployment reservation scheme waste a lot of resources and single-objective deployment algorithm is not comprehensive, a virtual machine resource allocation algorithm based on virtual machines group multi-objective genetic algorithm is proposed. The algorithm is divided into group coding and resources coding. Resources coding integrated coding according to the history resource need of virtual machines to physical machine and integrate number of physical machine and resource need of physical machine occupied by virtual machine through improved crossover and mutation operations. The experimental results show that the algorithm is effective to reduce the number of physical machine and resource utilization of physical machine, saving energy as much as possible.

Keywords: Cloud computing, Resource allocation, Virtualization, Energy-saving, Genetic algorithms

Hudai Fu, Jingang Gao Multi circle detection by using evidence accumulation
Computer Modelling & New Technologies 2014 ***18****(11) 285-289*

The traditional circle detection algorithm has complex computation, large memory space occupation and other deficiencies. It has low detecting efficiency and not suitable for the multi circle detection. So it proposes a multi circle detection method based on global search and evidence accumulation in the paper. The evidence accumulation and the weighted average are combined in the algorithm. The pseudo centre processed during the evidence accumulation are classified and analysed. Three kinds of pseudo centres are eliminated by class. Finally, the other circle parameters are calculated. It is proved that the algorithm has high precision, high efficiency, and low sensitivity to the defect of local information. Also the detecting time will not increase with the number of circles. The

multi circle detection effect is obviously superior to the traditional randomized circle detection algorithm.

Keywords: Multi Circle Detection, Global Search, Evidence Accumulation

Ye Chen, Xiaoqun Qin Applications of dynamic adaptive bee colony algorithm in multi-threshold image segmentation

Computer Modelling & New Technologies 2014 ***18****(11) 290-295*

Artificial bee colony (ABC) is an evolutionary computation method, which is inspired from the specific collaborative social group behaviour among the individual bees in the colony and which is a heuristic optimization algorithm based on population search strategy. This paper has proposed a quick dynamic adaptive bee colony algorithm, which analyses the performances of the artificial bee colony algorithm and it designs a multi-threshold image segmentation method realizing a dynamic adaptive artificial bee colony (DAABC) with multi-threshold OTSU as the fitness function. The main characteristics of this method include: reducing the noise interferences in the multi-threshold image segmentation; effectively narrowing down the search range of the threshold; guaranteeing the quickness of the segmentation speed; determining the search range of the reconnaissance ants with adaptive dynamic control and accelerating the convergence speed of bee colony algorithm. The experimental results demonstrate that the method in this paper is better than the image segmentation method based on particle swarm optimization (PSO) and artificial fish swarm algorithm (AFSA).

Keywords: Artificial Bee Colony Algorithm, Multi-Threshold, Image Segmentation, OTSU

Yu Liu, Lei Chen, Shihong Chen Alignment-based approximate SPARQL querying on linked open data

Computer Modelling & New Technologies 2014 ***18****(11) 296-303*

With the growth of Linked Open Data, more and more applications are developed to take full advantage of its massive data. However, all these applications face an inevitable problem - how to retrieve information from these datasets with different schemas, which results in that a query for a dataset may get none answer from other datasets. To solve this problem, ontology alignment has been adopted in some Linked Open Data querying systems. In this paper, we follow this idea and make further efforts to find more approximate answers by employing relations and probability values in the result of ontology alignment. The fundamental of our method is the similarity between entities, which is used to evaluate the similarity of rewritten query relative to original query. In order to facilitate user to query other dataset with original query, an algorithm for alignment-based approximate querying is proposed. In experiments, the SPARQL queries for DBpedia are rewritten on the basis of alignment result between DBpedia and YAGO. The results of experiments show that alignment-based approximate querying can not only retrieve approximate results, but also overcome the problem caused by imprecise result of ontology alignment, which is very common for most of alignment techniques.

Keywords: Linked Open Data, Approximate Querying, Ontology Alignment, SPARQL

Hui Guo, Jie He A fractal de-noising algorithm based on least absolute deviation method

Computer Modelling & New Technologies 2014 ***18****(11) 304-310*

Against the shortcoming that the traditional method of fractal image compression coding has inferior decoding quality on the original image subject to salt-and-pepper noise interference, this paper raises a least absolute deviation (LAD) method to be applied in fractal image compression, which can replace the method of least square error in computing contrast and brightness adjustment value and solve the L1-norm recursive problem using weighted median. The experimental result indicates that the LAD method has a very good anti-noise effect on the outliers introduced by salt-and-pepper noise.

Keywords: Fractal Image Compression, Least Square Error Method, Least Absolute Deviation Method, Salt-and-Pepper Noise

Ye Chen, Xiaoqun Qin, Xinmin Zhou Single threshold segmentation for noisy image based on fuzzy ant colony algorithm

Computer Modelling & New Technologies 2014 ***18****(11) 311-316*

Firstly, this paper pre-processes the image to be segmented through grey-scale morphological method. Then, based on the in-depth analysis of basic ant colony algorithm, it explains the shortcomings of this algorithm; proposes the improved strategy of ant colony algorithm, namely fuzzy ant colony algorithm, which designs the fitness function of artificial ant colony algorithm with minimum cross entropy and applies the improved fuzzy ant colony algorithm in the spatial-domain noisy image single segmentation. Finally, starting from the segmentation results and convergence, it compares the performances of the improved ant colony algorithm and the basic ant colony algorithm, GA

algorithm and AFS algorithm.

Keywords: Noisy Image, Ant Colony Algorithm, Histogram Feature, Threshold Segmentation, Cross Entropy

Jiali Tang, Chenrong Huang, Yijun Liu, Honghui Fan, Jianmin Zuo Face super-resolution algorithm based on SVM-improved learning

Computer Modelling & New Technologies 2014 ***18****(11) 317-322*

As many other inverse problems, human face image super-resolution is an ill-posed problem. The problem has been approached in the context of example-based superresolution learning. However, these methods need to run though all the sample set, which results in high calculation load and image degradation because of mis-matching. In this paper, we propose a new face image superresolution algorithm based on Support Vector Regression (SVR) pre-classified learning. A Principal Component Analysis (PCA) based pre-process is used to select a subset of samples. Then the best-matching sample images are trained to ensure the content relevance between the sample patch and the input low-resolution image. Further improvement involves a combination of classification and SVR-based techniques. Therefore, experiment results show that the proposed algorithm gets better reconstruction performance and faster program running speed.

Keywords: Face Super-resolution, Support Vector Machine (SVM), Principal Component Analysis (PCA), Example-based Algorithm

Jingang Gao, Shuang Zhang, Hua Wang Application of halcon in the image analysis of dry cutting gear meshing region

Computer Modelling & New Technologies 2014 ***18****(11) 323-327*

The master-slave bevel gear pair of some automobile rear axle is used to be as the research object, to obtain the image of the gear meshing contact region. Firstly, get the grey image of HSV space from RGB image to HSV space by the pre-processing algorithm; secondly, segment features of image accurately by the adaptive threshold segmentation algorithm; thirdly, fill the holes in the image of meshing region perfectly by proposed holes repair algorithm; fourthly, mark all connected domains of image with 4 -adjacent points domain labelling algorithm; finally, the Marking image area, width and the ratio of width and height features are selected to extract the image of the gear meshing region to obtain the geometry information of the gear meshing region. Research results show: contact centre E/L should be in 45% ~ 50%, and the length of the contact region B/L should be in 45% ~ 55%.

Keywords: HALCON, dry cutting gear, contact region, Image processing

Wenju Wang, Liujie Sun, Zhang Xuan Editing method of virtual human motion path based on motion cycle step-length

Computer Modelling & New Technologies 2014 ***18****(11) 328-336*

An editing method of virtual human motion path based on movement cycle step-length is presented: it uses the Cardinal spline interpolation to extract the original path and equably sets the path editing control points; alters the positions of path editing control points to generate a new path and adjusts the position and orientation of the original motion sequence frame's root joint nodes on the new path; automatically judges any path curve segment that generates the problems of footstep disharmony after editing; for every incorrect curve segment, deletes the frames of general motion cycle step-length between the two path editing control points that contain incorrect curve segment based on the definition of motion cycle step-length and adjusts the spacing and moving direction of root nodes in every frame on the paths between the two control points; finally does constraint re-establish to all floor frames in the new path. The experimental results show that: the motion path editing method is simple and easy to operate. The generated virtual human motion is natural and smooth. It can effectively eliminate the footstep sliding phenomenon to improve the reusability of motion capture data.

Keywords: Motion path editing, Cardinal spline interpolation, Motion cycle step, Constraint reestablishment, Footskate cleanup

Yumei Ning, Zhenguo Ding, Ping Zeng Early fault warning mechanism based-on association rules in server clusters management system

Computer Modelling & New Technologies 2014 ***18****(11) 337-343*

With the rapid development of Internet and business internal information technology, the problem of server cluster fault early warning becomes important in server clusters management. To solve the problem, a server cluster fault early warning system (SCFEWS) based on event association rules is presented. This system is mainly based on the fault event association tree. First obtain event relationship and association rules. Construct early warning events

associated tree with system logs and association rules, and store it in binary tree linked list. Then by using the warning event filtering algorithm, redundancy fault early warning events are filtered out and only source event for early-warning notification are kept. Experiment shows the proposed algorithm can effectively improve the accuracy of fault location for server clusters management.

Keywords: Server Cluster, Early Warning, Fault Association Tree, Warning Event Filtering Algorithm

Wensheng Wang, Huifeng Xue, Feng Zhang A new dynamic regulation UIPO model of groundwater based on cloud computing and hydroinformatics
Computer Modelling & New Technologies 2014 ***18****(11) 344-351*

In order to solve the problem of excessive mining of the coal resources in Yulin mine area caused enormous damage to groundwater and in turn threatening the regional sustainable development, proposed a new dynamic regulation UIPO model of groundwater, which combined hydroinformatics, cloud computing and multi-source data fusion algorithm, and the mine hydrogeological spatial database, the visualization of 3D geological model and groundwater dynamic evolution model are created. Simulation results show that the UIPO complete with these models and with groundwater - ecological environment - economy system model all connected data analysis for decision support system and with complete hydrogeological and spatial process method by means of big data, can simulate the temporal and spatial variations of groundwater resources, forecast future impact on Yulin mine area groundwater for sake of large-scale exploitation.

Keywords: hydroinformatics, cloud computing, groundwater, dynamic regulation, big data

Zuowen Wang Composition of web applications in clouds environment
Computer Modelling & New Technologies 2014 ***18****(11) 352-357*

In recent years, for the advances of Cloud Computing technologies, cloud applications have been popularity for their rich set of features. The advantages of cloud applications include that users can utilize them in a low cost, threshold, and risk way; these applications can be quickly deployed on the clouds without duplication of work such that developers can focus on enhancing their QoS to improve core competitiveness. Therefore, their practical use on business with promising values can be expected. As such, cloud applications are recognized as a trend for the next generation of business applications, and hence how to migrate these on-premise applications to the clouds becomes a desired field in the literature. For this need, we present an ontology-based method for the composition process that specifically addresses the cloud features and the composition of on-premise applications into the clouds. In particular, for enabling the selection of desired clouds, the method imposes semantic ontologies on the specifications of the candidate clouds from which the desired ones can be effectively selected. For illustration, the method is applied to the composition of a CSS application to its cloud version.

Keywords: cloud computing, cloud application, semantic ontology, service composition

Chunyan Han OSTU image segmentation algorithm of fruit fly optimization algorithm
Computer Modelling & New Technologies 2104 ***18****(11) 358-363*

Traditional OSTU algorithm has the disadvantages of a large amount of calculation and low calculating speed. Based on the combination of Fruit Fly Optimization Algorithm and OSTU algorithm, an image segmentation algorithm is created from Fruit Fly Optimization Algorithm to improve OSTU, stressing the basic principles and calculation procedures of this revised algorithm. In order to verify the validity of this algorithm, the work compared the quality of image segmentation, segmentation speed and algorithm stability of 4 sets of standard test images. The simulation results show that the segmentation speed of revised OSTU algorithm is much faster than that of traditional OSTU algorithm when Fruit Fly Optimization Algorithm is applied to improve OSTU algorithm. Meanwhile, the quality of image segmentation is also more stable under the same condition of time limit.

Keywords: OSTU algorithm, image segmentation, Fruit Fly Optimization Algorithm, grey value, stability

Lili Jing, Yang Nie, Lifang Zhao Colour recognition system based on TCS3200D
Computer Modelling & New Technologies 2014 ***18****(11) 364-368*

The colours of objects can be converted into pulses with different frequencies by TCS3200D. Colour recognition system was designed through the combination of Single Chip Microcomputer (SCM) and PC on the basis of the colour vision principle of TCS3200D. Fully utilizing TCS3200D, the system integrates the advantages of photodiodes and converter of light intensity to frequency to simplify circuit. The effects of factors, such as illuminant, orientation and the surfaces of objects, on the measurement were eliminated by the white balance adjustment. Additionally, colour vision errors were effectively reduced by the correction of the measured data with

ANFIS and subtractive clustering.

Keywords: TCS3200D, colour recognition, ANFIS

Chun Zhang Distributed ontology based information retrieval using semantic web
Computer Modelling & New Technologies 2014 ***18****(11) 369-374*

In recent years, user demand for integrated searches over several independently operating semantic web systems have been increasing rapidly. Integrated semantic searches enable more meaningful results to be generated because information with similar meanings in diverse areas and domains is likely to be used for inference. However, it is not easy to integrate physically independent, distributed, and heterogeneous database systems to provide a single, integrated semantic web system to end-users. In this paper, we propose a novel system that integrates heterogeneous semantic web systems based on schema mapping. The user can generate only one SPARQL query using the integrated schema without the necessity of checking the schemas of individual systems each time thereby reducing additional costs to generate queries for individual systems. Furthermore, the user is not required to collect individual query results manually after performing a query and additional costs for establishing systems can be reduced because no change in existing system structures is required. If currently established systems are expanded by adding the schema structures of other ontology systems, the cost to establish another integrated retrieval system can be saved. To evaluate the effectiveness of our approach, we have implemented a prototype that integrates two national information retrieval systems.

Keywords: integrated information retrieval, ontology, schema mapping, semantic web

Yuwen Huang Research on cost-sensitive ensemble classification for mining imbalanced massive data streams
Computer Modelling & New Technologies 2014 ***18****(11) 375-382*

The existing classifiers for massive data streams do not consider the imbalance distribution and cost factors, so this paper proposes the approach of the cost-sensitive ensemble classification for imbalanced massive data streams (CECIDS). Firstly, this paper gives the construction method for cost-sensitive ensemble SVM Classification, which is integrated by the classifiers with oversampling, sub-sampling and reconstituted sample space. Secondly, we propose a classifier method BL_KNNModelб which is based on KNNModel algorithm for imbalanced massive data streams. BL_KNNModel can detect the concept drift streams by using the variable windows size, which has lower time complexity. At last, the cost-sensitive ensemble classifier for imbalanced massive data streams is given, which has the virtue of high classification and lower time complexity. In addition, the cost-sensitive ensemble SVM algorithm is used to handle the confused instances. The experiments using both synthetic and real datasets show that compared to the other classification algorithms for imbalanced data streams, CECIDS has higher evaluating indicator and more excellence integrated learning curve.

Keywords: Imbalanced data streams, Ensemble classification, Cost-sensitive SVM

Xiaoming Dong Accurate self-localization of mobile robots based on vision sensors
Computer Modelling & New Technologies 2014 ***18****(11) 383-387*

Robot localization is a challenging problem in indoor environment since no GPS information is available. In this paper, an algorithm was proposed for accurate localization, which designed a delicate way to extract the feature points at first, then the position of the robot was determined using the relation of the features in different images, finally, the Kalman filer was designed to decrease the error caused by robot's moving. Experiments showed the accuracy and robustness of this algorithm.

Keywords: mobile robot, multi-cues fusion, vision sensor, Kalman filer, localization

Zhihui Zhu, Ting Liu, Lirong Xiong, Meihu Ma Identification of the hatching egg before the incubation based on hyperspectral imaging and GA-BP network
Computer Modelling & New Technologies 2014 ***18****(11) 388-393*

The removal of the unfertilized egg from the hatching egg before the incubation could improve the efficiency of incubation. The identification of the unfertilized and fertilized eggs by hyperspectral imaging technology combined with GA-BP algorithm was proposed. The comparative analysis for the unfertilized and fertilized eggs was implemented by different pretreatment and principal component. In order to improve the performance of BP neural network, GA algorithm was used to optimize the network parameters. The application of GA-BP network was established the qualitative detection model. The results of the study showed that the MSC + SD pretreatment method was the most suitable for the model. The determination coefficient was 0.95, which indicated the optimized network

model had a good generalization ability and high prediction precision with unfertilized eggs accuracy being 93%, fertilized eggs accuracy being 94%, the overall accuracy being 93.5%. The results indicated that the method of non-destructive identification for fertilized and unfertilized eggs based on hyperspectral imaging technology combined with GA-BP algorithm was feasible.

Keywords: hyperspectral imaging technology, before incubation, hatching eggs, identification, GA-BP

Haiyan Zhuang, Gang Wang Mining multiple level association rules under weighted concise support framework

Computer Modelling & New Technologies 2014 ***18****(11) 394-400*

Association rules tell us interesting relationships between different items in transaction database. Traditional association rule has two disadvantages. Firstly, it assumes every two items have same significance in database, which is unreasonable in many real applications and usually leads to incorrect results. Secondly, traditional association rule representation contains too much redundancy which makes it difficult to be mined and used. This paper addresses the problem of mining weighted concise association rules based on closed itemsets under weighted support-significant framework, in which each item with different significance is assigned different weight. Through exploiting specific technique, the proposed algorithm can mine all weighted concise association rules while duplicate weighted itemset search space is pruned. As illustrated in experiments, the proposed method leads to better results and achieves better performance.

Keywords: weighted concise association rule, transaction database, closed itemset, support-significance

Hui Bu, Ran Liu Data oriented workflow using semantic technologies

Computer Modelling & New Technologies 2014 ***18****(11) 401-407*

Scientific workflows are a topic of great interest in the Grid community that sees in the workflow model an attractive paradigm for programming distributed wide area Grid infrastructures. Scientific workflows have recently emerged as a new paradigm for scientists to formalize and structure complex and distributed scientific processes to enable and accelerate many scientific discoveries. In contrast to business workflows, which are typically control flow oriented, scientific workflows tend to be dataflow oriented, introducing a new set of requirements for system. In this paper, we consider a general workflow setting in which input data sets are processed by a graph of transformations to produce output results. Our goal is to perform efficient selective refresh of elements in the output data, i.e., compute the latest values of specific out-put elements when the input data may have changed. The data provenance is investigated to be used to enable efficient refresh. The proposed approach is based on capturing one level data provenance at each transformation when the workflow is run initially. Then at refresh time provenance is used to determine (transitively) which input elements are responsible for given output elements, and the workflow is rerun only on that portion of the data needed for refresh. The reported preliminary experimental results are developed on the overhead of provenance capture, and on the crossover point be-tween selective refresh and full workflow computation development.

Keywords: scientific workflows, scientific workflow management system, semantic technologies

Shan Feng A boundary knowledge field based data mining method

Computer Modelling & New Technologies 2014 ***18****(11) 408-414*

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. Knowledge is the source of getting and keeping the strength of competence, knowledge becomes an important strategic resource at the age of knowledge economy. The way to facilitate knowledge transfer smoothly plays a critical role in determining the competence of an organization or organizational system. To explore the mechanism of knowledge transfer, the characteristics of sticky transfer (flow) is compared with the flow characteristics of viscous fluid, and the knowledge on the theories of field in physics, viscous fluid mechanics and boundary layer is employed and analysed in this paper. Firstly, the concept of boundary layer in the knowledge field is proposed to analyse the difference of knowledge stickiness in the knowledge field and describe the contradictory relationship between knowledge stickiness and liquidity. Secondly, based on the analysis of the knowledge transfer in the boundary layer, the dynamic mechanisms of sticky knowledge transfer are analysed from the three aspects of the knowledge potential difference force, the viscous force and the extern driving force. The rotation mechanism of knowledge in the knowledge field is discussed, and the dynamic model of the boundary layer of knowledge field is built. Finally, the phenomenon of knowledge flowing into and flowing out of boundary layer is discussed, and the knowledge transfer conservation equation in the boundary layer is constructed to describe the updating and appreciation of knowledge in the boundary

layer of knowledge field.

Keywords: data mining, knowledge transfer, knowledge field, boundary layer

Lijuan Liao Design of real-time clock based on ARM embedded system
Computer Modelling & New Technologies 2014 ***18****(11) 415-418*

Real-time clock of ARM processor has disadvantages such as dependence on the processor, low interrupt level and unadjustable accuracy. Linux embedded system based on ARM processor is provided with independent real-time clock using X1227 real-time clock chip. In addition, the work presents transplant of uCLinux system in S3C2440A, procedure of system boot loader and frame of Linux character device driver, thus achieving design of RTC driver based on I2C protocol of uCLinux system.

Keywords: S3C2440A, RTC, X1227

Xiaolei Li Spatial and temporal mining method using GPS data
Computer Modelling & New Technologies 2014 ***18****(11) 419-425*

Geographic information has spawned many novel Web applications where global positioning system (GPS) plays important roles in bridging the applications and end users. Learning knowledge from users' raw GPS data can provide rich context information for both geographic and mobile applications. However, so far, raw GPS data are still used directly without much understanding. Spatial-temporal data analysis plays an important role in many applications, including transportation infrastructure, border security and inland security. To analyse the moving patterns of vehicles on a road network, a measure for determining the similarity of vehicle trajectories with respect to space and time has to be defined. Although previous research has addressed the trajectory similarity problem, most of the studies focus on Euclidian distance instead of network distance. This paper deals with the variations in applying a spatial-temporal similarity measure with given Points of Interest (POI) and Time of Interest (TOI), treating spatial similarity as a combination of structural and sequence similarities that is evaluated using the techniques of dynamic programming. The similarity set thus formed will be used by the remote database to broadcast trigger-based messages to participating vehicles in a neighbourhood for future route- and information-sharing activities. The performance of the scheme is evaluated using experiments on standard real-life data.

Keywords: K-means, clustering algorithm, error rate, iteration, reduction, stable

Wenhong Wu, Hengmao Niu An improved watershed algorithm for image segmentation
Computer Modelling & New Technologies 2014 ***18****(11) 426-431*

Watershed transform is a key operator in image segmentation algorithms. However, the computation load of watershed transform is too large for real-time applications. Previously published watershed segmentation algorithms required at least three global synchronization points: minima detection, labelling and flooding. This paper presented an algorithm of watershed transformation based on opening-closing operation and distance transform. It improved the classical watershed segmentation algorithm based on distance transform, overcoming over segmentation. The experiment result demonstrated that this method for segmentation inherits the advantage of watershed algorithm based on distance transform that it successfully segment out each dowel in the image bringing convenience to computer vision and auto-counting of dowels. It also overcame over-segmentation existed in traditional watershed segmentation preserving the original edges of each dowel in the image completely. This algorithm can be combined with any image segmentation algorithm to give more precise segmentation results. An example is also shown by combining a background registration and change-detection-based segmentation algorithm with Watershed. This new video segmentation algorithm can give accurate object masks with acceptable computation complexity.

Keywords: watershed transform, image segmentation, automatic segmentation method

Ying Fang, Heyan Huang A hybrid multi-class text categorization based on SVM-DT
Computer Modelling & New Technologies 2014 ***18****(11) 432-438*

How to improve the text categorization efficiency as well as keeping high speed is a research problem. Several factors are effected the processing of the decision tree construction, such as, the degree, the balancing degree, the constructing way, the group number and the division degree between groups etc. Considered the various roles between the above factors, a comprehensive algorithm to construct the SVM-DT (Support Vector Machine - Decision Tree) is proposed. In this method, three conditions are considered respectively. The text categorization experiments on massive corpus demonstrate that the algorithm can improve the efficiency in some degree and decrease the training and testing time largely at the same time. The algorithm to construct the SVM-DT is feasible

and adaptable.

Keywords: Text Categorization; Support Vector Machine; Decision Tree; Multi-class category; Corpus; Positive Sequence Tree

Xuewen Yu Design of data acquisition node based on CAN bus
Computer Modelling & New Technologies 2014 ***18****(11) 439-443*

Taking single chip microcomputer (SCM) STC90C58AD as control core, a universal CAN bus data acquisition node was designed using CAN bus controller SJA1000 and CAN bus transceiver PAC82C251. Also, the hardware circuit and software design of CAN bus communication were given. Before data collection, current signal has been I-V transformed, and voltage signal has been filtering processed and amplified. Such program is suitable for multi-sensor field data collection in industry.

Keywords: CAN bus, SJA1000, data acquisition

Jiong Cai Reducer optimization design based on chaotic particle swarm optimization (CPSO)
Computer Modelling & New Technologies 2014 ***18****(11) 444-446*

For the optimization design of two-stage gear reducer, an optimization mathematical model is built in this work to determine the objective functions and constraints. And chaotic particle swarm optimization (CPSO) is utilized to optimize these functions and constraints. Algorithm simulation is carried out based on CSPO algorithm steps, and the results are compared with particle swarm optimization (PSO). Simulation indicates that CSPO can optimize the results of PSO and achieve faster convergence rate.

Keywords: chaos particle swarm algorithm, optimization design, mathematical model, reducer

Lifang Yang, Lin Liu TV image enhancement technology based on particle swarm optimization
Computer Modelling & New Technologies 2014 ***18****(11) 447-453*

Television images can be blurred and indistinct by noises in the acquisition and transmission process. Traditionally, control parameters of fuzzy enhancement algorithm are manually controlled, which leads to poor enhancement effect and efficiency. In this work, particle swarm optimization (PSO), due to its fewer parameters and global optimization capability, is combined with fuzzy enhancement algorithm for the optimization of fuzzy enhancement parameters. Simulation results show that PSO can make television images clearer and highlight in certain features, thus improving the visual effect of television images.

Keywords: particle swarm optimization, fuzzy enhancement, image processing, fitness function

Daode Zhang, Yanli Li The automatic cutting algorithm of the plastic film based on image processing
Computer Modelling & New Technologies 2014 ***18****(11) 454-458*

This paper proposes an algorithm based on image processing to solve the dislocation problem when cutting the edge**,** which is in the light of the uneven edge and the wrong side phenomenon. First of all, the images of the plastic bag edge captured by the camera are pre-treated**;** then the image is enhanced by fuzzy contrast enhancement algorithm**,** which is aim to extract the creases contour; finally the edge of the image is detected used by the Canny operator to get the crease on the edge of plastic bag, the cutter position can be obtained by offset some distance from the creases. The research results show that this algorithm is very effective to solve the dislocation problem when cutting the plastic bag edge automatically.

Keywords: fuzzy contrast enhancement, the Canny operator, edge detection, plastic film cutting

Rui Li, Qingshan Feng, Maolin Cai, Haijun Li, Chenghai Liu, Xiaoming Zhao The method of measurement for buried pipeline centerline based on data fusion
Computer Modelling & New Technologies 2014 ***18****(11) 459-464*

For the problem of long-distance oil and gas pipeline centerline measurement, a method of multi-sensor data fusion using the pipe centerline calculation is presented. The error model is set up by the system of navigation and nonlinear dynamic systems. Using the data of IMU and odometer to calculate the information of pig navigation. All of the errors were calculated by Kalman filter for estimation and compensation. It is concluded that the error for horizontal is 0.35m, the error for vertical is 0.74m for comparison of same pipeline centerline in different inspection time. One feature point is dug to verify the accuracy of the inspection which error is less than 1m. This method is effective for the buried pipeline to perform safely.

Keywords: in-line inspection of long distance pipeline, sensor fusion, inertia navigation, extended Kalman filter, error correction

Weiqing Li, Zehui Lu, Shihong Shen A hand gesture interaction system based on Kinect
*Computer Modelling & New Technologies 2014 **18**(11) 465-469*

We introduce a hand gesture interaction system using Kinect, which takes advantage of real-time dynamic motion capture, image recognition and so on, so that people can interact with computer by natural hand gestures. Five kinds of gesture are defined and can be recognized by the system. Kinect-based hand movements and gesture recognition algorithm is studied. A method for hand area image segmentation from the depth map is proposed, using the information of hand joints in skeleton map. We realized a hand gesture recognition algorithm with SVM and tested its stability and robustness. Finally, experimental results verified the feasibility of the algorithm, and a hand gesture interactive demonstration is implemented.

Keywords: hand gesture recognition, Kinect, SVM, interaction

Yujia Li, Jingbing Li Robust volume data watermarking based on perceptual hashing
*Computer Modelling & New Technologies 2014 **18**(11) 470-476*

This paper proposed a perceptual hashing algorithm of robust blind watermarking method for volume data. Which address the problems of authentication and protection of personal information. The scheme obtains the feature vectors of volume data and quantizes them to generate the hash value of the volume data. By combining the concept of zero-watermarking, the algorithm for watermarking of volume data that is robust to geometric attacks. The experimental results demonstrate that the proposed algorithm has good invisibility and robustness

Keywords: watermarking, perceptual hashing, volume data, DFT

Fan Wu, Jingbing Li The text image watermarking using arnold scrambling and DFT
*Computer Modelling & New Technologies 2014 **18**(11) 477-481*

With the popularization of Internet and the development at full speed of the multi-media technology, the copyright protection of digital works has already become the hot issue at present. Generally speaking, image, audio and video watermarking is comparatively similar in algorithm realization, with their redundancy, in which we can embed watermark. But other than the aforementioned, there is no redundancy to transfer secret info in text document. Nowadays, to embed watermark in text documents are limited to methods such as shifting the line and word, amending the characters' traits and disposing in the level of semantics. All these algorithms are not robust or lack of concealment, generally not serving the turn of Chinese text documents riffles. Based on the studies of the document digital watermarking methods and techniques, this dissertation presents that the problems of existed documents watermarking algorithms can be solved by Arnold Scrambling and DFT technique. The experimental results show that the scheme has strong robustness against common attacks and geometric attacks.

Keywords: Arnold scrambling, digital watermarking, DFT, zero-watermarking, text image

Xue-Chen Wang, Xiao-Guang Yue, Mostafa Ranjbar, Sanjay Kumar Boddhu, Maia Viera Cañiv Opposite degree algorithm and its application in engineering data processing
*Computer Modelling & New Technologies 2014 **18**(11) 482-485*

In order to analysis and predict data, a new intelligent algorithm (opposite degree algorithm) is used for actual engineering example. The algorithm is based on concept of prior numerical, posterior values and opposite degree in the nature. The human's languages have positive words and negative words. The matrix method can calculate the opposite degree and predict data by considering data's relationship and opposite degree. At the same time, relevant results are obtained through the opposite degree calculation by using the data of coal and gas outburst. After the comparison of the actual results, the accuracy of prediction is 100%. The preliminary validation of opposite degree algorithm shows that the algorithm is basically feasible and effective. If this algorithm can be improved, it is expected to be applied in practical fields more widely.

Keywords: opposite degree, algorithm, engineering data processing, coal and gas outburst, prior value, posterior value, prediction method

Hui Wang, Tao Zheng, Weiwei Zhang A strategy of attribute reduction based on partition
*Computer Modelling & New Technologies 2014 **18**(11) 486-491*

The attribute reduction is an important pre-processing step for data mining. In order to avoid striking equivalence classes repeatedly for positive region or information entropy reduction it is proposed to calculate attribute reduction by constructing partition directly. At the same time the judgments of the absolute reduction and the relative reduction based on the equivalent division are proved. And the data description quality for the relative reduction has been

defined. It is shown that striking minimum relative reduction of decision table is in the cost of the relative decline of description quality for classification.

Keywords: attribute reduction, partition, rough sets, data mining

Chao Chen, Jian Tang, Zuguang Jin A path planning algorithm research for seeing eyes robot based on visibility graph algorithm

Computer Modelling & New Technologies 2014 ***18****(11) 492-497*

This paper presents a new indoor path planning algorithm for seeing eyes robot using the RFID system. Through combination of UHF radio frequency identification system and low radio frequency identification system to realize accurate positioning for robot. New algorithm combines the idea of visibility graph algorithm with A^* search algorithm, which not only improves the efficiency of searching but also guarantees the feasibility of path planning at the same time. The simulation verify the effectiveness and feasibility of the method.

Keywords: path planning, seeing eyes robot, visibility graph algorithm

Meng Xu, Qingzhong Li, Lizhen Cui A framework to support flexible application collaboration in cloud computing

Computer Modelling & New Technologies 2014 ***18****(11) 498-504*

With the development of Internet, software has more and more collaborative trends. Collaboration application is becoming a new hotspot. Development of collaborative application process is complicated and need many technologies to realize it. It is difficult for an ISV (Independent Software Vendor) or a business department to provide all application modules of a whole collaboration application. Moreover it is difficult to integrate heterogeneous application module developed by different organizations (ISV or business department). According to the problems mentioned above, this paper proposes a PaaS platform framework to build collaboration application dynamically and flexibly. Such PaaS platform can provide the ability of application delivery deployment for organizations and provide the ability of collaboration application customization for end users.

Keywords: cloud computing, PaaS, collaboration application, process

Shaohua Tao, Meilian Li, Zhili Zhang Synchronization optimization under symmetry network

Computer Modelling & New Technologies 2014 ***18****(11) 505-509*

Rich symmetries have been found in many real networks, which is an extensive structural of networks. We study the relationships between synchronization and symmetry of network. One of fascinating problems related to symmetry network is how to optimization phase synchronization by employing symmetric structure. For this purpose, we optimize the network structure by symmetry properties of network, which is statistics the orbits of network. Nodes in the same orbit have similar properties, and then we reduce the network size by nodes in same orbit. We simulate the BA and SW model. The simulations result show that symmetry structure can optimize the network topology and can enhance phase synchronization of network.

Keywords: synchronization, symmetry network, orbit

Xinyuan Cao, Mingsheng Chen, Bingbing Chen, Liangliang Cheng, Qi Qi Solution of compressed sensing for wide angle EM scattering analysis based on MFIE

Computer Modelling & New Technologies 2014 ***18****(11) 510-515*

Fast analysis of electromagnetic scattering problems over a wide incident angle is always a difficult problem in computational electromagnetics. Up to the present, almost all of the traditional numerical methods need to solve one discrete angle after another to finish calculating this kind of problem. In this paper, we propose a new method, which can fix it effectively by applying compressed sensing into method of moments for magnetic field integral equation. The theory and calculation process of the solution are described in detail in the paper, and by numerical experiments of different three dimensional objects, the accuracy and the efficiency of the algorithm are also discussed.

Keywords: compressed sensing, method of moments, magnetic field integral equation

Zili He, Tingchun Shi, Biaobiao Gao The influence and improvement of scanning speed to the line width in fused deposition manufacturing

Computer Modelling & New Technologies 2014 ***18****(11) 516-520*

During the process of fused deposition manufacturing, the extruding speed and the scanning speed of nozzle don't match in some path of large change in direction may lead the extrusion of material more or less than demand too

much which causes accumulation of material or the lack of line width. The method of controlling FDM's forming accuracy which is under the restrictions of the mechanical structure of the work platform is studied. By adjusting the extruding speed and measuring the line width on the work platform the width of forming line in process can be controlled in allowed range. The experiment results indicate that through controlling the speed of extruding motor the extrusion of material is controlled efficiently and the accuracy of molding is improved.

Keywords: movement theory of nozzle, scanning speed, extruding motor controlling, measure of line-width

Haifei Zhang Query semantic data from relational database: an on-demand mapping approach
Computer Modelling & New Technologies 2014 ***18****(11) 521-524*

One of the tasks for semantic web is to integrate large amounts of current information in relational database, which behind Web into machine-understandable RDF data model to form a "web of data". So relational database semantic query namely RDF access to relational database is an important issue in semantic web research. To realize the query is to build mapping relation between relational database schema and ontology. However, there is natural isomerism between them. The traditional method to eliminate the isomerism is to convert relational database schema into a similar ontology form and then to build all concepts and attributes mappings between conversion ontology and input ontology. This paper realized an on-demanding mapping method when users request query, avoided building all concepts and attributes mappings between conversion ontology and input ontology and improved mapping efficiency.

Keywords: relational database, semantic data, conversion ontology, input ontology, on-demanding mapping

Xu Bing, Yizhi Zhang The study and design on the campus network surveillance
Computer Modelling & New Technologies 2014 ***18****(11) 525-528*

In order of improving the efficiency of network administration, one of the most useful methods is to inspect and measure of the network traffic. By introducing technology relevant to network flow monitor and analysing monitor system of campus network flow, this paper brings forth collection of Campus network flow and statistic of network flow, adopting Visual C++ 6.0 technology to design structure of this plan. This paper also points the key technology and means to realize monitor system of campus network flow and monitor system of campus network flow. It is reliable and extensible to for realization of system to improve administrative function. It is easy to realize the function of collection of campus network flow and flow statistic.

Keywords: network monitor, network flow, collection of flow, data statistic, VC++ 6.0

Fen Guo, Huaqing Min, Ming Yin A self-aware strategy for virtual machines placement on clouds
Computer Modelling & New Technologies 2014 ***18****(11) 529-535*

Cloud computing is a new computing service mode, and virtualization is a key technology of it. A self-aware strategy (SAST) for Virtual machines (VMs) management on clouds is proposed which is multi-attributed weighted on the resources. It manages the virtual resource basing on the requests of users and the real-time state of the system dynamically. It consists of three phases: (1) monitoring the cloud performance including VMs and Physical Machines (PMs), with the data standardized; (2) measuring the cloud load balance value with the attribute weighted measurement model; (3) using the placement algorithm to choose the best appropriate PM to place the VM requested. The main contribution of the paper is that a cloud load balance measurement model is introduced and a VM scheduling strategy is proposed which includes the VMs placement optimization algorithm and the VMs dynamic migration algorithm. The SAST is tested on the simulation platform comparing with other traditional ones. As a result, we concluded that it guaranteed the SLA and achieved better load balance of cloud. And at the same time, it minimized the number of the started PMs on clouds to reduce energy consumption.

Keywords: cloud computing, virtualization, placement, scheduling

Xiaorong Cheng, Sizu Hou Searching security policy with acyclic directed graphs
Computer Modelling & New Technologies 2014 ***18****(11) 536-539*

To improve the efficiency of policy searching, a method based on the use of a weighted directed graph is studied. Regarding security states as vertices and trigger conditions as edges, the security policy knowledge base can be described as an acyclic weighted directed graph. Firstly, the graph is divided into some areas with just has an initial state node and a termination state node. Secondly, weights for each edge are set according to trigger condition frequencies, and then the optimal path from the initial state node to the termination state node is found using the A* algorithm. Finally, all state nodes are reordered on the basis of their optimal path to build an adjacency matrix and conduct depth-first traversal to search policies. Experiments showed that this method improved policy search

efficiency.

Keywords: security policy search, acyclic directed graph, A* algorithm, adjacency matrix

Sizu Hou, Xiaorong Cheng Design and implementation of an event correlation model in a network security linkage system

Computer Modelling & New Technologies 2014 ***18****(11) 540-544*

Based on the shortcomings of poor compatibility, weak practicality, and low accuracy in current linkage systems, this paper designed a gradation of event correlation model with real-time response mechanism. It will be analyzing technology in data mining association rules is introduced to analyse the processing of security incidents. Then the system through the analysis of a large number of real-time data collecting all kinds of security devices found hidden in the data and related information to improve detection precision and safety accident treatment work. At last, apply this model into the system, to demonstrate the effectiveness of the model and priority.

Keywords: linkage system, event correlation model, correlation rule, security event

Yueqiu Jiang, Yujun Wang, Hongwei Gao, Shuang Ma Research on key technologies of medicine grain defect detection system based on machine vision

Computer Modelling & New Technologies 2014 ***18****(11) 545-551*

In the process of medicine grain production may generate many kinds of defects. If these unqualified medicine granules are not timely detected, it will not only affect the company's reputation but also the health of the patient. This paper mainly studied how to detect the unqualified medicine grain base on machine vision. It mainly consists of three kinds of common defects segmentation and defect area calculation. Firstly, preprocess the medicine grain image for the following procedures. Secondly, obtain the defect region by improved segmentation algorithm, in order to deal with three different drugs grain defects this paper improved three segmentation algorithms, for the damaged tablets propose a local edge detection algorithm that based on grey level difference, for the irregular-shaped tablets adopts the ellipse detection algorithm, which based on Hough transform technology, for the Capsule has air uses the multi-scale Canny edge detection operator. Finally, adopt Chain code contour tracking algorithm and Three-point method calculate the area of the damaged tablets, determine whether the tablets meet the requirements. Experiments show that the system can detect unqualified medicine granule quickly and accurately, it is of great practical value.

Keywords: image processing, median filtering, defect detection, region segmentation, area calculation

Jiangbo Zheng Research on a regional innovation system: viewed from the degree distribution of complex networks theory

Computer Modelling & New Technologies 2014 ***18****(11) 552-557*

A Regional Innovation Systems is viewed as a special type of complex network in this paper, and the complex features of the entities in the system as well as the interactions between the entities are discussed in detail. Based on the Degree Distribution of Complex Networks theory, firstly this paper proves the feasible degree distributions in practical networks through mathematical reasoning and analysing. Then, an empirical case of degree distribution characteristics for the Regional Innovation System is tested by mean of calculating the Correlation Coefficient with the statistical data of Guangdong Province. Coupling Complex Networks Theory with these data and the practical conditions of Guangdong province, this paper sheds light on the insight of the calculating results. Such research shows that the methodologies and conclusions of this paper are proper.

Keywords: regional innovation system, degree distribution, complex networks, data calculation, Pearson correlation coefficient

Weiqing Li, Ranran Xu, Mengyu Yuan A GPU-based simulation system for infrared images of deep space targets

Computer Modelling & New Technologies 2014 ***18****(11) 558-564*

In study of deep space targets recognition, infrared images of deep space targets are needed for repeat testing and evaluating. Since the limitation of deep space flight experiments, it is difficult to obtain sufficient infrared images under different conditions. Infrared image simulation technology is brought up to solve this problem efficiently. The principle of deep space targets infrared imaging was studied. Based on the infrared sensor's optical properties, a hierarchical imaging model was built. The infrared camera and all the effects were simulated respectively, including motion trail of target and space objects, blurring, dispersion, blind elements, and noise. A mixed noise model was introduced by combining the random noise and Perling noise model. In the image simulating process, Graphic Processing Unit was used to produce noise image in real time. According to the reference photo of infrared sensors,

infrared simulated images were evaluated using histogram distribution, the trend of intensity, and Signal to Noise Ratio, and the results show these images satisfied targets recognition algorithm.

Keywords: deep space targets, infrared optical properties, infrared imaging, graphic processing unit, Perling noise

Nan Yao, Kaisheng Wang, Jin Yu Analysis based on computer graphic design and visual communication design

Computer Modelling & New Technologies 2014 ***18****(11) 565-568*

The development and application of computer graphic design and visual communication design greatly changes people's life. Designers can use design tools, which are highly-advanced and professional to create a broader scope of design and design theme works. In our daily life, computer graphics technology is widely applied into various fields, such as the military, medical, communication, art etc. Computer graphics and image technology has been emphasized in the visual field, which improves the overall effect and the level of visual communication design to a large extent. This article mentioned the definition of the technology of computer graphics and the visual communication, then described the significance of realizing the visual communication design elements, and hoped the results will be helpful to relative fields.

Keywords: computer graphics, visual communication, the overall effect, the significance

Xiehua Yu Content of smart wireless sensor network security and its network security policy

Computer Modelling & New Technologies 2014 ***18****(11) 569-573*

Wireless sensor network is generally composed by plenty of micro-sensors that arranged on designated area. The supervision of these sensor nodes is used to finish the collecting, disposing and uploading of vast information. However, the security of wireless sensor network has many problems, since the sensor node itself exists plenty of limitation. Aiming at the problems of smart wireless sensor network security as well as the analysis of wireless sensor network node easy been attacked, wiretapped and forged without safeguard, this paper put forward a security policy of smart wireless sensor network. Simple and useful intrusion detection policy was realized from a series of improvement of LEACH protocol of low-energy self-adaptation cluster routing protocol.

Keywords: smart wireless sensor network security, secure routing, key management, intrusion detection

Hong Yang Duality for multi-objective semi-infinite programming with $K-(F_b,\rho)-$ convexity

Computer Modelling & New Technologies 2014 ***18****(11) 574-577*

In this paper, some nonsmooth generalized convex functions called uniform $K-(F_b,\rho)-$ convex function, uniform $K-(F_b,\rho)-$ pseudoconvex function, uniform $K-(F_b,\rho)-$ quasiconvex function are defined using $K-$directional derivative and $K-$subdifferential. Nonsmooth multi-objective semi-infinite programming involving these generalized convex functions is researched, some Mond-Weir type duality results are obtained.

Keywords: Nonsmooth, Multi-objective Semi-infinite Programming, Mond-Weir type duality, Uniform $K-(F_b,\rho)-$ Convex Function

Yihe Liu, Shuang Zhang, Yuping Qin A role-based security information flow model in grid environment

Computer Modelling & New Technologies 2014 ***18****(11) 578-583*

Security is an important component of a grid, and it directly affects the development of the grid and the practical application of the grid system software. According to the practical application problem (namely realization of the role-based management) in the role management occurring in the information application system of my school, a role-based security information flow model is proposed from the point of view of guaranteeing the information security. In this paper, the object concept in the general network environment is expended, and the organization security classification of an object and relation between the security classification and the role set are used to classify the security and define the strategy for information flow, finally a security information flow model based on the grid environment is presented. The safe classify of the object is divided by the related information of role set, the organization security classifications, and classifications etc. At the same time, the information flow role is described. A new secure information flow model based on rules and grid environment is described using these methods. It is proven from strict mathematical justification that the new model satisfies properties of the finite lattice and least upper bound operator, and it is reasonable and safe. Furthermore, it is an extension of the BLP model and the role-based information flow model as well as extension of the security information flow model in the general network environment. Therefore, it is significant to the study of grid security.

Keywords: Grid Security, BLP model, Rule, Information flow model

Operation research and decision making

Xiaohong Kon, Lixia Chan, Junpeng Xu Fashion colour forecasting based on BP neural network
Computer Modelling & New Technologies 2014 ***18****(11) 584-591*

Fashion colour forecasting is a hot issue in fashion industry and also a hard problem because of much uncertain information. Utilizing strong mapping capability of BP Neural Network (BPNN) for nonlinear function, this paper investigated the forecasting model of fashion colour. Based on colour data of recent several years, the forecasting model for future colour trend is discussed and built. The historical data are input to train the Neural Network weights and the different parameters of BPNN were investigated to find how to affect the forecasting performance. The results demonstrate the algorithm is very efficient in colour forecasting and can approximate nonlinear relationship of fashion colour very close.

Keywords: Fashion colour, BP neural network, Trend forecasting, Network parameters

Jinghai Yin, Zhendong Mu Gender impact on the identification based on EEG
Computer Modelling & New Technologies 2014 ***18****(11) 592-599*

To study the genders impact on identification, this paper analysis the electroencephalograph (EEG) of eight male subjects and seven female subjects. In order to reduce the noise signal interference, the high pass and low pass were used to cut extra frequencies, and in order to prominent the feature signal, the power spectrum method was used to convent the time domain signal to frequency domain, and then fisher distance was used to extraction the feature. All EEG signal was acquired by neurescan, and the EEG signal was evoked by VEP method used subjects photo. The experiment was divided into three models: all subjects were the same sex, added some opposite sex, added some stranger. The analysis results show, to model 1, the correct recognition rate for male subjects, average is 88.50, and this for female, average is 92.51%; the false recognition rate for male subjects, average is 30.84%, and this for female, average is 27.67%,this result indicates VEP can be used as identification tool, the results of model 2 and model 3 show weather opposite sex or stranger should affect the correct recognition rate, but to male subject, the opposite sex effect is greater than stranger, to female, the result were reversed. The results also show noise photo affected female lower than male.

Keywords: Genders Impact, EEG, gender, Fisher Distance

Zhengmei Lin, Jiwei Yao Sport service evaluation of urban community based on fuzzy comprehensive evaluation
Computer Modelling & New Technologies 2014 ***18****(11) 600-604*

Fuzzy comprehensive evaluation refers to a comprehensive assessment method based on fuzzy mathematics through quantifying the factors difficult to quantify with obscure boundaries and using the principle of fuzzy relation synthetic. Fuzzy comprehensive evaluation can comprehensively evaluate the object system involving fuzzy factors and is widely applied in the fields of economy and society. This paper takes new public service theory as the research perspective; designs evaluation index system which can comprehensively reflect sport development level of urban community applies fuzzy comprehensive evaluation to evaluate sport development level of urban community in Hunan province and provides the evaluation process. The evaluation results show leadership organization and team construction of the urban community fail to develop to certain level and that the health index of community residents is slightly low. This paper presents suggestions to improve this situation.

Keywords: Fuzzy Mathematics, Fuzzy Comprehensive Evaluation, Evaluation Index System, Urban Community, Sport Service

Feng Jiang, Min Gao, Hui Xia, Qingyu Xiong, Junhao Wen An evaluation approach based on word-of-mouth for trust models in recommendation systems
Computer Modelling & New Technologies 2014 ***18****(11) 605-609*

Recommendation systems have been recognized as an effective approach to heavy information load. In recommendation systems, trust/reputation has attracted increasing attention because it helps to improve the precision of recommendation and the robustness of systems to shilling attacks. Recommendation system oriented trust models, mostly rating-based, used to build the reputation and trustiness among users. They are often evaluated in terms of how accurately they help to predict user ratings and how robustly they resist shilling attacks. However, those evaluation techniques disregard the trust values themselves: how accurately they calculate the trust values themselves is not measured. To solve the problem, in this work, we propose an approach to measure the trust values based on

electronic word-of-mouth (eWoM) theory. The eWoM believes a user is reliability if he is of good public praise. In our approach, firstly, according to eWoM, the reliability value of a user can be judged by other users' votes - whether the user's ratings or feedbacks are positive or negative. Secondly, the trust values of users can be calculated by a trust model. Finally, we compare trust values and reliability values. As a case study, we propose a simple rating-based trust model and then evaluate the trust model based on the proposed evaluation approach and Amazon dataset.

Keywords: Recommendation Systems, Evaluation Approach, Trust Model

Huajie Chen, Mingchang Liu Application of principal component analysis in fire risk prediction of stadium
Computer Modelling & New Technologies 2014 ***18****(11) 610-615*

The security of stadium is one of the problems increasingly concerned by people. Among various security problems, the fire risk of stadium is the most important. By using principal component analysis (PCA) and combining the influencing factors of risk prediction of stadium, several influencing factors of the fire risk of stadium are evaluated and analysed, then the uppermost principal component factors are selected, the inconspicuous components are eliminated, the influence of relevant factors on the fire risk prediction of stadium is analysed, and the correlation of various indicators with fire risks of stadium is understood to find the potential best security management prediction system, for the purpose of taking pertinent prediction measures of fire accident risks to overcome the unfavourable factors in the security management, and ensure the safe and normal operation of stadium as well as the personal and property security of sporters.

Keywords: Principal Component Analysis, Stadium, Fire Risk, Prediction

Ronghui Hu Application of factor analysis in risk evaluation of basketball arena project construction
Computer Modelling & New Technologies 2014 ***18****(11) 616-620*

In basketball arena project construction, the project safety is an important topic. Aimed at the current situation of basketball arena project construction, the risk factors in the project construction are evaluated from the aspect of factor analysis method and the data in the process are processed according to relevant theory of factor analysis method and by virtue of SAPP software, so as to verify the scientific reasonableness of the method. In addition, in the process of evaluation, the subjectivity is linked with the objectivity in combination with the reality to provide main basis for the safety risk problems in the basketball arena project construction, which is beneficial for the safety management in the project construction. The application of factor analysis method in the basketball arena project construction can provide some new thoughts for other analysis methods, so as to enhance the theoretical research basis for the factor analysis method.

Keywords: Factor analysis method, Basketball arena, Project construction risk, Evaluation

Hao Zhang Short-term traffic flow forecast of highway network based on chaos time series method
Computer Modelling & New Technologies 2014 ***18****(11) 621-625*

With the construction of the highway network and the growing traffic flows, demand for real-time control and guiding service have become increasingly prominent. As for short-term forecast of highway network, it is not only the basis and foundation of the real-time control and guiding service for traffic flows, and the precise forecast result will have a magnificent impact on improving the traffic capacity and service levels. This paper builds a short-term traffic flows forecast model for highway network based on chaotic time series analysis and prediction theory. The forecast of the traffic flows in given areas can be calculated. Results show that this model is feasible and has a high accuracy.

Keywords: highway network, traffic flow, chaos theory, time series, prediction

Sujiao Liu Optimization of industrial structure configuration based on fruit fly optimization algorithm
Computer Modelling & New Technologies 2014 ***18****(11) 626-630*

Establishing the mathematical optimization model of Shangqiu's agricultural structure from three aspects - economy, ecology and society, the agricultural industry structure in Shangqiu is regarded as the research object, utilizing fruit fly optimization algorithm to solve the mathematical model. Simulation results show that fruit fly optimization algorithm can solve the optimal solution for Shangqiu's agricultural industry structure to achieve the maximum benefits, thus providing the decision-making basis for adjustment and development of Shangqiu's agricultural structure.

Keywords: fruit fly optimization algorithm, industrial structure, optimization model, population size, iterations

Shuwen Ma Investigation of the strategic alignment in public sector organisations using knowledge based

strategy
*Computer Modelling & New Technologies 2014 **18**(11) 631-636*

Ageing workforce is one of the critical challenges a public sector organisation is facing and will face more terribly. The cost of knowledge loss can have huge impact on the bottom line of business. A major issue facing Public Sector Organisations (PSOs) in recent times has been the increasing pressure to demonstrate value from investments in Information Technology (IT). One omnipresent but often overlooked solution is Strategic Alignment (SA) between corporate business objectives and IT initiatives as achieving SA remains one of the more enduring challenges for organisations. Moreover, deriving value from IT investments through SA requires both IT and business executives to foster synergies between their respective areas, which contributes to increased organisational performance. Conversely, business-IT misalignment may eventually lead to failure in achieving business goals. Shared Domain Knowledge (SDK) is a key factor within the social dimension of SA and is concerned with the level of knowledge business and IT have of each other's to each other's missions, objectives, and plans. This paper presents findings from a study which investigated the influence of SDK on SA within organisations in the Australian public sector. The developed research model examined SDK between business and IT professionals as a factor that would potentially influence SA. The findings suggest that increased levels of SDK between professionals from the business and IT domains leads to more efficient SA in PSOs.

Keywords: shared domain knowledge, strategic alignment, public sector, organisational performance

Gang Lu Evaluation of e-commerce website based on fruit fly algorithm optimization RBF algorithm
*Computer Modelling & New Technologies 2014 **18**(11) 637-644*

The feature and various index properties of E-commerce website are considered as a whole by applying the expert grading method, with the construction of the multi-index hierarchical structure of an E-commerce website competitiveness index evaluation and the establishment of an E-commerce website competiveness index evaluation index system. The competiveness level of the website is quantified after calculating the competiveness index of the E-commerce website. On this basis, this work adopted Radial Basis Function (RBF) neural network algorithm to perform evaluation research on the competiveness index of E-commerce website. Aiming at the problems exist in the evaluation research, this work tried to use Fruit Fly Optimization Algorithm (FOA) to perform improvement on the RBF neural network algorithm. Through the simulation and comparison of practical examples, FOA-RBF algorithm is obviously better than RBF neural network algorithm when the E-commerce website competiveness index is calculated and evaluated, thus the validity and reliability of calculating method presented in this work are verified.

Keywords: e-commerce, neural network, radial basis function, fruit fly optimization algorithm, expert grading method

Zihui Yang Web search engine based trend analysis of electronic commerce market
*Computer Modelling & New Technologies 2014 **18**(11) 645-650*

Google Trends offer weekly free information about internet searches. Users can see and download search volume patterns for the search term and the information is available by the category and by the location of those making the search. Also, Google provides "Hot searches" and "Top charts" including top and rising searches that include the search term. All these information is up to date as it provides weekly figures for a period up to and including the current week. In here, the proposed methods present a predictive model for Electronic Commerce market using the searched data in Google search engine (Google Trend data). Through the predictive model for the market and analysis of the Google Trend data, the proposed methods can get an efficient and meaningful result for the Electronic Commerce market, also we can enforce the marketing on highly ranked countries and cities. Those are very useful information for the Electronic Commerce manufacturers.

Keywords: electronic commerce, search engine, trend analysis

Hong-Liang Qiu Impacts of tourist destination image on place attachment and tourist loyalty
*Computer Modelling & New Technologies 2014 **18**(11) 651-656*

Although the importance of tourist destination image as a tool to enhance tourist loyalty is commonly acknowledged, prior research on the relationship between tourist destination image and tourist loyalty is not in-depth. Drawing on place attachment theory, a model depicting the relationship among tourist destination image, place attachment and tourist loyalty is constructed. Using the sample of 337 inbound tourists from Japanese and Korean and the structural equation modeling method, the empirical results reveal that: 1) Landscape image, merchandise image and facility image have positive effects on affective image. Landscape image and merchandise image significantly and directly affect place attachment while partially mediating the effect of affective image. Service image has a direct effect on place attachment. Facility image has an indirect effect on place attachment. Facility image significantly and directly

affects tourist loyalty while partially mediating the effect of place attachment. 2) Affective image has a direct effect on place attachment and it is an antecedent of tourist loyalty while completely mediating the effect of place attachment.

Keywords: tourist destination image, cognitive image, affective image, place attachment, tourist loyalty

Lei Hong Deployment algorithm based on dynamic multi-populations particle swarm optimization for wireless sensor networks

Computer Modelling & New Technologies 2014 ***18****(11) 657-662*

Aiming at improving coverage rate and reducing coverage holes of wireless sensor networks, this paper proposes a deployment algorithm based on dynamic multi-populations particle swarm optimization. K-Means clustering algorithm is employed to divide the network into several sub-populations dynamically, which could weaken particles on the pursuit of local optima, realize the improvement of basic PSO (Particle Swarm Optimization) algorithm, and solve the "premature" problem of basic PSO algorithm effectively. In addition, it also accelerates the convergence of the algorithm. Simulation results show that this deployment algorithm can improve the network coverage rate effectively. Comparing with the conventional particle swarm optimization algorithm, its coverage rate is increased by 3.66%.

Keywords: deployment, particle swarm optimization, *k*-means, wireless sensor networks

Bao Wu Financial contagion dynamics and fragility assessment of industrial complex network

Computer Modelling & New Technologies 2014 ***18****(11) 663-668*

The paper proposed a mathematical modelling of financial contagion dynamics that is tightly linked to systematic risk of industrial complex network. And the paper provides a practical method to assess fragility of industrial complex network in the context of financial contagion. To examine its function, an experimental analysis based on real data set of a Chinese textile industrial network is conducted. The experimental analysis shows that the method proposed in the paper is effective ad reliable, and is capable to assess fragility of industrial complex network in the context of financial contagion.

Keywords: financial contagion, fragility, industrial complex network, assessment method

Han Jingwei, Owen Liefung Yue Study of collaborative relations through "Daily Maersk" service in China

Computer Modelling & New Technologies 2014 ***18****(11) 669-674*

Maersk Line has recently introduced a daily service in Asia-Europe route. This new service will bring in a revolutionary impact on the current fixed schedule shipping service in China. There are different views from China market on such daily service since the daily Asia-Europe service route calls mainly China's ports. Through the study of the 'daily Maersk' from 5 perspectives, namely shipping market demand, the required basic port infrastructure, relevant Government policies and regulations, operational aspects and low carbon emission, it is possible to reveal a kind of 'collaborative relations' in the China fixed schedule shipping market. The analysis adopts a qualitative approach via a questionnaire survey of selected experts. With the support of a theoretical mathematical model, it is possible to quantify the collaborative relations in different development phases in China. The study concludes that there are 3 broad phases of the development of collaborative relations and the current daily service emerges in the 'new development era' phase. It is because there are sufficient freight and adequate port infrastructure. The daily service outperforms the general industry practice under the "Rotterdam Rules" on carrier's responsibility and liability and injects contemporary logistics management concepts in the industry. Moreover, it is in the same pace of current China's requirements on energy saving and carbon emission reduction targets.

Keywords: 'Maersk', collaboration, collaborative relations, Asia-Europe

Zhuo Yang, Hong-liang Qiu Empirical analysis on influencing factors of capital structure of China's real estate listed company: evidence from Chinese listed company

Computer Modelling & New Technologies 2014 ***18****(11) 675-681*

Capital, the guarantee for normal operation of an enterprise, is crucial to production. Therefore, optimizing the capital structure has been an important task in the development of enterprises. Combining with VAR model, this paper selects 7 factors, which are profitability, operation ability, the current debt servicing ability, development ability, tax, strategic position and the assets structure, so as to make an exploratory study on the capital structure of Chinese real estate industry, based on the data from June 2002 to December 2012. The study shows that the Liquid ratio and the gross profit are of great effects on the capital structure of a listed company.

Keywords: capital structure, listed company, the real estate, impulse responses

Liming Yang, Yanwen Wang, Xiuju Gao Research on the performance measurement model of knowledge management based on Grey relational analysis
Computer Modelling & New Technologies 2014 ***18****(11) 682-688*

Performance measurement of knowledge management is a decision-making analysis project that involves multiple complex factors, levels and fuzzy uncertain information. On the basis of analysis of influence factors of knowledge management performance measurement, the study established an evaluation index system of enterprise knowledge management performance. Meanwhile, by combining grey relational analysis method and Euclidean distance measurement, a performance measurement model of knowledge measurement was established. Via standardization of different types of evaluation indexes of knowledge management performance, Euclidean distances of standardized evaluation indexes of knowledge management performance and the grey relational coefficients based on Euclidean distances were established respectively. Then the weighted grey correlations of evaluation indexes of knowledge management performance were obtained. In this way, evaluation analysis of enterprise knowledge management performance was realized. Finally, the model and algorithm was tested with a case study. The result proves that the method of combining grey relational analysis and Euclidean distance is efficient and has its application value in performance evaluation of knowledge management.

Keywords: knowledge management, performance measurement, Grey relational analysis, euclidean distance, model

Yaqin Lu, Cunzhi Tian, Yi Wu Corporate growth, liquidity assets value and financing decision
Computer Modelling & New Technologies 2014 ***18****(11) 689-694*

Based on the external financing analysis framework under asymmetric information, this paper analyzes the influence of corporate growth on liquid assets value and financing decision. Both of the theory and numerical calculations show that liquid assets value would increase with the increasing in corporate growth. For the reinvestment decision in the case of higher reinvestment demand, if the price of liquidity assets is higher than value, it is optimal not to reinvest; if the price is lower than value, it is optimal to reinvest; if the price is equal to value, there is no difference found in reinvesting or not.

Keywords: liquidity shocks, liquid assets value, corporate growth, financing decision

Xiaoyuan Geng, Yongde Wang Identification and application of investors' risk appetite-based on the analysis of risk allocation of China multi-layer capital market system
Computer Modelling & New Technologies ***18****(11) 695-705*

In order to measure investors' risk appetite more accurately, from the focus on the investors' demand for the capital market, this article deduces the utility level of investors in the capital market by the inverse of the investor demand (only when the demand function satisfies integrablility, then it will be deduced the utility function inversely), and thus measure and identify the investors' risk appetite. While based on this theory approach, the paper empirically analyses the risk allocation of China multi-layer capital market, and the results show that: risk allocation of China multi-layer capital market system is non Pareto efficient, the risk allocation of each market does not meet the structure of multi-layer capital market established, but these problems can be improved by adjusting the market trading mechanisms.

Keywords: Stochastic demand; Integrability problem; Risk appetite; Risk allocation

Yan Li, Fuyu Wang The Decision of Scrap Reverse Logistics Operation Mode for Steel Enterprises Base on Evolutionary Game Theory
Computer Modelling & New Technologies 2014 ***18****(11) 706-712*

This paper discusses how steel enterprises choose the appropriate scrap reverse logistics operation mode from self-operation mode and the third party mode. To solve the problem, evolutionary game theory is used to research the game relation of cooperation between steel enterprises and the third -party enterprises. Firstly, the replication dynamic equations of both players are build based on payoff matrix .Secondly, the evolutionary stable strategies are acquired by stability analysis on evolution dynamic process of the game two players. Finally, combined with numerical simulation, some factors that impact stable strategy choice are analyzed such as initial state of system, extra income and risk cost of cooperation, invested initial cost and loss for cooperation. The conclusion provides theoretical reference for steel enterprises selecting scrap reverse logistics mode.

Keywords: scrap reverse logistics mode;evolutionary game theory;replication dynamic equation;steel enterprises

Biao Song A Scientific and Research Performance Evaluation model of Institutes of Higher Learning Based on Multilevel Fuzzy Comprehensive Decision Analysis
*Computer Modelling & New Technologies **18**(11) 713-717*

Scientific and research performance is an important part of the evaluation on the overall strength and the ability to achieve sustainable development of institutes of higher learning. With the purpose to deal with the multi-attribute, multilevel and fuzziness, this paper constructs a scientific and research performance evaluation system for institutes of higher learning and proposes an evaluation model based on multilevel fuzzy comprehensive decision analysis. After indicators are standardized, we can get the fuzzy nearness by fuzzy comprehensive decision analysis and Analytical Hierarchy Process for evaluating various schemes. It provides support for the analysis of teaching ability, research ability and the sustainable development of institutes of higher learning.

Keywords: Higher school; scientific and research; fuzzy decision analysis; performance evaluation; model

Liao Yiqin, Chen Xun Social welfare, climate change and strategy selections for developing countries
*Computer Modelling & New Technologies **18**(11) 717-723*

This paper introduces external effect of carbon emission in social welfare function, constructs two-stage trade game among three countries, analyzes partially equilibrium output of the three countries and then discusses the influences of different carbon tax policies on social welfare in each country. The study shows it is more effective for developing countries to adopt strategies about founding free trade area and domestic carbon tax collection, up to a higher social total welfare lever, when they face carbon border tax adjustments(BTAs) from developed countries under Nash game conditions. By further studying, the efficiency of domestic carbon tax policy depends on the carbon intensity relation of each country; a higher relative intensity of carbon abroad decreases the negative external effect value caused by the carbon emissions.

Keywords: border tax adjustments, social welfare function, game theory

Guishen Yu An evaluation model of sustainable development of sports tourism industry based on matter-element theory
*Computer Modelling & New Technologies 2014 **18**(11) 724-731*

The sustainable development of sports tourism industry is concerned with many factors. Its evaluation is a complex system engineering. This paper studies the complexity and diversity of factors that influence the sustainable development of sports tourism industry and proposes an evaluation model of sustainable development based on matter-element theory. An indicator system is put in place. Evaluation indicators of classical field matter-elements model, section domain matter-element model and evaluation objects matter-element model are constructed based on matter-element theory. Different methods of calculating extension degree are adopted according to characteristics of the evaluation objects matter-element model to calculate the comprehensive extension degree between evaluation objects matter-element model and classical field matter-elements model. This extension degree refers to the layer of sustainable development capability of evaluation objects. It will provide strategic support for the development of sports tourism industry. Case study has proved that the model and the algorithm are effective.

Keywords: sports tourism industry; sustainable development; matter-element theory; extension degree; evaluation model

Xiqin Wen, Kaibo Wu, Anhua Peng A method of relative Grey relation degree combined with combination weights for materials selection
*Computer Modelling & New Technologies 2014 **18**(11) 732-738*

The quality and cost of a product rely heavily on suitable material selection, and therefore the ability to select the most appropriate material for a given application is the fundamental challenges faced by the design engineer. The general grey relational analysis (GRA) has three weaknesses, (i) the weight determination depends only on expert judgments, (ii) the qualitative indexes are simply quantified with exact numbers, and (iii) the general GRA only takes into account the relationship between the imaginarily best material and the candidate materials. Weights were determined by combining subjective and objective weights based on maximum deviation, the qualitative indexes were fuzzily quantified through trapezoidal fuzzy numbers (TFNs), and then ranked alternatives according to relative grey relation grade. The illustrative example showed that the results matched well with that using WAA and TOPSIS, proved the proposed method reasonable and trustworthy. And therefore the proposed method possesses important application values.

Keywords: material selection, combination weights (CW), relative Grey relation analysis (RGRA), fuzzy numbers, maximum deviation

Jian Zhang, Zhongwei Shen, Hao Shen An analysis of landscape characteristics of urban rivers – with the case of Chengdu
*Computer Modelling & New Technologies 2014 **18**(11) 739-749*

Urban river landscape is a key ecological regulating system in cities, attracting increasing attention from people. Our research conducted a systematic and comprehensive research of landscape characteristics of main rivers in Chengdu from some new perspectives, including the city in the large, city location and Single River. This research is intended to better understand the landscape structure along main rivers in Chengdu, to facilitate the evaluation of macro-regional ecological environment, and to provide references for the study of regional ecological environment.

Keywords: urban rivers, landscape characteristic, city location, tree crown sizes, canopy contribution rate, landscape area

Zhenhua Liu Analysis of flood disasters from 206 BC to 1949 in China
*Computer Modelling & New Technologies 2014 **18**(11) 750-757*

Flood disasters produces not only direct economic losses but also serious environmental problems, the government attaches great importance to flood disasters prevention from ancient times to the present. A variety of analytic methods are adopted, for example mathematical model analysis, chart analysis, qualitative and quantitative analysis. Some mathematical models are constructed by means of statistical analysis of historical data for flood disasters, such as mathematical model for flood cycle each dynasty. There are the following results. First, there are 1037 times flood disasters from 206 BC to 1936, once every 2.07 years. According to analysis of mathematical model for flood cycle each dynasty from 206 BC to 1936, the flood cycle is the declining concave curve, except for the Five Dynasties. Second, in modern history, the most number of catastrophic floods is the Yellow River reaching 8 times, the Yangtze River 5 times. Finally, water pollution caused by floods can not be ignored. In 2012 92 percent of untreated rural sewage was discharged dispersedly.The vast majority of pollution load on the ground runs into directly riverway through stormwater runoff. Water wells are easily contaminated for germs and parasite. We should pay attention to water pollution caused by floods and combine organically flood control with water environment protection.The government coordinates closely ideas, rule of law, technology and investment with rural residents for governing rural sewage to reduce the risk of water pollution caused by floods.

Keywords: flood disasters, flood cycle model, qualitative and quantitative analysis, rural sewage, water pollution

Mei Sun, Hui Feng, Siyuan Tang, Ziqiang Luo Application of fuzzy mathematics models in hospital management evaluation
*Computer Modelling & New Technologies 2014 **18**(11) 758-763*

To formulate a hospital management evaluation system and conduct empirical research, Delphi, analytic hierarchy process, and fuzzy comprehensive evaluation are used to build a hospital management evaluation system and conduct hospital management empirical research. And then a hospital management evaluation system is built. This system contains five first-level indices, namely, administrative management, human resource management, medical management, financial management and logistical support, and 23 second-level indices. Empirical research shows that the comprehensive evaluation of hospital management is above average and that service quality and infrastructure construction have the highest and lowest evaluation scores, respectively. This evaluation system is an effective tool for studying hospital management. Chinese hospital management requires further improvement, especially in terms of infrastructure construction.

Keywords: fuzzy mathematics, hospital management, evaluation system, analytic hierarchy process

Jian Zhang, Zhongwei Shen, Hao Shen Chengdu River status and cause analysis
*Computer Modelling & New Technologies 2014 **18**(11) 764-770*

Focusing on the basic model of the riverside landscape of Chengdu's five major rivers, this paper conducts contrastive analysis on the model characteristics of urban riverside landscape, location of the city where it locates and the urban construction status and explains in detail about the relation between the formation reasons of urban riverside landscape and the level of urbanization development. It discusses the effect of urban development and residents demand on urban riverside landscape design so as to perfect the design thought about it.

Keywords: cause analysis, landscape model, correlation analysis, land property, river status

Wu Xiao, Guanghua Yang, Yaoqi Yang Coal mining subsidence data extraction and verification in a high groundwater area based on Landsat-8 imagery and subsidence prediction
*Computer Modelling & New Technologies 2014 **18**(11) 771-775*

Coal is the main energy resource in China, with its extraction and utilization playing an important role in national economic development. However, coal mining may be causative with respect to critical land subsidence and damage to land. The eastern plain coal mining region of China represents an example of overlapping crop cultivation and coal extraction, and is considered a coal mining area characterized by a high water table. Accordingly, declining ground elevation and seasonal water logging of land due to mining subsidence have become major concerns within the region. Based upon the existing procedure for land reclamation planning, both land damage boundaries and land damage magnitude were determined for the region via subsidence prediction and vertical displacement, respectively. In the current study, a coal mine in Shandong province was employed as a case study area with the following work phases implemented: 1) subsidence prediction was implemented, with land damage magnitude demarcated via the proposed standard; 2) mining induced water area and wetland was extracted via use of remote sensing; 3) comparative analyses of the aforementioned methodologies were undertaken, with a revised methodology proposed for effective provision of improved land damage demarcation. Results indicate that land subsidence at Dongtan coal mines during May 2013 was 1616.70 hm^2, of which categorically mild, moderate, and severe lands were 22.54 hm^2, 257.67 hm^2, and 436.49 hm^2, respectively.

Keywords: Landsat-8, high groundwater, mining subsidence, land damage information, subsidence prediction, damage assessment

Wei Wang, Xiaodan Huang Computational model of implicit interaction for entertainment

Computer Modelling & New Technologies 2014 ***18****(11) 776-784*

Implicit interaction between human and computer is worthy of being researched, especially in the entertainment application. The reason is that in order to be more natural, computers need to interact and collaborate with persons actively. For this purpose, a computational model of implicit interaction is proposed and applied to a computer for entertainment. Firstly, emotional Hidden Markov Model (eHMM) as a part of the computational model of implicit interaction is researched. Then, three parts of ACT-R cognitive architecture are integrated into it to apply for entertainment. Finally, some experiments are carried out with styles of game process recording. Results indicate that the proposed model is helpful to make computers more active and adaptive to persons by adjustting entertainment process, which illustrates a good prospect of application.

Keywords: implicit interaction, ACT-R, affective computing, entertainment

Jingtao Liu Construction and effect evaluation of the college English multimedia teaching system based on the Blended-learning model

Computer Modelling & New Technologies 2014 ***18****(11) 785-790*

As network technology develops, an increasing number of new technologies and concepts have entered the traditional college teaching system. To improve the education quality and level of domestic college English, as well as to meet the demand for the reform of domestic college English education, the Blended-learning model is introduced to construct a new multimedia teaching system based on the traditional multimedia teaching model of college English. Moreover, a questionnaire analysis is conducted among all staff members that participate in the new multimedia teaching system. Questionnaire analysis indicated that the construction of the Blended-learning model-based traditional multimedia teaching system of college English proposed in this study is practically operable. The effect of this model is significantly better than that of the traditional English teaching model.

Keywords: Blended-learning model, college English, multimedia teaching, system construction

Jianbing Liu, Fang Guo Construction quality risk management of projects on the basis of rough set and neural network

Computer Modelling & New Technologies 2014 ***18****(11) 791-797*

Construction quality associated with the life of construction enterprises. Risk assessment of construction quality referred to a comprehensive evaluation of the degree of risk confronting construction units during the construction process. Construction units provide a decision-making basis for employers and supervisors, who considered the evaluations of construction quality and accidents in their decisions. The risk factors of construction project quality were classified into personal risk and material risk and machinery and equipment risk and method risk and environment risk. On the basis of these factors, we constructed an index system of project construction quality risk. The risk evaluation model of project construction quality was constructed on the basis of rough sets and neural networks. Finally, a case study residential building projects in the Ganzhou Development Zone and research tools of Rosetta based on rough sets and MATLAB7.0 based on neural networks were used to test model accuracy and reason ability. Empirical results showed that the model has great practical significance.

Keywords: construction quality risk assessment, projects, rough set, neural networks

Lingyun Zhou, Youheng Huang Dynamic modelling of container transport modes between inland terminals and seaports

Computer Modelling & New Technologies 2014 ***18****(11) 798-801*

Container transport has played an important role in international trade. This paper mainly focuses on the container transport modes selection between inland terminals and seaports through a comprehensive comparison of multiple decision factors. According to the multi-object fuzzy mathematics decision-making theory and the container transport processes, the main factors influencing the container transportation mode decision are analyzed. Moreover, a fuzzy decision model of container transport modes between inland terminals and seaports is built by introducing varied weight factors. Thereafter, the proposed model for the final selection of container transport modes between Changsha city and Shanghai harbor is demonstrated through an illustrative example. The results of this example indicate that the model can reflect dynamically importance degrees of related decision factors and different demands of decision makers, and this approach provides a more accurate, effective, and systematic decision support tool for the optimized intermodal mode selection between inland terminals and seaports.

*Keywords: c*ontainer transport, transport mode, route optimization, fuzzy decision making

Xiaohua Ma Location of competitive basketball athletes based on RSR comprehensive evaluation method

Computer Modelling & New Technologies 2014 ***18****(11) 802-806*

Rank sum ratio (RSR) positions each research index to superior or inferior index, conducts rank allocation of these indices in all evaluation objects according to rank principles, sorts them, and carries out grading according to the RSR value. In this way, the superior and inferior indices can be clearly determined. For complex open sports, good sport decisions can make each member give play to their strong points to achieve good results. However, most existing research methods belong to non-quantitative research methods. Research on the location of competitive basketball athletes is rare. This study adopts RSR comprehensive analysis method to assess examination indices for 15 members in a basketball team of a college to select the members that are most suitable for five locations in basketball competitions and corresponding benches and candidates. A general method is provided to decide the location of commercial basketball athletes.

Keywords: sport decision-making RSR comprehensive evaluation method, commercial basketball athlete

Wenxin Xu, Jiwei Yao Management information system for college track and field games on the basis of infrared radio-frequency technique

Computer Modelling & New Technologies 2014 ***18****(11) 807-813*

College track and field sports involve numerous participants and events. Moreover, recording and announcement jobs are complex. The requirements for accuracy and precision are high. The traditional manual management mode cannot adapt to development needs. Based on the advantages of Radio Frequency Identification (RFID) such as no-barrier read, remote penetration, speed scanning, large memory space, anti-pollution capacity, durability, diversified shapes, reusability, and good security, an RFID-based infrared radiofrequency technique is designed to achieve automatic identity verification for athletes and to construct a complete score management information system. An automatic check mode is used to identify athletes by combining an ID card, a card reader, and a computer. This mode can rapidly identify the personal information of athletes and prevent cheating. This system has important significance for improving management level and reducing manpower.

Keywords: RFID, track and field games, management information system, card reader

Guo-yi Chen Models for measuring and predicting value creation during merger and acquisitions: a study of bank industry

Computer Modelling & New Technologies 2014 ***18****(11) 814-819*

This paper employs event study methodology with a 36-day event window to assess the value effects of the US bank mergers occurring between 1994 and 2003. A 38-transaction sample is chosen from the top fifty US bank mergers (according to the assets of targets) occurred during the period from 1994 to 2003. Through analysis, result indicates that the average cumulative abnormal return of the bidders in the chosen sample is negative (-0.99%), while the targets and combined firms are both positive (15.07% and 2.57% respectively). Significance testing also verified that the negative bidder return is confirmed to be insignificant, whilst the positive return of the target and combined firm are both significant. Combined together, It indicates that the 3,517 US bank mergers occurred between 1994 and 2003 create insignificantly negative value for bidders, whilst benefit the target and the integrated banks with significant positive

gains.

Keywords: mergers, acquisitions, event study, value effects, measurement model

Fumeng Gao Structural equation model of college foreign language writing and classroom teaching quality from perspective of teacher evaluation

Computer Modelling & New Technologies 2014 ***18****(11) 820-823*

Existing teaching quality evaluation systems have insufficient comprehensive indices reflecting teaching quality: index repetition, complication, quantification difficulty, and difficulty in judging the relationship. To make classroom teaching quality evaluation consistent with teaching effect, element weight is confirmed based on a structural equation model. The present paper takes a college foreign language writing class as the object of study and sets up a structural equation model for classroom teaching quality evaluation. Effect coefficients of each variable are calculated, and the weight of each element of classroom teaching quality evaluation is then obtained. Such method avoids artificially scoring elements during the confirmation of element weight, reduces evaluation subjectivity, and makes the evaluation results more accurate, rational, and credible.

Keywords: structural equation model, teacher evaluation, classroom teaching, teaching quality

Chen Yan, Lan Nan, Liu Yunlang, Huang Rong The empirical analysis on mode of developing the rural areas with the aid of the urban areas

Computer Modelling & New Technologies 2014 ***18****(11) 824-831*

Developing the rural areas with the aid of the urban areas, which was summarized through the long practice in China, is a new mode of promoting the coordinated development of urban-rural areas. The evaluation of the impact of this mode on promoting the coordinated development of urban-rural areas has great significance for reference for the urbanization development in other countries around the world. This paper takes Chenggong New District in Yunnan, China as an example, calculates the coordination degree of its urban and rural development in 1999, 2004 and 2009 with the use of urban and rural development coordination degree model, then makes clustering analysis of the urban and rural development coordination degree in various districts in Chenggong New District in 2009 with the use of Ward's method. Results indicate that a) the urban and rural development coordination degree of Chenggong New District is 0.359 in 1999, 0.545 in 2004, and 0.504 in 2009. The increase in the coordination degree year by year shows the mode of developing the rural areas with the aid of the urban areas has obvious effects on the coordinated development of urban-rural areas; b)the mode of developing the rural areas with the aid of the urban areas has obvious coordination impact on the aspects of infrastructure construction and basic livelihood guarantee, but has not enough coordination impact on the aspects of environmental governance, medical treatment and education; c) when the mode of developing the rural areas with the aid of the urban areas is applied, the coordinated development of industries and especially the agricultural industrialization development should be paid attention to.

Keywords: the mode of developing the rural areas with the aid of the urban areas, the coordinated development of urban-rural areas, Chenggong New District, China

Qi Yue, Yuhua Li Two-sided matching considering the preferences of agents and intermediary

Computer Modelling & New Technologies 2014 ***18****(11) 832-835*

This paper proposes a novel method for solving the two-sided matching problem, in which the preferences provided by agents are ordinal numbers, and the preferences provided by intermediary is the expense standard on ordinal numbers. In this paper, the description of the considered two-sided matching problem is given. Then the concepts of satisfaction degrees and expense are introduced. Furthermore, a multi-objective optimization model can be set up consider satisfaction degrees of agents and expense of intermediary. The method of weighted sums based on membership function is used to convert the multi-objective optimization model into a single-objective model. The matching result is obtained by solving the model.

Keywords: two-sided matching, ordinal number, satisfaction degree, expense

Zhaoxia Si, Ningli Chen VAR dynamic analysis of the impact of population structure on urban residential land price: the case of Zhengzhou

Computer Modelling & New Technologies 2014 ***18****(11) 836-842*

Based on the vector autoregression model, this paper focuses on Zhengzhou and uses the 1994–2013 population structural variables and relevant data on commercial and residential land price to analyze the dynamic relationship between population structure and residential land price through impulse response function and variance decomposition. Results show a co-integration between Zhengzhou residential land price and three variables, namely,

urbanization rate, per capita disposable income, and household size. The short-term variation in land price is mainly caused by the residential land price itself, and the three population substructural variables show long-term effects on the land price with a certain time lag. Among these variables, per capita disposable income has the shortest positive effect on residential land price, whereas urbanization rate has the longest and most remarkable positive effect.

Keywords: population, land price, VAR model, Zhengzhou

Xue Wu Research on the application of ARMA model in China's trade economic development
Computer Modelling & New Technologies 2014 ***18****(11) 841-847*

The slowing down development of China's foreign trade brings significant impact to its regional economy, which makes many middle and small-sized export enterprises facing with business crisis and even bankruptcy. In this paper, a deep analysis on the development status of China's trade economy was conducted with ARMA model. The short-term development trend and causes of China's trade economy were predicted and analyzed. Results demonstrated that China's imports will decrease while export will keep increasing. The balance of trade in China will decrease and the total volume of foreign trade will decline. This indicates the depression of China's trade economy, which is against China's economic development. To improve the export structure and facilitate the rapid development of trade economy, China is suggested to adopt various measures, such as optimizing fiscal policy, promoting technology upgrading, enhancing export management and perfect import-export infrastructures. Meanwhile, China shall expand domestic demand vigorously to offset impact of trade economic slowdown on regional economy.

Keywords: ARMA model, trade economy, application research, decision reference

Yanpeng Feng, Hongyan Zheng, Bo Wu Transmission power control strategy based on partially observable Markov processes for IEEE802.11 WLAN
Computer Modelling & New Technologies 2014 ***18****(11) 848-854*

With the limitation of independent channels in IEEE802.11 WLAN, co-channel AP interferes with each other in repeat coverage area. As traditional AP launched a fixed power in sending data packages, which ignores the differences and mobility of STA. According to the mobility of STA and the partially observable feature of AP-STA link, this paper analyses the internal relations among the link state, transmission power and channel interference of co-channel AP, and proposes a single-link transmission power control (TPC) algorithm based on the Partially Observable Markov Decision Processes (POMDP). Firstly, the single AP-STA link POMDP-TPC model is constructed, and the neural network learning model is established to describe the observation function of POMDP-TPC. Secondly, the algorithm constructs reachable belief searching trees to obtain the approximate optimization, which implements the dynamic creating and on-line updating of the power consumption policy. Finally, under the experiment environment in OPNET and IEEE802.11b, the results demonstrate the algorithm can reduce the AP power consumption efficiently and improve the network throughput greatly.

Keywords: wireless local area networks, access point, POMDP, transmission power control, anti-interference

Tan Li, Weiwei Wu An analysis framework for building commodity futures market simulation model based on heterogeneous traders
Computer Modelling & New Technologies 2014 ***18****(11) 855-860*

Futures price volatility is always the hot topic for academic researchers and traders in futures market. For exploring the rules of futures price fluctuation, we try to develop a new analysing framework from the angle of the heterogeneous traders. We describe heterogeneous traders as four respects: trading motive, predict styles, risk-return preference, reaction speed, which can accurately describe the heterogeneous traders in futures market. According to the categories of heterogeneous traders, we construct traders' internal model, which is used in a commodity futures market simulation model to verify the framework, and the results show the framework is useful.

Keywords: bounded rationality, commodity futures markets, simulation model, heterogeneous traders

Jingxian Tang, Yan Yu, Haitao Zhou GEM-based evaluation of competitiveness of enterprise cluster
Computer Modelling & New Technologies 2014 ***18****(11) 861-864*

Enterprise cluster development can improve development capacity of enterprises and industrial competitiveness by lowering cost, stimulating innovations, increasing efficiency and intensifying competition. It is one key industrial development trend in the future and an important consideration of enterprises in choosing regions. Recently, industrial cluster in China has achieved primary development. Provinces and cities are putting great efforts to creating industrial parks of Industrial Cluster. This plays an important role in promoting regional economic development and regional industrial competitiveness. In this paper, a competitiveness evaluation model of enterprise

cluster was established using the GEM model. It was applied to the furniture enterprise cluster in Guangdong Province. Accuracy, time effectiveness, advantages and disadvantages of the model was analyzed through the case study. Advantages and disadvantages of regional enterprise cluster can be identified from competitiveness analysis of the model. Countermeasures to these disadvantages were suggested.

Keywords: GEM model cluster competitiveness

Yuejun Liu, Fang Fang, Yali Zhang Research on contract management evaluation of construction company based on fuzzy comprehensive evaluation model
Computer Modelling & New Technologies 2014 ***18****(11) 865-873*

Contract management is the core of project management of construction companies. Establishing a perfect contract management evaluation system of construction companies is of important significance to enhance their contract signing and performance as well as competitiveness in international market. Based on the overall contract management of construction companies, this paper discussed how to establish an evaluation system. Combined with fuzzy comprehensive evaluation, the established evaluation system was verified by an empirical analysis.

Keywords: contract management capability, evaluation model, fuzzy comprehensive evaluation, empirical analysis

Hailei Zhao, Dehuan Jin Research on the risks of financial derivatives and risk control from the perspective of the financial crisis
Computer Modelling & New Technologies 2014 ***18****(11) 874-879*

After the economic crisis in Wall Street in 2008, the risks of financial derivatives and risk control have received great attention from countries in the world. The research on the risks of financial derivatives and risk control can effectively prevent the financial crisis. Through the extraction of the characteristics of risks of financial derivatives, the paper expounds the application of sensitivity analysis and in the risk measurement financial derivatives. The result shows that despite certain achievements of China's financial derivatives in the development process, there is still much room for improvement. Based on the result and with reference to the practice of the UK, US and Japan in the risk prevention and control of financial derivatives, the paper proposes some suggestions.

Keywords: financial derivatives, risk prevention and control, VAR model, financial crisis

Xueying Zheng A study on the relationship between the developments of China's trading economy and environmental factors
Computer Modelling & New Technologies 2014 ***18****(11) 880-884*

Trade environment refers to the social and economic environment facing trading economy, and it can directly affect the pace of trade development and trade structure. Establishing a good trade development environment can optimize trade structure, and promote the sustained and healthy economic development. This paper firstly analyzes the current environment of China's trade development with relevant theories, then analyzes the relationship between environmental factors and the development of trading economy based on indexes, builds error correction model, and uses the model to make in-depth analysis on the impacts of environment on the development of trading economy. According to the study results, the ratio of trade dependency over trade surplus exerts side effects on the development of trading economy, while trade openness exerts positive effects on the development of trading economy. In terms of import, the development of China's trading economy shows excessive dependence on foreign energy and other resource-based products, this seriously affects the stable development of economy. In terms of export, some Chinese companies that have overcapacity show over-reliance on international markets, thus reducing the competitiveness of export products, and leading to the situation where China's trading economy is vulnerable to fluctuations of international markets. Based on this, a method is proposed in the paper to improve China's trade environment.

Keywords: error correction model, trading economy, environment; impact

Pingping Qiao Empirical analysis of Chinese cultural products trade based on the gravity model
Computer Modelling & New Technologies 2014 ***18****(11) 885-889*

Based on date of bilateral cultural trades between China and 26 other countries (regions), the paper uses a gravity model to perform an empirical test on the influencing factors and export potential of cultural products. The research shows that Chinese economic size, importing country's economic scale, GDP per capital, trade openness and preferential trade arrangements have a positive effect on the export of Chinese cultural products. Spatial distance and cultural distance from China have a negative effect on exports of Chinese cultural products. The effect of cultural

distance is bigger than that of spatial distance. It is very important for promoting the exports of Chinese cultural products to strengthen communication and overcome "cultural discount". In the 26 samples, the trades of cultural products between China and 7 countries (regions) are sufficient, and the trades of cultural products between China and other 19 countries (regions) are insufficient. China has a great potential for cultural products trades, but China should develop it targeted.

Keywords: gravity model, cultural products trade, influencing factors, export potential

Xuewen Jin Research on the influential factors of China's logistics demand based on the econometric model

Computer Modelling & New Technologies 2014 ***18****(11) 890-894*

Logistics demand reflects the pace of economic development, and the pace of economic development also affects the change of logistics demand in turn. In fact, logistics demand is affected by many factors. Through the theoretical analysis, the paper points out that the influential factors on logistics demand mainly include four aspects, namely, economic development level, regional industrial structure, macro-economic policy and economic system, and consumption level and concept. Then, the econometric model is established with the data in different periods of time based on those four aspects to make an in-depth analysis of the application of the model. At the same time, in order to promote the development of the logistics industry, improve logistics demand and meet the development demand of the regional economy, it puts forward reasonable suggestions.

Keywords: logistics demand, econometric model, influential factors

Hong Wang The review and research on agile oriented method n the pilot industry system

Computer Modelling & New Technologies 2014 ***18****(11) 895-901*

Agile methodology has getting wide recognition within the software industry due to its flexibility and ability to cope with rapid changes in the software development environments. It is however comes with a number of demands that must be complied with. This paper presents a pilot systematic literature review (SLR) study on the limitations of Agile methods in the industry based on primary research. In this study, conference and journal papers in the IEEE, published between 2007 and 2012 were investigated. 29 papers were found as the most relevant. While the SLR findings have brought to both limitations in the implementation and in the Agile methods, the former becomes the most addressed issues. The result revealed that high dependency on people/personnel, organizational dependency, as well as high impact on organizational structure and culture as the three most repeatedly addressed factors. While these three factors are mutually related, people factor especially upper level management strong involvement and support can be regarded as a primary necessity in the Agile implementation. In spite of apparent emphasis on people critical function stated in the Agile principles, plus the excellent rules in the gist of the principles, the problem still arose when it comes to the implementation part. This indicates the need for future work on proper guidelines for management, given that existing guidelines for Agile adoptions and implementations are general and less focus given to upper level managers.

Keywords: Agile, limitation, human factor, pilot management, systematic review

Jianda Zhu Study on GIS based intelligence transportation

Computer Modelling & New Technologies 2014 ***18****(11) 902-907*

Recently a vast amount of geographic information system (GIS) data make road networks around the world as polylines with attributes. In this form, the data are insufficient for applications such as simulation and 3D visualization—tools which will grow in power and demand as sensor data become more pervasive and as governments try to optimize their existing physical infrastructure. With the increased dissemination and computing power of mobile devices, it is now possible to execute distributed artificial intelligence applications for various situations: intelligent routing using algorithms, planning, is tribute optimization of traffic lights. This paper reports on our development of a GIS-based traffic network analysis system, named GIS-based Transport Decision Support System, which provides a graphical analysis platform to transportation planners and researchers for transportation network analysis. The system has the functions of designing traffic networks on digital maps and doing traffic equilibrium analysis, as well as a novel function to integrate local detailed structures of intersections into global networks. The latter is particularly useful for the analysis of large traffic network where the detailed local network structures of some intersections have to be taken into account. The system links great volumes of traffic data and geography information data accumulated for visualization traffic analysis. We added information data on the following: road structure data, zone geography information data and node geography information data. The system also enabled us to extract traffic data by road section and by specific condition.

Keywords: K-means, clustering algorithm, error rate, iteration, reduction, stable

Ting Hao, Xi Zhao Financial management based decision making in the big data era
Computer Modelling & New Technologies 2014 ***18****(11) 908-913*

Financial management are derived for the profitability analyses of demand-side management (DSM) alternatives. The present value of cost and equivalent uniform annual cost models are selected to determine the least-cost solution, while the net present value, pay back year and benefit/cost ratio models are proposed for the execution of cost–benefit analyses. In a market economy, Market orientation is the only correct choice for economic development strategy, so it is also significant for financial management innovation. While most studies in the past decade focused on the consequences of fund in financial management, few have investigated antecedents to market orientation concept. This paper derives two fuzzy financial profitability models, namely, a least cost solution model and a cost-benefit analysis model, to evaluate the fuzzy financial profitability of load management alternatives. A straightforward vertex parameters' fuzzy mathematics operation using the function principle is de-rived as an alternative to the traditional extension principle and is applied to evaluate a number of different financial decision indexes. The developed models represent readily implemented possibility analysis tools for use in the arena of uncertain financial decision-making.

Keywords: financial management, decision making, finance decision making, big data

Shixiong Wang, Yi Jiang Extending opinion dynamics model for collective online behaviours analysis
Computer Modelling & New Technologies 2014 ***18****(11) 914-920*

In online social networks, opinion dynamics generally lead to different types of collective online behaviour such as consensus, polarization and fragment. Then an open problem arises: how are different typical collective online behaviours emerged from the behavioural decisions of individual and interactions among individuals during the process of opinion dynamics? This work examines the process of opinion dynamic in online social networks and different types of interactions among individuals on this process. An opinion-driven dynamics model, which combines a social network-based opinion dynamics model with generative individual behaviour, is proposed by adding antagonistic responses to the DW model. The proposed model integrates three types of interactions and setting up two thresholds to characterize individual behaviour. The behavioural component utilizes an initiation threshold such that if an individual's opinion exceeds this threshold, the individual will initiate the behaviour. In order to verify the effectiveness of the model, simulations are presented to examine how different typical collective behaviours emerge. As a result, we find that opinion dynamics with different threshold lead to different types of collective online behaviours. The openness of individuals to a differing opinion is the key factors to consensus or fragment.

Keywords: opinion dynamics, consensus, antagonism, fragment, online social network

Aiyun Guo A student profile model based online English learning
Computer Modelling & New Technologies 2014 ***18****(11) 921-926*

Many Chinese universities have begun reforms to enhance educational competitiveness in our globalizing economy. This study aims to ascertain the status of English education and English-medium instruction at a Chinese engineering school and to offer workable suggestions for English communication training for Chinese graduate engineering students. Colleges and universities across China are adopting bilingual education to meet the need for well-rounded personnel with sound knowledge in specialized areas and competency in foreign languages. The development and difficulties of current bilingual education in China are discussed. First, the short-term English word context is generated to identify related concepts of the word. Second, the user context is generated based on the click through data of users. Finally, a forgetting factor is introduced to merge the independent user context in a user session, which maintains the evolution of user preferences. It is significant to the reform from teacher-centered to student-centered teaching mode. It is helpful to the cultivation of the students' collaborative ability and spirit, and has important theoretical significance and practical value.

Keywords: English education, online learning, student profile

Jianxun Wang, Qingyun Zhou Industrial design based on computer aided simulation
Computer Modelling & New Technologies 2014 ***18****(11) 927-933*

Recently, new technologies have emerged in industrial automation platforms. A rapid modelling and simulation environment is required to integrate these new technologies with existing devices and platforms to reduce the design effort and time to market. System-level modelling is a popular design technique that provides early simulation,

verification, and architectural exploration. However, integration of real devices with system models is quite challenging due to synchronization and hard real-time constraints in industrial automation. Simulations are software tools approximating and predicting the behaviour of real industrial plants. Unlike real plants, the utilization of simulations cannot cause dam-ages and it saves time and costs during series of experiments. A shortcoming of current simulation models is the complicated runtime integration into legacy industrial systems and platforms, as well as ad-hoc design phase, introducing manual and error-prone work. This paper contributes to improve the efficiency of simulation model design and integration. It utilizes a semantic knowledge base, implemented by ontologies and their mappings. The integration uses the Automation Service Bus and the paper explains how to configure the runtime integration level semantically.

Keywords: assembly systems, estimation, part feeding, pose statistics, simulation

Jie Guo Intelligent human resources systems in the information technology era
Computer Modelling & New Technologies 2014 ***18****(11) 934-939*

Effective human resource management facilitates the success of an organization and the progress of a society. We describe an evolutionary computer model that simulates different modes of interaction between people and their environment. A two-level genotype-phenotype structure is used to represent the characteristics of an individual. The environment is modelled as a two-dimensional array of regions in which each region is characterized by a set of regional features and organizational culture (e.g., leadership strategies). Human resource decisions are subject to limitations, because they always depend on human knowledge, judgment and preference. Decision support applications can be used to provide fair and consistent decisions, besides to improve the effectiveness of decision making processes. This study consists of three parts; the first part is to understand the IDSS concepts, applications and related research in human resources decision making application known as HR DSS. The second part is to identify the potential intelligent techniques that can be used in HR DSS application, and the third part is to suggest the HR DSS framework that is related to human resource decisions. Finally, the paper proposed the HR DSS framework and the potential intelligent techniques that can be used to develop the IDSS application in any phases of decision making processes.

Keywords: human resource decision support, human resource management systems, human resource strategy

Ting Hong PCA-based analysis on factors of English translation ability
Computer Modelling & New Technologies 2014 ***18****(11) 940-944*

This work studied the relationship between English proficiency and translation ability as well as that between English and translation teaching among English majors. PCA is utilized to quantify and analyze the relationship between translation ability and English skills such as listening, reading, error correction and writing. Then we obtain the quantitative relation between translation ability and its factors, providing decision basis for the improvement of English translation skills and teaching.

Keywords: PCA analysis, translation ability, English teaching, quantitative analysis

Jiale Tian Collaborative mechanism of project management based on complex system theory
Computer Modelling & New Technologies 2014 ***18****(11) 945-950*

Although it is arguable that humans have been studying complex systems for thousands of years, the modern scientific study of complex systems is relatively young in comparison to conventional fields of science with simple system assumptions, such as physics and chemistry. The history of the scientific study of these systems follows several different research trends. The project management community is actively demonstrating substantial interest in the development of viable methods to assess and improve project management maturity. There is little empirical evidence on the benefits of deploying a project management office (PMO) and/or conducting project reviews. The increasing complexity of exploratory activities in pharmaceutical innovation makes less likely that a project can stand alone. Project managers not only resort to in-house innovation but also external sources to propel a central project. This paper introduces the notion of a quality function for individual tasks and uses the functional form of the bivariate normal, to model quality at the task level. Using real data from two case studies, a translation agency and a software development company, the quality function is specified and incorporated into a mathematical programming model that allows quality to be explicitly considered in project planning and scheduling. An alternative model formulation leads to the creation of quality level curves that enable managers to evaluate the nonlinear tradeoffs between quality, time, and cost for each of the example projects. The results of these analyses lead to specific decisions about the planned values for these three fundamental dimensions at the task level and provide insights for project planning and scheduling that can be gained through improved understanding of the choices and tradeoffs.

Keywords: project management, project planning, project scheduling, quality management

Wentao Xiao, Gongfa Li, Honghai Liu, Guozhang Jiang, Ze Liu, Disi Chen, Weiliang Ding, Wei Miao, Zhe Li Soft sensor system of coke oven flue temperature based on CBR and PCA-RBFNN
Computer Modelling & New Technologies 2014 ***18****(11) 951-958*

The key process indicator - coke oven flue temperature – is difficult to detect online with instruments in the coke oven heating process, thus an intelligent forecasting model is developed which is composed of four parts: the data gathering and handling unit, the optional forecasting unit, the online amendment unit and the effect evaluation unit. The optional intelligent forecasting model and its corresponding algorithm are established for different categories of practice operating conditions. In normal operating condition, the nearest neighbor clustering algorithm based on the principal component analysis and neural network with the radial basis function is selected. In unconventional operating condition, the case-based reasoning technology is selected. The models of different conditions are validated and applied according to the actual data in a steel enterprise coke production, the results show that the established forecasting model can reflect different practice conditions and meet the real-time control requirements.

Keywords: intelligent prediction, neural network, principal component analysis, neighbor clustering algorithm

Hongli Xiao Intelligent transportation systems based on computer aided simulation method
Computer Modelling & New Technologies 2014 ***18****(11) 959-964*

Road transportation is one of the main sources of greenhouse gas emissions, which lead to global warming and climate change. Promoting the decarbonisation of this sector through more efficient and greener mobility is a challenging task that can be achieved by intelligent transportation systems (ITS) enabled by vehicular communications. Intelligent Transportation Systems (ITS) have been developed for more than ten years in China. Furthermore, a new generation Intelligent Transportation Systems should be launched to meet the requirement of rapid development of transportation in China. For the last two decades, intelligent transportation systems (ITS) have emerged as an efficient way of improving the performance of transportation systems, enhancing travel security, and providing more choices to travellers. A significant change in ITS in recent years is that much more data are collected from a variety of sources and can be processed into various forms for different stakeholders. This paper presents an overview of the background, concepts, basic methods, major issues, and current applications of Parallel transportation Management Systems (PtMS). In essence, parallel control and management is a data-driven approach for modelling, analysis, and decision-making that considers both the engineering and social complexity in its processes. The developments and applications described here clearly indicate that PtMS is effective for use in networked complex traffic systems and is closely related to emerging technologies in cloud computing, social computing, and cyber physical social systems.

Keywords: agents, intelligent transportation systems (ITS), mobile agent systems, multiagent systems

Xing Xu, Yun Zhao, Jia Hu, Xinli Wu Research on the robustness of supply chain network with uncertainty
Computer Modelling & New Technologies 2014 ***18****(11) 965-970*

A new method is proposed to enhance robustness of complex supply chain network structure in uncertain conditions of market, costs and others. The researcher constructs the topology of a supply chain network firstly, and then explains the concept of central nodes of network and robust supply chain network based on scale-free network model from the perspective of quality control and places research emphasis on the method of enhancing the robustness of complex supply chain network structure. Finally, the supply chain network of auto parts enterprises in Zhejiang Province is cited as an example to verify the validity of the method.

Keywords: supply chain network, network quality control, scale-free network, robustness

Benxue Wang Housing price forecast based on rough-set extreme learning machine
Computer Modelling & New Technologies 2014 ***18****(11) 971-975*

The work, based on various factors affecting housing price in 31 provinces cities as research object, firstly adopted rough set theory to reduce those factors. Then, the main reduced influence factors were used as the input of extreme learning machine. On such basis, the housing price forecast model based on rough-set extreme learning machine was ultimately established. According to the simulation results, the algorithm in this work has good prediction effect, and its prediction precision is higher than that of BP neural network and RBF neural network. Therefore, this algorithm, with a certain practical and theoretical value, can be promoted to other areas for predication and classification.

Keywords: rough set theory, neural network, extreme learning machine, real estate, forecast error

Xiaoyan Li, Min Zhang Prediction of employment figures in the three main industries based on extreme learning machine
*Computer Modelling & New Technologies 2014 **18**(11) 976-980*

Extreme learning machine, with a fast speed of training, achieves globally optimal solutions and excellent generalization ability. This work is based on the production value and employment figure of the three main industries in China during 1996-2012 as research objects. An ELM employment figure prediction model was constructed with production value and the employment figure of the three main industries respectively as the input and output of extreme learning machine. Simulation test results proved a good effect and high accuracy of ELM employment figure prediction model. Besides, the comparison of ELM, BP and RBF algorithms further proved the effectiveness and precision of ELM, which has certain practical application value.

Keywords: three main industries, extreme learning machine, production value, neural network, training samples, test samples

Lihua Yu, Qingsheng Xie Research on quality information integrated management for complex precision parts in multi-varieties and small batch manufacturing mode
*Computer Modelling & New Technologies 2014 **18**(11) 981-985*

There were many problems such as quality information interaction low, real-time monitoring and tracing of production quality difficulty in workshops for multi-varieties and small-batch complicated precision parts. These problems restricted seriously the improvement of product quality. Therefore, a quality information integrated operation mode was built, combining with advanced management thoughts of lean management and integration. The main characteristics of quality information integrated management process were analysed. At last, the mode was applied in a manufacturing enterprise in CASTC 061 base for testing its effects. Results show that it is successful and satisfactory in practical application effect.

Keywords: workshop, integration, multi-varieties and small-batch, quality control

Yanli Yang, Xianyu Wang Incentive analysis of labour dispatch under asymmetric information theory
*Computer Modelling & New Technologies 2014 **18**(11) 986-990*

Under the circumstances of the fierce market competition, great importance is attached to labour dispatch as an innovative employment way since it saves more labour costs for enterprises. However, there exists obvious asymmetric information among the dispatched employees, employers and accepting entities. This paper makes an analysis of the incentive offered by accepting entities to the dispatched employees and employers under the circumstances of asymmetric information and symmetric information. It verifies through models that in the case of symmetric information accepting entities can offer incentive to the dispatched employees and employers so as to achieve Pareto optimality by means of linear contracts. In case of asymmetric information, the expected revenue of accepting entities will be influenced by the abilities, degree of risk aversion and effort costs of the dispatched employees and employers, which is of great reference value to the enterprise practice.

Keywords: Labour Dispatch, Symmetric Information, Asymmetric Information, Risk Aversion, Effort Costs

Qiang Li Establishment of comprehensive index model based on raster data and its application in the study of regional differentiation
*Computer Modelling & New Technologies 2014 **18**(11) 991-997*

The small grid calculation model was used as the theoretical basis; A variety of meteorological data in 1971-2007 of Chinese Loess Plateau, remote sensing image (TM), DEM and other maps in 2000 were used; factor analysis, raster calculator, variance analysis and other methods in ArcGIS were used; the mathematical model of the latitude and longitude related to the main meteorological elements of the loess plateau and comprehensive index model for regionalization were obtained; the north and south boundaries of the Loess Plateau were determined.

Keywords: factor analysis, analysis of variance, raster data, comprehensive index model, boundaries, loess plateau

Yinglan Fang, Bing Han, Binghui Chen Data mining applications research based on ISBN management
*Computer Modelling & New Technologies 2014 **18**(11) 998-1003*

ISBN management systems existed irregularities publishing. How to summarizes the current book publishing rule according to publish information, it can better grasp the overall book market trends based on existing books. This paper introduced data mining ideas to the book publishing field through the study of real-name system to apply business processes. There were two data mining models. One was association rule analysis model based on subject field to book type distribution. It analyzed book type using classic Apriori association rule analysis algorithms to

identify book publishing hot and the overall trend of publishing business. It can effectively regulate press book publishing behaviour and has great significant to the China's publishing industry healthy and orderly development.

Keywords: data mining, association rule analysis, ISBN

Zhihang Tang, Zhonghua Wen Recommendation system based on collaborative filtering in RapidMiner
Computer Modelling & New Technologies ***18****(11) 1004-1008*

Recommender systems facilitate decision-making processes through informed assistance and enhanced user experience. To aid in the decision-making process, recommender systems use the available data on the items themselves. Personalized recommender systems subsequently use this input data, and convert it to an output in the form of ordered lists or scores of items in which a user might be interested. These lists or scores are the final result the user will be presented with, and their goal is to assist the user in the decision-making process. The application of recommender systems outlined was just a small introduction to the possibilities of the extension. Recommender systems became essential in an information- and decision-overloaded world. They changed the way users make decisions, and helped their creators to increase revenue at the same time. Bringing recommender systems to a broader audience is essential in order to popularize them beyond the limits of scientific research and high technology entrepreneurship. The recommender systems will assist you in reaching quality, informed decisions.

Keywords: recommender systems, collaborative-based systems, nearest neighbour

Yu Wang A study of region division based on spatial units fusion of clustering algorithm
Computer Modelling & New Technologies 2014 ***18****(11) 1009-1013*

Region division is the foundation of regional socio-economic development planning. Traditional regional division is only built upon the information of attributes in each spatial unit. However, the spatial relationships and their spatial interaction between units are ignored. In the study, based on the spatial linkage theory, the principle and steps of spatial units fusion of clustering algorithm is proposed. According the spatial range of Chinese railway as a case, determine the spatial units and the linkage model was built based on the proximity and attribute characters, the experiments of regional division for Chinese railway are achieved. The experiment shows the results are highly accordant with the real situations and have proved the feasibility of the algorithm.

Keywords: spatial linkage theory, region division, spatial units fusion of clustering algorithm, Chinese railway

Tao Meng A FAHP-based comprehensive evaluation on rural supermarket service quality: a case study of Jiangsu province
Computer Modelling & New Technologies 2014 ***18****(11) 1014-1019*

With the changes of market supply and demand, as well as increasing business competition, service quality has become a key factor constraining rural supermarket survival and development. This article targets at the rural supermarkets in Jiangsu Province. In the context of constructing an indicator system of rural supermarket service quality evaluation, it acquires data by questionnaires, uses AHP to establish the weights, and applies the multi-level fuzzy AHP to assess rural supermarket service quality. It concludes that there are some spaces for further improvement of rural supermarket service quality and proposes relevant solutions.

Keywords: FAHP, rural supermarkets, service quality

Jun Li Face recognition algorithms for embedded entrance guard system
Computer Modelling & New Technologies 2014 ***18****(11) 1020-1025*

Biometric identification technology is used to identify individuals based on their unique physiologica1 characteristics under sampling and measurement. The hardware platform of our entrance guard system is based on SAMSUNG S3C6410 embedded development board with ARM11 processor, and we adopt ordinary camera as facial image capture device, 7-inch touch-screen LCD display as an input, electronic door locks as executing components by the onboard port control. The software platform is based on Linux operating system to design and optimize face recognition algorithm, which is combined with local binary pattern (LBP) algorithm, principal component analysis (PCA) algorithm, and ridge regression algorithm. Also, the face recognition system is divided into two parts, namely, a face database training part and a face recognition part. Precisely, the face database training part acquires face information through the processes of face detection, feature extraction and stores the face information to serve as the detection basis of the face recognition part; then the face recognition part is connected with the entrance guard system and sends out an instruction for executing related action to the entrance guard system according to the face recognition result. Specially, the embedded face recognition entrance guard system is contactless and easy to collect

the facial database, less in power consumption, low in cost, and has easy installation and stable performance.

Keywords: face recognition, local binary patterns, principal components analysis, embedded system

Xin Ma, Fuxiaoxuan Liang, Wenbin Wang Control particle swarm optimization for unit commitment problem under emissions reduction

Computer Modelling & New Technologies 2014 ***18****(11) 1026-1033*

The control particle swarm optimization (CPSO) algorithm is introduced to solve the unit commitment problem under the background of emissions reduction. Because the standard particle swarm optimization algorithm is easy to fall into local optimal solution. The closed loop control concept and feedback mechanism of classical control theory are posited, each particle is considered as controlled object to meet the changing needs in searching process, while dynamically adjust the inertia weight by proportion-Integra-derivative (PID) controllers according to the adaptation value of each step. These strategies greatly ensure the diversity of particles and improve the global search ability of the algorithm. The simulation results show that CPSO algorithm can reduce the dimension of the problem and ensure the feasibility of the particle in the optimization process, while it also has good convergence characteristics and global search ability.

Keywords: power system, emissions reduction, unit commitment, control particle swarm optimization

Tao Yi, Shanshan Cui, Yi Zhang Research on the cost estimation of transmission line project with case-based reasoning

Computer Modelling & New Technologies 2014 ***18****(11) 1034-1041*

Investment estimation phase of power transmission line project is an important stage of project cost control, the traditional unit investment forecasting methods, however, cannot meet the need of engineering rapid estimation. The research objective of this paper is to forecast the unit investment of the Transmission Line Project applying the case-based reasoning system. Based on the factor analysis tools of multivariate statistical software SPSS, the correlation coefficient matrix of characteristics factors are identified, and weight coefficient of characteristics factors is calculated. Based on the formula of European weighted distance, the similarity is calculated and case similarity is retrieved. Finally, the model is constructed to predict the unit investment of Transmission Line project, and the correctness and usefulness of the model is verified. It is shown that this case-based reasoning system of unit investment can achieve real-time updating of cost index database. The primary contribution of this research is the combination of the cost forecast and Case-based reasoning system of artificial intelligence. It is expected that this work will provide a reference and guidance for transmission line project investment decision in estimating phase.

Keywords: case-based reasoning, transmission line projects, forecasting

Qingbin Han Comparison on turnovers of agricultural products futures based on wavelet packet transform method

Computer Modelling & New Technologies 2014 ***18****(11) 1042-1047*

In order to distinguish the different patterns and evolving trends on turnovers of agricultural products futures between Zhengzhou Commodity Exchange and Dalian Commodity Exchange. A time-frequency analysis approach, i.e. wavelet packet energy spectrum, is investigated in this paper. Firstly, wavelet packet energy spectrum is briefly introduced. Secondly, two different non-stationary signals of turnover of agricultural products futures from 2009 to 2013 coming from Zhengzhou Commodity Exchange and Dalian Commodity Exchange are described in wavelet packet energy spectrum. First, reconstructed coefficients of main analysis wavelets packet of signals of turnover of agricultural products futures of Zhengzhou commodity exchange and Dalian Commodity Exchange are calculated. The obtained frequency band power ratios are used to show the different characteristics and evolving trends on turnovers of agricultural products futures between Zhengzhou Commodity Exchange and Dalian Commodity Exchange. With these results, the two signals are distinctly different from each other. It is proved that the technique of wavelet packet energy spectrum is effective for the purpose of distinction of turnover of agricultural products futures in commodity exchanges.

Keywords: turnover amounts, time-frequency analysis, non-stationary signals, wavelet packet energy spectrum, frequency band power ratios

Yuan Qin Research on Cooper ethical decision model based on Chinese opera

Computer Modelling & New Technologies 2014 ***18****(11) 1048-1051*

The effectiveness of Chinese opera decision directly affects the quality of traditional opera. In the era of rapid

development of internet technology today, the transmission mode of opera has changed. Its spreading range is wider, and its propagation speed is faster. For not considering the ethical factors, traditional opera decision is easy to have decision deviation. With reference to the theory of Cooper ethical decision model, the basic framework of Chinese opera ethical decision is constructed, including its definition, criterion, flow block diagram, etc.

Keywords: Chinese opera decision, ethical decision, Cooper's ethical decision model

Dongling Sun A construction of an autonomous vocabulary self-study system aided by the web pages
Computer Modelling & New Technologies 2014 ***18****(11) 1052-1055*

The thesis conducted a survey to confirm the hypothesis that vocabulary is the stepping stone in English Study to have an innovative application of modern educational technology to enrich the ways, methods and devices of enlarging the students' vocabulary. From the survey, it is easy to see that the hypothesis is true and the most students of the researcher felt vocabulary affected their motivation and interest in English study. So it is necessary to enlarge the vocabulary of the students in order to promote their application ability of foreign language. Vocabulary study is a process of self-study, and based on the theory of English study, the researcher has designed an autonomous vocabulary self-study system aided by the web to give flexible help for the students. The research is at its designing period and needs further assessments with a case study.

Keywords: construction, vocabulary, autonomous, system, web

Bin Ge, Kai Wang, Yue Han A design for simulation model and algorithm of rail transport of molten iron in steel enterprise
Computer Modelling & New Technologies 2014 ***18****(11) 1056-1061*

An emulation method can be chosen according to the characteristics in rail transport of molten iron in steel companies and a simulation model can be established based on the basic conditions for model, distributed tank mode and path selection, etc. By studying the automatic collision avoidance algorithm, a method of shortest path optimization for rail transport scheduling of molten iron was proposed based on ant colony algorithm. Simulation results show the validity and the feasibility of the algorithm. Programs and strategies of implementing visual simulation platform are proposed laying the theoretical foundation of further research and application for rail transport scheduling model of molten iron and intelligent optimization algorithm.

Keywords: dynamic routing, selection model, ant colony optimization algorithm, collision avoidance algorithm, mode of allocated tank, rail transport of molten iron

Hairui Wang, Ya Li The design and implement of alarm processing system for large-scale railway maintenance equipment diesel engine combustion control based on multi-agent
Computer Modelling & New Technologies 2014 ***18****(11) 1062-1067*

The Alarm Processing System reference model for large-scale railway maintenance equipment diesel engine combustion control, which is based on multi-agent, is an important system for high speed and automation. Alarm Processing System are vital aspects in M large-scale railway maintenance equipment diesel engine combustion control, in this sense, alarm processing system should support decision-making tools as known as decision support system(DSS), new maintenance approaches and techniques, the enterprise thinking and flexibility. In this paper a Multi-agents based alarm processing System reference model for large-scale railway maintenance equipment diesel engine combustion control is presented which combines the existing models and multi-agents. This model is based on a generic framework using multi-agent systems for the diesel engine combustion control the diesel engine combustion control on-line monitoring system. In this sense, the alarm processing system is viewed like a feedback control process and the actions are related to the decision-making in the scheduling of the preventive maintenance task and the running of preventive and corrective specific maintenance tasks. The result of an evaluation of the Multi-agents based alarm processing system reference model for the diesel engine combustion control are presented. This new model is compared to some important existing models and applied to a real investigation.

Keywords: diesel engine combustion control, alarm processing systems, multi-agent system, automation

Ming Pu, Hairui Wang Fault diagnosis expert system for large-scale railway maintenance equipment based on BP neural network
Computer Modelling & New Technologies 2014 ***18****(11) 1068-1072*

According to the characteristics of the neural network and expert system, a fault diagnosis method for large-scale railway maintenance equipment based on Back Propagation (BP) neural network and expert system is proposed. Fault diagnosis for large-scale railway maintenance equipment on BP neural network and expert system model are

constructed. A weak of collection and expression of knowledge on traditional expert system is overcome. Availability of the method based on BP neural network system and expert system is verified by experimental results of large-scale railway maintenance equipment fault.

Keywords: BP neural network; neural network-knowledge base; expert system; fault diagnosis

Jiahua Zhou Innovation model for college ideological and political education in network environment
Computer Modelling & New Technologies 2014 ***18****(11) 1073-1077*

Ideological and political education is impacted by the rapid development of network technology, computer technology and high-speed communication technology, in aspects of concept, content, ways and means, and can not meet requirements of current network age development. Therefore, the ideological and political education is placed in network technology to form new pattern. This paper, first, made brief overview of the concept and characteristics of ideological and political education in network environment, then summarize college ideological and political education at home and abroad. Afterwards, it analyzed the challenges and opportunities after combined network technology and college ideological and political education, then proposed relative solutions in the perspectives of concept, content, means and position. At last, it put forward the construction of shared community pattern for ideological and political education in network environment, consisting of the organization structure and operation mechanism of shared community.

Keywords: network platform, college ideological and political education, teaching model research, innovation

Juan Zhang, Guanghui Chen, Boqin Liu Management on the recombination of manufacturing whole process logistics operation based on MES
Computer Modelling & New Technologies 2014 ***18****(11) 1078-1081*

With the constant intensification of market competition, manufacturing enterprises are racking their minds to lower the cost so as to obtain more profit. As the enterprise's third profit source, logistics has attached more and more enterprises' attention. According to the relevant data statistics, in the current production system, trucking expense accounts for 30% to 75% of total expense. However, after optimal management of logistics, it can save 15% to 30% of expense for enterprises. This paper discussed the current situation and research results of the workshop logistics of machine manufacturing, and analyzed the loopholes existed in the whole process of traditional logistics management of manufacturing. It also established a kind of manufacturing whole process logistics lean management operation model based on MES, which provided a kind of feasible management mode for machinery industry to lower the logistics cost of the whole manufacturing process.

Keywords: MES, manufacturing whole process, logistics lean management, operation model

Jingdong Hao Design and construction of college English teaching model based on multi-dimensional integrated technology
Computer Modelling & New Technologies 2014 ***18****(11) 1082-1084*

How to enhance language teaching efficiency, especially college English is always the focus of people. People pay more and more attention to the progress of the CAT with the development of linguistics and improvement of computer and network technology, at the same time, the combination of language teaching and computer science is also a major breakthrough in the history of the development of language teaching. This paper used the C3I-CAT integrated technology into college English curriculum design and research and development of independent study system, and developed more college English courseware and application methods that is matching to the need of teaching practice, thus to improve the learning efficiency of college English course.

Keywords: college English course, information technology, integration, design of model

Ying Bai Comprehensive evaluation of education mode of university English model and swot analysis
Computer Modelling & New Technologies 2014 ***18****(11) 1085-1088*

College English is a compulsory course for non-English majors, which is an important public class. The pros and cons of university English education are directly related to the improvement of comprehensive quality of college students. What the specific requirements of implementing quality education for college English education and how to further strengthen and improve the university English education are hot topics and have been widely concerned. This paper analyzes the traditional mode of university English education quality, and set up a comprehensive evaluation mode of education model, so as to provide some help for the university English reform.

Keywords: university English education, character building education, analysis of advantages and disadvantages, comprehensive evaluation

Yan Zhang Research of golden week's tourism flow based on tourism system model visual
*Computer Modelling & New Technologies 2014 **18**(11) 1089-1093*

Travel system is a set of various travel issues; some negative effect of tourism factors in the overall tourism system will affect the tourists' decision. Since the "golden week" holiday, the growing numbers of residents traveller, form a tourist flow for a period of time, while promote domestic economic growth, accompanied by some negative effects, such as poor quality of "golden week" travel, poor tourist satisfaction, cause certain environmental pollution and so on, all aspects of the social concern is caused. This article make Huangshan scenic spot and the ancient villages in southern Anhui province as an example, based on tourism system vision, on the basis of the tourism golden week to explore, set up and analyze the "golden week" tourism flow system model. Classified tourism flow mobility factors by using the model, analyzing influence factor of golden week tourism, according to affect the positive and negative effects of tourism flows, come up with some suggestions and improvement measures, in order to improve the management and planning of tourism, meet the needs of the national holiday.

Keywords: travel system, golden week, tourist flow, Huangshan, villages in southern Anhui province

Pingping Xiao Establishment and implementation of network congestion control algorithm based on real-time streaming transmission
*Computer Modelling & New Technologies 2014 **18**(11) 1094-1098*

In order to improve the network performance of computer and avoid the occurrence of network congestion better, this paper analyzed the problems faced by the integration of real-time streaming transport and network and their solutions in the perspective of the characteristics of real-time streaming transport. As to the congestion control of streaming media, TFRC algorithm was analyzed emphatically. Based on TFRC, network congestion was improved, monitored and predicted with parameter of real-time cache length; sending rate was corrected based on cache length when the network was saturate, in order to avoid congestion in time and improve fairness; at last, the test comparison proceeded by network simulation platform NS2. The result indicated that, the improved TFRC was fairer; meanwhile, it showed more friendliness to TCP.

Keywords: real-time streaming transport, congestion control, TFRC, cache length

Jindong Chen Analysis evaluation and quantitative research on economic law behaviour cost and market efficiency
*Computer Modelling & New Technologies 2014 **18**(11) 1099-1103*

Fiscal expenditure, as the material support of economic law behaviour, plays a critical function in regulating resource distribution, promoting economic development and guaranteeing social far equity. Analysis evaluation of fiscal expenditure is obtained through connotation of fiscal expenditure and investigation evaluation. This paper thought that, analysis evaluation of economic law behaviour cost referred to make comprehensive evaluation in aspects of economy, efficiency and benefits on fiscal expenditure process and result according to specific principles and using scientific and normative evaluation method and technology, with fiscal expenditure as management objectives. The results it obtained will be mostly reflected on market effect. Through the quantitative research on fiscal expenditure and economic benefits evaluation in our country, this paper made digitalization and objectification analysis on the comparison of economic law behaviour cost and market benefit, studied the relationship between fiscal expenditure scale and market economic, further estimated the optimal fiscal expenditure scale in our county, in order to provide basis and rule for the formulation of economic law behaviour cost based on quantitative analysis.

Keywords: economic law behaviour, fiscal expenditure, market benefits, quantitative analysis

Yanli Yang, Xianyu Wang Incentive system of skill-oriented dispatched employees based double moral hazard
*Computer Modelling & New Technologies 2014 **18**(11) 1104-1108*

As market competition gets increasingly fierce, the elastic way of employment arouses great concern among more and more enterprises, which results in rapid development of labour dispatch. Labour dispatch is different from the traditional way of employment and it involves the three parties of the employer, the accepting entity and the dispatched employee with the striking feature of asymmetric information. From the perspective of the employer, this paper analyses the incentives of the three parties in labour dispatch while taking the employer as the first principal and the accepting entity as the second principal in consideration of the factors that the intermediary agent's training will influence the quality of the employee. It verifies through models the sharing ratio paid by the accepting entity to the employer, the sharing ratio paid by the employer to the dispatched employee and the fixed wage paid by the accepting entity to the dispatched

employee in expectation of offering concrete suggestions to the practice of enterprises.

Keywords: Labour Dispatch, Agent's Effort, Double Programming, Sharing Ratio

Zhihang Tang, Zhonghua Wen Comparative study of personalizing recommender systems based on shopping system

*Computer Modelling & New Technologies **18**(11) 1109-1114*

Making choices is an integral part of everyday life; Recommender systems facilitate decision-making processes through informed assistance and enhanced user experience. To aid in the decision-making process, recommender systems use the available data on the items themselves, Personalized recommender systems subsequently use this input data, and convert it to an output in the form of ordered lists or scores of items in which a user might be interested. These lists or scores are the final result the user will be presented with, and their goal is to assist the user in the decision-making process. Recommender systems facilitate making choices, improve user experience, and increase revenue, therefore should be easily accessible for deployment to interested parties. The implementation of recommender systems in RapidMiner has been additionally simplified through the Recommender Extension.

Keywords: shopping system, personalized recommender systems, RapidMiner, decision-making

Caihong Zhang Analysis on Ideological and Political Teaching Curriculum Design and Information Transfer Based on Complex Adaptive System Theory

*Computer Modelling & New Technologies 2014 **18**(11) 1115-1118*

In the field of technology, based on scholar's research results of instructional system design complexity, this paper discussed the development of instructional system design thought. It also analyzed the basic framework of ideological and political teaching curriculum design based on complex adaptive system.

Keywords: system, complexity, instructional design

Liming Xiao, Rui Jing The Attractiveness and Motivation of China's International Investment Based on Positive Investment Inertia

*Computer Modelling & New Technologies 2014 **18**(11) 1119-1122*

Nowadays, economic globalization is the main trend of world economic development. Each country takes the participation in international investment as a strategy. This paper analyzes China's international investment from the view of investment inertia. Though the attractiveness and motivation of China's investment keeps increasing, the integral development level is low which stays in the stage of zero inertia. The total amount of China's investment is increasing. But the invitation of investment is unreasonable; the short-term investment in foreign country prevails. We seek the strategic asset of developed country eagerly, and prefer to seek the market and resource of developing country.

Keywords: international investment; inertia; attractiveness; motivation

Yue Zhang Operation Path and Strategy of Logistics Enterprise Supply Chain Logistics Management under Horizontal Integration

*Computer Modelling & New Technologies 2014 **18**(11) 1119-1122*

This paper put forward the significance of horizontal integration service provided for society by logistics enterprise, and discussed in detail about how to realize horizontal integration service and what kind of operation model can be adopted by relevant sub-service. Researches on these contents not only have certain effect on correlated theoretical research, but also can provide certain guidance for logistics enterprise. This research aimed to promote the compound integrated service ability and international competiveness through horizontal integration management mode.

Keywords: Logistics, horizontal integration, operation and management mode

Weidong Wang An entropy method-based index system for the competiveness of industrial cluster – a case study on the typical clusters in Zhejiang province in China

*Computer Modelling & New Technologies 2014 **18**(11) 1128-1131*

Basing on the previous research results, this study constructed an index system for the competiveness of the industrial cluster using analytic network process (ANP) method. Moreover, it employed the entropy method in objective assignment method to assign weights to the indexes and conducted empirical analysis by exampling the typical clusters in Zhejiang province. The results showed that the industrial concentration degree, specialization degree (location quotient), and Herfindahl-Hirschman index (HHI) took relatively high proportions in the indexes

concerning the competitiveness of industrial clusters. This study also drawn an important conclusion, namely, high industrial concentration degree was conducive to improve cluster competitiveness and reduce cluster risk, while lower industrial concentration degree facilitated the formation of high overall cluster competitiveness.

Keywords: ANP, entropy method, Industrial cluster competitiveness

NATURE PHENOMENA AND INNOVATIVE ENGINEERING

Wei Huang Wireless vehicle detection node based on tunnelling magneto resistance sensor

Computer Modelling & New Technologies 2014 ***18****(11) 1128-1133*

A wireless node based on tunnelling magneto resistance sensor was designed for large scale vehicle detection in Intelligent Transport System. With regard to the sensor's characteristic of high resistance, rapid response and high linearity, the signal acquisition and regulation circuits were designed to meet the requirement of geomagnetic measure for three dimension axis of the sensor. A vehicle noise pre-detection unit was implemented to wake up the microcontroller from sleep state before vehicle enter the detection area. Low power chips were considered and the all sensor units were power supplied by MCU, cooperated with the improved power-efficient ATA algorithm, the power consumption was minimized. Experimental results showed that the designed node was capable of capturing the magnetic feature of different vehicle types and on line vehicle flow detecting for long time.

Keywords: tunnelling magneto resistance, geomagnetic induce, vehicle detection, ATA, WSN

Wan Ning Mechanical characteristics of multi-span concrete continuous girder oblique bridge based on ANSYS

Computer Modelling & New Technologies 2014 ***18****(11) 1134-1138*

In order to adapt to environment and reduce the damage of construction for environment, the principle that environment adapts to construction is updated for bridge design, therefore lots special-shape bridges including oblique bridge emerge. However, there is no accurate and developed theory for computing the mechanical characteristics of oblique bridge. Generally, equivalent straight bridge is applied for computing. By performing finite element modelling, the internal force and displacement of oblique bridge and equivalent straight bridge with load were compared. The results showed that due to the special structure of oblique bridge, torque is generated under load, accompanying with coupling effect of bend-twist. Therefore, the mechanical characteristics analysis of straight bridge can not accurately explain that of oblique bridge.

Keywords: ANSYS, oblique bridge, computer simulations, finite element modelling

Xiangyue Yuan, Zhongjia Chen Design of new type biomass pellet forming machine with plunger roller and ring-die at room temperature

Computer Modelling & New Technologies 2014 ***18****(11) 1139-1145*

Bio-energy has become the fourth major energy resource after oil, coal, natural gas energy on account of the properties of green, clean, and renewable. And biomass densification is an important technology which makes the biomass resource be low cost and high value. But for current biomass forming equipment, there exist common shortcomings of high energy consumption, low efficiency, serious wear of forming parts, and etc. In this paper, a forming mechanism of a new type biomass pellet machine with plunger roller and ring-die was proposed, which using the mesh extrusion pressure between plungers and forming cavities (or called holes) to make the loose biomass material densification. It could avoid direct contact and intense friction of forming parts and biomass materials out of the holes. So the energy consumption of the materials in ineffective forming process was reduced and the wear of forming parts was effective alleviated. Then the new type biomass pellet forming machine was designed according to the requirements by Solidworks after force analysis and theoretical calculation, and a prototype was built with the whole power 22 kW and production capacity 450 kg/h. Compression experiments were done finally on the prototype. It was shown that the forming quality of solid pellet fuel tended to be stable after 30 minutes' running and the pellet density could reach more than 0.8 g/cm^3. So it represented that the designed pellet machine had a good practical operation and met the design demands.

Keywords: forming technology of biomass densification, pellet forming mechanism with plunger roller and ring-die, pellet fuel

Yong Zhang, Lixia Li Reaction mechanism of carboxymethyl starch-based wood adhesive

Computer Modelling & New Technologies 2014 ***18****(11) 1146-1151*

The technology involved in manufacturing carboxymethyl starch-based adhesive was studied with the aim of enhancing its water-resistance. The optimum formula was based on previous research using a quadratic orthogonal

rotation combination design to synthesise the carboxymethyl starch-based wood adhesive. Compared with the original starch, carboxymethyl starch improves the performance of starch adhesive due to its advantageous properties such as hydrophilicity and stability during freezing and thawing etc. Infrared spectroscopy and scanning electron microscopy were used to analyse the synthesis of the carboxymethyl starch-based adhesive. Many functional groups could be identified by their characteristic Fourier transformation infrared vibration frequency and this was used to explain the structural changes. At the same time, the microstructure of the adhesive was analysed by scanning electron microscopy (SEM), and porous, round or elliptical granules in the carboxymethyl starch were observed. An SEM of the main starch agent revealed a network structure; the granules did not aggregate together to form large groups due to the polyvinyl alcohol that prevented granule agglomeration. However, the SEM of the starch adhesive showed a smoother and more compact network structure, due to the large number of strong chemical bonds that were formed by the association of isocyanate with the main starch agent.

Keywords: carboxymethyl starch, wood adhesive, optimisation, mechanism

Cai Lipeng Experimental study on heat-preservation wall materials made from waste foam
Computer Modelling & New Technologies 2014 ***18****(11) 1152-1155*

In this paper, a new energy-saving-and-environmental-protection heat-preservation wall material is studied. This material is mainly made by waste foam, cement and sand in the structure of "sandwich" - cement foamed plastic is in the middle of cement mortar. Experimental study reveals that the heat-preservation block of this material has lighter self-weight and higher compressive strength and better performance in heat preservation, sound insulation and frost resistance, that is to say, its overall property is obviously superior to that of others such as aerated concrete block. This block consumes a large amount of waste foam that is difficult to dispose, which conducive to environmental protection and energy is saving. It can bring favourable social, economic and environmental benefit. Besides, less energy is consumed in the production of this material and it is easy to realize volume production, so this material is a new wall material of energy saving and environment protection.

Keywords: waste foam, heat preservation, wall material, new technologies

Cai Lipeng Experimental study on heat-preservation wall materials made from waste foam
Computer Modelling & New Technologies 2014 ***18****(11) 1152-1155*

In this paper, a new energy-saving-and-environmental-protection heat-preservation wall material is studied. This material is mainly made by waste foam, cement and sand in the structure of "sandwich" - cement foamed plastic is in the middle of cement mortar. Experimental study reveals that the heat-preservation block of this material has lighter self-weight and higher compressive strength and better performance in heat preservation, sound insulation and frost resistance, that is to say, its overall property is obviously superior to that of others such as aerated concrete block. This block consumes a large amount of waste foam that is difficult to dispose, which conducive to environmental protection and energy is saving. It can bring favourable social, economic and environmental benefit. Besides, less energy is consumed in the production of this material and it is easy to realize volume production, so this material is a new wall material of energy saving and environment protection.

Keywords: waste foam, heat preservation, wall material, new technologies

Wang H F, Fan Y H, Yang L The analysis of vibrations induced by variation section vortex in tension leg platform for a floating wind turbine
Computer Modelling & New Technologies 2014 ***18****(11) 1156-1162*

During the previous decade, several offshore wind-farms were constructed for offshore wind energy generation showed promise as a source of green energy. However, there are several challenges to be met in the design and construction of the foundations for offshore wind turbines. The fatigue load plays an important and crucial role in the design of the supporting structures. In this paper, the vortex-induced vibrations of the tension leg platform were studied. Two types of structures namely; cylindrical Tension Leg Platform (TLP) and variable cross-section TLP were designed and studies on them were conducted to compare the advantages and drawbacks resulting from vortex-induced vibrations. Both uniform and shear flow were considered to simulate water flow through the structures. The variable cross-section TLP, which possesses outstanding mechanical properties, gave lower vortex shedding frequency compared to the cylindrical TLP for the same velocity. This is the objective desired in vortex-induced vibration.

Keywords: vortex-induced, tension leg platform, variation section

Yang Zhongguo Study on the device of multipath household heat metering
Computer Modelling & New Technologies 2014 ***18****(11) 1163-1166*

One of the most important contents of reforming heating system is that more attention will be paid to residential energy conservation and improvement of heating facilities in order to charge heating fee by heat consumption. But the out-dated technology prohibits the widespread use of multipath household heat metering. Since the heating system in China is still confronted by the substantial energy waste, the jumble of pipelines and difficult management of the system etc, a company in Liao Ning has developed a new kind of water supply system on multipath household heat metering. The new system blazes new trails on household heat metering, solving the problem of traditional pipelines. The paper uses the computational method of hydrodynamics and provides the experimental platform of pipe resistance measurement and its theoretical parameter for the system. The study shows that resistance coefficient of the experimental pipeline corresponds with the theory, and multipath household heat metering has favourable prospects.

Keywords: converter multipath, household heat metering, pipe resistance measurement

Lv Jingxia, Gao Lin, Wen Jian Research on the relationship between moisture content and the dielectric constant of the tree trunk by the radar wave

Computer Modelling & New Technologies 2014 ***18****(11) 1167-1171*

Radar wave technology offers a noninvasive, quick determination technique and has potential for the applications of the non-destructive detection (NDT) for the trees trunk and wood based materials. The precision of NDT determination by radar wave is influenced the wood dielectric constant which is closely related to the moisture content of the trees trunk. For our investigations we used TRU trees radar wave to detect the trees trunk. Four typical trees trunks were selected: polar, willow, pine, eucalyptus. Different trees trunks moisture content and dielectric constant were tested under the radar wave frequency respectively. Models of the relationship between moisture content and dielectric constant of the trees trunks were established for improving the accurate of radar wave NDT.

Keywords: radar wave, dielectric constant, moisture content, tree trunk, non-destructive detection

Wang Zijian, Wu Liming Physical parameter identification and wave force inversion research of bridge pier structure

Computer Modelling & New Technologies 2014 ***18****(11) 1172-1180*

According to total compensation composite inversion algorithm, physical parameter identification and inversion of wave time interval are done under the condition of input information of isolated pier in wave force part being unknown. Based on the characteristics of large-diameter piers in shallow water which are affected by wave force and combining with “statistical average method” in probability theory, structural physical parameter identification and wave inversion calculation are done under the condition of unknown wave force. Analysis of numerical values shows that results of structural physical parameter identification cater for accuracy requirements which is feasible for inversion method of parameter identification of large-diameter piers in shallow water with unknown input information thus providing a new method for structural physical parameter identification research of river-spanning and sea-crossing piers affected by wave force and providing references for engineering application.

Keywords: bridge pier structure, parameter identification, unknown input, load inversion, time domain correlation

Kang Yiting, Feng Yali, Lin Jinggao, Zhang Wenming Load spectrum for creep-fatigue life prediction of viewport used in human occupied vehicle

Computer Modelling & New Technologies 2014 ***18****(11) 1181-1186*

Acrylic plastic, polymethylmethacrylate (PMMA), are widely employed as the material for the viewports of human occupied vehicle (HOV) which usually dives into deep sea. The service life of viewport is critical to the reliability and safety of HOV. In order to predict life of viewport in design stage, mathematical statistics method is applied to establish the load spectrum for viewport. It is found that ALVIN (America) HOV’s dive-depth data is in a skewed distribution, and a piecewise function combined Gumbel and Weibull distributions is proposed for data fitting. HOV undertakes long-term and cyclic load in service, which will cause damage on viewport, so a creep-fatigue load spectrum is established and applied for JIAOLONG (China) HOV’s viewport, which integrates both the dive-frequency of each depth range and the duration that maximum stress acts. The proposed method for determination of creep-fatigue load spectrum could thus be considered to be employed for failure analysis and life prediction of modern HOV’s viewport.

Keywords: viewport, human occupied vehicle, distribution function, creep and fatigue failure, load spectrum

Cheng Xinmin, Li Jing, Jiang Yunliang, Gao Ge Application of GMM-UBM with an embedded AANN in the acoustic emission signal recognition

*Computer Modelling & New Technologies 2014 **18**(11) 1187-1191*

In this paper, we propose to recognize the Acoustic Emission (AE) signal, by using a Gaussian Mixture Model-Universal Background Model (GMM-UBW) with an embedded AANN. The AANN based GMM-UBW combines the learning ability of neural network and strong distribution capabilities of GMM. And, it trains the model parameters alternatively in order to approach the maximum likelihood. For illustrating the effectiveness of the proposed recognition method for the AE signal, an experiment is conducted. In the experiment, three cases of AE signal are considered, namely with no rub impact, slight rub impact and serious rub impact. The experimental results reveal that the AANN based GMM-UBW outperforms the GMM, with respect to the recognition rate, for any case of AE signal. For the case of slight rub impact, the GMM-UBW and that with embedded AANN both have the worst recognition performance, among the three cases. And, the proposed method has the biggest improvement for this case.

Keywords: acoustic emission; Auto-Associative Neural Network; Gaussian Mixture-Universal Background Model

Wang Xiaorong, Liu Chenpeng, Liu Bo, Tang Wenxian, Zhang Jian Research on method to promote performance of multiple injections diesel engine

*Computer Modelling & New Technologies 2014 **18**(11) 1192-1199*

Based on certain type of two-stroke, low-speed diesel engine for marine propulsion, we, in our research project, developed a 3-D model of in-cylinder combustion processes in a diesel engine. Thus, we were able to compare simulated in-cylinder pressure against the experimental pressure under single injection with a maximum error tolerance of 0.36%, the results of which imply that the two are basically consistent. Subsequently, based on the same injection timing, a comparison was made between the performance of a diesel engine using single injection and one using dual injection. The results indicate that the additional injection can lower NOx by58.9% and soot by 30.1%. Moreover, the piston's workload is reduced by 10.1%. Through calibrating injection quality centre, adjusting injection timing of secondary injection reasonably, we gained the method improving performance of combustion and secondary injection diesel engine. By verifying, the method can promote working ability of secondary injection diesel engine effectively. After adjusting injection timing, working ability of secondary injection is as well as it of single injection, but emissions reduce sharply, thereinto, NOx reduced 12.7%, soot reduced 41.1%.

Keywords: Marine low-speed two-stroke diesel engine; injection timing; double injection; diesel performance

Wang Jun, Wang Guoqing, Li Ming, Yu Wenhui, Tian Hao An improved hand vein image acquisition method based on the proposed image quality evaluation system

*Computer Modelling & New Technologies 2014 **18**(11) 1200-1204*

In the part of hand vein image acquisition of vein recognition, we basically cannot get the images that are suitable for later process due to both the subjective factors and the objective factors. After a deep analysis on the main reasons including the different thickness of hands and mass of things on the hand, we propose the design of hand vein acquisition device based on the adaptive modulation control. The key part of the design is the proposed double-processing image quality assessment system: the first part of which involves the process of quality assessment based on the effective information of multidimensional histogram and the normalized information entropy; the second part involves adding the grey image variance and quantities of cross points with the proper weights, then two process will be made according to the final score: if the score is lower than the set threshold then image will be discarded, and on the other hand the brightness degree of the LED groups of the device will be adaptively adjusted according to the D-value of the score and the threshold so that we can get the high-quality hand vein image for later feature extraction and matching and get a higher recognition rate.

Keywords: hand vein recognition; adaptive control; two-set quality evaluation system; D-value

Yang Jingfeng, Li Yong, Zhang Nanfeng, He Jiarong, Xue Yueju Pure electric buses status information compression and transmission methods basis on optimized Huffman coding algorithms

*Computer Modelling & New Technologies 2014 **18**(11) 1205-1212*

To strengthen safety supervision, satellite positioning information is needed to be uploaded as well as battery status information of pure electric buses. For the unbalanced coverage of mobile communication base station, data loss and incomplete data are prone to emergence when a large amounts of real-time data transmission through 3G/GPRS network in the poor communication environment of subregion, which is difficult to achieve large-scale farmland information collection and transmission simultaneously. Therefore, optimized Huffman coding compression algorithm for text-oriented information is promoted in this paper. The whole process test results of data exchange and decompression show that, the data compression algorithms can realize data compression effectively. The data sampling period is relatively smaller and its adjacent data is closer, the higher compression rate can be obtained by

corresponding compression algorithms, and it can the basically ensure the decompression data without distortion. This will improve the efficiency of information transmission of transportation, as well as to ensure the integrity of information, which is of great significance to realization of transportation energy conservation.

Keywords: vehicle information, data compression, pure electric buses, battery status information

Du Huiqiu, Wang Jingbo, Lin Zhong-hai, Wang Pingjian, Yan Long Allowed capacity calculation of wind farm based on probabilistic constraint

Computer Modelling & New Technologies 2014 ***18****(11) 1213-1218*

This paper presents a method for penetration capacity calculation of wind farm based on probabilistic constraint. According to the stochastic programming theory, under a confidence interval, a mathematical model that considers the randomness of wind speed is established based on the static safety and power quality constrains. The solving method is proposed according to particle swarm optimization algorithm and stochastic simulation technique too. According to a real power system in Yunnan, the penetration capacity of wind power is calculated under different confidence interval, and the feasibility and effectiveness of this method are verified too. The related factors of affecting the penetration capacity of wind power are analysed, which provide the reference for the planning and design of wind farm.

Keywords: penetration capacity of wind power, chance constraint, stochastic simulation, confidence interval, particle swarm optimization

Zhu Lida, Chen Hongyu, Tian Zhongzi, Wang Hai Parametric design and simulation analysis of turbine blade

Computer Modelling & New Technologies 2014 ***18****(11) 1219-1224*

To improve design quality and efficiency of turbine blade, the geometric model of blade is parameterized by using the quintic polynomial method in this paper. The geometric shape and performance of blade mainly depend on blade profile and cascade with parametric model. First of all, the blade profile lines are designed and generated in different parameters based on the theory of quintic polynomial. The effect of relative parameters on the blade profile is analysed conveniently and rapidly. Then, the geometric shape of blade is generated in the arranged blade profiles according to blade cascade position and consistency. The 3D model of blade is created by the parametric blade profile and cascade. It is shown that parametric model of blade will provide the reference and foundation for application research.

Keywords: turbine blade, quintic polynomial, parametric design, 3D model

Xia Shulan, Wang Jilin A speech emotion enhancement method for hearing aid

Computer Modelling & New Technologies 2014 ***18****(11) 1225-1228*

In this paper, emotional perception of the hearing-impaired patients for hearing aid is investigated, and a speech enhancement algorithm is proposed, which is text-independent and requires less and non-parallel training data. In addition, the conversion of prosodic and spectral parameters is also studied. The Eigenvoice Gaussian mixture model (EV-GMM) is used to transform the F0s and spectral parameters, which is built using multiple pre-stored sources emotional and target neutral speech sentences. In the training and testing stages, the duration modification is utilized to improve the performance of EV-GMM training and converted output quality and an adaptive median filter is proposed to smooth the trajectory of the converted speech. Perceptual and objective experiments are presented, simulation results corroborate the effectiveness of the proposed algorithms.

Keywords: hearing aid, emotional speech enhancement, EV-GMM, duration modification

Sun Yan, Lang Maoxiang, Wang Danzhu Optimization and integration method for railway freight stations based on a hybrid neural network model

Computer Modelling & New Technologies 2014 ***18****(11) 1229-1237*

Given to the current problems existing in the operation of railway freight stations and the entire railway freight transport network, in order to integrate the railway freight stations and optimize the traditional railway freight transport mode, we first propose a strategy on the optimization and integration for railway freight stations, then design a hybrid neural network model to recognize the operating performance of each railway freight station by classifying them into four ranks based on the proposed strategy. The characteristic of the proposed model is its combination of the respective advantages of unsupervised learning algorithm based neural network and supervised learning algorithm based neural network. Finally, an empirical study from Hohhot Railway Administration is given to verify the feasibility of the proposed model. The simulation results of the empirical study indicate that (1) the accurate recognition of training samples has significant influence on the classification result; (2) the proposed model can recognize the operating performance of the railway freight stations under relatively high accuracy.

Keywords: Railway Freight Station, Optimization and Integration, Hybrid Neural Network Model, SOM Neural Network, Probabilistic Neural Network

Yu Song, Hua Liu *PH* sensor design based on potentiometric analysis method
Computer Modelling & New Technologies 2014 ***18****(11) 1242-1246*

In order to accurately, continuously detect solution *pH* on line, regarding C8051F020 as the control centre, the output voltage of glass electrode is amplified through high input impedance circuit utilizing potentiometric method to obtain related voltage signal of *pH* value. Standard solution and temperature sensor ADT7301 are utilized for system calibration and moving average filter of the detected *pH* values, thereby achieving real-time online detection of *pH* value. Detection results are accurate and reliable; low power consumption and miniaturation also facilitate rapid collection and analysis of data under different environments.

Keywords: pH detection, C8051F020, glass electrode

Chen Weihua Digital controller for electric vehicle synchronous motor rotor
Computer Modelling & New Technologies 2014 ***18****(11) 1247-1251*

At present, the number of electric vehicle is increasing, and, in case of big vehicle speed change, the vehicle motor cannot be used efficiently. In order to resolve this problem, it is necessary to improve the topological structure and control strategy, and design a new converter. In this paper, we apply a two-channel synchronous Buck converter. The main circuit is to achieve the maximum power output by the stage regulation, while the improved converter topology realizes the overall system function. The experimental results show that the new converter has superior performance in big vehicle speed change, and has the high quality energy output. This research has an important practical significance to improve the utilization of renewable energy.

Keywords: power system, electric vehicle, digital controller, inverter, converter

Peng Zhangming, Cai Huihua, Sun Qingchen, Chen Guojin, Su Shaohui Effect factor research on monitoring wear of piston ring based on magnetoresistive sensor
Computer Modelling & New Technologies 2014 ***18****(11) 1252-1255*

Piston ring is one of the important parts in diesel engine, and its excessive wear reduces the sealing performance of the combustion chamber, the method on monitoring wear of piston rings with the magneto-resistive sensor is researched from the simulation viewpoint, the research shows that the amplitudes reduce with the increase of the piston ring wear. The theoretical basis for monitoring piston ring wear with a MR sensor is provided.

Keywords: piston ring wear, magneto-resistive sensor, diesel engine

Liao Xuechao Steel sheet location tracking and automatic sorting self-adaption control model
Computer Modelling & New Technologies 2014 ***18****(11) 1256-1260*

The sorting control system is the critical procedure in the final product line of cold rolling mill. Taking the tinning steel sheet sorting control system of shearing production line of WISCO's cold rolling plant as an example, the process of steel sorting control system is introduced. The running principle and control concept of steel sheet position tracking and sorting are particularly presented. The practical application proves that the system can position the tinning steel sheet which is moving in high speed, which is of high controllability and reliability.

Keywords: sorting control, position tracking, magnet system, self-adapt regulating, dropping posture control

Dong Lei, Li Weimin, Zhao Weiguo, Chen Yunfei Probabilistic neural Network with statistical feature for fault diagnosis of permanent magnet motor
Computer Modelling & New Technologies 2014 ***18****(11) 1261-1265*

Permanent magnet motors are very important components in commercially available equipments and industrial applications due to high reliability and robust performance, and it is important to take an appropriate and effective approach to diagnose fault for them. The implementation of probabilistic neural network (PNN) with the statistical features for permanent magnet motor is developed in this paper, and the statistical features are determined according to the stator current characteristics of motor to effectively reduce dimensionality of sample space. The experimental results demonstrate that, compared with RBF network, the proposed method is more effective in identifying various types of faults.

Keywords: Permanent Magnet Motor, Probabilistic Neural Network, Statistical Feature

Tan Xiaoling, Tan Zefu, Qu Juan, Xi Guangwen SVM -Based evaluation model for college laboratory learning

*Computer Modelling & New Technologies 2014 **18**(11) 1266-1270*

Evaluation for laboratory learning is based on different factors, while each factor is varied by individuals. Hence it is difficult to express the quantitative nonlinear functional relationship among the evaluation indexes. With limited sample, Support Vector Machine (SVM) could be generalized by compromising between model's complexity and learning ability. That is its advantage on the evaluation of small sample, nonlinear and multi-indexes. It is a good try to apply Support Vector Machine (SVM) to laboratory learning evaluation. With Support Vector Machine (SVM), the relationship between the learning quality and evaluation indexes could be revealed. Experiments show that Support Vector Machine (SVM) model is with high prediction accuracy, faster speed and simple algorithm. It is suitable and more reasonable for laboratory learning evaluation.

Keywords: Support Vector Machine (SVM), evaluation model, laboratory learning quality

Zhang Yalong, Ogura Hisakazu, Ma Xuan Usefulness of lethal chromosomes in genetic algorithms solving the constrained optimization problems

*Computer Modelling & New Technologies 2014 **18**(11) 1271-1274*

The infeasible solutions are often generated in population as evolutionary computation solving the combinatorial optimization problems. The number of infeasible solutions impacts the performance of the evolutionary computation searching the optimal solution, in the worst case the algorithm ceases to run. In genetic algorithms, encoding of infeasible solutions is referred to as lethal chromosomes. In this study, we discover a property of lethal chromosomes that: although lethal chromosomes carry out the infeasible solutions in genetic algorithms, their statistical property implies an underlying similarity with the exact solution of the optimization problems. Hereby we propose an operation using statistical property of lethal chromosomes to handle with the lethal chromosomes themselves. Simulation experiments on a large number of test cases demonstrated that it can improves obviously the performance of genetic algorithms to use the statistical property of lethal chromosomes.

Keywords: Genetic Algorithm, Infeasible Solution, Lethal Chromosome, Evolutionary Computation

Zhang Min, Fan Qiang, Zhang Fucang, Li Xia, Xue Xuzhang, Wang Guodong Crop canopy temperature model of ditch-cultivated based on artificial neural network

*Computer Modelling & New Technologies 2014 **18**(11) 1275-1280*

Aiming at the mechanism model are influenced by multiple random factors, this paper establishes canopy temperature models based on BP network and RBF network respectively. The models take the temperature, humidity, illumination, soil temperature and ditch depth in the closed greenhouse as input neurons and takes canopy temperature as the output neuron. The results show that both models can well predict ditch-cultivated crop canopy temperature. The mean error between the simulation value and measured value of BP network model is 0.8408℃, and root-mean-square error of 0.5789℃. Actual output and expected output of RBF network model differ little, mean error of 0.2236℃ and root-mean-square error of 0.3496℃. In contrast, RBF network model can more accurately predict crop canopy temperature of ditch-cultivated than BP network model.

Keywords: Ditch-cultivated, Canopy Temperature Model, Artificial Neural Network

Wang Liying, You Lisha Analysis on vibration of small-scale hydroelectric generating unit

*Computer Modelling & New Technologies 2014 **18**(11) 1281-1285*

An analytical model for the vibrations of operating conditions of small-scale hydroelectric generating unit is developed based on frequency domain and time domain. Firstly the vibration of unit 2 in Xida hydropower station is tested by using intelligent data logger, where the DASP10 software is used to collect the data and analyse them; and then the data are analysed by the time domain analysis, shaft centreline orbit analysis and auto-spectrum analysis respectively. Finally some instructive conclusions on the exceedance of shaft degrees and the overweight phenomena are drawn, which may assist an overhaul to raise the operating efficiency and the power output.

Keywords: Hydroelectric Generating Unit, Vibration, Time Domain Analysis, Shaft Centreline Orbit, Auto-spectrum Analysis

Zhu Shisong, Zhang Haiyan, Zhang Cuiyun, Li Changqing The application of fusion structure in the coal mine safety state evaluation

*Computer Modelling & New Technologies 2014 **18**(11) 1286-1290*

Coal mine safety is a very important link in the current coal mine production. Currently, there are a lot of kinds of

methods for the evaluation of coal mine safety environment in order to solve the evaluation of the coal mine safety status. However, the complexity and diversity of coal mine environment cause that there is no method can evaluate coal mine safety status. So a new method is proposed, which called two-stage fusion structure according to China's natural environment of the underground coal mine, is used to evaluate the condition of coal mine safety. Firstly, it gets various parameters for affecting the safety of underground using the first level evaluation method; secondly, the second level fusion method is used to realize the coal mine security situation assessment. The experimental results show that this method has better performance more than other methods in evaluating coal mine safety.

Keywords: the first level fusion algorithm, fusion decision, mine safety, better performance

Liu Shuchong, Chen Xun Seismic signals wavelet packet de-noising method based on improved threshold function and adaptive threshold

Computer Modelling & New Technologies 2014 ***18****(11) 1291-1296*

Wavelet analysis is one of the effective method to improve the signal to noise ratio and resolution of seismic data, a wavelet packet seismic signal denoising method based on a new threshold function and adaptive threshold was put forward according to the distortion problem of traditional threshold function denoising method, which make up the defects of traditional thresholding method. Wavelet packet decomposition techniques was used for seismic wave signal denoising processing, and the synthetic seismic signals and actual seismic data was done wavelet packet decomposition processing through MATLAB, better removing high frequency random noise to retain the useful signals. Experimental results showed that the method can effectively remove noise and improve seismic resolution, with better denoising effect.

Keywords: seismic signal, adaptive threshold, denoising, resolution

Xia Jiening, Chen Zhigao Study on automatic identification method of X-ray fluorescence spectrum

Computer Modelling & New Technologies 2014 ***18****(11) 1297-1305*

The way to improve the accuracy and reliability of automatic unscrambling and identification technology on X-ray fluorescence spectrometer spectrum is studied in this essay. Accordingly, two different automatic identification methods based on Fast Fourier Transform and Wavelet Transform are presented. By the tool LabVIEW, such two methods are applied to the qualitative analysis on X-ray fluorescence spectrums, and the features of such two methods are compared. Based on the experiments and analysis on amount of samples, it can be concluded that the automatic identification method based on the Wavelet transform theory is better than the other method for the former has a better local resolution. Therefore, the characteristic values of the singular points are more clearly recognized by the method based on the Wavelet transform. Through the study in this essay, theories on automatic identification are enriched, which set a foundation for further studied in future.

Keywords: X-ray fluorescence spectrometer, spectrum identification, LabVIEW, fast Fourier transform, wavelet transform

Cui Fan, Wu Zhiyuan, Chen Jia, Wang Lei Fine detection of mine collapse column by anti-explosive ground penetrating radar

Computer Modelling & New Technologies 2014 ***18****(11) 1306-1311*

This work discussed the technique for fine detection of mine collapse column by anti-explosive ground penetrating radar (GPR) and its feasibility. In stage of theoretical research, numerical forward modelling was used to study the response characteristics of collapse column on radar profile. In actual exploration, the spatial distribution pattern of collapse column was accurately stored by installing complex environment observing system in the pit. In processing and explanation, the disturbance types of mine were comprehensively analysed, and various measures were taken in the suppression of interference signals to obtain better exploration effects. The study showed that fine detection of collapse column can be realized by anti-explosive ground penetrating radar under the complex conditions in the pit.

Keywords: Anti-explosive ground penetrating radar (GPR), collapse column, fine detection

Chen Disi, Li Gongfa, Liu Honghai, Jiang Guozhang, Liu Jia, Liu Ze, Ding Weiliang, Miao Wei, Li Zh Analysis of thermal-mechanical coupling and structural optimization of continuous casting roller bearing e

Computer Modelling & New Technologies 2014 ***18****(11) 1312-1315*

The continuous casting roller bearing was one of the important parts of continuous caster [1], its working temperature was relativity high, the working load was also very large and the working condition was complex. Since the external heat of casting roller bearings was much higher than internal, the continuous casting roller bearing was different from the general high-speed rolling bearing with heavy load. If the external heat that the bearing suffered could not dissipate in time, the extreme high working temperature might accelerate the failure of the bearing and severely reduce the

productivity of the caster. To optimize the thermal structure of the continuous casting roller bearing, the thermal coupling analysis of the bearing should be conducted. Firstly, the stress field of the working continuous casting roller bearing should be analysed by ANSYS and the three-dimensional geometric CAD model and the CAE model of it should be established. Then find out the location where the bearing suffered the largest force by stress analysis, in the case of bearing block with cooling water, load the working temperatures to the bearing, after analysing, the result showed that the external temperature was the main contributor of internal stress, and the rolling element was the part inside of the bearing which suffered the largest stress. To optimize the structure of the continuous casting roller bearing, the rolling element was turned into axial hollow structure, which could reduce the extreme stress of the rolling element bearing. By analysing all the maximum thermal stress of rolling element in different feasible hollowness, and finally the most suitable hollowness could be found out. The analysis results above showed that the optimization of the rolling element structure of the bearing could effectively reduce the internal stress and improve reliability of bearings in high working temperature. The conclusion of this study was significant in bearing optimizing or designing.

Keywords: thermal stress, coupling, structural optimization, continuous casting roller

Cheng Fuwei, Li Gongfa, Liu Honghai, Jiang Guozhang, Liu Ze, Chen Disi,Miao Wei, Li Zhe, Ding Weiliang Temperature data acquisition and remote monitoring of ladle based on LabVIEW

Computer Modelling & New Technologies 2014 ***18****(11) 1320-1325*

Ladle is an important container parts, plays a vital role to storage and operation of steel smelting. The length of the ladle service life directly affects the efficiency of steel production and the cost of production. The service life of the ladle is mainly affected by thermal mechanical stress of ladle lining, thermal mechanical stress is mainly caused by the severe changes of temperature, in order to fully understand the working state of ladle to ensure the safety of the ladle, the ladle temperature monitoring is particularly important. Ladle design of remote monitoring system uses the structured and modular design thought, on the basis of using the LabVIEW virtual instrument platform; realize the collection of ladle temperature data and remote monitoring. Establish ladle remote client monitoring system by TCP communication protocol, realize the ladle monitoring networked. This system uses the real-time waveform display and friendly human-computer interaction interface, greatly enhanced the real-time and visibility of the ladle remote detection system.

Keywords: the ladle, LabVIEW, remote monitoring list

Zhao Liang, Yu Tao A new decoupled sliding mode control approach for the linear motion of a spherical rolling robot

Computer Modelling & New Technologies 2014 ***18****(11) 1326-1329*

This paper investigates the dynamics and control aspects of the linear motion of a pendulum-driven spherical rolling robot. The dynamic model is deduced for the linear motion of a spherical robot by using the Euler-Lagrange formulation. By appropriate definitions the equations of motion for the robotic system are transformed into the state space form. A novel decoupled sliding mode control approach is proposed to achieve set-point regulation of the linear motion. This approach consists of the construction of a cascade sliding mode controller and the design of a nonlinear reaching law by using a switching component that dynamically adapts to the variations of the controlled system. The asymptotic stability of the robotic system is verified through Lyapunov analysis, and the validity of the proposed approach is illustrated through numerical simulations.

Keywords: spherical robot, linear motion, dynamic model, decoupled sliding mode control, exponential reaching law

Bian Xiangjuan, Gong Youping, Zhen Liyun MEMS model order reduction method based on SPRIM

Computer Modelling & New Technologies 2014 ***18****(11) 1330-1334*

According to large scale MEMS united constraints equations, the paper investigates model order reduction (MOR) techniques based on Structure Preserving Reduced-order Interconnect Macro modelling (SPRIM) method, which can be used to generate computationally efficient solutions for multiphasic MEMS simulation. To united constrained model, the high dimensionality of the original system state space is mapped to a suitable low-dimensional subspace, obtained a low-dimensional state sub-space model. A improved algorithms(SPRIM) from Arnoldi algorithms are implemented to extract low dimensional Krylov subspaces from Unified state subspace models for model order reduction, reduced order electro thermal–mechanical models are generated for a MEMS micro beam using the developed programs. Developed programs automatically generate compact structure preserving models and can be used to significantly improve the computational efficiency without much loss in accuracy and model stability for coupled-field MEMS simulation.

Keywords: MEMS (Micro-Electrical Mechanical System), model order reduction, sub-space model, model reflection

Chen Jianjun Wireless monitoring system for temperature and humidity based on ZigBee
*Computer Modelling & New Technologies 2014 **18**(11) 1335-1339*

Traditional methods of environmental monitoring have the shortages including difficult network layout and low intelligent of node. A monitoring system for temperature and humidity was designed based on ZigBee (wireless network), with SHT11 (temperature and humidity integrated sensor) and CC2430 (wireless radio frequency module integrated with MCU). Data of temperature and humidity, collected at some acquisition terminals in SHT11 and CC2430, were displayed on monitoring host via centre node. By using solution of single chip ZigBee, the system was optimized with high precision, low power dissipation and simple equipment.

Keywords: CC2430 wireless network; monitoring temperature and humidity

Zou Jiazhu, Yuan Fengwei, Deng Qian Analyses on flow and heat transfer performance and of heat exchanger with continuous helical baffles
*Computer Modelling & New Technologies 2014 **18**(11) 1340-1345*

A numerical simulation for heat exchanger with continuous helical baffles was carried out. The study focuses on the effects of helix angle on flow and heat transfer characteristics, and heat exchanger performance is evaluated by entropy generation number based on the analysis of the second law of thermodynamics. The results show that both the shell-side heat transfer coefficient and pressure drop decrease with the increase of the helix angle at certain mass flow rate. The latter decreases more quickly than the former. The tangential velocity distribution on shell-side cross section is more uniform with continuous helical baffles than with segmental baffles. The axial velocity at certain radial position decreases as the helix angle increases in the inner region near the central dummy tube, whereas it increases as the helix angle increases in the outer region near the shell. The heat exchange quantity distribution in tubes at different radial positions is more uniform at larger helix angel.

Keywords: continuous helical baffles, heat exchanger, helix angle, entropy generation, numerical simulation

Zhang Jifei Exploratory development of ancient village based on the low carbon travel development model
*Computer Modelling & New Technologies 2014 **18**(11) 1346-1349*

This paper summarized basic theory of low carbon travel and the functional mechanism of the low carbon economy to improve the village tourism development by analysing its conception, and development status and existing problem of the ancient village tourism in the low-carbon tourism development mode of our country nowadays is put forwarded. Combining with actual situation, this paper aims to propose the emission reduction countermeasures in allusion to these problems above, in order to provide certain reference to the development of ancient villages low carbon tourism.

Keywords: Low carbon concept; Low carbon travel; Ancient village travel; Emission reduction countermeasure

Nan Bo, Wu Yue Numerical simulation on winding CFRP pipe axial compression stability
*Computer Modelling & New Technologies 2014 **18**(11) 1350-1355*

Because of the requirements of strict weight, airship structures are usually built by high strength and light-weight CFRP composite truss, CFRP composite tube is the basic unit of it. Winding CFRP tube ultimate bearing capacity has been further improved by combinations of various fiber directions, while stable bearing capacity is an important indicator that could influence the bearing characteristics. It concludes the destruction and deformation characteristics, as well as the relationship between stability factor Φ and slenderness λ, through analysing different CFRP tubes of slenderness ratio with The arc-length method, and drawing the Load-displacement curve, which provided a theoretical basis to better the truss design, and discussed the laws of how component defects affect its stability capacity via a large number of parameters.

Keywords: carbon fiber-reinforced plastic (CFRP), the arc-length method, stability coefficient, slenderness ratio

Luo Xin, Liu Xuezheng, Tao Ran, Shi Youqun Content-based retrieval of music using mel frequency cepstral coefficient (MFCC)
*Computer Modelling & New Technologies 2014 **18**(11) 1356-1361*

In the last few years, with the growing of multimedia in Internet, MP3 music become one of the most popular types. Some of the MP3 music collections available are approaching the scale of million tracks and this has posed a major challenge for searching, retrieving, and organizing music content. In this paper, we proposed a method to retrieve the MP3 lossy compression format music by using MFCC features. The Kullback-Leibler Divergence and Earth Mover's Distance (EMD) are used to compute music similarity. Experiments show that the retrieving probability of our design can achieve high recall values of 95% out of a total of 1951 tracks in the database.

Keywords: content-based music retrieval, MP3, MFCC, kullback-leibler divergence, EMD

Guo Yali, Han Yan, Liu Linmao Inversion strategy research of travel time tomography with sparse rays
*Computer Modelling & New Technologies **18**(11) 1362-1367*

The incomplete data travel time tomography with sparse rays can result in the ill-posed inverse problem in practical engineering, so the inversion strategy is very important in order to obtain reasonable inversion result. In this paper, the generalized inverse theory is taken and the influences are discussed which the system layout, initial model and prior information will impose on the inversion. The indexes of system optimal layout, the selection principle of initial model and regularization methods are presented in this paper. A velocity model of explosion is imitated and the inversion results are compared. A conclusion can be gained that system optimal layout, initial model rational selection and regularization methods utilization can help to improve inversion precision farthest in practical project.

Keywords: sparse rays, inversion strategy, regularization methods

Qiao Wei, Sun Jie, Du Qingfu Diffusion limited aggregation of magnetic particles with exponential decreasing interactions in three-dimensional space
*Computer Modelling & New Technologies 2014 **18**(11) 1368-1373*

Using the Monte Carlo simulation, we investigate the three-dimensional fractal growth of a magnetic diffusion-limited aggregation (MDLA), which consists of magnetic particles interacting with an exponential potential βCe-αr. The cluster morphology, fractal dimension and magnetic susceptibility of this MDLA are analysed with respect to the range facto rα and the coupling energy βC. In the case of long-range ferromagnetic interaction, our results show that the cluster morphology grows to be a hexagonal symmetry as the coupling energy increases, which is different from the two-dimensional simulation. For a proper coupling energy, the fractal dimension takes the maximal value and the cluster morphology becomes more compact. In the case of short-range interaction, the critical value of the cluster specific magnetic moment is much larger than the simulation result in the MDLA with the interacting potential of power law.

Keywords: diffusion limited aggregation, magnetic particle, exponential decreasing potential

Zhang Yuekan, Liu Peikun, Xiao Linjing, Yang Xinghua, Yang Junru Numerical simulation of multiphase flow inside hydrocyclone based on CFD
*Computer Modelling & New Technologies 2014 **18**(11) 1374-1379*

This paper applied computational fluid dynamics (CFD) method to investigate the internal multiphase flow filed in a 75 mm hydrocyclone. The Reynolds stress model (RSM) and VOF model were employed in the numerical simulation. This study discussed the velocity and pressure distribution in the hydrocyclone, and analysed the formation and development mechanism of air core. The numerical simulation results showed that the flow field was very unstable in the region of the air core. The axial velocity gradient reached its maxima, and the turbulent fluctuation was strongest in the simulation region. This study provided theoretical basis on further research of the air core effect on separation efficiency and pressure drop.

Keywords: numerical simulation, multiphase flow, hydrocyclone, CFD

Lin Yisong, Zhuang Mingjie The complexity analysis and performance comparison of MIMO systems based on antenna selection techniques
*Computer Modelling & New Technologies 2014 **18**(11) 1380-1386*

MIMO systems, also known as multiple-input multiple-output, can effectively improve the capacity and reliability of wireless communication. Antenna selection (AS) is a low-cost low-complexity attractive approach in MIMO systems that capture many advantages of these systems. In this paper, we comprehensively review and describe various kinds of AS schemes in MIMO systems. First, we discuss the techniques of antenna selection from the perspective of different channel environments. Analysis results show that exploiting the AS technique can avoid sending redundant information and improve the channel capacity in the low rank and correlated channel conditions. Then the applications of AS systems in spatial diversity and spatial multiplexing are considered. In order to reduce the complexity of AS algorithm, we propose a novel joint transmit and receive AS scheme in MIMO systems. The results of simulations demonstrate that the proposed AS method of performance outperforms other methods, and the proposed algorithm can significantly reduce computational complexity compared to the optimum algorithm. Finally, we summarize some conclusions about the antenna selection.

Keywords: MIMO, antenna selection, spatial diversity, spatial multiplexing

Zhang HongTao, Zhao Wei Design and realization of atmospheric pressure altitude measuring system with

temperature compensation based on FPGA
*Computer Modelling & New Technologies 2014 **18**(11) 1387-1392*

Starting from the analysis of the measuring principle of atmospheric pressure altitude and the necessity of the pressure sensor temperature compensation, this paper presents the design and realization of an atmospheric pressure altitude measuring system with high performance which use FPGA as the data processor and with the pressure sensor temperature compensation design. Article discusses in detail the hardware circuit design of the measuring system, and the internal structure of FPGA software design. At last, the results of system test verify the feasibility and effectiveness of the atmospheric pressure altitude measuring system. Because of the FPGA's characteristics of high reliability, strong data processing ability, high speed and so on, and the effective combination of temperature compensation design, the atmospheric pressure altitude measuring system has the advantages of high measuring precision, high reliability, good real-time and low power consumption.

Keywords: atmospheric pressure altitude, FPGA, temperature compensation algorithm, measuring system

Chen Xiaojie, Qiu Chengjun, Liu Hongmei, Qu Wei, Liu Yibo Design and implementation of a vibration cantilever energy harvester with suspended piezoelectric beam
*Computer Modelling & New Technologies 2014 **18**(11) 1393-1398*

The frequency response and energy conversion efficiency are the critical issue for vibration energy harvester. In this paper, a vibration energy harvester with suspended piezoelectric beam is proposed. This harvester structure is composed of a discrete support beam in the bottom and a suspended piezoelectric beam on top. This suspended beam structure is beneficial to the higher response frequency and the energy conversion efficiency by applying the tension on the shear mode piezoelectric layer. An analytical bending model of suspended piezoelectric beam is developed, based on the actual deformation: rotational and translational movement modes. The characteristic of suspended piezoelectric beam is also analyzed by Finite element method. Finally, a 1000:1 prototype is fabricated and measured. The optimums acquired experiment results show that it is matching well with the model. The maximum of output voltage is 6.5 V.

Keywords: energy harvester, vibration, suspended piezoelectric beam, piezoelectric layer

Ye Pengchao, Guan Xiaochun, Chen Xiaojing Prediction of capacitance of electrolytic capacitor with ripples
*Computer Modelling & New Technologies 2014 **18**(11) 1399-1405*

The electrolytic capacitor is one of the most critical components in the switching power supply, in actual use of the power supply, the ripple is an important factor that leads to failure of the electrolytic capacitor. The research on the ripple's effect on the service life of the electrolytic capacitor is significant to the stability of the switching power supply. This paper designed a controllable circuit generating ripples of different phases, tested the electrolytic capacitors' changes in capacitance with ripples of different phases, analysed the reason of the changes in the capacitance, selected the fastest change, and used three prediction models of support vector regression, radial basis function and kernel-based partial least-squares algorithm to predict the change trends of capacitance of electrolytic capacitor with ripples of this phase. The results showed that compared with radial basis function and support vector regression model kernel-based partial least-squares algorithm can more accurately predict the service life of the electrolytic capacitor.

Keywords: electrolytic capacitor, ripples, different phases, prediction of capacitance, capacitor life

Chen Yang, Zhou Bin, Yin Jingwei Quadratic polynomial fitting of total energy of null subcarriers in underwater acoustic OFDM communication
*Computer Modelling & New Technologies 2014 **18**(11) 1406-1409*

Doppler scale estimation is one key module in underwater acoustic (UWA) orthogonal frequency division multiplexing (OFDM) communication. In this paper, null subcarrier based blind Doppler scale estimation is considered, which is an extremal problem. The cost function is constructed of the total energy of null subcarriers through DFT. The frequencies of null subcarriers are calculated according to non-uniform Doppler shifts at each tentative scaling factor. Then it is proofed that the cost function can be fitted as a quadratic polynomial near the global minimum. So the accurate location of the global minimum can be achieved through polynomial interpolation. This theory is a basement of developing new methods of Doppler scaling estimation and is verified through an experiment conducted in shallow water.

Keywords: orthogonal frequency division multiplexing (OFDM), underwater acoustic (UWA) communication, polynomial interpolation, Doppler scale estimation

Liu Haohua, Wan Zhou, Li Yiyang, Chen Zhunen Research of the vibrating infrasonic sensor based on Fiber Bragg Grating

*Computer Modelling & New Technologies 2014 **18**(11) 1410-1415*

Current research on infrasound more extensive and in-depth, the sensor is essential prerequisite for infrasonic detection, in order to overcome some lack of current infrasound sensor for detecting the infrasound, paper innovative design a vibrating wire sensor structure, and make it capable of receiving a full rang infrasound, range adjustable, easy installation, etc. and use fiber Bragg grating which has obvious advantages in terms of low-frequency detection as the sensing element. Test the performance of the sensor and the results show that the average of the sensor is 6.11%, within 1 Hz-20 Hz range, the sensor has good linearity and small error, has practical value.

Keywords: Fiber Bragg grating; vibration wire; infrasonic sensor

Gao Ang, Hu Yansu, Duan Weijun The error analysis of TDOA based ultrasonic position

*Computer Modelling & New Technologies 2014 **18**(11) 1416-1420*

How to enhance the accuracy and range is a typical issue in the ultrasonic position technology. The position accuracy depends on the hardware architecture and the detecting methods. This paper designs a hardware platform called Hexagon-Buck with six transceivers around. For the inherent property of TDOA (Time of Difference of Arrival) ranging, the RF synchronous time and embedded software processing time are two inevitable factors that leads to the distance error. At the same time, the ultrasonic velocity is sensitive to the environment temperature. The paper considers both the temperature compensation and Least Squares Linear Regression to enhance the ranging accuracy. By taking the physical test data to identify the linear regression model, we can estimate the real distance. The validity of the method has been proved by the final experiment.

Keywords: ultrasonic ranging, TDOA, least squares linear regression, hexagon-buck

Fan Zheng, Zheng Xianfeng The study and design of inter-harmonic detection device based on quasi-synchronous technique in power system

*Computer Modelling & New Technologies 2014 **18**(11) 1421-1424*

On the basis of analyzing the characteristics of harmonic signals, the inter-harmonic detection algorithm of quasi-synchronous sampling is introduced in this paper. This algorithm pre-processes the original sampled signal and separates all the harmonics and inter-harmonics of the reconstructed signal with designed comb FIR filters, to restrain the mutual interference between them, and finally achieve accurate measurement of harmonic and inter-harmonic in power system. At last, the validity of this algorithm is verified with harmonic and inter-harmonic signal generated from standard signal source.

Keywords: quality of electric energy, inter-harmonics, quasi-synchronization, comb FIR filter

Chen Xiaojing, Liu Ke, Ye Peng An optimization method of signal de-noising in discrete wavelet transform based on generalized cross-validation

*Computer Modelling & New Technologies 2014 **18**(11) 1425-1431*

A method for automatically selecting the asymptotical optimal parameters is presented for signal de-noising in discrete wavelet transform. The parameters of wavelet de-noising were first encoded. A generalized cross-validation algorithm was then used to select these parameters automatically. The parameters that obtained the smallest generalized cross-validation were asymptotically optimal. Simulation signals with different features and range signal-to-noise ratios were used to demonstrate the optimality of the proposed method. In addition, the Raman spectrum of edible oil and nuclear magnetic resonance spectrum of quinine and Boc-protected proline were employed as real-world data to validate the proposed method. The proposed method achieved superior performances in both real-world data and in artificial simulation.

Keywords: signal de-noising, wavelet transform, generalized cross-validation, parameter optimization

Tu Chunjuan, Guo Xuhong Research on the machinability when dry turning hardened steel with ceramic tool

*Computer Modelling & New Technologies 2014 **18**(11) 1432-1437*

The experiments of dry turning hardened steel with ceramic tool (CC6050) were carried out. The cutting forces of ceramic tool with different cutting parameters were measured by Kistler cutting force acquisition system, the prediction model of the surface roughness was built by Particle Swarm Optimization (PSO), the wear morphology of the tool and chip characteristics were observed by scanning electron microscope (SEM), analyzed the wear zone of the tool by the EDAXPV990(EDS). Results indicated that the cutting depth was the dominant factor affecting the cutting force, next was feed speed, the minimum was cutting speed. The feed speed was the dominant factor affecting

the surface roughness, next was cutting depth, the minimum was cutting speed. The crater wear was shown on the rake face and the evenly distributed strip shaped wear was shown on the flank of the CC6050 tool, the main wear mechanism was abrasive wear, diffusion wear, adhesive wear and oxidation wear. The shape of chip was saw-tooth, the rough slice layers with periodic flow were shown on the upside of the chip and the streaks were evenly distributed on the underside of the chip. The cutting speed and feed speed were larger, the serrated level of the chip was higher and the angle of shear was larger.

Keywords: Ceramic Tool, Dry Turning, Hardened Steel, Machinability

Wenlong Du Design of reflective optical fiber displacement sensor using double optical paths
Computer Modelling & New Technologies ***18****(11) 1438-1442*

A finite element three-dimension space model is built by ANSYS based on the first multi-tower self-anchored suspension bridge in China. After that, the work aims at developing finite element analysis using backward and forward methods, discussing internal force of main bridge components in different construction stages. Then, linear change law of the main cable in different load conditions is analysed, and mechanism behaviour of this new structure in the whole construction is acquainted. Therefore, the work provides key force analysis for multi-tower self-anchored suspension bridge, making a good example for design and construction of similar bridges.

Keywords: Multi-tower Self-anchored Suspension Bridge, Mechanism Behaviour, Backward Analysis, Forward Analysis, Finite Element

Jiang Fei Optimization of immune particle swarm algorithm and application on wireless sensor networks
Computer Modelling & New Technologies 2014 ***18****(11) 1443-1448*

As a new network technology, wireless sensor network (WSN) have been applied to military, intelligent transportation, environmental monitoring and other fields. Localization is one of the important support technologies of Wireless Sensor Networks. Location information is important to network monitoring. It is meaningless, if there is no location information. We need to adopt a certain mechanism and algorithm to implement the localization of Wireless Sensor Networks. Based on the analysis of the features of Wireless Sensor Networks on range-free positioning algorithm and DV-Hop positioning algorithm error sources, this paper focuses on the improvement research on DV-Hop positioning algorithm. Inspired by biological immune system and mechanism, this paper introduces the immune information processing mechanisms in the immune system to the particle swarm optimization algorithm and thus gets an immune particle swarm optimization (IPSO) algorithm. By applying into the running of DV-Hop positioning algorithm, the paper proposes a DV-Hop improved algorithm, which is, the WSN positioning algorithm based on IPSO algorithm. Simulation experiments show that the improved algorithm can significantly reduce positioning errors to improve positioning accuracy.

Keywords: wireless sensor networks, immune particle swarm optimization, DV-Hop positioning algorithm