



COMPUTER MODELLING
AND
NEW TECHNOLOGIES

2014
VOLUME 18 NO 8

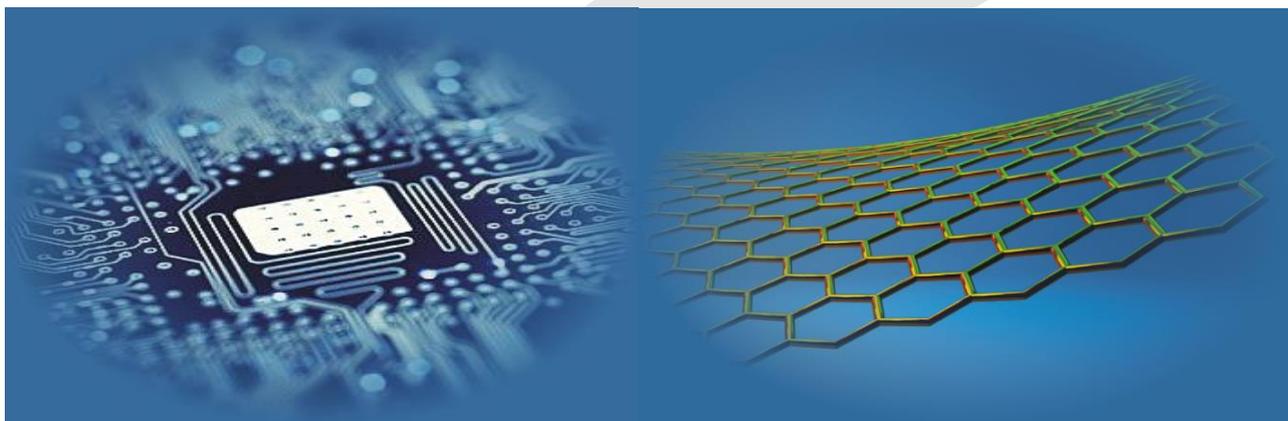
ISSN 1407-5806 ISSN 1407-5814 on-line

Latvian Transport Development and Education Association

Computer Modelling and New Technologies

2014 Volume 18 No 8

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



Riga – 2014

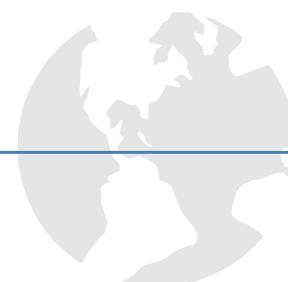
EDITORIAL BOARD

Prof. Igor Kabashkin	Chairman of the Board , <i>Transport & Telecommunication Institute, Latvia</i>
Prof. Yuri Shunin	Editor-in-Chief , <i>Information Systems Management Institute, Latvia</i>
Prof. Adolfas Baublys	<i>Vilnius Gediminas Technical University, Lithuania</i>
Prof. Stefano Bellucci	<i>Frascati National Laboratories – National Institute of Nuclear Physics, Italy</i>
Dr. Brent Bowen	<i>Embry-Riddle Aeronautical University, United States of America</i>
Prof. Olgierd Dumbrajs	<i>University of Latvia, Solid State Physics Institute, Latvia</i>
Prof. Pavel D'yachkov	<i>Kurnakov Institute for General and Inorganic Chemistry, Russian Academy of Sciences, Russian Federation</i>
Prof. Dietmar Fink	<i>University of Mexico, United Mexican States</i>
Prof. Alytis Gruodis	<i>Vilnius University, Lithuania</i>
Prof. Arnold Kiv	<i>Ben-Gurion University of the Negev, Israel</i>
Prof. Vladimir Litovchenko	<i>V. Lashkaryov Institute of Semiconductor Physics of National Academy of Science of Ukraine, Ukraine</i>
Prof. Sergey Maksimenko	<i>Institute for Nuclear Problem, Belarus State University, Belarus</i>
Prof. Ravil Muhamedyev	<i>International IT University, Kazakhstan</i>
Prof. Eva Rysiakiewicz-Pasek	<i>Institute of Physics, Wroclaw University of Technology, Poland</i>
Prof. Michael Schenk	<i>Fraunhofer Institute for Factory Operation and Automation IFF, Germany</i>
Prof. Kurt Schwartz	<i>Gesellschaft für Schwerionenforschung mbH, Darmstadt, Germany</i>
Contributing Editor	Prof. Victor Gopeyenko, <i>Information Systems Management Institute, Latvia</i>
Literary Editor	Prof. Tamara Lobanova-Shunina, <i>Riga Technical University, Latvia</i>
Technical Editor , secretary of Editorial Board	MSc Comp Nataly Burluckaya, <i>Information Systems Management Institute, Latvia</i>

Journal topics:	Publisher	Supporting Organizations
<ul style="list-style-type: none"> mathematical and computer modelling computer and information technologies natural and engineering sciences operation research and decision making nanoscience and nanotechnologies innovative education 	Latvian Transport Development and Education Association	Latvian Academy of Sciences Latvian Operations Research Society Transport and Telecommunication Institute, Latvia Fraunhofer Institute for Factory Operation and Automation IFF, Germany International IT University, Kazakhstan

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

EDITORIAL CORRESPONDENCE	COMPUTER MODELLING AND NEW TECHNOLOGIES, 2014, Vol. 18, No.8 ISSN 1407-5806, ISSN 1407-5814 (on-line: www.cmnt.lv)
Latvian Transport Development and Education Association 68 Graudu, office C105, LV-1058 Riga, Latvia Phone: (+371) 29411640 E-mail: yu_shunin@inbox.lv http://www.cmnt.lv	Scientific and research journal The journal is being published since 1996 The papers published in Journal 'Computer Modelling and New Technologies' are included in: INSPEC , www.theiet.org/resources/inspec/ VINITI , http://www2.viniti.ru/ CAS Database http://www.cas.org/ EI Compindex



Content

Editors' Remarks	5
-------------------------	---

Mathematical and Computer Modelling

Dan Li, Ting Yang, Guangsheng Chen	A AHP-based method to solve contradiction matrix with multiple engineering parameters	7
Xiaowei Niu, Liwan Chen, Qiang Chen, Hui Xie, Hongbing Li	An improved eigenstructure method for estimating DOA in the presence of parameters uncertainties	14
Weipeng Jing, Yaqiu Liu	Multiple DAGs reliability model and fault-tolerant scheduling algorithm in cloud computing system	22
Xiaonan Zhang, Xiaoyong Lu	Multi-state system reliability assessment based on Bayesian networks	31
X G Hong, H Liu, Y Xiao	Fast fractional-pel interpolation algorithm of H.264 based on CUDA	39
Li-jun Deng, Jian Liu	New approach for ventilation network graph drawing based on Sugiyama method and GA-SA algorithm	45
Zeyu Sun, Shouying Li	Multicast routing algorithm based on cloud computing strategy in wireless sensor networks	50
Yuekan Zhang, Peikun Liu, Linjing Xiao, Xinghua Yang, Junru Yang	Computational simulation of the effects of vortex finder diameter on the air core in a hydrocyclone separator	57
Jianhou Gan, Bin Wen, Jinxu Li	Research on ontology mapping of tourism information resources based on description logic	63
Jing Sheng, Hai-Fei Long	Parametric modelling and simulation on oblique cutting based on MSC.Marc	68
Dongwen Zhang, Xinguo An	H^∞ fault-tolerant control for nonlinear singular system via a fault diagnosis observer	73
Jian Xu	Equilibrium distributions of the queue length in M/M/c queueing system	80
Juanjuan Suo, Huimin Dong	Novel method for quality assessment of computational translation	85
Zheng-Mao Zhou, Shun-Hong Zhong, Ming Cai	An efficient and flexible modelling approach for multi-DSP system	91
Chenxia Suo, Yong Yang	A study on application of judgment matrix intelligent correction method in satisfaction evaluation	100

Information and Computer Technologies

Ying Wang, Zhongmin Wang, Sheping Zhai	Spectral colour calibration for multi-ink printer	108
Tong Qin, Xinran Liu	A trust-based resource selection algorithm in Cloud Computing	115
Binghua Cheng, Fei Shao	Optimal routing strategy on weighted networks	122
Baoquan Jin, Xiaohui Hao, Hongjuan Zhang, Yan Gao	The power supply design software model for mine based on VBA technology	127
Yan Gu, Yiqiang Wang, Xiuhua Yuan, Xiaoqin Zhou, Bangcheng Zhang	Software reliability allocation model of CNC system based on software architecture	131
Jiajing Wang, Shuhong Jiao, Zhenyu Sun	A novel unsupervised segmentation for remote sensing image using MRF	137
Shengang Hao, Zhang Li, Gao Xin	Formal resource request representation for remote environment control system	143
Chang-jun Han	Study on load capacity-based cascading failure model in the computer network	148
Chunya Tong	A novel edge detection method based on 2-D Gabor wavelet	153
Jian Yang	A strategy for fault management in LDC wireless sensor network	158

Operation Research and Decision Making

Huaping Zhang	A performance evaluation model of green supply chain based on fuzzy analysis method of multi-attribute decision-making	164
Jiaojin Ci	A performance evaluation model of supply Chain based on extension correlation function	170
Ling Hou, Meng Li, Dongyan Chen	Sourcing and pricing strategy research of competition supply chain under supply disruption	176
Jinshuan Peng, Lei Xu	Lane changing intent identification based on logistic regression model	186
Yujing Wang, Yicheng Jiang, Shouqiang Kang	The application of time domain and frequency domain statistical factors on rolling bearing performance degradation assessment	192
Hao Xu	Evolution of the Nanjing urban green land based on GIS analysis	199
Yaoting Chen, Junyu Dai	A comparative analysis on circulation efficiency of different banana circulation modes in Zhangzhou city based on DEA model	205
Tao Shi, Zhijie Xing	A Comprehensive FAHP evaluation model on domestic sports economy development mode	210
Xinling Du, Lingyi Meng, Yujian Wu	A study on the determinants of e-commerce customer satisfaction	216
Liping Fu, Juan Li, Zuting Zheng	A study on the efficiency of public culture service based on DEA cross evaluation	223
Lixian Jing, Juan Li, Yi An	Application of CCD-model-based DEA analysis method in research on agricultural economic growth	231
Yan-Liang Liu	Impact analysis of trade openness on China's economic growth	236
Chong Qian, Zhan-ao Wang	Influence of China's trade imbalance on economy in the background of great nation	242

Cai-yun Gao, Xi-min Cui	Nonlinear time series of deformation forecasting using improved BP neural networks	249
Chunling Chen, Long Wang, Tongyu Xu, Jiawei Qi	Relative humidity prediction of northern greenhouse environmental factors on the basis of a radial basis function neural network	254
Zhang Jilin	Research on stock analysis methods based on fluid mechanics	259
Jing Xu, Guanxin Yao	Rural logistics service providers pricing and competition-cooperation research considering the 3PL accessibility	269
Yangcheng Hu	Stakeholder orientation and financial performance of NPOs: development and testing of a mediating model	275
Yong-shi Hu, Yue Yu, Ming-xing Xu	Study on distribution and economic growth based on feder-model: evidence from Fujian province in China	280
Yan Zhang	Study on the land desertification early-warning system of Xinjiang in China	286
Chao Yu, Zongshan Pu, Lei Chen	Study on the location choice of Chinese outward foreign direct investment	292
Zhang Lei	The collaborative optimization of uncertain supply chain network under multi-generation co-existence	298
Hao-ran Shi, Xiaoqing Liu, Yao Yang	On dynamic iterative algorithm and the loss of newsvendor problem	306
Mei Yang	A market segmentation model of enterprise marketing based on an improved grey correlation analysis methods	312
Tian Jia	Study on enterprise extension marketing model based on extension engineering methods	316
Ziya Wang, Ran Li	An improved Grey prediction model and the application in college sports information management system	320
Ran Li	An improved method of controlling bullwhip effect and the analysis of the bullwhip effect	327
Nature Phenomena and Innovative Engineering		
Jian Chu, Guoyu Wang, Shan Xu	An extension evaluation model of the operation state of aero engine	333
Hualong Xie, Nan He, Fei Li	The bionic design, virtual prototype modeling and motion simulation of biped robot with heterogeneous legs	339
Mei Liu, Songling Wang, Zhengren Wu	Influence of liquid physical properties on liquid film flow characteristics of uneven wall	347
Danling Wang, Yanfei Wang	Multifractal analysis on gene and PPI networks	353
Yong Yang, Weiwei Zhu	Calculation of microstress in machining distortion of titanium alloy monolithic component based on x-ray diffraction experiment	358
Xianguang Kong, Yihui Li, Lei Yin, Xiaowen Wang	Research and implement of CATIA parametric modelling-based cutter information integration in VERICUT	363
Jian-Zhang Wu, Yu Xiao, Wei Gao	Bilinear model for ontology mapping	368
Gewei Tan, Wei Lin	The two-step motion compensation combined squint wavenumber domain algorithm based on fractional Fourier transform	373
Chao Wang	Research on grid replacement technology and two-dimensional isothermal simulation on melt flow of blown film	379
Liyang Liu, Chao Li, Qi Zhang	Analysis of cylindrical cam molded surface	384
Lifang Ma, Yaxin Wang, Yang Liu, Shizhong Zhang, Yu Chen	Comparative study on prosthetic socket materials	387
Ling Wang, Peng Guo, Lin Liu	Extraction of soil salinization information from the Manas river basin based on TM Images	393
Jianping Li, Tongxiao Shang, Yixiao Guan	Volume integral equation-based electromagnetic inversion of 3D complex resistivity bodies	398
Authors' Index		404
Cumulative Index		405



Editors' Remarks

A Code of Morals

by Rudyard Kipling

Now Jones had left his new-wed bride to keep his house in order,
 And hied away to the Hurrum Hills above the Afghan border,
 To sit on a rock with a heliograph; but ere he left he taught
 His wife the working of the Code that sets the miles at naught.

And Love had made him very sage, as Nature made her fair;
 So Cupid and Apollo linked, per heliograph, the pair.
 At dawn, across the Hurrum Hills, he flashed her counsel wise
 At e'en, the dying sunset bore her husband's homilies.

He warned her 'gainst seductive youths in scarlet clad and gold,
 As much as 'gainst the blandishments paternal of the old;
 But kept his gravest warnings for (hereby the ditty hangs)
 That snowy-haired Lothario, Lieutenant-General Bangs.

'Twas General Bangs, with Aide and Staff, who tittupped on the way,
 When they beheld a heliograph tempestuously at play.
 They thought of Border risings, and of stations sacked and burnt
 So stopped to take the message down -- and this is whay they learnt

"Dash dot dot, dot, dot dash, dot dash dot" twice. The General swore.
 "Was ever General Officer addressed as 'dear' before?
 "My Love,' i' faith! 'My Duck,' Gadzooks! 'My darling popsy-wop!
 "Spirit of great Lord Wolseley, who is on that mountaintop?"

The artless Aide-de-camp were mute; the gilded Staff were still,
 As, dumb with pent-up mirth, they booked that message from the hill;
 For clear as summer lightning-flare, the husband's warning ran:
 "Don't dance or ride with General Bangs -- a most immoral man."

At dawn, across the Hurrum Hills, he flashed her counsel wise
 But, howsoever Love be blind, the world at large hath eyes.
 With damnatory dot and dash he heliographed his wife
 Some interesting details of the General's private life.

The artless Aide-de-camp were mute, the shining Staff were still,
 And red and ever redder grew the General's shaven gill.
 And this is what he said at last (his feelings matter not):
 "I think we've tapped a private line. Hi! Threes about there! Trot!"

All honour unto Bangs, for ne'er did Jones thereafter know
 By word or act official who read off that helio.
 But the tale is on the Frontier, and from Michni to Mooltan
 They know the worthy General as "that most immoral man."

Rudyard Kipling (1809-1849) ♦

This 18th volume No.8 presents actual papers on main topics of Journal specialization, namely, **Mathematical and Computer Modelling, Computer and Information Technologies, Operation Research and Decision Making and Nature Phenomena and Innovative Engineering.**

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

EDITORS



Yuri Shunin



Igor Kabashkin

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



A AHP-based method to solve contradiction matrix with multiple engineering parameters

Dan Li, Ting Yang, Guangsheng Chen*

College of Information and Computer Science Engineering, Northeast Forestry University, Harbin, China

Received 1 May 2014, www.tsi.lv

Abstract

Currently, in the use of TRIZ contradiction matrix table, users need to manually find optimization parameters and deterioration parameters of the invention for the corresponding inventive principles. When many parameters are queried, the user is hard to get the statistics, which are most likely to correspond to the invention and have to rely on tedious accumulative calculation to predict the most likely corresponding inventive principle. In this paper, we aimed to apply the analytic hierarchy process (AHP) to predict the inventive principle; on the basis of the successful cases data, we can take advantage of AHP for the statistics and projections of 40 invention principle through the optimization parameters and deterioration parameters chosen. In this way, we ranked 40 invention principles by the use of probability to give users inventive principles of efficient prediction results and provide the user with a practical guide at the same time.

Keywords: TRIZ, analytic hierarchy process, contradiction matrix table, multiple engineering parameters

1 Introduction

TRIZ is the meaning of the theory of inventive problem solving, it is spelled by the Russian first letter of the words meaning that the theory of inventive problem solving (*Teoriya Resheniya Izobretatelskikh Zadatch*) composed. In the United States and Europe, it can also be abbreviated as TIPS.

The Russian Theory of TRIZ was originally proposed by *Altshuller* from 1946. This method solves technical problems and offers innovative product structures by employing a knowledge base built from the analyses of approximately 2.5 million patents, primarily on mechanical design [1]. TRIZ theory reveals the inherent laws of the invention, it focuses on clarifying and emphasizing contradictions existed in the system, and ultimately achieves the ideal solution completely. It is based on the laws of technology evolution to research the whole process of design and development, rather than random. Through years of verification, the improved use of TRIZ theory can greatly speed up the progress of invention and help people to invent high-quality innovative products. TRIZ consisted of many innovation tools. The basic constituents of TRIZ are the contradiction matrix, effect database, laws of evolution, ideal final result, substance field resources and ARIZ algorithm [2-4].

Contradiction matrix, which consists of 39 engineering parameters and 40 inventive principles, can effectively resolve the conflicts between customer requirements. Effect database is a knowledge database system consisting of physical, chemical, and geometrical effects and rules for solving problems. Among these TRIZ tools, contradiction

matrix is the most commonly used in TRIZ innovation. During the process, firstly, we should find the improved engineering parameter to optimize the system, but with this direction, a worsening engineering direction is created, so a worsening engineering parameter needs to be chosen. In the next step, with the two parameters, two or three inventive principles are matched in the contradiction matrix. The design engineer can solve the engineering innovative design problem with one of these inventive principles. But in the actual complicated system, improving engineering parameter is easily found toward the direction of the system improvement. On the other hand, a worsening engineering parameter is quite difficult to be chosen. The user may choose several worsening engineering parameters for innovative design problem. A proper principle selection is a very important issue for this process, but a multiplied likelihood of inventive principles is generated with this method. Due to the difficulty for users to decide which principle is the fittest. The decision-maker with TRIZ experience may need a large amount of data for analysis and many factors should be considered for selection of the proper principle. Or they can only rely on cumbersome cumulative calculation to predict the most likely principles of the invention.

AHP (Analytic Hierarchy Process AHP for short), proposed by T.L. Saaty, a professor of USA strategist in early 1970s, is a simple, flexible and practical multi-criteria decision making approach for the quantitative analysis of qualitative problem. The characteristics of the analytic hierarchy process are the basis for the analysis of complex decision problems in nature, its inherent relationship between factors and so on. It just makes use

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

of quantitative information to make decision during the process of mathematical thinking, and then provides solutions for complex decision problem of multi-objective, multi-criteria or no structural characteristics. It is particularly suitable for the measurement occasions where decision result is far more difficult to direct accurate.

This paper was organized as follows: The second part of the system, the article expounded the theory of TRIZ contradiction matrix and its solving process. The third part described in detail the principle of analytic hierarchy process (AHP) and basic steps. In the fourth part introduced the case analysis, and through the corresponding contradiction matrix to solve application software based on AHP analytic hierarchy process (AHP) to solve practical problems. The last part summarized the relevant results and development prospects.

2 Contradiction matrix

Through the study of a large number of invention patents, Altshuller summed up 39 common parameters, which are generally physical, geometrical and technical performance parameters, and usually used in the engineering field representation of system performance, where he extracted the most important TRIZ, with widespread use of 40 inventive principles. Altshuller linked the 39 general engineering parameters and 40 inventive principles organically to establish correspondence of the organized contradiction matrix of 39X39. The first column indicates the matrix optimization parameters for improvement, while the first line indicates deterioration parameters which will bring in determining the optimal parameters. After confirm optimization parameters and deterioration parameters, users can find a serial number corresponding innovation yards in the matrix table, the principle constitutes a collection of contradictions possible solutions.

When using the contradiction matrix and 40 inventive principles to solve practical problems, users should first determine the function of technology systems and raise the problem to be solved, and then convert into the universal significance of specific issues and deterioration and optimization parameters. After that, users should determine the corresponding matrix table to resolve conflicting principles of the invention to identify the solution of the problem with practice analysis. As shown in Figure 1:

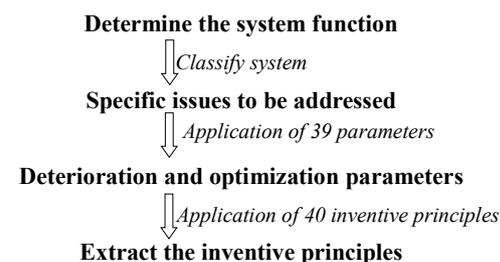


FIGURE 1 Application steps of contradiction matrix

In order to make the method to be more normative and manoeuvrability, several steps are given as follows:

- Determine the main function of the technical system.
- Decompose technical system in detail, divided into system levels, lists the super system, system, subsystem parts at all levels and all kinds of auxiliary functions.
- Describe the actual concrete problems existing in the technical system.
- Abstract the actual problem, apply the 39 contradictions matrix and determine the technical characteristics of the system, which should be improved.
- Screen designed systems to identify and deteriorate properties. Improvements in enhancing characteristics, while the other is bound to bring deterioration of one or more characteristics. Because of deteriorated parameters are not yet occurred and often, so when screening and determining the characteristics of deterioration requires "bold vision, careful verification".
- Query contradiction matrix table through the determined parameters.
- Find contradiction matrix table to get the recommended principles of the invention Sort Code.
- Find 40 Sort Code in accordance with the principles of the invention directory, and get the serial number and name of the principles of the invention.
- According to the invention of the principle of the serial number and name, corresponding to find article 40 invention principles and examples for invention principle of explanation.
- Apply the recommended principles of the invention one by one to specific issues, and explore how to apply each principle and implementation on specific issues.
- If the principles of the invention do not apply to specific problems, users need to redefine the project parameters and contradictions, then apply and search a contradiction matrix table again.
- Filter out the best solutions into the product design stage.

3 Analytic hierarchy process (AHP)

3.1 INTRODUCTION TO AHP

AHP (Analytic Hierarchy Process AHP for short) is an American Operations Research Professor TL Saaty proposed in the early 1970s, the AHP is a problem of qualitative quantitative analysis of a simple, flexible and practical method of multi-criteria decision-making. It addresses how to determine the relative importance of a set of activities in a multi-criteria decision problem [8]. It is characterized by dividing the complex problems in a variety of factors into interconnected orderly levels and streamlines, then making it principled. According to a subjective judgment of a certain objective reality structure

(mainly pairwise comparisons), AHP quantitatively describes the importance of a hierarchy of elements of pairwise comparisons by analysing the expert opinions, the objective judgment results together directly and efficiently. Then, by using mathematical method to calculate, the weights reflect the relative importance of the order of elements in each level. However, users should calculate the relative weights of all the elements right and sort based on all levels of the total order; and then establish the judgment matrix, by calculating the eigenvalues and eigenvectors of the judgment matrix, finally obtain the weight of the different options and provide evidence for the fittest. The AHP method is based on three principles: structure of the model; comparative judgment of the alternatives and the criteria; synthesis of the priorities. The method was introduced to China in 1982, with its combination of qualitative and quantitative characteristics of the decision-making to deal with a variety of factors, as well as the advantages of their system which is simple and flexible. So far AHP has been widely appreciated and applied in many study areas for complicated decision-making, especially in various fields of social economy, such as energy systems analysis, urban planning, economic management, research and evaluation.

3.2 THE BASIC STEPS OF THE AHP

3.2.1 Hierarchy model

Firstly a complex decision problem is structured as a hierarchy, analysed the problem profoundly. On the basis of in-depth analysis of the practical problem, the various factors related to different properties are divided into a number of levels from top to bottom, with a layer of the factors belonging to the upper layer of the element or elements affected, while the next control factors underlying layer or by action of factors. The objectives, criteria and alternatives are arranged in a hierarchical structure similar to a family tree. The uppermost layer of the target, usually only one factor, the program or object is generally lower layer, the intermediate may be one or several levels, typically as a criterion or indicator layer. Criteria (for example, more than nine) should be further decomposed sub-standard level when the number is large. The factors identified include stratification: Top (aim to solve the problem); the lowest level (for a variety of measures to address the problem, programs, etc.). The various factors in the same layer are basically relatively independent when compared with each other and should be considered in the appropriate level. Express clearly the relationship of these factors with the hierarchical structure. Usually, the hierarchy can be divided into the goal-guidelines or indicators-program [4]

3.2.2 Multiple pairwise comparison matrix

After the problem has been decomposed, and the hierarchy is constructed, prioritization procedure starts in order to

determine the relative importance of the criteria within each level. The criteria is based on related factors between two layers of the pairwise comparison, and n criteria can be summarized in an $(n \times n)$ evaluation matrix A in which every element a_{ij} ($i, j = 1, 2, 3, \dots, n$) is the quotient of weights of the criteria.

Based on pairwise comparison method and 1~9 scales, construct the comparison array. As the disposable element of the second layer is 40, far more than 9, so this article took the elements itself weight to construct paired comparison array.

While compare the importance of i and j with upper layer of some factor relative, a_{ij} is used to describe the quantified relative weights. A total of n elements in comparison, it is called a paired comparison matrix. Pairwise comparison matrix values of a_{ij} refer to Satty's proposal, which is according to the following scale assignment. The value of a_{ij} is among 1~9 and its inverse.

- $a_{ij} = 1$, i is as important as j ;
- $a_{ij} = 3$, i is more important than j slightly;
- $a_{ij} = 5$, i is more important than j ;
- $a_{ij} = 7$, i is much more important than j ;
- $a_{ij} = 9$, i is extremely important than j ;
- $a_{ij} = 2n$, $n = 1, 2, 3, 4$, the importance of i and j is between $a_{ij} = 2n - 1$ and $a_{ij} = 2n + 1$;
- $a_{ij} = \frac{1}{n}$, $n = 1, 2, 3, \dots, 9$, if and only if $a_{ij} = n$. Features:

paired comparison matrix. $a_{ij} > 0$, $a_{ij} = 1$, $a_{ij} = \frac{1}{a_{ji}}$,

when $i = j$, $a_{ij} = 1$.

3.2.3 Calculate the weight vector

In order to extract useful information from the judgment matrix and understand the regularity of things, we need to calculate the weight vector of judgment matrix to provide scientific basis for decision-making. Calculate the relative weight of factors in each judgment matrix according to its principles. It means calculating the largest eigenvalue and eigenvectors for each pairwise comparison matrix; and then test the consistency. If passed, the eigenvectors are equal to the weight vector. Accurate calculation of the largest eigenvalue and eigenvectors are too complex. Therefore, in this paper, we apply the simplified calculation method that any column vector of the consistent array is eigenvectors. The column vector of a reciprocal matrix of good consistency approximates to its eigenvectors. Therefore, the arithmetic average of the column vectors is available.

3.2.4 Consistency test

When determining an order of the matrix, it is often difficult to construct a matrix of conformance. However, the consistency of judgment matrix deviation condition should have a degree, therefore, we must determine

whether it is an acceptable matrix for identification, which is the connotation of the consistency check. We use consistency index, random consistency index and consistency ratio to test for consistency of each pairwise comparison matrix calculated maximum eigenvalue and the corresponding eigenvector. Calculate the lowest level of the target portfolio weight vector, and do a combination of consistency test according to the formula, if the test is passed, according to a combination of the weight vector, the results can be expressed in decision-making, or we need to rethink or re-construct a larger model that consistency ratio pairwise comparison matrix.

Only through consistency test, we can think the judgment matrix is reasonable logically. The consistency test is calculated as follows:

$$CR = CI / RI . \tag{1}$$

CI - consistency index, which is used to measure a paired comparison matrix inconsistent degree of indicators, can be calculated as:

$$CI = (\lambda_{\max} - n) / (n - 1), \tag{2}$$

where λ_{\max} is the largest eigenvalue of judgment matrix, n is the number of pairwise comparison factor. The smaller the value of the CI is, the greater the consistency indicates.

RI is the random consistency index and related to the order number judgment matrix. Generally, the order of the matrix increases, the greater possibility of deviation from the greater consistency of the random will be. It can be referred to the look-up table as follows in Table 1:

TABLE 1 Value of RI

Matrix order	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

CR is the consistency proportion. The consistency of judgment matrix can be acceptable when $CR < 0.10$, otherwise will be required to make modifications.

4 Analytic hierarchy process specific application

4.1 CONSTRUCTION OF HIERARCHY MODEL

For predicting the possibility of the inventive principles in the mechanical field, we first predicted the possibility of the principles of the invention as the first layer--the target layer; then stratified factors involved, the first layer consists of two factors, the principle of occurrences and principle applied probability, they are the second layer, i.e. the criterion level, $O = \{c_1, c_2\}$; the second layer comprises of 40 factors, respectively, acts as TRIZ 40 principles of the invention, they are the third layer - Option layer, $C_1 = \{P_1, P_2, \dots, P_{40}\}$, $C_2 = \{P_1, P_2, \dots, P_{40}\}$. As shown in Figure 2:

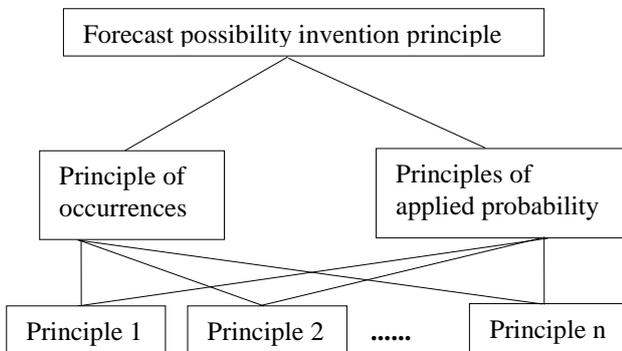


FIGURE 2 Hierarchy model of AHP

4.2 CONSTRUCT PAIRWISE COMPARISON MATRIX

According to the questionnaire with a number of success cases and the collation of data, we have the initial data, $P_i = \{x, y\}$, x represents the principle of occurrences, y represents the principle applied probability.

$P_1 = \{29, 0.75\}, P_2 = \{22, 0.50\}, P_3 = \{23, 0.75\}, P_4 = \{21, 0.75\}, P_5 = \{20, 0.75\}, P_6 = \{16, 0.75\}, P_7 = \{21, 0.75\}, P_8 = \{14, 0.50\}, P_9 = \{14, 0.50\}, P_{10} = \{25, 0.75\}, P_{11} = \{18, 0.50\}, P_{12} = \{14, 0.50\}, P_{13} = \{13, 0.50\}, P_{14} = \{20, 0.75\}, P_{15} = \{20, 0.25\}, P_{16} = \{13, 0.25\}, P_{17} = \{21, 0.50\}, P_{18} = \{20, 0.75\}, P_{19} = \{22, 0.25\}, P_{20} = \{10, 0.25\}, P_{21} = \{15, 0.25\}, P_{22} = \{22, 0.50\}, P_{23} = \{14, 0.50\}, P_{24} = \{26, 0.75\}, P_{25} = \{16, 0.50\}, P_{26} = \{16, 0.75\}, P_{27} = \{19, 0.50\}, P_{28} = \{31, 0.50\}, P_{29} = \{16, 0.50\}, P_{30} = \{16, 0.50\}, P_{31} = \{14, 0.25\}, P_{32} = \{13, 0.50\}, P_{33} = \{9, 0.25\}, P_{34} = \{9, 0.25\}, P_{35} = \{13, 0.25\}, P_{36} = \{14, 0.25\}, P_{37} = \{17, 0.50\}, P_{38} = \{5, 0.25\}, P_{39} = \{10, 0.25\}, P_{40} = \{16, 0.50\}.$

$$C_1 = \begin{pmatrix} 1 & P_{1,x}/P_{2,x} & P_{1,x}/P_{3,x} & \dots & P_{1,x}/P_{40,x} \\ P_{2,x}/P_{1,x} & 1 & P_{2,x}/P_{3,x} & \dots & P_{2,x}/P_{40,x} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{40,x}/P_{1,x} & P_{40,x}/P_{2,x} & P_{40,x}/P_{3,x} & \dots & 1 \end{pmatrix},$$

$$C_2 = \begin{pmatrix} 1 & P_{1,y}/P_{2,y} & P_{1,y}/P_{3,y} & \dots & P_{1,y}/P_{40,y} \\ P_{2,y}/P_{1,y} & 1 & P_{2,y}/P_{3,y} & \dots & P_{2,y}/P_{40,y} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{40,y}/P_{1,y} & P_{40,y}/P_{2,y} & P_{40,y}/P_{3,y} & \dots & 1 \end{pmatrix}.$$

The data into C_1 and C_2 , we have:

$$C_1 = \begin{pmatrix} 1 & 29/22 & 29/23 & 29/16 \\ 22/29 & 1 & 22/23 & 22/16 \\ \vdots & \vdots & \vdots & \vdots \\ 16/29 & 16/22 & 16/23 & 1 \end{pmatrix},$$

$$C_2 = \begin{pmatrix} 1 & 0.75/0.50 & 0.75/0.75 & 0.75/0.50 \\ 0.50/0.75 & 1 & 0.50/0.75 & 0.50/0.50 \\ \vdots & \vdots & \vdots & \vdots \\ 0.50/0.75 & 0.50/0.75 & 0.50/0.75 & 1 \end{pmatrix}.$$

4.3 CALCULATE THE WEIGHT VECTOR AND TEST CONSISTENCY

Because C_1 satisfy $C_{ij} \times C_{jk} = C_{ik}$, where $i, j, k=1,2,3, \dots, 40$, C_{ij} , C_{jk} , C_{ik} , are respectively the i row j column element, and the j row, column k element, the i row element of the column k of C_1 , so C_1 is consistent array.

Normalize the column vector of C_1 , we have A_1 :

$$A_1 = \begin{pmatrix} P_{1 \cdot x}/PX & P_{1 \cdot x}/PX & \dots & P_{1 \cdot x}/PX \\ P_{2 \cdot x}/PX & P_{2 \cdot x}/PX & \dots & P_{2 \cdot x}/PX \\ \vdots & \vdots & \vdots & \vdots \\ P_{40 \cdot x}/PX & P_{40 \cdot x}/PX & \dots & P_{40 \cdot x}/PX \end{pmatrix}.$$

The arithmetic average of A_1 , obtain the weight vectors W_1 :

$$W_1 = \begin{pmatrix} P_{1 \cdot x}/PX \\ P_{2 \cdot x}/PX \\ \vdots \\ P_{40 \cdot x}/PX \end{pmatrix},$$

wherein $PX = \sum P_i \cdot x \quad i = 1, 2, 3, \dots, 40$.

Values into W_1 , we have:

$$W_1^T = [0.0422, 0.0320, 0.0334, 0.0305, 0.0291, 0.0233, 0.0305, 0.0203, 0.0218, 0.0363, 0.0262, 0.0203, 0.0189, 0.0291, 0.0291, 0.0189, 0.0305, 0.0291, 0.0320, 0.0145, 0.0218, 0.0320, 0.0203, 0.0378, 0.0233, 0.0233, 0.0276, 0.0451, 0.0233, 0.0233, 0.0203, 0.0189, 0.0131, 0.0131, 0.0189, 0.0203, 0.0247, 0.0073, 0.0145, 0.0233].$$

Since C_2 satisfy $C_{ij} \times C_{jk} = C_{ik}$, wherein $i, j, k=1,2,3, \dots, 40$, C_{ij} , C_{jk} , C_{ik} , are respectively the i row j column element, the j row, column k element, the i row element of the column k of C_2 , so C_2 is consistent array

Normalize column vector of C_2 , we have A_2 :

$$A_2 = \begin{pmatrix} P_{1 \cdot y}/PY & P_{1 \cdot y}/PY & \dots & P_{1 \cdot y}/PY \\ P_{2 \cdot y}/PY & P_{2 \cdot y}/PY & \dots & P_{2 \cdot y}/PY \\ \vdots & \vdots & \vdots & \vdots \\ P_{40 \cdot y}/PY & P_{40 \cdot y}/PY & \dots & P_{40 \cdot y}/PY \end{pmatrix},$$

The arithmetic average of A_2 , obtain the weight vectors W_2 :

$$W_2 = \begin{pmatrix} P_{1 \cdot y}/PY \\ P_{2 \cdot y}/PY \\ \vdots \\ P_{40 \cdot y}/PY \end{pmatrix},$$

where in $PY = \sum P_j \cdot y \quad j = 1, 2, 3, \dots, 40$.

Values into W_2 , we have:

$$W_2^T = [0.0395, 0.0263, 0.0395, 0.0395, 0.0395, 0.0263, 0.0395, 0.0263, 0.0263, 0.0395, 0.0263, 0.0263, 0.0263, 0.0395, 0.0132, 0.0132, 0.0263, 0.0132, 0.0132, 0.0132, 0.0132, 0.0263, 0.0263, 0.0395, 0.0263, 0.0395, 0.0263, 0.0263, 0.0263, 0.0263, 0.0132, 0.0263, 0.0132, 0.0132, 0.0132, 0.0263, 0.0132, 0.0132, 0.0132, 0.0132].$$

4.4 CALCULATE THE WEIGHT OF COMBINATION VECTOR AND TEST THE CONSISTENCY

We believe that the principle of occurrences and principles applied probability are equally important to predict the possibility of the invention principles, effect rate of O_1 , O_2 are 0.5.

Values into O , we have:

$$O = \begin{pmatrix} 1 & 0.5/0.5 \\ 0.5/0.5 & 1 \end{pmatrix},$$

O to satisfy $O_{ij} \times O_{jk} = O_{ik}$, wherein $i, j, k=1,2$, O_{ij} , O_{jk} , O_{ik} are respectively the i row of the j column element, the j row of the k column element of i row of k column element of O , so O is consistent array.

Normalize column vector of O , we have A_3 :

$$A_3 = \begin{pmatrix} O_1/(O_1 + O_2) & O_1/(O_1 + O_2) \\ O_2/(O_1 + O_2) & O_2/(O_1 + O_2) \end{pmatrix}.$$

The arithmetic average of A_3 , obtain the weight vectors W_3 :

$$W_3 = \begin{pmatrix} O_1/(O_1 + O_2) \\ O_2/(O_1 + O_2) \end{pmatrix},$$

values into the W_3 , we have:

$$W_3^T = [0.5, 0.5].$$

Finally, the combination of the weight vector W :

$$W = W_3^T \cdot [W_1, W_2],$$

$$W = \begin{pmatrix} [O_1/(O_1 + O_2)] \cdot P_{1 \cdot x}/PX + [O_2/(O_1 + O_2)] \cdot P_{1 \cdot y}/PY \\ [O_1/(O_1 + O_2)] \cdot P_{2 \cdot x}/PX + [O_2/(O_1 + O_2)] \cdot P_{2 \cdot y}/PY \\ \vdots \\ [O_1/(O_1 + O_2)] \cdot P_{40 \cdot x}/PX + [O_2/(O_1 + O_2)] \cdot P_{40 \cdot y}/PY \end{pmatrix},$$

where $PX = \sum P_i \cdot x; PY = \sum P_j \cdot y; i = 1, 2, 3, \dots, 40$.

Values into the W , we have:

$$W^T = [0.0409, 0.0292, 0.0365, 0.0350, 0.0343, 0.0248, 0.0350, 0.0233, 0.0241, 0.0379, 0.0263, 0.0233, 0.0226, 0.0343, 0.0212, 0.0161, 0.0284, 0.0212, 0.0226, 0.0139, 0.0175, 0.0292, 0.0233, 0.0387, 0.0248, 0.0314, 0.0270, 0.0357, 0.0248, 0.0248, 0.0168, 0.0226, 0.0132, 0.0132, 0.0161, 0.0168, 0.0255, 0.0103, 0.0139, 0.0248].$$

5 Features of the system

Based on the above technical contradiction matrix computer knowledge representation, this paper proposed a search algorithm based on the principle of AHP, and studied the software application of computer aided innovations and researches in China and oversea. By analyzing the advantages and limitations of existing soft wares, we finally established multiple parameters based on AHP technology invention principle contradiction auxiliary innovation software.

The use steps of the software are as follows:

Firstly, enter a problem description and industry, and then select the optimized parameters. For example, the problem description: cell phones, industry: electronic, optimized parameters: the weight of the stationary object, the length of a stationary object, as shown in Figure 3.



FIGURE 3 Optimization parameter selection interface

Secondly, click [Submit] to get into the degradation parameter selection interface, choose degradation parameters, for example: Control complexity, design complexity, the degree of automation and productivity, as shown in Figure 4.



FIGURE 4 Degradation parameter selection interface

Then, click the [Submit] to get into the principles of the present invention interface, as shown in Figure 5.

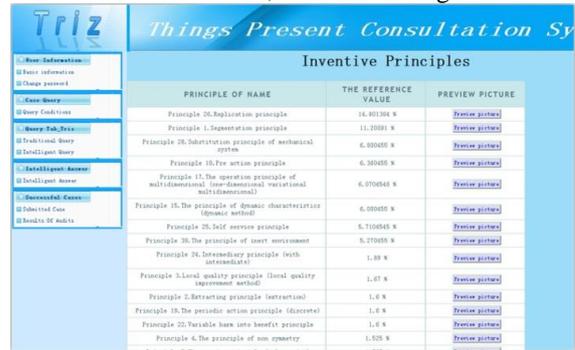


FIGURE 5 Invention principles interface

At last, click the [Preview] to get into the corresponding principle picture according to the principle of 1 as an example, as shown in Figure 6.

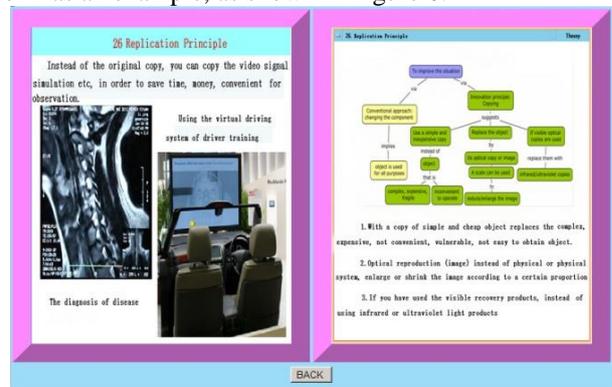


FIGURE 6 Corresponding principle picture

6 Conclusions

The TRIZ innovation system is a complex one. In different industries, the prediction of the principles of the invention to solve the problem is difficult to achieve. The reason may be that various factors affect each other and affecting factors are so many. Therefore, based on the scientific and practical principles, it should be combined with the characteristics of the TRIZ innovation system, using the Uncertain AHP to analyse the predictions of the inventive principles and determine the weight of impact factors, then obtain the predictive value of the 40 inventive principles. The purpose of this paper was to apply the analytic hierarchy process(AHP) for predicting the inventive principle ,namely that, on the basis of the successful cases data, we can take advantage of the analytic hierarchy process(AHP) for the statistics and projections of 40 invention principle through the optimization parameters and deterioration parameters chosen .In this way, we can rank 40 invention principles by the use of probability to give users inventive principles of efficient prediction results and provide the user with a practical guide at the same time. With the rapid development of TRIZ innovative system, the predictive value of the 40 inventive principles in different industries will be more and more accurate, the significance is increasing.

Acknowledgments

DL12EB01-03 and Heilongjiang Natural science fund in China Nos. F201116.

This work has been supported by the Fundamental Research Funds for the Central Universities Nos.

References

- [1] Li T-S, Huang H-H 2009 Applying TRIZ and Fuzzy AHP to develop innovative design for automated manufacturing systems. *Expert Systems with Applications* **36** 8302–12
- [2] Saaty, T L 1980 The analytic hierarchy process. *New York: McGraw-Hill*
- [3] Dagdeviren M, Yavuz S, Kilinc N 2009 Weapon selection using the AHP and TOPSIS methods under fuzzy environment *Expert Systems with Applications* **36** 8143–51
- [4] Albayrak, E, Erensal Y C 2004 Using analytic hierarchy process (AHP) to improve human performance. An application of multiple criteria decision making problem *Journal of Intelligent Manufacturing* **15** 491-503

Authors	
	<p>Li Dan, born on July 25, 1981, Dandong, China</p> <p>Current position, grades: lecturer of College of information and computer science engineering, Northeast Forestry University, Ph.D.</p> <p>University studies: Northeast Forestry University, in the College of information and computer science engineering (2000-2010).</p> <p>Scientific interest: computer aided innovation, TRIZ theory, forestry information.</p> <p>Publications: 10 papers, 5 patents.</p>
	<p>Yang Ting, born on February 14, 1992, Hubei, China</p> <p>Current position, grades: undergraduate student in the Northeast Forestry University.</p> <p>University studies: B.S. in major of computer science and technology, Northeast Forestry University in 2014.</p> <p>Scientific interest: computer aided innovation.</p> <p>Publications: 1 paper.</p>
	<p>Chen Guangsheng, born on May 2, 1969, Heilongjiang, China</p> <p>Current position, grades: professor of College of information and computer science engineering, Northeast Forestry University.</p> <p>University studies: Ph.D. in major of Wood science and technology, Harbin Northeast Forestry University in 2006.</p> <p>Scientific interest: computer aided innovation, TRIZ theory, forestry information.</p> <p>Publications: 12 papers.</p>

An improved eigenstructure method for estimating DOA in the presence of parameters uncertainties

Xiaowei Niu^{*}, Liwan Chen, Qiang Chen, Hui Xie, Hongbing Li

School of Electronic and information engineering, Chongqing Three Gorges University, Zip code, 404000, Wan Zhou, China

Received 1 March 2014, www.tsi.lv

Abstract

This paper presents an improved eigenstructure-based method for estimating the direction of arrival (DOA) of received signal in uniform circular-array, in the presence of sensor gain and phase uncertainties. A simple sensor gain and phase uncertainties calibration method, which does not require any prior knowledge of the DOAs, but also being capable of eliminating the DOA estimation ambiguity, is proposed. The performance of the proposed method is demonstrated by some representative computer simulation.

Keywords: eigenstructure, gain and phase uncertainty calibration, DOA estimation

1 Introduction

Direction of arrival (DOA) of multiple narrowband signals estimation has widely been discussed this decade. Existing DOA estimation algorithms, such as MUSIC [1], ESPRIT [2], Capon's beam former [3], subspace-based method [4] and parametric maximum-likelihood [5] are known to be highly sensitive to the errors in the array manifold. In these algorithms prior knowledge of the signals received by the sensor array from signal sources, is required. However, in practice there always exist various degrees of perturbation in sensor array. Therefore, it is required to calibrate sensors before the DOA estimation.

Recently, the estimation of the DOA of signal emitted by narrowband sources has been widely investigated using signal processing methods. In the method of Weiss and Friedlander [6] they proposed an algorithm to calibrate the sensor array, however this method suffers from suboptimal convergence problem. Paulraj and Kailath [7] proposed method of DOA estimation in the presence of sensor gain and phase uncertainty, based on a least-squares (LS) approach using a linear equispaced (LES) array. The method does not need calibrating sources, but it suffers from high computational requirement. The method in [8] used self-calibration algorithms based on least squares approach to compensate the problem of DOA estimation using LEA in the presence of phase errors. Moreover, the method does not require any prior knowledge about signal source direction. The authors in [9] developed an eigenstructure method for DOA estimation in the presence of sensor gain and phase perturbation, which compensates for the suboptimal convergence problem, which occurs in [6]. The method in [10] considered the problem of phase autocalibration for uniform rectangular array (URA). It solves the problem of ambiguity, which arises when the phase and the DOA parameters are identified together. In

[11] an iterative Maximum likelihood (ML) procedure was developed to estimate DOA using sparse sensor arrays composing of multiple widely separated sub arrays, which improve the performance achieved by Friedlander and Weiss in [12]. In [13] a MUSIC like algorithm was investigated for gain and phase estimation, assuming the source angle is known. In [14] the problem of gain and phase estimation, using the true covariance matrix was presented. The maximum likelihood calibration algorithm [15] was presented to compensate for the effect of mutual coupling, sensor gain, phase errors, and sensor position errors by estimating their calibration matrix using a set of calibrated sources at predetermined locations. Friedlander and Weiss [16] proposed an eigenstructure-based method to compensate for the mutual coupling and perturbation of gain and phase. Moreover, their method not required calibration sources. The method in [17] considered a problem of gain and phase estimation of (LEAs) based in different diagonal lines of covariance matrix.

The method in [9] studied DOA estimation problem in the presence of gain and phase uncertainties. It estimates the DOAs based on the eigendecomposition of a covariance matrix constructed from the dot product of the array output vector and its conjugate. However it has some drawbacks, such as; it is not applicable in Uniform linear arrays (ULA), and it gives ambiguous DOA estimation in circular array antennas. The proposed method solves the problem of DOA ambiguity in circular array antenna, but still not applicable in linear array antennas.

The rest of this paper is organized as follows: section II and section III describe the problem formulation and the calibration methods respectively. While section IV, discuss and present the results of computer simulations performed. Section V, gives the conclusion of the paper.

^{*} *Corresponding author* e-mail: nxw4525@126.com

2 Problem Formulations

Consider K narrowband far-field signals, $s_k(t)$ for $k=1,2,\dots,K$, with centre wavelength, λ , impinging on a planar array of M omnidirectional sensors labelled $1,2,\dots,M$, from directions θ_k for $k=1,2,\dots,K$, where sensor 1 is taken as the reference point and the coordinate of sensor m -th is denoted by (x_m, y_m) . The array output can be described as:

$$r_0(t) = \sum_{k=1}^K a(\theta_k) s_k(t) + n(t) = fGAs + n(t), \quad (1)$$

where

$$A = [a(\theta_1), a(\theta_2), \dots, a(\theta_K)], \quad (2)$$

$$a(\theta_k) = [1, e^{-j2\pi d_{m,k}}, \dots, e^{-j2\pi d_{M,k}}]^T, \quad (3)$$

$$l_{m,k} = x_m \cos \theta_k + y_m \sin \theta_k, \quad (4)$$

$$s = [s_1(t), s_2(t), \dots, s_K(t)]^T, \quad (5)$$

$$f = \text{diag}(e^{j\phi_1}, e^{j\phi_2}, \dots, e^{j\phi_M}), \quad (6)$$

$$G = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_M), \quad (7)$$

α_m and ϕ_m denotes the gain and phase uncertainties of Sensor m , respectively. $n(t)$ is the vector of additive Gaussian white noise and diag means diagonal matrix. In this paper, the superscripts “*”, T and H denote the conjugate, transpose, and conjugate transpose, respectively.

Assuming that the additive Gaussian white noise has a zero mean, σ_n^2 variance, and $R_s = E[ss^H]$, then the covariance matrix of array output will be:

$$R = fGAR_s A^H G^H f^H + \sigma_n^2 I. \quad (8)$$

Thus, the problem addressed here is as follows: the DOA and the corresponding array gain-phase uncertainties are estimated from a given array output vector.

3 Calibration Methods

Here we propose a simple method of estimating DOA with gain and phase uncertainties. The idea is similar to that in [9], but the proposed method has better performance, and no DOA ambiguity. For comparison purpose, we review the method in [9] first.

3.1 THE DOT PRODUCT METHOD

The dot product method in [9] is described as follows.

3.1.1 The gain uncertain estimation

The eigenvalue decomposing matrix is given as follows:

$$R = \sum_{m=1}^M \beta_m u_m u_m^H, \quad (9)$$

where β_m represents the eigenvalues in descending order, and u_m presents corresponding eigenvectors. Then the gain uncertain can be estimated as:

$$\hat{\alpha}_m = \text{sqrt} \left(\frac{R(m,m) - \hat{\sigma}_n^2}{R(1,1) - \hat{\sigma}_n^2} \right), \quad (10)$$

where

$$\hat{\sigma}_n^2 = \frac{1}{M-K} \sum_{m=K+1}^M \beta_m. \quad (11)$$

3.1.2 The DOA estimation

To get an unambiguous DOA estimation, we need to make the radius of circular array antenna less than or equal to $\lambda/4$, and the two directions $\hat{\theta}_k$ and $\hat{\theta}_n$ must not closed to each other.

Let $\text{Re}[\cdot]$ and $\text{Im}[\cdot]$ be the real and imaginary parts of a complex number, respectively.

For complex signals, the two dimensional spatial spectrum is defined as

$$p_c(\theta, \theta') = (V_c^H \text{Re}[a(\theta) \odot a^*(\theta')])^2 + V_c^H I_m [a(\theta) \odot a^*(\theta')]^2)^{-1}, \quad (12)$$

where V_c the noise eigenvector subspace of the dot product of the received signal with its conjugate, and $\|\cdot\|$ represents the 2-norm of a vector. Thus, the DOA pairs are given by:

$$(\hat{\theta}_k, \hat{\theta}_n) = \arg \max_{\theta' > \theta + \Delta\theta} p_c(\theta, \theta') \quad \text{for } k, n = 1, 2, 3, \dots, K; k \neq n \quad (13)$$

The subscript $\Delta\theta$ has little effect on the performance of the DOA estimation method, and it is convenient to set $\Delta\theta = 1^\circ$, while the subscript c implies complex.

For real signals, the two dimensional spatial spectrum is defined as follows:

$$p_r(\theta, \theta') = (\|V_r^H R_e[a(\theta) \cdot a^*(\theta')]\|^2)^{-1}, \quad (14)$$

$$(\hat{\theta}_k, \hat{\theta}_n) = \arg \max_{\theta' > \theta + \Delta\theta} p_r(\theta, \theta') \quad \text{for } k, n = 1, 2, 3, \dots, K; k \neq n \quad (15)$$

where the subscript r implies real.

3.1.3 The phase estimation

Using the notation $q = [\phi_1, \phi_2, \dots, \phi_M]$ and $F_k = \text{diag}\{a(\theta_k)\}$, the phase can be estimated as:

$$\hat{q} = \angle[z], \tag{16}$$

where $\angle[\cdot]$ represents the phase of complex number, and

$$z = \frac{Q^{-1}w}{(w^T Q^{-1}w)}, \tag{17}$$

$$Q = \sum_{k=1}^K F_k^H(\theta_k) V_o V_o^H F_k(\theta_k), \tag{18}$$

$$V_o = [u_{k+1}, \dots, u_M], \tag{19}$$

$$w = [1, 0, 0, \dots, 0]^T. \tag{20}$$

The proposed method in [9], when a uniform circular array with a radius more than $\lambda/4$ is used to estimate the 2-D spatial spectrum of the DOA, that is $(\hat{\theta}_1, \hat{\theta}_2)$, where $\hat{\theta}_1$ and $\hat{\theta}_2$ are not closed to each other, two pairs (one is a false peak while the other is the actual peak) of results were obtained, which means there exist DOA ambiguity. The authors in [9] stated that, to solve this problem you need to make the radius $\leq \lambda/4$ according to the theorem in [9]. Using the proposed method the radius of the circular array antenna can be equal to $\lambda/2$, without DOA ambiguity.

For an antenna consisting of four sensors with sensor coordinates: $(x_1, y_1) = (-l, 0)$, $(x_2, y_2) = (0, -l)$, $(x_3, y_3) = (l, 0)$ and $(x_4, y_4) = (0, l)$, where $l = \lambda/2$ and the DOAs of signal are defined in $[-90^\circ, 90^\circ]$. If the antenna is rotated by half the angle formed at the centre by two successive sensors (i.e, the coordinate become:

$$(x_1, y_1) = (-l/\sqrt{2}, -l/\sqrt{2}), \quad (x_2, y_2) = (l/\sqrt{2}, -l/\sqrt{2}), \\ (x_3, y_3) = (l/\sqrt{2}, l/\sqrt{2}), \quad (x_4, y_4) = (-l/\sqrt{2}, l/\sqrt{2}).$$

we found that, there will be no DOAs ambiguity.

To proof this consider four directions such that $(\theta_1, \theta_2) \neq (\theta_3, \theta_4)$, and $(\theta_1, \theta_2) \neq (\theta_4, \theta_3)$ from Equation (12), for complex-valued signals, the assumption that (θ_1, θ_2) and (θ_3, θ_4) are ambiguous (i.e. $p_c(\theta_1, \theta_2) = p_c(\theta_3, \theta_4)$) leads to following four different cases:

$$a(\theta_1) \odot a^*(\theta_2) = a(\theta_3) \odot a^*(\theta_4) \tag{21}$$

$$a(\theta_1) \odot a^*(\theta_2) = j(a(\theta_3) \odot a^*(\theta_4)) \tag{22}$$

$$a(\theta_1) \odot a^*(\theta_2) = -a(\theta_3) \odot a^*(\theta_4) \tag{23}$$

$$a(\theta_1) \odot a^*(\theta_2) = -j(a(\theta_3) \odot a^*(\theta_4)) \tag{24}$$

For sensors 2, and 4

$$\exp\left(\frac{-j2\pi l_{2,1}}{\lambda}\right) \odot \exp\left(\frac{j2\pi l_{2,2}}{\lambda}\right) = \exp\left(\frac{-j2\pi l_{2,3}}{\lambda}\right) \odot \exp\left(\frac{j2\pi l_{2,4}}{\lambda}\right), \tag{25}$$

$$\exp\left(\frac{-j2\pi l_{4,1}}{\lambda}\right) \odot \exp\left(\frac{j2\pi l_{4,2}}{\lambda}\right) = \exp\left(\frac{-j2\pi l_{4,3}}{\lambda}\right) \odot \exp\left(\frac{j2\pi l_{4,4}}{\lambda}\right), \tag{26}$$

For θ_k we have $l_{2,k} = x_2 \cos \theta_k - y_m \sin \theta_k$ and $l_{4,k} = -x_4 \cos \theta_k + y_4 \sin \theta_k$. It follows from Equations (25) and (26) that:

$$x_2(\cos \theta_1 - \cos \theta_2) - y_2(\sin \theta_1 - \sin \theta_2) = x_2(\cos \theta_3 - \cos \theta_4) - y_2(\sin \theta_3 - \sin \theta_4) + \lambda q_1, \tag{27}$$

$$-x_4(\cos \theta_1 - \cos \theta_2) + y_4(\sin \theta_1 - \sin \theta_2) = -x_4(\cos \theta_3 - \cos \theta_4) + y_4(\sin \theta_3 - \sin \theta_4) + \lambda q_2, \tag{28}$$

where q_1 and q_2 are integers.

Since

$$x_2 = x_3 = y_3 = y_4 = l/\sqrt{2}, \quad x_1 = x_4 = y_1 = y_2 = -l/\sqrt{2}$$

then the above two equations become:

$$(\cos \theta_1 - \cos \theta_2) + (\sin \theta_1 - \sin \theta_2) = (\cos \theta_3 - \cos \theta_4) + (\sin \theta_3 - \sin \theta_4) + \frac{\lambda q_1}{x_2}, \tag{29}$$

$$(\cos \theta_1 - \cos \theta_2) + (\sin \theta_1 - \sin \theta_2) = (\cos \theta_3 - \cos \theta_4) + (\sin \theta_3 - \sin \theta_4) + \frac{\lambda q_2}{y_4}. \tag{30}$$

Since

$x_k = \lambda/2$, $y_k = \lambda/2$, $|\theta_k| \leq 90^\circ$, $(\theta_1, \theta_2) \neq (\theta_3, \theta_4)$, and $(\theta_1, \theta_2) \neq (\theta_4, \theta_3)$, n_1 and n_2 are zero. From Equations (29) and (30), using sum-to-product identities, we can get:

$$\frac{\sin \frac{\theta_1 + \theta_2}{2}}{\cos \frac{\theta_1 + \theta_2}{2}} = \frac{\sin \frac{\theta_3 + \theta_4}{2}}{\cos \frac{\theta_3 + \theta_4}{2}}, \tag{31}$$

$$\left(\sin \frac{\theta_1 - \theta_2}{2}\right)^2 = \left(\sin \frac{\theta_3 - \theta_4}{2}\right)^2. \tag{32}$$

From Equations (31) and (32), we have

$$\theta_1 = \theta_3 + 2(n_1 + n_2)\pi, \tag{33}$$

$$\theta_2 = \theta_4 + 2(n_1 - n_2)\pi, \quad (34)$$

where n_1 and n_2 are integers. Since $|\theta_k| \leq 90^\circ$, $(\theta_1, \theta_2) \neq (\theta_3, \theta_4)$, $(\theta_1, \theta_2) \neq (\theta_4, \theta_3)$, Equations (33) and (34) cannot be simultaneously true. It follows that Equation (21) does not hold.

For sensors 2-4, Equations (22) leads to:

$$2\pi \left(\frac{x_2(\cos\theta_1 - \cos\theta_2) - y_2(\sin\theta_1 - \sin\theta_2)}{\lambda} \right) = 2\pi \left(\frac{x_2(\cos\theta_3 - \cos\theta_4) - y_2(\sin\theta_3 - \sin\theta_4)}{\lambda} \right) - \frac{4\pi n_1 + \pi}{2}, \quad (35)$$

$$2\pi \left(\frac{x_3(\cos\theta_1 - \cos\theta_2) + y_3(\sin\theta_1 - \sin\theta_2)}{\lambda} \right) = 2\pi \left(\frac{x_3(\cos\theta_3 - \cos\theta_4) - y_3(\sin\theta_3 - \sin\theta_4)}{\lambda} \right) + \frac{4\pi n_2 + \pi}{2}, \quad (36)$$

$$2\pi \left(\frac{-x_4(\cos\theta_1 - \cos\theta_2) + y_4(\sin\theta_1 - \sin\theta_2)}{\lambda} \right) = 2\pi \left(\frac{-x_4(\cos\theta_3 - \cos\theta_4) + y_4(\sin\theta_3 - \sin\theta_4)}{\lambda} \right) + \frac{4\pi n_3 + \pi}{2}, \quad (37)$$

where q_1 , q_2 and q_3 are integers. Since $x_2 = x_3 = y_3 = y_4 = l/\sqrt{2}$ and $x_1 = x_4 = y_1 = y_2 = -l/\sqrt{2}$, it's clear that Equations (35) and (37) conflicts with Equation (36). It follows that Equation (22) does not hold. Similarly, neither of Equations (23) nor (24) holds.

Therefore, in the case of complex-valued signals $p_c(\theta_1, \theta_2) \neq p_c(\theta_3, \theta_4)$ i.e. (θ_1, θ_2) and (θ_3, θ_4) are unambiguous. For real-valued signals, the assumption that (θ_1, θ_2) and (θ_3, θ_4) are ambiguous, i.e. $p_c(\theta_1, \theta_2) = p_c(\theta_3, \theta_4)$ leads to the following two different cases:

$$\text{Re}[a(\theta_1) \odot a^*(\theta_2)] = \text{Re}[a(\theta_3) \odot a^*(\theta_4)], \quad (38)$$

$$\text{Re}[a(\theta_1) \odot a^*(\theta_2)] = -\text{Re}[a(\theta_3) \odot a^*(\theta_4)]. \quad (39)$$

For sensors 2 and 4, Equation (38) leads to:

$$(\cos\theta_1 - \cos\theta_2) + (\sin\theta_1 - \sin\theta_2) = \pm [(\cos\theta_3 - \cos\theta_4) + (\sin\theta_3 - \sin\theta_4)] = \frac{\lambda q_1}{x_2}, \quad (40)$$

$$(\cos\theta_1 - \cos\theta_2) + (\sin\theta_1 - \sin\theta_2) = \pm [(\cos\theta_3 - \cos\theta_4) + (\sin\theta_3 - \sin\theta_4)] = \frac{\lambda q_2}{y_4}, \quad (41)$$

where q_1 and q_2 are integers, we obtain the same result as Equations (33) and (34). Thus, it follows that Equation (38) does not hold.

For sensors 2 and 4, Equation (39) leads to:

$$2\pi \left(\frac{x_2(\cos\theta_1 - \cos\theta_2) - y_2(\sin\theta_1 - \sin\theta_2)}{\lambda} \right) = \quad (42)$$

$$2\pi \left(\frac{x_2(\cos\theta_3 - \cos\theta_4) - y_2(\sin\theta_3 - \sin\theta_4)}{\lambda} \right) + 2\pi q_1 + \pi,$$

$$2\pi \left(\frac{x_3(\cos\theta_1 - \cos\theta_2) + y_3(\sin\theta_1 - \sin\theta_2)}{\lambda} \right) = \quad (43)$$

$$2\pi \left(\frac{x_3(\cos\theta_3 - \cos\theta_4) + y_3(\sin\theta_3 - \sin\theta_4)}{\lambda} \right) + 2\pi q_2 + \pi,$$

$$2\pi \left(\frac{-x_4(\cos\theta_1 - \cos\theta_2) + y_4(\sin\theta_1 - \sin\theta_2)}{\lambda} \right) = \quad (44)$$

$$2\pi \left(\frac{-x_4(\cos\theta_3 - \cos\theta_4) - y_4(\sin\theta_3 - \sin\theta_4)}{\lambda} \right) + 2\pi q_3 + \pi,$$

where q_1 , q_2 and q_3 are integers.

Since

$x_2 = x_3 = y_3 = y_4 = l/\sqrt{2}$ and $x_1 = x_4 = y_1 = y_2 = -l/\sqrt{2}$, it's clear that Equations (42) and (44) conflict with (43). It follows that (39) does not hold. Therefore, in the case of real-valued signals $p_c(\theta_1, \theta_2) \neq p_c(\theta_3, \theta_4)$, i.e. (θ_1, θ_2) and (θ_3, θ_4) are unambiguous.

The proposed method used the rotation technique to avoid DOA ambiguity. The steps of the proposed method can summarized as follows:

- 1) Rotate the uniform circular array antenna with angle $\delta/2$, so that $a(\theta_k)_{rotate} = a(\theta_k - \delta/2)$, where δ is the angle formed at the centre by two successive sensors.
- 2) Estimate the gain uncertain using Equation (10).
- 3) The DOAs of signals are estimated from Equation (13) in case of complex-valued signals or by Equation (15) in case of real-valued signals.
- 4) Estimate the phase uncertain using Equation (16), based on the estimated DOA.

4 Computer simulations

Consider a uniform circular array (UCA) composed of seven sensors, with the first sensor located at the origin. Two far-field narrowband signals are impinging on the array from directions $\theta_1 = 30^\circ$ and $\theta_2 = -20^\circ$, respectively. The range of the DOAs of the signals are defined in $[-90^\circ, 90^\circ]$. The unknown gain $\{a_m\}_{m=1}^M$ and unknown phase $\{\phi_m\}_{m=1}^M$ of the sensors are generated by $\alpha_m = 1 + \sqrt{12}\sigma_a C_m, \varphi_m = \sqrt{12}\sigma_\phi b_m$, where C_m and b_m are independent and identically distributed random variables distributed uniformly over $[-0.5, 0.5]$, while σ_a and σ_ϕ are the standard deviations of α_m and ϕ_m , and are equal

0.1 and 40° respectively. Assuming the powers of the signals at different directions are equal. Let $\Delta\theta=1^\circ$ for complex signals and 5° for real signals. The SNR and the sample number are 30 dB and 500, respectively. Figure 1 illustrates uniform circular array of seven sensors with radius $\lambda/2$, in which the first sensor is located at the origin. Its two-dimensional spatial spectrum $p_c(\theta_1, \theta_2)$ in the case of complex-valued signals is shown in Figure 2.

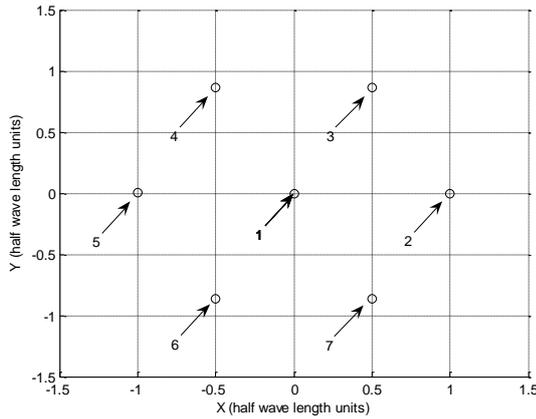


FIGURE 1 Uniform circular array of seven sensors with radius $\lambda/2$, the first sensor in the origin

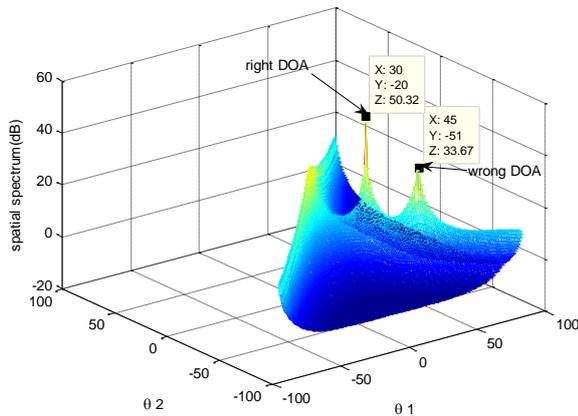


FIGURE 2 Two-dimensional spatial spectrum $p_c(\theta_1, \theta_2)$ in the case of complex-valued signals for Figure 1, $\theta_1 = 30^\circ$ and $\theta_2 = -20^\circ$

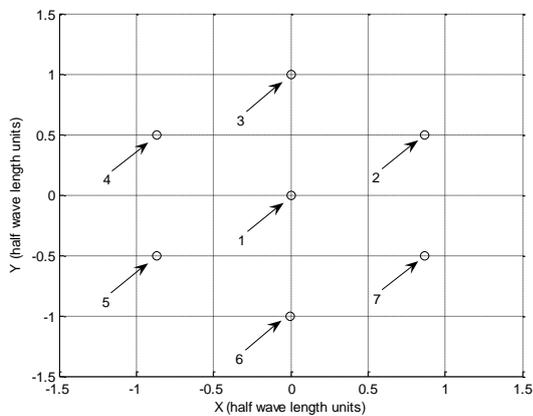


FIGURE 3 shows the uniform circular array antenna in Figure 1 after the rotation, $\delta/2 = 30^\circ$

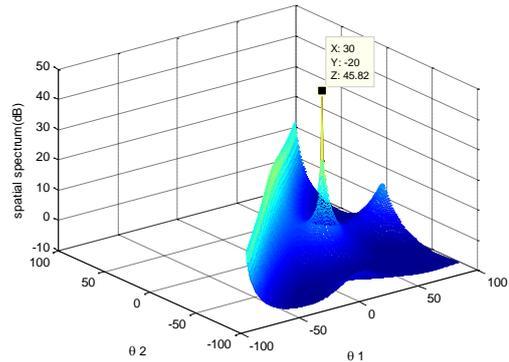


FIGURE 4 Two-dimensional spatial spectrum $p_c(\theta_1, \theta_2)$ in the case of complex-valued signals for Figure 3, $\theta_1 = 30^\circ$ and $\theta_2 = -20^\circ$

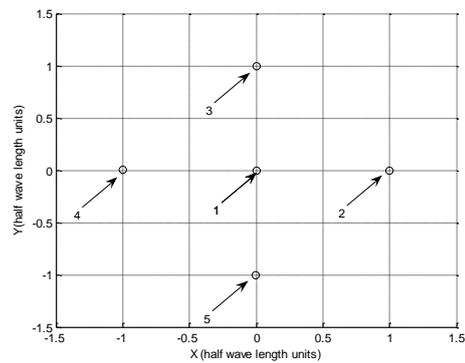


FIGURE 5 Uniform circular array of five sensors with radius $\lambda/2$, first sensor in the origin

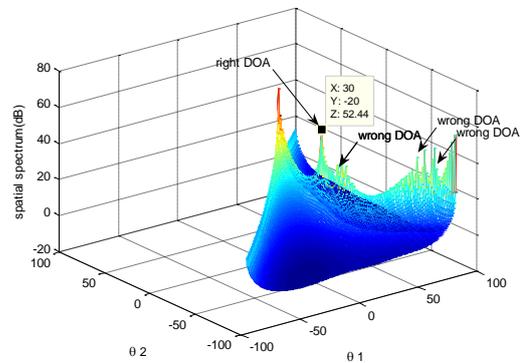


FIGURE 6 Two-dimensional spatial spectrum $p_c(\theta_1, \theta_2)$ in case of complex-valued signal for Figure 5, $\theta_1 = 30^\circ$ and $\theta_2 = -20^\circ$

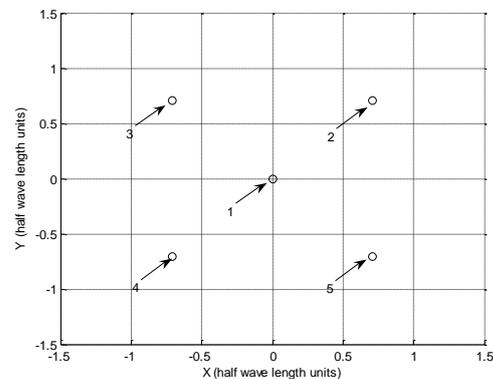


FIGURE 7 Shows the antenna in Figure (5) when rotated, $\delta/2 = 45^\circ$

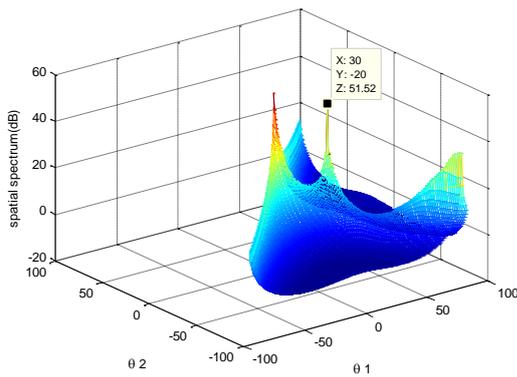


FIGURE 8 Two-dimensional spatial spectrum $p_s(\theta_1, \theta_2)$ in case of complex-valued signals for Figure 7, $\theta_1 = 30^\circ$ and $\theta_2 = -20^\circ$

From Figure 2, it is clear that there exists a false peak at point $(45^\circ, -51^\circ)$ beside the actual peak at point $(30^\circ, -20^\circ)$. This is because all the inter-sensor spacing is greater than $\lambda/4$ resulting in DOA estimation ambiguity.

We can avoid this false peak by rotating the antenna by $\delta/2 = 30^\circ$. Figure 3 shows the uniform circular array antenna in Figure 1 after the rotation. Figure 4 shows the corresponding spatial spectrum in the case of complex-valued signals. In Figure 4, it can be observed that there is only one peak in $(30^\circ, -20^\circ)$ and the false peak has been eliminated. This is due to the fact that rotating the antenna by an angle equal to half of the angle formed at the centre by two successive sensors leads to unambiguous DOA estimates, using the new DOA estimation method.

Figure 5 shows the case of a five sensors circular array antenna with radius $\lambda/2$ with first sensor at origin. Figure 6 shows its spatial spectrum in the case of complex-valued signals. From Figure 6 we can observe that, there are many peaks. However, by rotating the antenna by an angle of 45° we obtain only one peak as demonstrated next.

Figure 7 shows the antenna in Figure 5 when rotated, such that $\delta/2 = 45^\circ$. Figure 8 shows the corresponding spatial spectrum in the case of complex-valued signals. Figure 8 shows that there is only one peak in $(30^\circ, -20^\circ)$ and the false peaks are omitted.

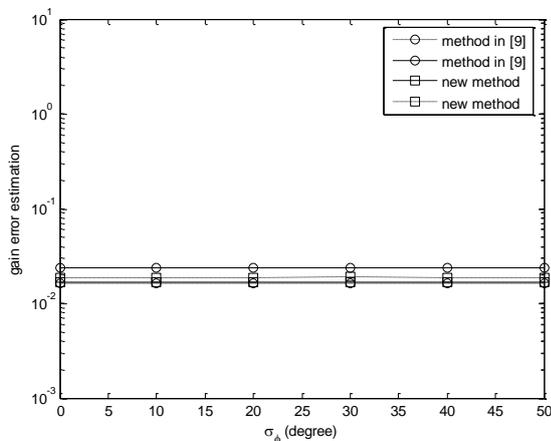


FIGURE 9a ARMSE of gain error estimates versus σ_ϕ (the dashed and solid plots represent the cases of complex-valued and real-valued signals, respectively)

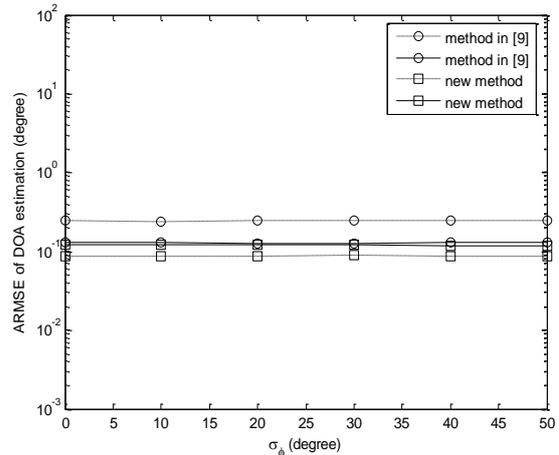


FIGURE 9b ARMSE of DOA error estimates versus σ_ϕ (the dashed and solid plots represent the cases of complex-valued and real-valued signals, respectively)

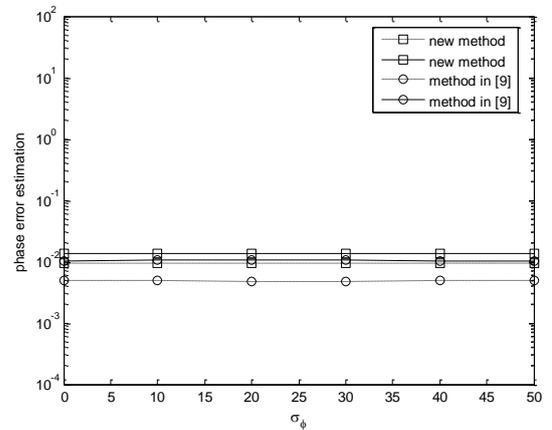
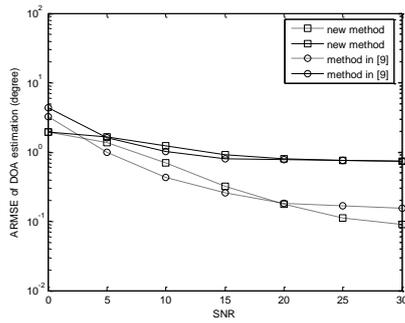


FIGURE 9c ARMSE of phase error estimates versus σ_ϕ (the dashed and solid plots represent the cases of complex-valued and real-valued signals, respectively)

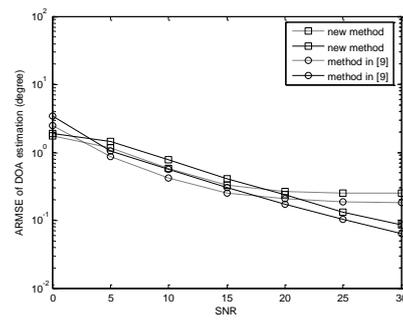
In order to examine the effect of phase error, SNR, and the number of snapshot, two signals were impinged from direction, -20° and 20° respectively. In each case the number of samples was 200 while all other simulation parameters were the same as in the previous experiment.

The effect of phase error was studied based on 500 experiments. The average root mean square error (ARMSE) [18] curves of the gain error, the DOA error, and the phase error estimates versus the standard deviation of the phase error σ_ϕ are shown in Figures 9(a-c) respectively. The Figures show that the performance of both the new method and that of the method in [9] are independent of phase errors and are approximately equal to each other.

Similarly, the effect of SNR based on σ_ϕ experiments. Figures 10(a) and (b) show the ARMSE of the DOA estimate versus the SNR, when σ_ϕ equals 5° and 50° respectively. Figure 10a shows that the performance of the high phase error work as same as small error.

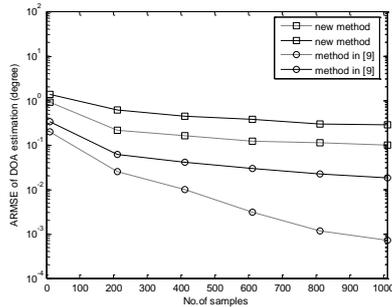


a) ARMSE of DOA error estimates versus $\sigma_\phi = 5^\circ$

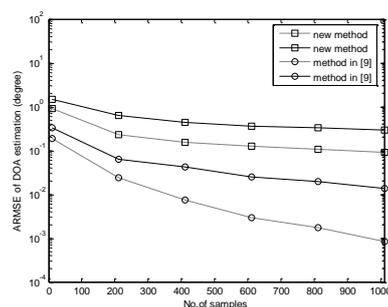


b) ARMSE of DOA error estimates versus $\sigma_\phi = 50^\circ$

FIGURE 10 ARMSE of DOA error estimates versus SNR, σ_ϕ (the dashed and solid plots represent the cases of complex-valued and real-valued signals, respectively)



a) ARMSE of DOA error estimates versus $\sigma_\phi = 5^\circ$



b) ARMSE of DOA error estimates versus $\sigma_\phi = 50^\circ$

FIGURE 11 ARMSE of DOA error estimates versus number of samples, σ_ϕ (The dashed and solid plots represent the cases of complex-valued and real-valued signals, respectively)

Also, the effect of sample number based on 500 experiments. Figures 11a and 11b show the ARMSE of DOA estimates versus the number of samples when σ_ϕ equals 5° and 50° respectively. Figure 11a shows that the new method has worse performance than method in [9]. Also the new method and the method in [9] perform better as the number of samples increases. Figure 11b shows that the performance of both methods remain unchanged as the phase error increases.

5 Conclusion

In this paper, we considered the problem of DOA estimation in the presence of gain and phase errors. By improving the dot product method, we propose a method for simultaneously estimating the DOA and gain-phase errors. Therefore, the proposed method overcomes the disadvantage of the Dot product method, which fails in UCA with radius more than $\lambda/4$. The problem of DOA ambiguity can be solved by rotating the antenna with an

angle equal to half the angle formed at the centre by two successive sensors. In addition, the proposed method is independent of phase errors, and its performance is almost the same as that of the Dot product method. The disadvantage of the new method is that it is only applicable in UCA, and not applicable in ULA (since there is no angle between two successive sensors).

6 Acknowledgements

The authors would like to thank the laboratory's colleagues of UESTC Ensemble for acting support, for providing us their data, Professor Chen for help to teach me how to use the software of data analysis and processing. Also thank professor Xie for helping improve the paper.

This work was also supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ131110), and Key Laboratory of signal and information processing Chongqing Three Gorges University.

References

- [1] Schmidt R O 1986 *IEEE Transactions on Antennas Propagation* 34(3) 276-80
- [2] Roy R. and Kailath T 1989 ESPRIT –*IEEE Transactions on Acoustics, Speech, Signal Processing*, July 1989 37(7) 984-95
- [3] Capon J 1969 *Proceedings of the IEEE* 57(8) 1408-18
- [4] Birot G, Albera L, Chevalier P 2010 *IEEE Transactions on Signal Processing* 58(8) 4144-55
- [5] Stoica P, Sharman K C 1990 *IEEE Transactions on Acoustics, Speech, Signal Processing* 38(7) 1132-43
- [6] Weiss A J, Friedlander B 1990 Eigenstructure Methods for Direction Finding with Sensor Gain and Phase Uncertainties *Circuits Systems Signal Processing* 9(3) 271-300
- [7] Paulraj A, Kailath T 1985 *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'85* 10 640-3
- [8] Wylie M P, Roy S, Messer H 1994 *IEEE Transactions on Signal Processing* 42(12) 3449-59
- [9] Liu A, Liao G, Zeng C, Yang Z, Xu Q 2011 *IEEE Transactions on Signal Processing* 59(12) 5944-56

[10]Heidenreich P, Zoubir A.M, Rubsam M 2012 *IEEE Transactions on Signal Processing* **60**(9) 4683-93

[11]Vorobyov S A, Gershman A B, Wong K M 2005 *IEEE Transactions on Signal Processing* **53**(1) 34-43

[12]Friedlander B, Weiss A J 1995 *IEEE Transactions on Signal Processing* **43**(7) 1557-67

[13]Soon V C, Tong L, Huang Y F, Liu R *IEEE Transactions on Signal Processing* **42**(4) 973-6

[14]Fuhrman D R 1994 *IEEE Transactions on Signal Processing* **42**(1) 77-7

[15]Ng B C, See C M S 1996 *IEEE Transactions on Antennas Propagation* **44**(6) 827-35

[16]Friedlander B, Weiss A J 1991 *IEEE Transactions on Antennas Propagation* **39**(3) 273-84

[17]Li Y, Er M H 2006 *IEEE Transactions on Signal Processing* **54**(2) 712-23

[18]Liu A, Liao G, Ma L, Xu Q 2010 *IEEE Geoscience Remote Sensing Letters* **7**(4) 731-5

Authors	
	<p>Xiaowei Niu, born in 1978, Luoyang, Henan, China</p> <p>University studies: B.E. degree in electrical engineering, in 1998. M.E. degree in signal processing from the Southwest University, Chongqing in 2008.</p> <p>Scientific interest: signal processing and detection theory, intelligence optimal algorithm, and pattern recognition theory including affective computing.</p>
	<p>Liwan Chen, born in 1964, Kaixian, Chongqing, China</p> <p>University studies: B.E. degree in electrical engineering, in 1984. M.E. degree in electrical engineering from Chongqing University, in 2007.</p> <p>Scientific interest: electrical engineering and signal processing.</p>
	<p>Qiang Chen, born in 1979, Hechuang, Chongqing, China</p> <p>University studies: B.E. degree in electrical engineering, in 1998. M.E. degree in signal processing from Xidian University, in 2011.</p> <p>Scientific interest: electrical engineering and electronic design automation</p>
	<p>Hui Xie, born in 1969, Wanzhou, Chongqing, China</p> <p>University studies: B.E. degree in electrical automation, in 1989. M.S. degree in Theory and New Technology of Electrical Engineering from the Chongqing University, Chongqing in 2007.</p> <p>Scientific interest: industrial intelligent control, device and signal processing, detection theory.</p>
	<p>Hongbing Li, born in 1981, Tongnan, Chongqing, China</p> <p>Current position, grades: PhD. student majoring in control theory and control engineering in Chongqing University.</p> <p>University studies: B.E. degree in Electronic information engineering from Chongqing Three Gorges University in 2003. M.S. degree in signal and information processing from Chongqing University of Technology, Chongqing, in 2011.</p> <p>Scientific interest: intelligent signal processing, wireless sensor networks.</p>

Multiple DAGs reliability model and fault-tolerant scheduling algorithm in cloud computing system

Weipeng Jing^{1, 2*}, Yaqiu Liu^{1, 2}

¹The College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China

²Heilongjiang province engineering technology research centre for forestry ecological big data storage and high performance (cloud) computing, Harbin 150040, China

Received 1 March 2014, www.tsi.lv

Abstract

In this paper, in order to provide the reliable scientific workflow scheduling problem for cloud computing, a dynamic of RANK-Hierarchical algorithm is put forward which taking account of communication contention as well as supporting task dependencies (CCRH). A communication contention model is first defined, as soon as the earliest completion of the primary and backup task is deduced, besides the executive processor is limited, use dynamic hierarchical method and calculate of each DAG unfair degree factor for multiple DAGs scientific workflow. It can deal with the problem that multiple DAGs workflow comes at different time and have various kinds of structure. Both the theory and experiments have proved the algorithm not only improve the scheduling fairness of multiple DAGs workflow but also shorten the average execution Makespan effectively while meeting reliability constraints and meanwhile the produce well robustness.

Keywords: cloud computing, multiple DAGs, RANK-Hierarchical, reliability

1 Introduction

Cloud computing as a new computing model gets more and more attention. It is integrated by a variety of distributed computing, storage and application resources, meantime, realized multi-level virtualization and abstraction. It is efficient to make large-scale network resources form provide to user in reliable way [2]. Cloud computing as the next generation computing model plays an important role on scientific computing and commercial computing, it has been concerned by current academia and the business community. Now some typical cloud computing has been appeared, such as Google Cloud [1], Microsoft Cloud [14], Amazon EC2 [15] and IBM Cloud [16], these systems are committed to achieve web search, social network based on cloud computing.

In these fields of scientific computing applications, such as high-energy physics, astronomy, polymer materials, earth sciences, forestry resources and so on, due to huge task of data need to deal with, cloud computing system can provide powerful computing support. Great relevance and priority constraint relationship that may exist between the type of application computing tasks, so it should be on-demand dynamically provision, configuration, reconfigure and deprivation computing resource services in the cloud computing environment to achieve cloud computing scientific workflows high scalability and availability. The aim of resource scheduling is to achieve calculation, collection of storage resources and scheduling tasks to meet the relationship between

spatial and temporal effectively. The Traverna [23], ASKALON [19], VGrADS [32], Pegasus [24] respectively to achieve distributed computing, scheduling management of storage resource.

In recent years, due to the dynamic expansion of cloud computing, high availability, resources assigned according to the need, some projects used cloud computing platform manage scientific workflow have been emerged. Such as the Amazon EC2 [15] can provide scalable, reliable, service-on-demand computing and storage services on scientific computing applications. Literature [30,31] describes the scientific workflow applied on the Amazon cloud platform's runtime and energy costs; addition, the ASKALON [19] and VGrADS [32] have been started to support scientific workflow applied on cloud computing platform.

In heterogeneous distributed environments (cloud computing, grid systems), use the DAG to describe task relationship for scientific workflow applications. The DAG workflow scheduling algorithm is divided into static scheduling algorithm and dynamic scheduling algorithm. The static scheduling algorithm is that assumption the overall structure and precedence constraints are known, execution time of the task can be calculated. So resources are allocated before the execution of the task, then no longer be adjusted. The dynamic scheduling algorithm can allocate resource dynamically based on workflow changed in the task execution process. Literature [3, 4] proves the static scheduling algorithm is better than the dynamic scheduling algorithm in different angles.

* *Corresponding author* e-mail: 39750600@qq.com

In the cloud computing scientific workflow applications, which faced mass intensive data processing performance:

1) in the scheduling process of cloud computing scientific workflow, existing multiple DAG submitted at the same time or submitted dynamically during calculation process, therefore it is required to the scheduling algorithm can meet the changes in dynamic environment;

2) in the enterprise-class workflow applications, it is need to set the trusted protection mechanism to tolerate failures when system is running;

3) the reasonable scheduling mechanism, it should to guard user that submit scientific computing request has little effect on data centre's load and position.

Therefore the dynamic scheduling algorithm which has a high reliability fault-tolerant ability is important in cloud computing to meet the users demand for dynamic tasks submitted. This paper provides solutions to the above questions by proposing an innovative dynamic of RANK-Hierarchical scheduling algorithms to maximize the performance and reliability.

The rest of this paper is organized as follows. In Section 2, we present system and mathematical models. As part of the system model, we design the processor model, task model, communication links model, the task priority to identify and underlying assumptions. In Section 3, we propose the multi-DAG scheduling algorithm CCRH. Simulation results show that proposed algorithms improve the performance compared to reference algorithms by varying number of DAGs parameters in Section 4. Prior related works are compared in Section 5 Finally, in Section 6, we conclude the paper by summarizing the comparison results and future work.

2 System model

In this section, we introduce a scheduling model in cloud computing provider, which consists of processor, tasks and communication link. In cloud computing, the parallel tasks of scientific workflow applications can use the weights of nodes and edges to represent a directed acyclic graph (DAG), the following is formal definition:

Definition 1: Node and weights of side of the DAG Figure can use the four-array $G = (V, E, w, c)$, which $V = \{v_1, v_2, v_3 \dots v_N\}$ represents the number of tasks, $E = \{e_{ij} | v_i, v_j \in V\}$ represents the communication edge combination of dependency between tasks, $w(v_i)$ represents computational cost of the task, $c(e_{ij})$ represents communication cost between v_i and v_j

Definition 2: The collection $\{v_x \in V : e_{xi} \in E\}$ represents the task of v_i 's predecessor node set, denoted by $pred(v_i)$. The collection $\{v_x \in V : e_{ix} \in E\}$ represents the task of v_i 's successor node set, denoted by $succ(v_i)$. If $pred(v_i) = \emptyset$, the task node v_i is the entry node,

expressed as v_{entry} . If $succ(v_i) = \emptyset$, the task node v_i is the exit node, expressed as v_{exit} .

Definition 3: cloud computing environment use a variety of heterogeneous computing platforms built environment, set of heterogeneous processors described as $P = \{P_1, P_2 \dots, P_M\}$, M , which indicates the number of heterogeneous processors. The processor P_k on the task v_i of primary's start time and completion time represent as $t_s^p(v_i, p_k)$, $t_f^p(v_i, p_k)$ respectively; the task v_j of backup's start time and completion time represent as $t_s^b(v_j, p_k)$, $t_f^b(v_j, p_k)$ respectively. Primary and secondary version of the task v_i scheduling processor represent as $P^p(v_i)$ and $P^b(v_i)$ respectively.

Definition 4: the cloud computing system is network structure of any interconnection. Denote any processor's communication between P_h and P_k as ℓ_{hk} .

In order to better illustrate the problem, we make the following assumptions:

1) The CPU time used by the task switching and process scheduler is negligible;

2) At the same time only exists one processor failure, it is impossible existing two processor failure at the same time. And fault according to the Poisson distribution;

3) The failure of the processor is fail-stop mode, the processor status is normal or failure to stop, while ignoring the fault detection time;

4) The multi-DAG workflow tasks arrive at any moment randomly in order to meet user's demand in cloud computing environment.

5) The communications link of cloud computing system is the arbitrary interconnection duplex structure, task communication only allow the same direction in the same time.

3 Multi-DAG scheduling algorithm CCRH

Priority of static scheduling method is the key to determine task priority, so computation of task priority has efficient impact on scheduling algorithm. The priority determination method of HEFT [18] algorithm is a widely typical algorithms applied to the actual. For example, the ASKALON [19] system also applied HEFT algorithm, and to prove the validity of scheduling DAG.

In the scientific workflow applications of cloud computing, as the DAG task reached dynamically, therefore a static priority method of calculation the multiple DAG mission priority cannot be used, this paper proposes a DAG scheduling algorithm mining dynamic and static, dynamic scheduling algorithm processes dynamically reached DAG task on the hierarchical, while static points to single DAG scheduling tasks in accordance with the static method.

3.1 CHECKING THE PDF FILE

Single DAG task priority uses static scheduling method, the algorithm set tasks v_i and v_j with dependent manner inspired by HEFT [18], and v_j run directly dependent on the operating results of v_i . Considering the computing and communication's general consumption computed tasks priority:

$$rank(v_i) = \overline{w(v_i)} + \max_{j \in succ(v_i)} \{\overline{c(e_{ij})} + rank(v_j)\}, \quad (1)$$

where $\overline{w(v_i)}$ represents average execution costs on all processors of task v_i , $\overline{c(e_{ij})}$ represents average communication cost between task v_i and v_j . The priority of task v_i is the largest value that plus direct successor task priority and communication with its own computational cost. Priority of all tasks is traversing the task graph upward from export task; the export task's priority is defined as:

$$rank(v_{exit}) = w(v_{exit}). \quad (2)$$

This paper considers the competition of communication link, the basic idea of the scheduling in the communication section is treat the computing nodes and communication in the DAG equally. Therefore, communication contention scheduling algorithm should not only consider the processor scheduling, but also consider the communication link scheduling among processors [25, 26]. In order to better solve the link communication competition, this article uses any cloud computing environment, network interconnect heterogeneous computing system communication path to find the shortest path search algorithm based on insertion strategy [26]. Defined $LST(e_{ij}, \ell)$, $LFT(e_{ij}, \ell)$ as e_{ij} communication start time and completion time in the communication link ℓ and: $LFT(e_{ij}, \ell) \geq LST(e_{ij}, \ell) + c(e_{ij})$.

3.2 PRIMARY AND BACKUP TASKING SCHEDULING

In order to improve the reliability of the cloud computing system, in this paper the primary and backup scheduling method is used to achieve fault tolerance which performs redundant tasks in the backup processor, while ensuring the real-time nature of the task. And in order to improve system performance, this paper uses overlapped primary and backup tasks to determine the earliest start time of primary and backup task.

3.2.1 The primary task scheduling

First the primary task is consider to schedule, according to the backup completion time of the set of predecessor

$pred(v_j)$, the start time of the primary task v_j has following three situations:

$$1) t_s^p(v_j, p) > \max_{v_i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\},$$

The start time of the primary task v_j is greater than the maximum of the latest completion time of the backup task set $pred(v_j)$ and data transmission time in the communication link, then if the processor where the primary of any task has failed, the task v_j can successfully receive the message which sent by all the predecessor task.

$$2) t_s^p(v_j, p) > \max_{v_i \in pred(v_j)} \{t_f^p(v_i, p), LFT(e_{ij}, \ell)\} \quad \text{and}$$

$$t_f^p(v_j, p) < \max_{v_i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\}.$$

The start time of the primary task v_j is less than the maximum of the latest completion time of the backup task in the task set $pred(v_j)$ and data transmission time in the communication link. In this case, if the start time of the backup task is less $\max_{v_i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\}$, then when the processor where the task v_i fails, the task v_j cannot successfully receive the message sent by the entire predecessor task, and cannot get the right results. Therefore, the start time of the backup task v_j is greater than $\max_{v_i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\}$ that meet the fault tolerance of the system.

$$3) t_s^p(v_j, p) > \max_{v_i \in pred(v_j)} \{t_f^p(v_i, p), LFT(e_{ij}, \ell)\},$$

$$t_s^p(v_j, p) < \max_{v_i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\} \quad \text{and} \quad \text{the completion time}$$

$$t_s^f(v_j, p) > \max_{v_i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\}.$$

The start time of the primary task v_j is greater than the maximum of the latest completion time of the backup task set $pred(v_j)$ and data transmission time in the communication link, and the completion time is less than the maximum completion time of the backup.

In CCRH, when the algorithm schedules the primary of the different DAG task, the primary task can be considered to be independent and non-priority task, and its constrains of independent scheduling is only with their own priority, looking for the processor of the earlier start to complete task based on scheduling processor queue.

3.2.2 The backup task scheduling

In this section, we analysis the earliest start time of executing the backup task v_j . First we define the constraints of scheduling the primary task. When the schedule of the primary task v_j meet the state (1) or (3), the start time of its backup task must meet:

$$t_s^B(v_j, p_h) > t_s^P(v_j, p_k). \quad (3)$$

When the schedule of the primary task v_j meet the state (2), the start time of its backup task must meet:

$$t_s^B(v_j, p_h) > \max_{v_i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\}. \quad (4)$$

In order to achieve the fault-tolerant of the system, the processor of Cloud computing environment be scheduled by the backup task v_j also need to meet the processor constraint: When the schedule of the primary task v_j meet the state (1), the processor scheduled by its backup task meet:

$$P^B(v_j) \notin \bigcup_{v_i \in pred(v_j)} P^P(v_i). \quad (5)$$

When the schedule of the primary task v_j meet the state (2) or (3), $pred(v_j)^2$ represents the task set that meeting the state (2) or (3) in set $pred(v_j)$, $pd(v_j)$ represents the task set that exist indirect and direct dependence and meet the state (2) or (3) of the task v_j :

$$pd(v_j) = \{pred(v_j)^2\} \cup \bigcup_{v_i \in pred(v_j)^2} pd(v_i). \quad (6)$$

So

$$P^B(v_j) \notin \{ \bigcup_{v_i \in pred(v_j)} P_{pd}^P(v_i) \} \cup \{ \bigcup_{v_i \in pred(v_j)} P^P(v_i) \}. \quad (7)$$

Here $P_{pd}^P(v_i)$ represents the processor where all primary task in $pd(v_i)$.

Lemma: the earliest start time of the backup task v_j is

$$\max_{i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\} \text{ and its backup task cannot be dispatched to the processor(the virtual machine) } \{ \bigcup_{v_i \in pred(v_j)} P_{pd}^P(v_i) \} \cup \{ \bigcup_{v_i \in pred(v_j)} P^P(v_i) \}.$$

Proof: assume that the start time of the backup task v_j is less than $\max_{i \in pred(v_j)} \{t_f^B(v_i, p), LFT(e_{ij}, \ell)\}$, then when the processor where the backup of $pred(v_j)$ complete time last failure, need to execute the backup task, v_j will not be able to receive the messages sent, and task v_j fail. Assume the backup task v_j is scheduled to the processor $\{ \bigcup_{v_i \in pred(v_j)} P_{pd}^P(v_i) \} \cup \{ \bigcup_{v_i \in pred(v_j)} P^P(v_i) \}$, then when the processor $P^P(v_j)$ fail, and the malfunction of its precursor node do not recover, then the backup task v_j will not run properly. So the assumption is not true.

The goal of scientific workflow task scheduling in cloud computing environment is getting the earliest

completion time (*Makespan*) of the task. The earliest completion time of all DAG tasks is the exit node completion time of the backup task.

$$Makespan = t_f^B(v_{exit}, p). \quad (8)$$

The earliest start time of CCRH looking for is calculating the earliest completion time of the backup task in scheduling strategy of the entire task.

3.3 MULTI-DAG HIERARCHICAL SCHEDULING

In DAG scheduling model of the cloud computing system, as the DAG workflow a will compete with other DAG workflow for the same set of computing resources, so the *Makespan* (the time from submit DAG a to finish the last task) of the workflow a is likely longer than the *Makespan* which it use the cloud computing environment separately, these two *Makespan* can be represented as $M_{multi}(a)$ and $M_{own}(a)$ separately. Literature [18] *Slowdown* is described this ratio: $Slowdown = M_{multi}(a) / M_{own}(a)$, so the inequities factor *Unfaines(s)* of a scheduling algorithm s is defined as :

$$Unfaines(s) = \sum_{\forall a \in A} |Slowdown(a) - AvgSlowdown|, \quad (9)$$

where A is a set of being given multi-DAG. *AvgSlowdown* is the average of *Slowdown* of all DAG,

$$\text{i.e. } AvgSlowdown = \frac{1}{|A|} \sum_{\forall a \in A} Slowdown(a), \quad |A| \text{ represent}$$

the base of set A , *Unfaines(s)* is an important indicator that be used to measure the unfair degree of multi-DAG scheduling algorithm.

In the literature [7], the method of the multi-DAG task scheduling is: sorting the new task and the remaining tasks in DAG ascending according to the weight. If the weight of the new DAG task is always less than that the remaining DAG tasks, then the new DAG task is not scheduled, which will lead that the new DAG task cannot be scheduled as the weight.

So this paper proposes the multi-DAG scheduling method based on layer, the basic principle is to stratify the every DAG arrived in cloud computing environment any time, and the every layer of the last DAG is merged to that one where the DAG task do not be performed. Then it sorts every layer ascending on the basis of the task weight. So it will avoid the problem of the time span increasing due to the remaining task of previous DAG not be scheduling.

The concrete steps are as follows:

1) To stratify each DAG task in the scientific workflow processing.eg: in the 0 moment DAG-A arrives, then DAG-A_i(i=1,2...m) represent the i-th layer of the DAG-A.

2) To calculate the priority weight of all the tasks in every DAG according to the formula.

3) If it is single DAG, sort the task descending base on the priority weight, and submit to the schedule queue, then schedule in turn. The order is the primary task first and then the backup. Otherwise go to step (4).

4) If it is multi-DAG, merge the first layer of the last DAG to the next layer of the current DAG task.eg: in the t moment, when DAG-B arrives, the DAG-A_i task is performing, then put the DAG-B₁ task to DAG-A_{i+1} layer.

5) The tasks of every layer sort ascending according to the weight, e.g.: the tasks of DAG-A_{i+1} sort ascending according to the weight, submit the schedule queue, then schedule the task in turn. The order is also the primary task first and then the backup.

6) Calculate unfair degree factor $Unfairness(s)$ according to the Equation (9), and sort ascending. Schedule task to the processor low $Unfairness(s)$ priority.

7) If the unfair degree factors $Unfairness(s)$ of the multiple DAG equal, then schedule tasks in turn according to the *Makespan*.

4 Test results and analysis

The simulation of the algorithm is compared with HEFT [18], BMCT [17] in the fairness of the fair factor scheduling, the scheduling time, the processor utilization and the task running time(the multi-DAG task of HEFT adopt the same way of stratifying), and compared with MaxAR [29] in robustness. In order to reflect the advantage of the algorithm in the scientific workflow better, we use four types of DAG task: random DAG task, FFT, Laplace and Fork-join, in which every type of the DAG contains 2-10 DAG task, and every DAG contains 10-50 task.

The experimental environment has Inter®Xeon E7420 2.13GHz, RAM 4G, the cloud computing environment of 1T hard disk. And use CCR to describe the ratio of communication and computing in DAG task graph, the value of the CCR select random number in 0.1-1.

4.1 THE FAIRNESS

Compare the fairness of CCRH, HEFT and BMCT algorithm for different scientific workflow DAG Figure. Figure 1 a-d represent the fairness of the three algorithms in the random DAG Task FFT, Laplace, Fork-join graph respectively. The HEFT and BMCT using the same layered approach, its fairness do not have much difference. As CCRH use dynamical method, its fairness has improved greatly, but do not appear larger hopping phenomenon.

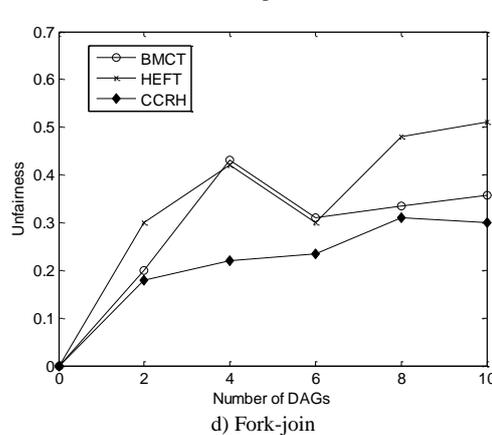
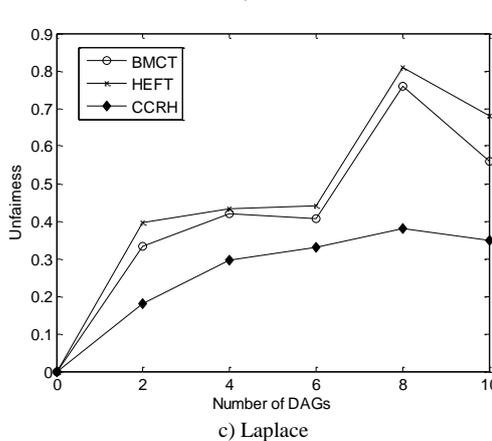
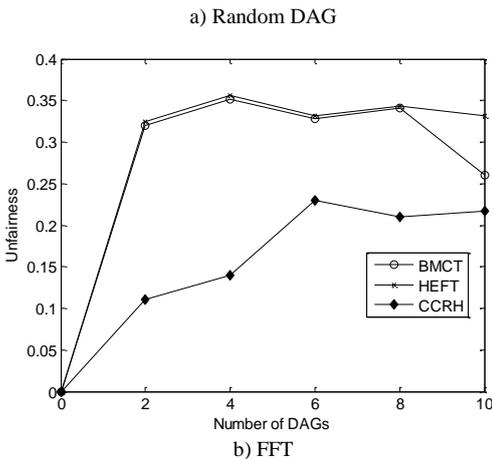
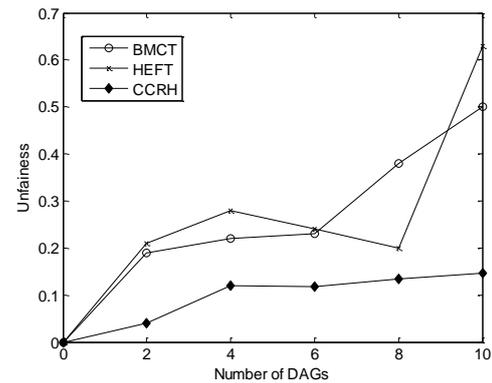


FIGURE 1 The Comparison of algorithm fairness

4.2 THE MAKESPAN ALGORITHM

Compare the *Makespan* of CCRH, HEFT and BMCT algorithm for different scientific workflow DAG Figure. Figures 2 a–d represent *Makespan* in random DAG Tasks, FFT, Laplace, Fork-join respectively on three algorithms. There is little difference can be seen from Figure 2 in the

three algorithms' *Makespan*. The main reason is that the three algorithms select a similar-priority comparison algorithm. BMCT is better than HEFT as BMCT considers communication constraints between tasks and as CCRH adopt the technology of primary and secondary version to improve system reliability, but its backup algorithm increases its *Makespan*.

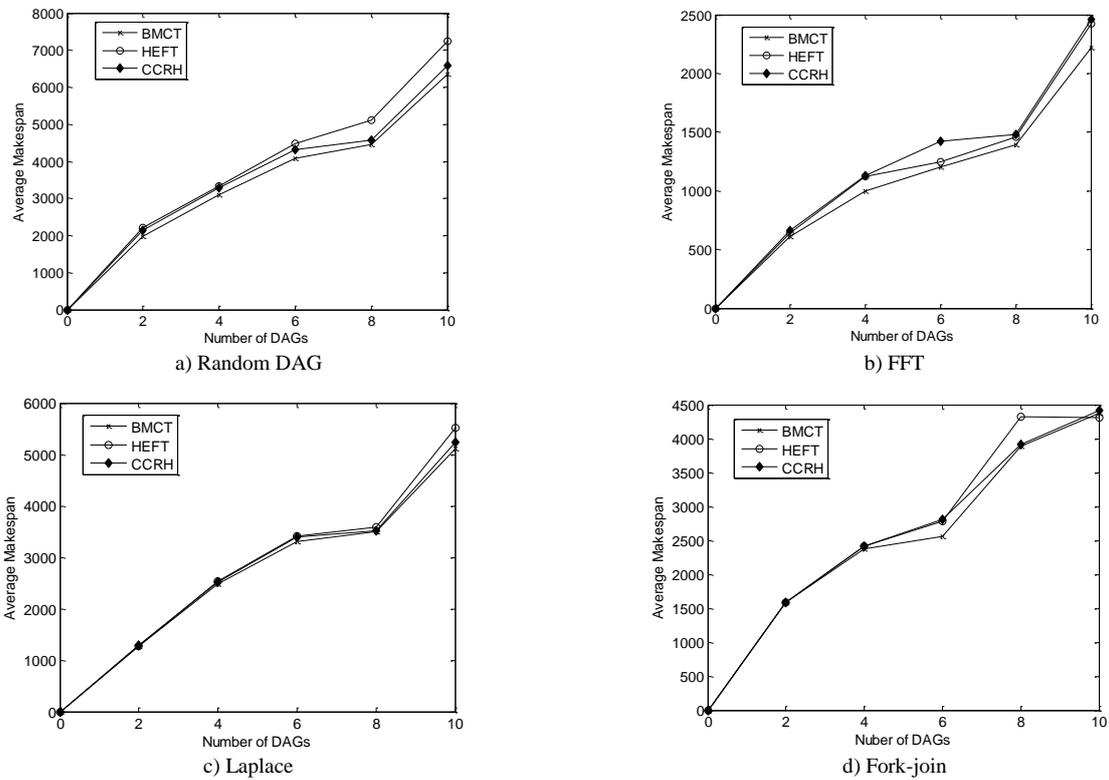


FIGURE 2 The Comparison of average Makespan algorithm

4.3 THE LARGE-SCALE DATA COMPUTING TIME

In order to better test the overall performance of algorithm, by analyzing running time of 100 DAG in randomly generated environment. It is can be seen that HEFT performance the best of three algorithms, BMCT is poor. The CCRH use the technology of primary and secondary versions to improve the reliability, meanwhile expense its running time.

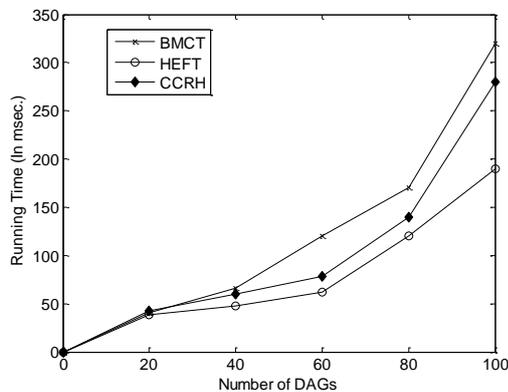


FIGURE 3 The Comparison of resource utilization

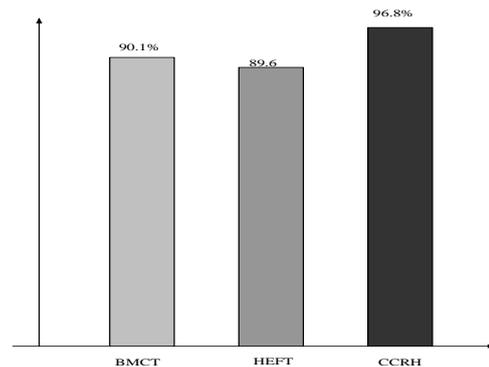


FIGURE 4 The large – scale data computing time

4.4 FOOTNOTES THE RESOURCE UTILIZATION

In order to better reflect utilization of the algorithm on cloud environment resources, we research the average utilization rate of the processor at four different scientific workflow loads. It is can be seen to from Figure.4 CCRH has a higher processing utilization. Thus CCRH have better benefits in the cloud environment that resource usage accounting

4.5 THE ROBUST OF ALGORITHM

In view of existing research for DAG task's scheduling algorithm, there is little on the robust of algorithm itself. The literature [28] tests 20 kinds of heuristic scheduling algorithm's robustness by introducing the standard deviation of Makespan. The literature and existing results assuming that the scheduler can obtain the information of the computing nodes at any time, and in practical applications, especially in large-scale cloud computing platforms, the collection of node information is affected by the competition in the communication link, node load factors, literature [29] proposed a novel robustness test method, which can be an measure the performance of algorithm effectively.

In order to reflect the scheduling model based on communication competitive advantage, in the simulation test, we assume that the scheduler compute nodes information in every 2 milliseconds. Testing the robustness of the algorithm on three parts: degradation approximation factor, average wait time and decay degree of critical path. Compare with the literature [29] MaxAR

algorithm. The approximation factor defined as: $\eta = \frac{t_{max}}{t_{min}}$,

here $t_{max} = \max_{i=1...N}^{k=1...M} t_f^p(v_i, p_k)$ is the maximum completion time of the backup task v_i in the processor set, $t_{min} = \min_{i=1...N}^{k=1...M} t_f^p(v_i, p_k)$ is the minimum completion time of the backup task v_i in the processor set; average waiting time of the critical path is $\overline{cpw} = \frac{1}{N} \sum_{i=1}^N (LFT(e_{ij}, \ell) - t_f^B(v_i, p_k))$; the average attenuation of the critical path is defined as follows:

$cps_i = 1 + \frac{LFT(e_{ij}, \ell) - t_f^B(v_i, p_k)}{t_f^B(v_i, p_k)}$. It can be seen from

Figure.5 due to use the tolerant mechanism of primary and secondary version, the CCRH performs the worst performance on approximate, but as CCRH can calculate the optimal start time of current task, thus the approximate factor attenuates in an acceptable range.

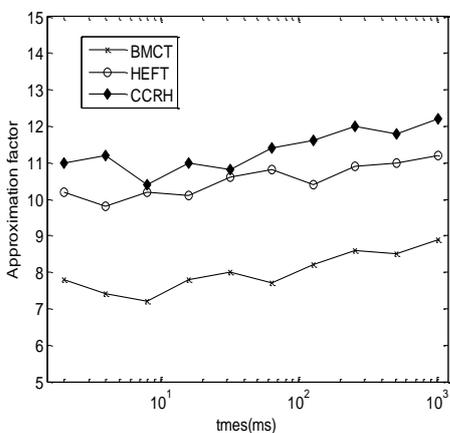


FIGURE 5 The comparison of approximate factor

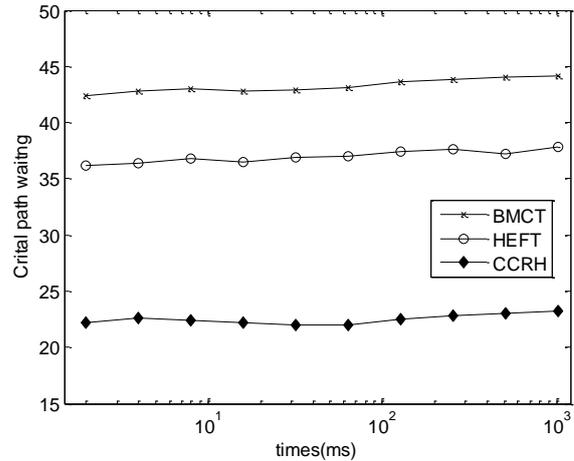


FIGURE 6 The average waiting time of critical path

Figure 6 shows the changes of critical path on average waiting time in the node information within the update interval. As CCRH fully considers link communications competition, its waiting time is more accurately reflect the algorithm to obtain the actual performance.

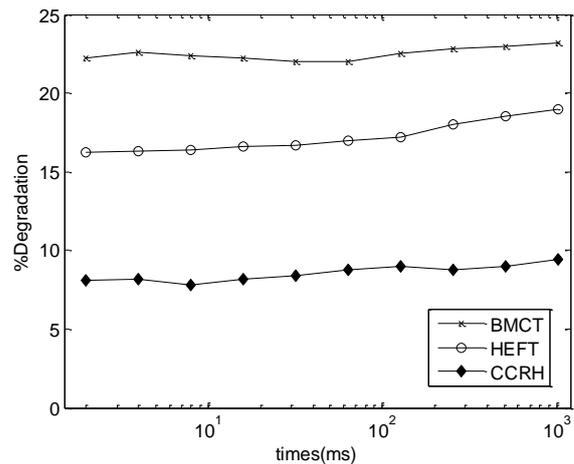


FIGURE 7 The average attenuation of critical path

It is can be seen from Figure.7 that the CCRH can calculate the optimal start time of the current task, the average attenuation of the critical path performs better. Meanwhile, in Figure 6 and Figure 7, with an increasing of node information's interval, so the algorithm robustness are affected, so choose the appropriate time to update the node information is the key to affect algorithm robustness.

5 Related work

The core idea of cloud computing is to manage and schedule a large number of computing resources connected by a network, and to constitute a pool of computing resources on-demand service to users. The key issues is how to schedule the resource fast and reasonable in cloud computing. So to schedule multiple DAG task is a way in effect of improving scientific calculation in scientific workflow applications of cloud computing.

Some of the relevant multi-DAG scheduling algorithms have been proposed, e.g.: literature [17] proposes a way of performing the multi-DAG task in turn, which result in producing a lot of idle wait time in the processor and prolonging the running time. Literature [5, 7, 12, 13] proposes a way of using a single DAG schedule way to schedule a complex DAG which is composed by multiple DAG. In [5], it proposes that multiple DAG merge into a complex DAG, then allocate resource for the complex DAG layer. In [7], it proposes a Planner-guided scheduling policy, which uses dynamic PANK-HYBD to schedule the priority for the multi-DAG tasks. These algorithms, however, does not consider the case that the DAG arrive at different times. In addition, literature [6] analysis [5, 7, 12, 13, 17] shows that the simple merger of DAG does not enhance the performance of algorithm significantly. Literature [13] proposes a multi-DAG scheduling algorithm that based on service time face to date-intensive applications. However this algorithm does not solve the problem which the running time increase because of the remaining tasks do not be scheduled in the former DAG. It is one of the keys in the cloud computing environment about scheduling problem that how to effectively solve the multi-DAG scheduling.

Moreover, the main goal of these algorithms is to explore the best completion time of the whole task, and ignore the task reliance. As the cloud computing is a new service model based on the large-scale low-cost service cluster, which hardware and software easily fail due to their own reasons or external factor. Literature [8] proposes to copy the task fully and define the position of this backup. In literature [9], an algorithm is proposed to meet the best Makespan and reliability, which improve the performance by putting the task to the computing nodes that has the smallest failure rate. Literature [10] proposes to improve the reliability by copy task based on literature [9], and schedule the task to the processor of load lightest. It proposes a fault-tolerant scheduling way of the priority constraint and the reliability cost driven, which emphasizes “the strong primary copy”, and demand that the task must received the result of its all predecessor node, so this algorithm only consider the predecessor node of the task, without considering the completion of all nodes task. In literature [8, 10, 11], it uses the copy way in compromising the reliability and system performance. However, these methods only judge the copy task itself, without calculate the start time of coping the task truly, which affects the algorithm performance.

References

- [1] <http://www.googlecloud.com/>
- [2] Boss G, Malladi P, Quan D, Legregni L, Hall H 2007 Cloud computing *IBM White Paper*
- [3] Wiczorek M, Prodan R, Fahringer T 2005 Scheduling of scientific workflows in the Askalon grid environment *SIGMOD Record* 3(34) 56-62
- [4] Mandal A, Kennedy K, Koelbel, C, Marin G, Mellor-Crummey J, Liu B, Johnsson L 2005 Scheduling strategies for mapping application workflows onto the grid *Proceedings of the 14th*

And these foregoing algorithms assume that the processors of any network are fully connected and it can receive the correlation information between scheduler and processor and between processors at any time. However, in practical applications, this assumption is untenable in the complex cloud computing environment. Its studies have shown that the scheduler algorithm considered the competition in the communication links can improve the accuracy grade effectively in Literature [27]. Literature [25] proposes a communication competition model in heterogeneous computing environment, and uses it to prove the validity of the scheduling algorithm, but this model is to consider the case of any network interconnection. In literature [26], it achieves the search and scheduler problem of processor in any interconnection network by the shortest path search algorithm in the communication competition model.

6 Contribution and future work

Against the reliable scheduling problem of scientific workflow in cloud computing system, this paper put forward a new method which use primary and secondary version to improve the system fault tolerance and dynamic hierarchical scheduling, the scheme has solved the problem when the multiple DAG task in quite different weights, the time span of DAG which arrived before will not be increased as the remaining tasks delays in scheduling. Simulation results show that in the premise of reliability requirements, the algorithm in fairness, Makespan, resource utilization, system run time showed better performance. The next step is to research in a given real cloud computing system architecture, how to solve the scheduling policies reliability under different failure probability, and through the different DAG scientific workflow load verify the algorithm's validity.

Acknowledgment

The work described in this paper is supported by the Fundamental Research Funds for the Central Universities (DL13CB05) and the Application technology research and development in Harbin (2013AE1CE007) and Technological innovation talent research project in Harbin (2013RFXXJ089).

International Symposium on High Performance Distributed Computing (HPDC 2005) North Carolina USA 125-34

- [5] Iverson M, Ozguner F 1999 Hierarchical, competitive scheduling of multiple dags in a dynamic heterogeneous environment. *Distributed Systems Engineering*, 1999 3(6) 112-20
- [6] Zhao H, Sakellariou R 2006 Scheduling multiple DAGs onto heterogeneous systems *Proceedings of the 15th Heterogeneous Computing Workshop (HCW)* Rhodes Island Greece

- [7] Yu Z, Shi W 2008 A planner-guided scheduling strategy for multiple workflow applications *Proceedings of the Parallel Processing – Workshops 2008 ICPP-W'08 International Conference Portland Oregon USA 2008* 1-8
- [8] Feng J, Humphrey M 2004 Eliminating Replica Selection—Using Multiple Replicas to Accelerate Data Transfer on Grids *Proceedings of the Parallel and Distributed Systems (ICPADS 2004)* Newport Beach CA USA 359-66
- [9] Dongarra J J, Jeannot E, Saule E, Shi Z 2007 Bi-objective scheduling algorithms for optimizing makespan and reliability on heterogeneous systems *Proceedings of the 19th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA 2007)* New York USA 280–8
- [10] Saule E, Trystram D 2009 Analyzing scheduling with ransient failures. *Information Processing Letters* **109**(11) 539-42
- [11] Qin X, Jiang H 2006 A novel fault-tolerant scheduling algorithm for precedence constrained tasks in real-time heterogeneous systems *Parallel Computing* **32**(5) 331-56
- [12] Höning U, Schiffmann W 2006 A meta-algorithm for scheduling multiple dags in homogeneous system environments *Proceedings of the 18th International Conference on Parallel and Distributed Computing and Systems (PDCS 2006)* Dallas Texas USA
- [13] Zhu L, Sun Z, Guo W, Jin Y, Sun W, Hu W 2007 Dynamic multi DAG scheduling algorithm for optical grid environment *SPIE 6784 Network Architectures, Management and Applications 2007* 67-84
- [14] <http://www.microsoft.com/azure>
- [15] <http://aws.amazon.com/ec2/>
- [16] <http://www.ibm.com/ibm/cloud/>
- [17] Sakellariou R, Zhao H 2004 A Hybrid Heuristic for DAG Scheduling on Heterogeneous Systems *Proceedings of the 13th Heterogeneous Computing Workshop(HCW 2004)* Santa Fe New Mexico USA
- [18] Topcuoglu H, Hariri S, Wu M 2002 Performance effective and low-complexity task scheduling for heterogeneous computing *IEEE Transactions on Parallel and Distributed Systems* **13**(3) 260-74
- [19] Wiczcerek M, Prodan R, Fahringer T 2005 Scheduling of scientific workflows in the Askalon grid environment *SIGMOD Record* **3**(34) 56–62
- [20] Pandey S, Wu L, Guru S, Buyya R 2010 A particle swarm optimization based heuristic for scheduling workflow applications in cloud computing environments *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications (AINA 2010)* Perth Australia 400-7
- [21] Salehi M A, Buyya R 2010 Adapting market-oriented scheduling policies for cloud computing *Proceedings of the 10th Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2010)* Busan Korea 2010 351-62
- [22] Casanova H, Legrand A, Zagorodnov D, Berman F 2000 Heuristics for scheduling parameter sweep applications in grid environments *Proceedings of the Heterogeneous Computing Workshop 2000* 349-63
- [23] Oinn T, Addis M, Ferri J, Mavin D, Senger M, Greenwood M, Carver T, Glover K, Pocock M R, Wipat A, Li P 2004 Tavern: A tool for the composition and enactment of bioinformatics workflows *Bioinformatics* **20**(17) 3045-54
- [24] Deelman E, Blythe J, Gil Y, Kesselman C, Mehta G, Patil S, Su M H, Vahi K, Livny M 2004 Pegasus: Mapping Scientific workflows onto the grid *Proceedings of the European Across Grids Conference Nicosia Cyprus* 11-20
- [25] Sinnen O, Sousa L A 2006 Toward a realistic task scheduling model, *IEEE Transactions on Parallel and Distributed Systems* **17**(3) 263-75
- [26] Tang X, Li K, Padua D 2010 Communication contention in APN list scheduling algorithm *Science in China (Series F Information Sciences)* **52**(1) 59-69
- [27] Macey B S, Zomaya A Y 1998 Performance evaluation of CP list scheduling heuristics for communication intensive task graphs *Proceedings of the First Merged International Parallel Processing Symposium and Symposium on Parallel and Distributed Processing March-April 1998* 538-541
- [28] Canon L-C, Jeannot E, Sakellariou R, Zheng W 2008 Comparative evaluation of the robustness of dag scheduling heuristics. In *Sergei Gorlatch, Paraskevi Fragopoulou, Thierry Priol editors, Integration Research in Grid Computing, Core GRID integration work-shop Hersonissos Crete Greece* 63–74
- [29] Hirales-Carbajal A; Tcherykh A; Yahyapour R 2012 Multiple Workflow Scheduling Strategies with User Run Time Estimates on a Grid *Journal of Grid Computing* **2012** **10**(2) 325-46
- [30] Juve G, Deelman R, Vahi K, Mehta G, Berriman B, Berman B P, Maechling P 2009 Scientific workflow applications on Amazon EC2 *Proceedings of the 5th IEEE International Conference on e-Science 2009* 59-66
- [31] Deelman E 2010 Grids and clouds: making workflow applications work in heterogeneous distributed environments *Int. J. High Perform. Comput.* **24** 284-98
- [32] Ramakrishnan L, Koelbel C, Kee Y-S, Wolski R, Nurmi D, Gannon D, Obertelli G, Yarkhan A, Mandal A, Huang T M, Thyagaraja K, Zagorodnov D 2009 VGrADS: enabling e-science workflows on Grids and clouds with fault tolerance *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* New York USA **47** 1–12

Authors



Weipeng Jing, born in January, 1979, Heilongjiang

Current position, grades: Lecture in Northeast Forestry University. Aa member of the CCF

Scientific interests: modelling and scheduling for distributed computing systems, system reliability estimation, fault tolerant computing and system reliability, distributed computing.



Yaqiu Liu, born in February, 1971, Heilongjiang

Current position, grades: Professor at the Northeast Forestry University.

University studies: M. Eng. in Control Theory and Engineering from Northeast Forestry University in 1999. PhD in Navigation, Guidance and Control from Harbin Institute of Technology in 2004.

Scientific interests: process control, distributed computing, cloud computing, intelligent control and soft computing, model reconstruction.

Multi-state system reliability assessment based on Bayesian networks

Xiaonan Zhang^{1, 2*}, Xiaoyong Lu²

¹College of Field Engineering, PLA University of Science and Technology, Nanjing 210007, China

²The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China

Received 1 March 2014, www.tsi.lv

Abstract

This paper considers a problem of multi-state system reliability modelling and assessment. By using the advantages of uncertainty reasoning and figurative expression of Bayesian network, a new method of modelling and assessment of multi-state system reliability based on BN is proposed to determine the nodes of BN and the multiple states of components of system, and to give the probability of each state and then utilizing conditional probability distributing (CPD) to describe the relationship among the component states, so as to express the states of correlated nodes and build a BN model of multi-state system. The model can clearly express the multiple states of system and component and the state probability, and also call directly calculate the system reliability on the basis of multiple state probabilities of component, thereby carrying out qualitative analysis and quantitative assessment of multi-state system reliability. By means of an example of multi-state radar system, we give the detailed multi-state system reliability analysis process based on BN. This paper not only proves the effectiveness of assessment of multi-state system reliability based on BN, but contributes to good help of complex system reliability, safety analysis.

Keywords: multi-state system, Bayesian networks, reliability assessment

1 Introduction

In traditional binary reliability framework, both systems and components can only take two possible states: completely working and totally failed. However, engineering systems typically have multiple partial failure states in addition to the above-mentioned completely working and totally failed states. Reliability analysis considering multiple possible states is known as multi-states reliability analysis. Multi-state reliability analysis recognizes the multiple possible states of engineering systems, and enables more accurate system reliability analysis.

Traditionally, system reliability has been analysed from a binary perspective assuming the system and its components can be in either of two states: completely functioning or failed. However, many systems that provide basic services, such as telecommunications, gas and oil production, transportation and electric power distribution, operate at various levels of performance as opposed to the binary perspective. These types of systems may provide a service or function at degraded component performance levels. Therefore, it is essential to model and analyse them accordingly. For these systems, multi-state system reliability methods have been proposed as a more appropriate modelling and computational approach.

The idea of multi-state system was first touched as early as in 1968 by Hirsch et al [1]. It was systematically introduced and studied in 1970s by Barlow and Wu [2],

EI-Neweihi et al. [3] and Ross [4] by considering a component or a system having more than two possible states. In their work, the primary concepts of multi-state reliability were studied, including system structure function, minimal cut (path) set, relevancy and coherency. The results by the early studies on multi-state reliability were generalized in the work of Griffith [5], Natving [6], Hudson and Kapur [7], and Block and Savits [8]. The early advances in multi-state reliability theory were summarized by EI-Neweihi and Proschan [9].

An important issue is how to model practical system in the multi-state context through careful analysis and definition. Many binary reliability models [10] have been extended to multi-state reliability models, such as the series-parallel system models [2, 11], the k-out-of-n system models [12], the weighted k-out-of-n system model [13], the network system models [14], etc. There might be more than one way to extend a binary reliability model to the multi-state context. For example, in Barlow and Wu's definition of multi-state series-parallel system [2], the state of a parallel subsystem is equal to the state of the best component. However, in Levitin's definition of multi-state series-parallel systems, the capacity of a parallel subsystem is equal to the sum of the capacities of its constituent components. Under traditional definition of multi-state k-out-of-n: G system [3, 15], the system is in state j or above when at least k components are in state j or above. Huang et al. proposed the model of generalized multi-state k-out-of-n: G system by allowing different

* Corresponding author e-mail: zxn8206@163.com

requirements of the number of components on different states [12, 16]. The model of multi-state consecutive system was also redefined [17]. The binary network reliability models have also been extended to multi-state versions by allowing the links and/or the nodes to have more than two possible states [14, 18, 19].

One way to analyse multi-state systems is using a binary variable to represent a single state of a component [20]. The problem is that there will be dependencies among variables that characterize the same component. The stochastic process approach is a more universal approach in modelling and evaluation of power systems [21]. Because the stochastic process approaches require equation solving whose computation burden can be significantly influenced by the number of components and the number of states, the stochastic approach can only be applied to relatively small systems. Levitin et al. developed the Universal Generating Function (UGF) approach to evaluate multi-state systems [14, 22], which can be used to deal with a wide range of multi-state systems. Like in the reliability evaluation of binary systems, Monte-Carlo simulation can be used for the evaluation of multi-state systems [18]. But compared to analytical algorithms, the main disadvantage of this approach is that it is not computationally efficient, especially for large systems with a large number of components.

With the concerning about the multi-state system reliability by related scholars, the multi-state system reliability theory has been some progress, but these methods there are some limitations, which is at the exploration preliminary stage.

In recent years, Bayesian network (BN) has found applications in, e.g., software reliability [13–16], fault finding systems [17–23], and maintenance modelling [24, 25]. One important feature that makes BN appealing is the possibility of combining different sources of information to provide a global safety assessment. Bouissou et al. [13] report on the experience of a hierarchical construction of a BN to combine different sources of evidence in the reliability analysis of complex software systems. On a similar line, Fenton et al. [14] showed that the robustness and well-founded underlying theory of BN can provide significant advantages. Wooff et al. [15] designed software tests using BN, and concluded that BN are well suited for these problems.

Because BN is good at analysing the uncertainty and correlation of random variables, BN technology applying in system reliability assessment can well make up for existing assessment methods. BN graphical expression function and conditional probability diagram (CPD) can make the relationship expression between systems and components more intuitive and clear. Some researchers constitute BN modelling framework which is particularly easy to use in interaction with domain experts, also in the system reliability field [25-29]. Common aims and goals are currently being recognized by researchers in classical reliability theory and the BN community, and examples of

fields of fruitful cooperation include probabilistic inference for fault detection and identification, monitoring, maintenance, and prediction. However, as far as the complex multi-state system reliability modelling and assessment, there is no systematic study and conclusion.

In this paper, by using the advantages of uncertainty reasoning and figurative expression of Bayesian network, a new method of modelling and assessment of multi-state system reliability based on BN is proposed to determine the nodes of BN and the multiple states of elements of system, and to give the probability of each state and then utilizing conditional probability distributing (CPD) to describe the relationship among the element states, so as to express the states of correlated nodes and build a BN model of multi-state system. The model can clearly express the multiple states of system and elements and the state probability, and also call directly calculate the system reliability on the basis of multiple state probabilities of elements, thereby carrying out qualitative analysis and quantitative assessment of multi-state system reliability. Analysis of practical examples proves the effectiveness of assessment of multi-state system reliability by using BN method.

This paper is organized as follows. In the next section, we introduce basic concepts of Bayesian Network theory. In the section 3, we present the two state system reliability modelling based on BN. Section 4 presents multi-state system reliability model. In section 5, by means of an example of multi-state radar system, we show the detailed multi-state system reliability analysis process based on BN. Finally, section 6 concludes the paper.

2 Bayesian Network theory

BN is probabilistic networks based on graph theory. Each node represents a variable and the arcs indicate direct probabilistic relations between the connected nodes. Variables are defined over several states. The BN allow taking into account time by defining different nodes to represent the variables at different time slices.

BN is directed acyclic graphs used to represent uncertain knowledge in Artificial Intelligence [15]. A BN is defined as a couple: $G((N,A),P)$, where (N,A) represents the graph; N is a set of nodes; A is a set of arcs; P represents the set of probability distributions that are associated to each node. When a node is not a root node, i.e. when it has some parent nodes, the distribution is a conditional probability distribution that quantifies the probabilistic dependency between that node and its parents.

In accordance with the definition of BN conditional probability:

$$P(A/B) = \frac{P(B/A)}{P(B)}, \quad (1)$$

where $P(B)$ is the prior probability, $P(A/B)$ for the posterior probability.

Supposed A is a variable, there are n states $a_1, a_2, \dots, a_i, \dots, a_n$, according to the total probability equation:

$$P(B) = \sum P(B / A = a_i)P(A = a_i), \quad (2)$$

the posterior probability $P(A/B)$ can be calculated.

With the conditional independence, BN can carry out two-way reasoning, not only forward reasoning, derived from the prior probability to posterior probability, which is from reasons to results, but also derived from the posterior probability to prior probability with the formula, which is from results to reasons.

3 Two state system reliability modelling based on BN

Application BN for systematic assessment in a straight form, we do not have to calculate the system minimal cut sets and minimal path sets, avoiding non-payment computing. In the case, that fault tree (FT) has been established and the FT can be directly mapped into BN.

FT is a kind of analysis method from the whole to part and from the top level to the down level according to varieties of fault reasons. The structure of BN model is correspondence with the FT; the difference is that BN makes the various fault reasons analysis, from the part to the overall, from the down to the top showing a branch shape. The establishment approach of two state systems BN models based on FT are as following:

1) Identify and model the relevant variables and their interpretation. Each basic event of FT corresponds to the root node in BN; each logic gate of FT establishes the corresponding middle node in BN; the same basic events appearing multiple times of FT can be expressed in a root node in BN.

2) Establishment a directed acyclic graph. According to the logic gate and the corresponding BN nodes, the directed arc which links the root with the leaves is expressed as parent and offspring relationship.

3) Giving the conditional probability of each variable, generating Conditional Probability Diagram (CPD). Corresponds to the FT, the priori probability of each root in BN is given. For each logic gate, the additional equivalent of CPD is given for the corresponding node. Based on the logic relationship of each gate, such corresponding CPD can be automatically generated.

Example: a system composed of three valves C_1, C_2, C_3 , Figure 1a is the system reliability block diagram. System function is defined as the passage fluid flow from A to B , normal state as a "pass", failure state "broken".

According to the reliability block diagram, FT is established in Figure 1b, in which T is expressed as the system failure event (top events), that X_i is expressed as the state of component i , M is an intermediate state of the event. According to the above rules, BN is shown in Figure 1c, the root node x_i is expressed as the basic event; that the leaf nodes t is expressed as the middle nodes; in CPD, 1 is expressed as fault, 0 is normal.

After the establishment of BN, we apply BN inference algorithm [23], such as the Equation (3) shows that:

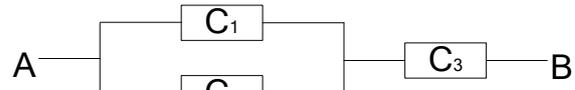


FIGURE 1a The system reliability block diagram.

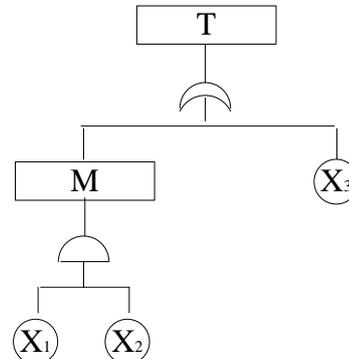


FIGURE 1b The system reliability FT

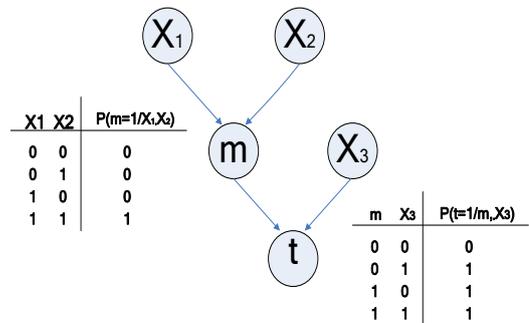


FIGURE 1c System reliability BN modeling

Figure 1 two state system reliability modelling based on BN:

$$\begin{aligned}
 P(t = 1) &= \sum_{X_1, X_2, X_3, m} P(X_1, X_2, X_3, m, t) = \\
 &\sum_{X_3, m} P(t = 1 / m, X_3) \sum_{X_1, X_2} P(m / X_1, X_2) P(X_1) P(X_2) = \\
 &\sum_{X_3, m} P(t = 1 / m, X_3) P(X_1 = 1) P(X_2 = 1) = \\
 &1 - (1 - P(X_1 = 1) P(X_2 = 1)) P(X_3 = 0).
 \end{aligned} \quad (3)$$

So the top event probability is that:

$$P(t = 1) = 1 - (1 - 0.0008) \times (1 - 0.01) = 0.010792.$$

The system reliability is:

$$R_s = 1 - P(t) = 1 - 0.010792 = 0.989208.$$

4 Multi-state system reliability model based on BN

Multi-state system is divided into discrete multi-state system and continuous multi-state systems. The systems and components state are limited or discrete, which is known as discrete multi-state system. For example, diode has open circuit, short circuit and working, which is three states system and if a system has the following four states:

- 1) The system working fine (perfect condition).
- 2) The system in degradation working state.
- 3) The system is not working because of the fault.
- 4) The system is not working because of maintenance program. We can use 0,1,2,3 expressing the four states system. In this paper, we study this kind of discrete multi-state system.

Multi-state reliability BN model is described by three different systems in the following. If the component has open circuit and short circuit, two failure modes, such systems also have two kinds of failure modes. In the same situation, the component has open circuit, short circuit and normal state, three states, the system composed of such components has the same three states.

In this paper we use this kind of three states as the example to study the multi-state reliability BN model.

4.1 PARALLEL SYSTEM OF TWO THREE-STATE COMPONENTS

The parallel system reliability block diagram is shown in Figure 2a. A open circuit component will not cause the system failure, and a short circuit component will cause the system failure. Figures 2b and 2c are system reliability block diagram in the two failure modes.

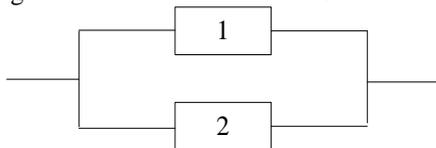


FIGURE 2a The system reliability block diagram

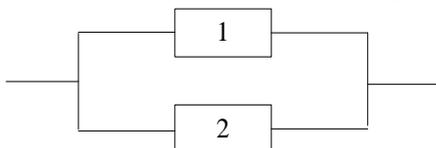


FIGURE 2b Open circuit failure mode



FIGURE 2c Short circuit failure mode

Figure 2 Parallel system of two three-state components.

According to literature [24]:

$$Q_0 = q_{01}q_{02}, Q_s = 1 - (1 - q_{s1})(1 - q_{s2}). \tag{4}$$

In the type, Q_0 is open circuit system failure probability, Q_s is short circuit system failure probability, q_{oi} is open circuit failure probability of component i ; q_{si} is short circuit failure probability of component i ; R_s is the working probability of component. Then the working probability of the system is that:

$$R_s = 1 - Q_0 - Q_s = (1 - q_{s1})(1 - q_{s2}) - q_{01}q_{02}. \tag{5}$$

If applying the method, when the component number increasing, not only the minimal path sets and the minimal cut sets is difficult to be got, but no doubt the above formula will become more complex computation.

Comparing the above method, we use BN reliability model to solve the problem in the following. It is shown in Figure 3. We use 0, 1, 2 to represent system and component open circuit, short circuit state and normal state respectively. P is system or component state probability; node a, b represent two state component; X is the system state.

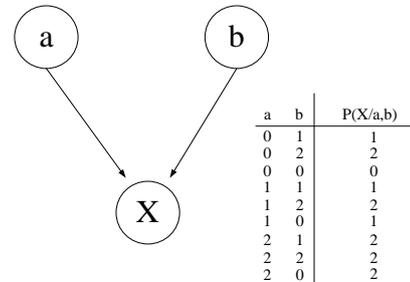


FIGURE 3 BN reliability model of parallel system

$$P(X) = \sum_{a,b} p(a,b, X) = \sum_a P(a) \sum_b [P(X/b)P(b)] = P(a)P(b). \tag{6}$$

$$P(a) = 1 \text{ or } P(b) = 1, P(X) = 1;$$

$$P(a) = 0 \text{ or } P(b) = 0, P(X) = 0;$$

$$P(a) \neq 1, P(b) = 2 \text{ or } P(a) = 2, P(b) \neq 1, P(X) = 2.$$

4.2 SERIES SYSTEM OF TWO THREE-STATE COMPONENTS

The series system reliability block diagram is shown in Figure 4a. A open circuit component will cause the system failure, and a short circuit component will not cause the system failure. Figures 4b and 4c are system reliability block diagram in the two failure modes.



FIGURE 4a The system reliability block diagram

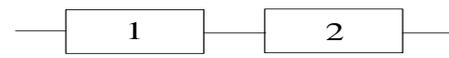


FIGURE 4b Open circuit failure mode

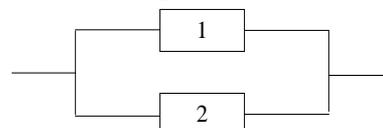


FIGURE 4c Short circuit failure mode

Figure 4 Series system of two three-state components. According to literature [24]:

$$Q_0 = 1 - (1 - q_{01})(1 - q_{02}), Q_s = q_{s1}q_{s2}, \tag{7}$$

$$R_s = 1 - Q_0 - Q_s = (1 - q_{01})(1 - q_{02}) - q_{s1}q_{s2}. \tag{8}$$

In the same way, we use BN reliability model to solve the problem in the following. It is shown in Figure 5. We use the CPD to analyse the node X.

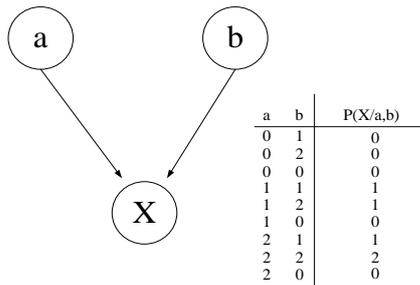


FIGURE 5 BN reliability model of series system

$$P(X) = \sum_{a,b} p(a,b,X) = \sum_a P(a) \sum_b [P(X/b)P(b)] = P(a)P(b), \quad (9)$$

$$P(a) = 0 \text{ or } P(b) = 0, P(X) = 0;$$

$$P(a) = 1 \text{ or } P(b) = 1, P(X) = 1;$$

$$P(a) \neq 0, P(b)=2 \text{ or } P(a) = 2 P(b) \neq 0, P(X)=2.$$

Figures 3 and 5 show, for the same number of multi-state components of the series system and parallel system, the reliability model based on BN is consistent in form, and only the CPD is different, so the multi-state system can be expressed by adjusting the CPD.

4.3 k-OUT-OF-n SYSTEM OF THREE-STATES COMPONENTS

k-out-of-n system can work if at least k components are working. In Figure 4, two-out-of-three of three-state component BN model is established. We use 0, 1, 2 to represent system and component state probability respectively. Nodes a, b, c represent the three basic event; node X represents system.

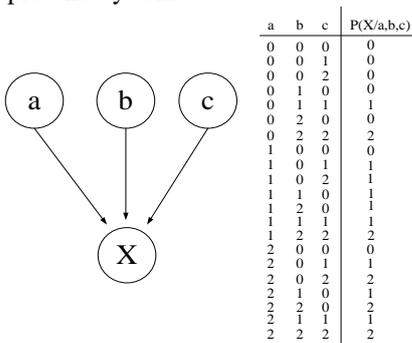


FIGURE 6 BN reliability model of two-out-of-three system

The main establishment steps of multi-state system reliability model based on BN network is as following:

- 1) Determine the BN network node. The network root represents basic events, the leaf node represents the system.
- 2) Determine the multiple states of discrete systems and components
- 3) Give the state probability of each component, which is usually given by actual test data.
- 4) Describe various components state relationship with the CPD; express the associated node state; establish of the BN model of system reliability.

Through the above three systems, we can know that multi-state system reliability model based on BN network has better visual image, and the state is expressed more clearly. Although as the number of system components increasing, CPD expression is more complex, but CPD of BN network is simple, regular, and suitable for programming.

5 Complex multi-state system reliability analyses

Figure 7 is a radar system, which consists of eight sub-system components, antennas X_1 , receiver X_2 , transmitter X_3 , actuators X_4 , display screen consisted of a color display instrument X_5 , and two series of black and white display instrument X_6 and X_7 made in parallel, signal processor X_8 , data processor X_9 , parallel data bus X_{10} and X_{11} , other subsystems can be considered as basic components. In which signal processor and data processor has a co-processing functions, in order to improve reliability, assuming that the two subsystems with memory and compensation, they can work in reduction success.

Supposed X_1, X_3, X_4, X_5, X_6 having all three kinds of states: 0 (failed), 1 (reduction success) and 2 (success); $X_2, X_7, X_8, X_9, X_{10}, X_{11}$ having only two kinds of states: 0 (failed) and 1 (success). In the following, we establish the multi-state fault tree and BN of this system separately. Through analysis, we can see that the method of BN has more analytical modelling than the traditional multi-state fault tree analysis methods.

5.1 ANALYSIS

Figure 8 shows the corresponding radar system multi-state fault tree, which gives the state space of middle events and top event.

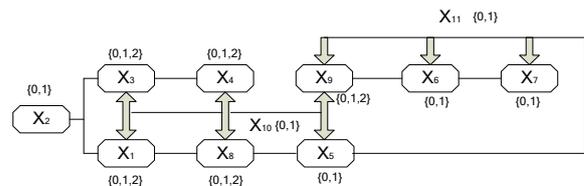


FIGURE 7 A radar system

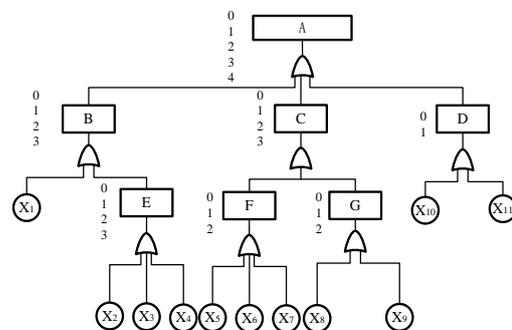


FIGURE 8 Multi-state system fault tree

The multi-state logic operator of middle events and top event is as follows:

TABLE 1 Logic operator

A	(B,C,D)
0	All vector including 0
1	111
2	121,131,211,221,311
3	231,321
4	331
B	(X1,E)
0	00,01,02,10,20,30
1	11
2	12,21,22,31
3	32
C	(F,G)
0	00,01,02,10,20
1	11
2	12,21
3	22
D	(X10,X11)
0	00
1	01,10,11
E	(X2,X3,X4)
0	All vector including 0
1	111
2	112,121
3	122
F	(X5,X6,X7)
0	000,010,001
1	011
2	100,101,110,111
G	(X8,X9)
0	00,01,10
1	02,11,12,20,21
2	22

TABLE 2 Root node conditional probability

	X1	X3	X4	X5	X6	
0	0.008	0.034	0.001	0.008	0.002	
1	0.042	0.058	0.033	0.09	0.04	
2	0.95	0.908	0.966	0.902	0.958	
	X2	X7	X8	X9	X10	X11
0	0.025	0.0017	0.0015	0.0015	0.021	0.021
1	0.975	0.9983	0.9985	0.9985	0.979	0.979

5.2 RESULTS

5.2.1 Top event probability

In the traditional multi-state fault tree analysis, top event probability need to calculate the system minimal cut sets and minimal path sets. Based on BN, we can calculate each node probability directly, avoiding non-payment computing. In this radar system, the top event A probability formula and procedure is as following:

$$P(A = i) = \sum_{B,C,D,\dots,X_{10},X_{11}} P(B, C, D, \dots, X_{10}, X_{11}, A = i) \quad (10)$$

$B, C, E \in \{0,1,2,3\}, D, F, G, X_1, X_3, X_4, X_5, X_6 \in \{0,1,2\}$
 $X_2, X_7, X_8, X_9, X_{10}, X_{11} \in \{0,1\}, i = 0,1,2,3,4$

TABLE 3 A probability in each state

A	0	1	2	3	4
probability	0.0675169	0.00000002	0.0167952	0.215148	0.70054

All operator are translated into CPD. The CPD E, A are as follows:

- $P(E=0/X2=0)=1$
- $P(E=0/X3=0)=1$
- $P(E=0/X4=0)=1$
- $P(E=1/X2=1, X3=1, X4=1)=1$
- $P(E=2/X2=1, X3=1, X4=2)=1$
- $P(E=2/X2=1, X3=2, X4=1)=1$
- $P(E=3/X2=1, X3=2, X4=2)=1$
- $P(A=0/B=0)=1$
- $P(A=0/C=0)=1$
- $P(A=0/D=0)=1$
- $P(A=1/B=1, C=1, D=1)=1$
- $P(A=3/B=2, C=3, D=1)=1$
- $P(A=3/B=3, C=2, D=1)=1$
- $P(A=4/B=3, C=3, D=1)=1$
- $P(A=0/else)=1$

Figure 9 shows the BN reliability model of this radar system. Table 2 gives the conditional probability of each root node.

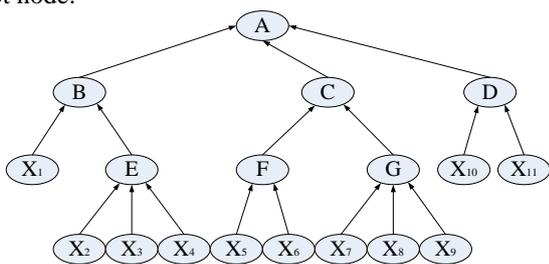


FIGURE 9 Radar system BN reliability model

Bayes Net Toolbox (BNT) has the corresponding procedure to solve the problem to simplified the computing. The BNT procedure is:

```

N=18;
dag=zeros(N,N)
X1=1; X2=2; X3=3; X4=4; X5=5; X6=6; X7=7; X8=8;
X9=9; X10=10; X11=11; E=12; F=13; G=14; B=15;
C=16; D=17; A=18
dag(X2,E)=1;
dag(X3,E)=1;
dag(X4,E)=1;
dag(X5,F)=1;
dag(X6,F)=1;
dag(X7,G)=1;
dag(X8,G)=1;
dag(X9,G)=1;
dag(X10,D)=1;
dag(X11,D)=1;
dag(X1,B)=1;
dag(E,B)=1;
dag(F,C)=1;
dag(G,C)=1;
dag(B,A)=1;
dag(C,A)=1;
dag(D,A)=1;
discrete_nodes=1:N;
node_size=2 *ones(1,N);
bnet=mk_bnet(dag, node_sizes, 'discrete', discrete_nodes);
bnet.CPD[X1]= tabular_CPD(bnet, X1, [0.008 0.042
0.95]);
    
```

```

bnet.CPD[X2]= tabular_CPD(bnet, X2, [0.025 0.975]);
bnet.CPD[X3]= tabular_CPD(bnet, X3, [0.001 0.033 0.966]);
bnet.CPD[X4]= tabular_CPD(bnet, X4, [0.008 0.042 0.95]);
bnet.CPD[X5]= tabular_CPD(bnet, X5, [0.008 0.09 0.902]);
bnet.CPD[X6]= tabular_CPD(bnet, X6, [0.002 0.04 0.958]);
bnet.CPD[X7]= tabular_CPD(bnet, X7, [0.0017 0.9983]);
bnet.CPD[X8]= tabular_CPD(bnet, X8, [0.0015 0.9985]);
bnet.CPD[X9]= tabular_CPD(bnet, X9, [0.0015 0.9985]);
bnet.CPD[X10]= tabular_CPD(bnet, X10, [0.021 0.979]);
bnet.CPD[X11]= tabular_CPD(bnet, X11, [0.021 0.979]);
bnet.CPD[B]= tabular_CPD(bnet, B, [1 0 0 0 0 1 1 1]);
bnet.CPD[C]= tabular_CPD(bnet, C, [1 0 0 0 0 1 1 1]);
bnet.CPD[D]= tabular_CPD(bnet, D, [1 0 0 0 0 1 1 1]);
bnet.CPD[E]= tabular_CPD(bnet, E, [1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1]);
bnet.CPD[F]= tabular_CPD(bnet, F, [1 0 0 0 0 1 1 1]);
bnet.CPD[G]= tabular_CPD(bnet, G, [1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1]);
bnet.CPD[A]= tabular_CPD(bnet, A, [1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1]);
engine=jtree_inf_engine(bnet);
evidence=cell(1,N);
evidence{A}=3;
[engine, lolik]=enter_evidence(engine, evidence);
evid.T
ans=0.215148
    
```

5.2.2 The importance

In the traditional multi-state fault tree analysis, the importance of components E_i need to get all the quality implication set, and then calculate the importance index E_i . However based on BN, the importance of components E_i can be directly calculated by the conditional probability components. Here we take the RAW (Risk Achievement Worth) as an example, and other type of importance can be calculated according to their definition.

Supposed the state space of system TE is $(0,1,\dots,M)$, the state space of component E_i is $(0,1,\dots,M_i)$, the RAW importance in state L can be calculated by type.

$$R_i(i, j) = \frac{1}{M_i + 1} \sum_{j=0}^{M_i} \frac{P(TE = l \setminus E_i = j)}{P(TE = l)}$$

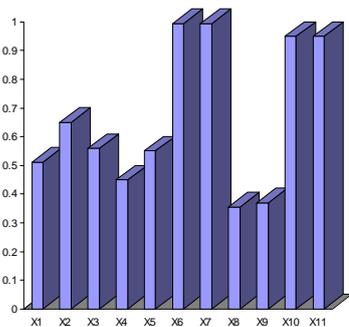


FIGURE 10 RAW importance in state 4

Figure 10 shows the various component RAW importance in state 4. We can be seen from Figure 9, components X_6, X_7 has the greatest importance, while X_8 has the smallest importance. So we can improve the system reliability based on the importance analysis.

5.2.3 Posterior probability

In addition, BN can get more rich information, such as the posterior probability. Supposed system failure, in order to diagnosing the fault and improving the system reliability, we need to compute all nodes posterior probability, while BN gives the fix computing and procedure, which need not compute all node combination, and get the best result.

Supposed system fault at the moment T , compute the posterior probability of component X_5 .

$$P(X_5 / A) = \frac{P(A = 1, X_5 = 1)}{P(A)} = \frac{\sum_{X_1, X_2, \dots, X_{11}} P(A = 1, X_1, X_2, \dots, X_{11} = 1)}{P(A)}$$

BNT procedure is as following:

```

evidence=cell(1,N);
evidence{A}=1;
[engine, loglik]=enter_evidence(engine, evidence);
marg=marginal_nodes(engine, X5 = 1 );
marg.T
ans=0.4235
    
```

Figure 11 shows probability distribution of each node changing when subsystem C transferring from state 3 into state 2. Obviously, we can get the information that it is probably that the state changing of component X_5 and X_6 lead to the subsystem C changing. Therefore if monitoring subsystem C changing from the intact state 3 to state 2, it should firstly investigate X_5 and X_6 in order to improve the system reliability.

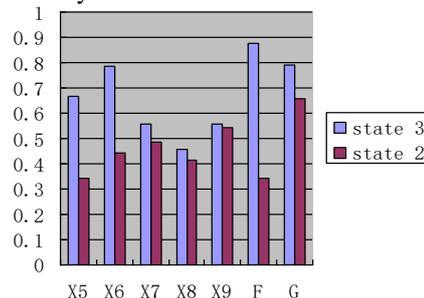


FIGURE 11 Each node probability distribution changing

6 Conclusions

- 1) According to the BN's two-way uncertainty logical reasoning ability and parallel computing characteristics, we study two state and multi-state system reliability modelling and assessment based on BN.
- 2) Multi-state system reliability modelling based on BN has the features of the good structure and hierarchy, and

the expression of CPD makes multi-state relationship between components and systems more simple and intuitive. Such as the same series and parallel systems with the same components, because BN form is the same, we can reflect the different system just by adjusting the CPD.

3) During computing the model's reliability, we do not find the system minimal cut sets or minimal path sets, so as to simplify the calculation and greatly enhance the computing accuracy and efficiency. The calculated results of the examples of show the effectiveness and advantage of multi-state system reliability modelling and evaluation based on BN.

4) At last, by means of an example of multi-state radar system, we give the detailed multi-state system reliability

analysis process based on BN. The topology of the BN is constructed; the conditional probability distributions and prior distributions are obtained according to multi-state logic operators. Analysis is performed on BN to obtain the probability of top event, importance measures of components and posterior probability and the corresponding formula and procedure is given. Through analysis, we can see that the method of BN has more analytical modelling than the traditional multi-state fault tree analysis methods. This paper not only proves the effectiveness of assessment of multi-state system reliability based on BN, but contributes to good help of complex system reliability, safety analysis.

References

- [1] Hirsch W M, Meisner M, Boll C 1968 Cannibalization in Multicomponent Systems and Theory of Reliability *Naval Research Logistics* **15**(3) 331-60
- [2] Barlow R E, Wu A S 1978 Coherent systems with multistate components *Mathematics of Operation Research* **3**(4) 275-81
- [3] El-Newehi E, Proschan F, Sethuraman J 1978 Multi-state coherent system *Journal of Applied Probability* **15** 675-88
- [4] Ross S 1979 Multivalued State Component Systems *Annals of Probability* **7** 379-83
- [5] Griffith W 1980 Multistate Reliability Models *Journal of Applied Probability* **17** 735-44
- [6] Natvig B 1982 Two suggestions of how to define a multi-state coherent system *Applied probability* **14** 391-402
- [7] Hudson J C, Kapur K C 1982 Reliability Theory for Multistate Systems with Multistate Components *Microelectronics and Reliability* **22**(1) 1-7
- [8] Block H W, Savits T H 1982 A decomposition for multistate monotone systems *Journal of Applied Probability* **19**(2) 391-402
- [9] El-Newehi E, Proschan F 1984 Degradable System – A Survey of Multistate System Theory *Communications in Statistics Theory and Methods* **13**(4) 405-32
- [10] Modarres M, Kaminskiy M, Krivtsov V 1999 Reliability Engineering and Risk Analysis: A Practical Guide *New York Marcel Dekker*
- [11] Levitin G 2005 Universal Generating Function in Reliability Analysis and Optimization *Springer-Verlag*
- [12] Huang J, Zuo M J, Wu Y H 2000 *IEEE Transactions on Reliability* **49**(1) 105-11
- [13] Li W, Zuo M J 2008 Reliability evaluation of multi-state weighted k-out-of-n systems *Reliability Engineering and System Safety* **93**(1) 160-7
- [14] Lisnianski A, Levitin G 2003 Multi-state System Reliability: Assessment, Optimization and Application *World Scientific Singapore*
- [15] Boedigheimer R A, Kapur K C 1994 *IEEE Transaction on Reliability* **43**(1) 46-50
- [16] Huang J Multi-state system reliability analysis Ph. D. Thesis University of Alberta 2001
- [17] Huang J, Zuo M J, Fang Z 2003 Multi-state consecutive k-out-of-n systems *IIE Transactions* **35**(6) 527-34
- [18] Ramirez-Marquez J E, Coit DW 2005 A Monte-Carlo simulation approach for approximating multi-state two-terminal reliability *Reliability Engineering System Safety* **87**(3) 253-64
- [19] Satitsatian S, Kapur K C 2006 *IEEE Transactions on Reliability* **55**(2) 199-206
- [20] Zang X Y, Sun H R, Trivedi K S 1999 *IEEE Transactions on Reliability* **48**(1) 50-60
- [21] Billinton R, Allan R 1996 Reliability evaluation of power systems *Plenum Press New York*
- [22] Levitin G, Lisnianski A, Ben Haim H, Elmakis D 1998 *IEEE Transactions on Reliability* **47**(2) 165-72
- [23] Boudali H, Dugan J B 2005 A discrete-time Bayesian network reliability modeling and analysis framework *Reliability Engineering and System Safety* **87**(3) 337-49
- [24] Jiang R, Zuo M 1999 Reliability model and application *Beijing China Machine Press (in Chinese)*
- [25] Yin X, Qian W 2009 Multi-state System Reliability Modeling and Assessment Based on Bayesian Networks *Chinese Journal of Mechanical Engineering* **45** 206-9
- [26] Qian W, Yin X, Xie L 2008 Discretized Modeling Process of Reliability of Multi-state Mechanical Systems *Journal of Northeastern University* **29** 1609-13
- [27] Tagaras G, Nenes G 2007 Two-sided Bayesian X control charts for short production runs. Bayesian Process Monitoring Control and Optimization, Colosimo B M, del Castillo E (eds.) *Chapman & Hall/CRC Press Inc London/Boca Raton* 167-86
- [28] Makis V 2008 Multivariate Bayesian control chart *Operations Research* **56**(2) 487-96
- [29] Nenes G, Panagiotidou S 2010 A Bayesian Model for the Joint Optimization of Quality and Maintenance Decisions *Quality and Reliability Engineering International* **27**(2) 125-248

Authors	
	<p>Xiaonan Zhang, born in 1982, Dandong of Liaoning Province, China</p> <p>Current position, grades: Lecturer at the PLA university of science and technology is engaged in the scientific research work of the Chinese people's liberation army.</p> <p>University studies: PhD at PLA university of science and military equipment.</p> <p>Scientific interest: weapons and equipment development, reliability engineering and design.</p> <p>Publications: More than 20 dissertations.</p>
	<p>Haiyong Lu, born in 1974, Haian of Jiangsu Province, China</p> <p>Current position, grades: Senior engineer of the institute of China electronics technology group.</p> <p>University studies: Master's at Zhengzhou university in 2002. Professional mechanical and electrical integration.</p> <p>Scientific interest: Military command and control software.</p> <p>Experience: Military electronic scientific research work, command and control software in system integration, satellite communications, system reliability, high theoretical level and rich experience in engineering practice, many national and military level scientific research projects.</p>

Fast fractional-pel interpolation algorithm of H.264 based on CUDA

X G Hong¹, H Liu^{1*}, Y Xiao²

¹*School of Information Engineering, Nanchang University, Nanchang 330031, China*

²*Institute of Computing Technology, Nanchang 330031, China*

Received 12 June 2014, www.tsi.lv

Abstract

H.264 video standard introduces fractional pixel motion compensation technology to obtain a more precise motion vector and a higher compression ratio. But, it increases the complexity of the motion compensation process at the same time. In order to solve the difficulties, we analysis the procedure of fractional-pel interpolation in H.264 and propose a fast fractional-pel interpolation algorithm based on CUDA. Experimental results show that the fast algorithm enables locating fractional pixel effectively and improves the speed of fractional pixel motion estimation. Compared with the CPU serial algorithm, the fast algorithm can significantly improve encoding rate almost four times in processing high-resolution video sequences.

Keywords: H.264, CUDA, interpolation algorithm, fractional-pel

1 Introduction

H.264 is a highly compressed digital video codec standard. It is proposed by the *Joint Video Team (JVT)*, which is grouped by the *Video Coding Experts Group (VCEG)* of ITU-T and the *Moving Picture Experts Group (MPEG)* of ISO/IEC [1]. Compared with other video coding standards such as MPEG-2, H.263, the H.264 standard has high compression ratio and high adaptability to the network. Because of that feature, the H.264 standard can be widely used in digital television, wireless video communication, video conference over IP and other multimedia services.

Nevertheless, the design of H.264 uses intra prediction in intra-frame, multiple frames reference capability, quarter-pixel interpolation, and flexible macro-block ordering in order to enhance *Motion Estimation (ME)* and *Motion Compensation (MC)*. But, it also increases the complexity of algorithm and the encoding computation. By analysing the H.264 encoding process, we conclude that the inter prediction takes more than 75% of the encoding time [2]. Moreover, in the process of inter prediction to get the fractional pixel has cost most of the calculation. So, in order to improve the computing speed of inter prediction, we must take an efficiently and fast interpolation algorithm to get the fractional pixel.

On the other hand, personal computers commonly equipped with GPU. Recently, the progress of GPU has caught a lot of attention; they have changed from fixed pipelines to programmable pipelines; the hardware design also includes multiple cores, bigger memory sizes and better interconnection networks which offer practical and acceptable solutions for speeding both graphics and non-graphics applications. GPU are highly parallel and are

normally used as a coprocessor to assist the Central Processing Unit (CPU) in computing massive data.

NVIDIA developed a powerful GPU architecture denominated Compute Unified Device Architecture (CUDA), which is formed by a single program multiple data-computing device. Hence, the fractional pixel interpolation algorithm developed in the H.264 encoding algorithm fits well in the GPU philosophy and offers a new challenge for the GPU.

This paper proposes a fast fractional pixel interpolation algorithm to accelerate the half-pixel and quarter-pixel in the process of inter prediction in the H.264 by using CUDA. The proposed algorithm efficiently develops the parallel computing of GPU to implement the parallel computing of fractional pixel interpolation. The remainder of this article is organized as follows: Section 2 briefly introduces fractional pixel interpolation process in H.264. Section 3 the fast fractional-pel interpolation algorithm processing based on CUDA is presented. Experimental results and compared to traditional approach are discussed in Section 4. Sections 5 show the conclusion and some ongoing work.

2 Overview of H.264 fractional-pel interpolation process

H.264 uses the 1/4 pixel precision to complete Motion Estimation (ME) and Motion Compensation (MC). Compared with 1/2 pixel precision in H.263, it can get more than 2dB coding gain [3]. The mainly process of fractional pixel interpolation algorithm shows as follows [4]:

I. Get the half-pixel: The half-pixel where between two integer-pixels (as b, h, m, s, etc. shown in Figure 1 (a)).

* *Corresponding author* e-mail: liuhao_ncu@sina.com

H.264 uses A sixth-order finite impulse response filter (FIR) to obtain the interpolated value by the neighbouring integer pixels. The weighted value of FIR is $1/32, -5/32, 5/8, 5/8, -5/32, 1/32$.

We can get the half-pixel value of b as follows:

$$b = \text{round}((E - 5F + 20G + 20H - 5I + J) / 32), \quad (1)$$

Similarly, the integer-pixel include of A, C, G, M, R, T through the sixth-order FIR can obtain half-pixel value of h. Once all of the adjacent integer-pixels (vertical or horizontal direction) get their half pixels, the remaining half pixels such as j can be calculated by six of vertical or horizontal half pixels. For example, the half pixel of j is calculated by value of cc, dd, h, m, ee, ff, which is shown in the Figure 1 (a).

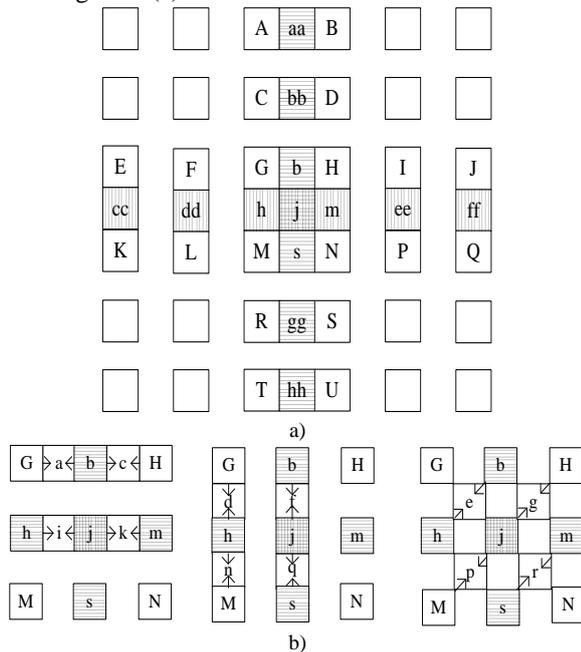


FIGURE 1 Fractional pixel distribution

II. Get the quarter-pixel: We use the linear Interpolation algorithm to obtain the value of quarter-pixel by the half pixel and integer pixel in the horizontal or vertical. The quarter-pixel distribution is shown in the Figure 1 (b). We can get the quarter-pixel value of a as follow:

$$a = \text{round}((G + b) / 2), \quad (2)$$

The remaining quarter-pixels such as e, g, p, r can be calculated by linear interpolation with a pair of half-pixel on the diagonal.

3 CUDA implementation of fractional-pel

3.1 CUDA PROGRAMMING MODEL

CPU and GPU work together in the CUDA programming model [5]. CPU is responsible for the serial parts of the code, and GPU is responsible for parts of the algorithm where is intensive and can be used to parallel Computing.

The structure of the CUDA programming model is shown in the Figure 2. Typically, the different of the implementation of the algorithm in GPU and CPU is the kernel function. Kernel function is part of a parallel program, which is used to parallel computing. The kernel function is used `__global__` to declare function type in the CPU and executes on the device in the GPU.

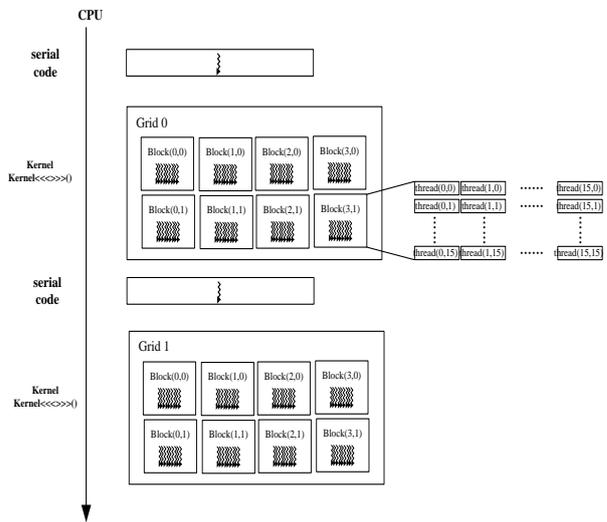


FIGURE 2 The structure of the CUDA programming model

3.2 ANALYSIS OF FRACTIONAL PIXEL INTERPOLATION

By analysing the H.264 encoding process [6-8], we conclude that the 4×4 sub-block includes three kinds of possible pixel value such as integer-pixel, half-pixel and quarter-pixel. As shown in Figure 3, the red point mark as integer pixel such as the point of 0. The white point is the fractional pixels. Among the white point, the point of 2,8,10 is half-pixel, the others is quarter-pixel. According to the predicted locations can be divided into the following six kinds of situations in the process of inter prediction.

I. The prediction point right on the integer-pixel where is region of 0. In this case, we should not need to calculate other fractional pixel. This is the simplest type of situation.

II. The prediction point right on the region of 1_2_3. In this case, we must calculate the corresponding fractional pixel by the interpolation algorithm.

III. The prediction point right on the region of 4_8_12. This case is similar to the second case so that we use the same method get the fractional pixel.

IV. The prediction point right on the region of 6_10_14. In this case, the middle of half-pixel should use the other value of half-pixel to calculate. So, we must get the corresponding half-pixel first, then uses this half-pixel and the others half-pixel to get the value of quarter-pixel.

V. The prediction point right on the region of 9_10_11. This case is similar to the fourth case so that we use the same method get the fractional pixel.

VI. The prediction point right on the region of 5_7_13_15. In this case, all of the pixels is quarter-pixel

and they must be calculated by linear interpolation with a pair of corner half-pixel. The computation in this case is quite large because all of the quarter-pixel by linear interpolation.

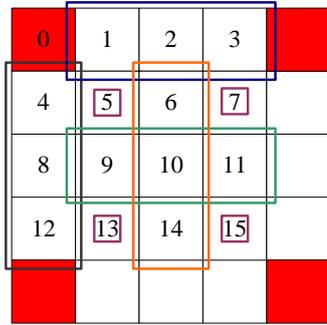


FIGURE 3 The 4×4 sub-block pixel distribution

From the analysis of the six cases, we can find that it needs according to different conditions to deal with the fractional pixel in H.264. In each condition it should use a large number of interpolations to derive the corresponding value of half-pixel and quarter-pixel. Because of this reason, the inter prediction occupies most of H.264 coding and decoding time. So, using the fast fractional-pel interpolation algorithm based on CUDA can accelerate fractional pixel positioning time so that it can improve encoding and decoding time and efficiency of H.264.

3.3 THE SPECIFIC IMPLEMENTATION PROCESS

Firstly, judging the region where the prediction point falls. Secondly, according to the prediction region get the corresponding half-pixel. Finally, according to the value of half-pixel get the corresponding quarter-pixel in the prediction region.

3.3.1 Judging the prediction region

In JM8.6 specification [9] defines a method named *get_block()*. In this method, there are two variables named *dx* and *dy* which are use to judge the location of prediction point. The variable of *dx* means the abscissa of the nearest integer-pixel on left. The variable of *dy* means the ordinate of the nearest integer-pixel on left. The specific conditions of the judgment are shown in the Table 1.

TABLE 1 Judgment of specific conditions

Judge conditions	Region of location
if (dx == 0 && dy == 0){	region of 0
if (dy == 0){ if ((dx&1) == 1){.....}}	region of 1_2_3
if (dx == 0){ if ((dy&1) == 1){.....}}	region of 4_8_12
if (dx == 2){ if ((dy&1) == 1){.....}}	region of 6_10_14
if (dy == 2){ if ((dx&1) == 1){.....}}	region of 9_10_11
others	region of 5_7_13_15

3.3.2 Calculate the fractional pixel in the corresponding region

In this step, we use the bilinear interpolation algorithm based on CUDA to get the value of half-pixel [10, 11]. Then parallel computing the corresponding quarter-pixel

by the calculated half-pixel. We use the prediction point right on the region of 5_7_13_15 as an example to describe the fast interpolation algorithm based on CUDA to get the value of fractional pixel.

I. As shown in the Figure 1 (a), the half-pixel of *b* can be calculated by the four integer-pixels on the diagonal (C, D, M, N). Calculation method as follows:

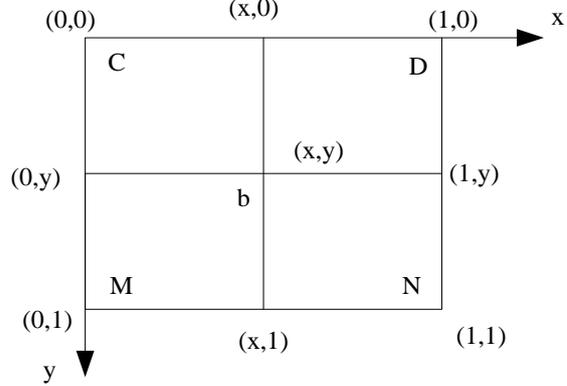


FIGURE 4 Bilinear interpolation algorithm

As shown in the Figure 4, the point of *b* is $F_{xy}(x, y)$, the value of integer-pixel{C: $F_{00}(0,0)$, D: $F_{10}(1,0)$, M: $F_{01}(0,1)$, N: $F_{11}(1,1)$ } is $(f_{00}, f_{10}, f_{01}, f_{11})$.

In the Y-direction, we use the linear interpolation the get the value of $F_{0y}(0, y)$:

$$f_{0y} = f_{00} + y(f_{01} - f_{00}). \quad (3)$$

In the X-direction, we use the linear interpolation the get the value of $F_{1y}(1, y)$:

$$f_{1y} = f_{10} + y(f_{11} - f_{10}), \quad (4)$$

Then, we use the value of $F_{0y}(0, y)$ and $F_{1y}(1, y)$ to get the value of $F_{xy}(x, y)$ in the X-direction:

$$f_{xy} = f_{0y} + y(f_{1y} - f_{0y}), \quad (5)$$

Substituting Equations (3) and (4) into Equation (5), combination and simplification can get as follow:

$$f_{xy} = (1-x)(1-y)f_{00} + y(1-x)f_{01} + x(1-y)f_{10} + (xy)f_{11}. \quad (6)$$

Set: $v(x_0) = 1 - x$; $v(y_0) = 1 - y$; $v(x_1) = x$; $v(y_1) = y$ then $v_{00} = v(x_0)v(y_0)$; $v_{01} = v(x_0)v(y_1)$; $v_{10} = v(x_1)v(y_0)$; $v_{11} = v(x_1)v(y_1)$. Then, Equation (6) can be simplified as follows:

$$f_{xy} = v_{00}f_{00} + v_{01}f_{01} + v_{10}f_{10} + v_{11}f_{11}. \quad (7)$$

II. Using Equation (7) to calculate the half-pixel. The basic computational unit of fractional pixel in the H.264 is the 4×4 sub-block [12]. We use a thread-block to calculate a 4×4 sub-block. For the resolution is $W \times H$ in a frame, the number of thread-block is shown as Equation (8):

$$n = \lceil W/4 \rceil \lceil H/4 \rceil. \quad (8)$$

The number of half-pixel, which should be calculated in a 4×4 sub-block is 33. So, we should distribute 33 threads in a thread-block and each thread is responsible for calculating the value of a half-pixel. The flow diagram is shown in the Figure 5.

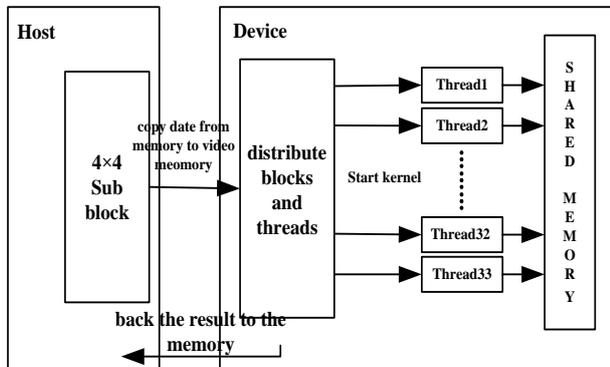


FIGURE 5 Flow diagram of calculate half-pixel based on CUDA

III. Calculate the quarter-pixel. There are two kinds of quarter-pixel that we should to calculate. One is used the integer-pixel and half-pixel by the linear interpolation to get (the point of a in the Figure 1 (b)). The other one is used a pair of half-pixel on the diagonal to get (the point of e in the Figure 1 (b)). The number of tread-block can be calculated by Equation (8). There are 120 quarter-pixels in a 4×4 sub-block, so, we distribute 120 threads in a thread-block and each thread is responsible for calculating the value of a quarter-pixel.

IV. The process of implementation on CUDA.

It can be seen from the CUDA programming model that the CPU serial code is responsible for data preparation and initialization of the device. The work which is shown as Host in the Figure 5 mainly includes memory allocation, video memory allocation, grid configuration, etc. The CPU will start the kernel function when the preparatory work is completed. The kernel function is used to parallel processing the data on the video memory by the GPU. To realize the algorithm proposed in this paper, we should create two kernel functions. One of the functions is used to calculate the half-pixel and the other one calculates the quarter-pixel. As the Device show in the Figure 5, the block in the kernel function creates threads to parallel computing [13] the value of half-pixel. Then, CPU executes its serial code to clean the kernel function and start next kernel function. At the same time, the CPU serial code will put the value of integer-pixel and half-pixel into video memory. After that, the kernel function will start the threads to parallel computing the value of quarter-pixel. The data will be copied from video memory to memory when fractional pixels have been calculated. Finally, the space of memory and video memory will be freed and exit the CUDA.

V. The pseudo-code, which describes implementation process is shown as follows:

```

Fractional_pel:
//allocation of video memory space
cudaMalloc()
//copy the data from memory to video memory
cudaMemcpy(cudaMemcpyHostToDevice)
dim3 block_half(n,1,1)
//calculate the number of thread, distribute 33 threads.
dim3 thread_half(33,1,1)
//obtain the value of half-pixel.
GPUhalf_pel<<<block_half,thread_half>>>()
//-----calculate the quarter pixel-----
dim3 block_quarter(n,1,1)
//calculate the number of thread, distribute 120 threads. dim3
thread_quarter(120,1,1)
GPUquarter_pel<<<block_quarter,thread_quarter>>>()
//obtain the value of quarter-pixel
cudaMemcpy(cudaMemcpyDeviceToHost)
//free the space of memory and video memory
cudaFree()

```

4 Experimental results and analysis

In this work, we use NVIDIA GeForce 9600 GSO as GPU platform which presents the characteristics depicted in Table 2. The CPU platform is AMD Sempron™ Processor 3200+ 1.80GHz. We use Microsoft Visual Studio 2008 and CUDA5.0 as the programming platform which runs on the operating system of Window 7. We use *cuda_Nsiht_Visual_Stuio_Edtion* as the performance analysis tool.

TABLE 2 GPU main features

Characteristic	GeForce9600 GSO
Compute capability	1.1
Global memory	512M
Number of multiprocessors	6
Number of cores	48
Constant memory	64KB
Shared memory per block	16KB
Registers per block	8192
Active threads per multiprocessor	768
Max threads per block	512
GPU Clock rate	1.5GHz

The performance evaluation of the fast fractional-pel interpolation algorithm of H.264 based on CUDA that we proposed based on JM8.6 encoder [9]. The test sequences are shown in Table 3.

TABLE 3 Test sequences

Test sequences	Resolution	Frames	Sample format
Carphone	QCIF(176×144)	150	YUV 4:2:0
Akiyo	QCIF(176×144)	300	YUV 4:2:0
Bus	CIF(352×288)	150	YUV 4:2:0
Foreman	CIF(352×288)	300	YUV 4:2:0
Blue_sky	1080P	150	YUV 4:2:0
Riverbed	1080P	300	YUV 4:2:0

For testing, QCIF (176×144), CIF (352x288) and 1080p are selected. These three kinds of video sequences have different resolution and different number of frames. The sample format of them is YUV 4:2:0. Firstly, the

proposed algorithm is put into JM encoder to replace the original algorithm, and then we realize the proposed algorithm on CUDA-based GPU. At last, we integrate the CPU implementation with the original algorithm and the GPU implementation with the proposed algorithm in JM encoder. The test sequences shown in Table 2 are used to test these two implementations. We get the encoding rate of two implementations. Table 4 lists the results of the comparison results of CUDA technology and CPU implementations. And Figure 6 shows the improvement of these video sequences.

TABLE 4 Encoding rate comparisons

Test sequences	Encoding rate(fps)		Speed Up
	CPU	CUDA	
Carphone	8.93	9.34	1.05
Akiyo	8.34	9.25	1.10
Bus	5.72	8.67	1.52
Foreman	4.89	8.56	1.75
Blue_sky	2.56	8.45	3.30
Riverbed	2.23	8.33	3.74

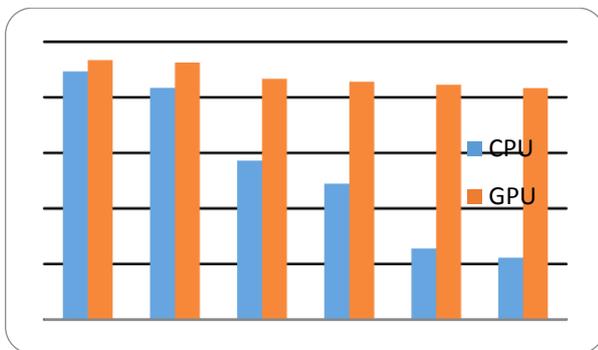


FIGURE 6 Comparison of encoding rate

From Table 4 and Figure 6, we can see the proposed algorithm for acceleration of the encoding rare in different resolution and the CUDA-based algorithm achieves higher efficiency than the CPU-based algorithm. For video sequences with the same resolution and different number of frames (*Carphone* and *Akiyo*, *Bus* and *Foreman*, *Blus_sky* and *Riverbed* shown in the Table 3), we can summarize that under the same resolution, the amount of encoding computation is not very large when video

sequence frames is less. In this case, encoder takes most time in data copies when it uses CUDA so that the acceleration effect is not obvious. However, with the increase of the video frames and amount of computation, the advantage of the GPU parallel processing can be reflected and the speedup also increases.

For video sequences with the same number of frames and different resolution (*Carphone* and *Foreman*, *Akiyo* and *Blus_sky*, *Bus* and *Riverbed* shown in the Table 3), we can summarize that the acceleration effect of low-resolution video sequences is not obvious. However, with the increase of the resolution of the video, there are more 4×4 sub-block will be divided in a frame of the video sequence. In this case, the number of the fractional-pel, which should be calculated will increases so that it increases the amount of computation of the encoder and the encoding rate is also decreased.

However, from the speedup shown in the Table 4, we can know that the acceleration effect is more obvious with the increase of the resolution. This is also reflected that the fast algorithm based CUDA can significantly improve encoding rate in processing high-resolution video sequences.

5 Conclusions

In this paper, we take the advantage of advantage of parallel computing in CUDA and propose to use CUDA technology to speed up the H.264 fractional-pel interpolation. The results show it is an effective approach to deal with this highly data-adaptive processing algorithm and it is can use to deal with high-resolution video sequences. H.264 standard is the most widely used standard and it is important to optimize its algorithm and execution time on a continuous basis. For future work, we will continue to optimize other modules and reduce complexity of the whole process based on CUDA technology.

Acknowledgments

This project is supported by the Graduate innovation fund project of Jiangxi Province, China 2013.

References

- [1] ITU-T RECOMMENDATION 2003 Advanced Video Coding for Generic Audiovisual Services *ISO/IEC 14496*
- [2] Blasi S G, Peixoto E, Izquierdo E 2013 Enhanced Inter-Prediction Via Shifting Transformation in the H. 264/AVC *Circuits and Systems for Video Technology* **23**(4) 735-740
- [3] Wedi T, Musmann H G 2003 Motion and aliasing-compensated prediction for hybrid video coding *Circuits and Systems for Video Technology* **13**(7) 577-86
- [4] Richardson I E 2004 H.264 and MPEG-4 video compression: video coding for next-generation multimedia *Wiley*
- [5] NVIDIA 2012 CUDA Compute Unified Device Architecture Programming Guide Version 5.0 *Applications* <http://developer.download.nvidia.com>
- [6] Fang Y, Zhou J 2006 Fast Fractional-pel Interpolation Algorithm of H.264 *Computer Engineering* (1) 076
- [7] Chen Z, Xu J, He Y, Zheng J 2006 Fast integer-pel and fractional-pel motion estimation for H.264/AVC *Visual Communication and Image Representation* **17**(2) 264-290
- [8] Wang Y J, Cheng C C, Chang T S 2007 A fast algorithm and its VLSI architecture for fractional motion estimation for H. 264/MPEG-4 AVC video coding *Circuits and Systems for Video Technology* **17**(5) 578-83
- [9] Joint Video Team Software 2010 JM8.6 *Applications* <http://iphome.hhi.de/suehring/tml/download/>
- [10] Gribbon K T, Bailey D G 2004 A novel approach to real-time bilinear interpolation *Field-Programmable Technology* 126-31
- [11] Liu J J, He Z, Chen L 2010 Bilinear interpolation of geomagnetic field *Computer Application and System Modeling (ICCAISM)* V2-665
- [12] Lu X, Tourapis A M, Yin P, Boyce J 2005 Fast mode decision and motion estimation for H.264 with a focus on MPEG-2/H. 264 transcoding *Circuits and Systems* 1246-9
- [13] Farber R 2011 CUDA application design and development *Elsevier*

Authors	
	<p>Xianggong Hong</p> <p>Current position, grades: Associate professor, Nanchang University. University studies: MS degree in School of Electronics and Communications Engineering at Nanchang University in 2009. Scientific interest: image processing technology, video communication technology.</p>
	<p>Hao Liu</p> <p>Current position, grades: Master of Signal and Information Processing, Nanchang University. University studies: Nanchang University in School of Information Engineering (2008-2015). Scientific interest: communication and information system, video communication technology.</p>
	<p>Yun Xiao</p> <p>Current position, grades: natural science researcher of Institute of Computing Technology, Jiangxi. University studies: MS degree at School of Computer Science from East China, Jiaotong University. Scientific interest: data mining, database development.</p>

New approach for ventilation network graph drawing based on Sugiyama method and GA-SA algorithm

Li-jun Deng*, Jian Liu

College of Safety Science and Engineering, Liaoning Technical University, Liaoning Fuxin 123000, China

Received 1 June 2014, www.tsi.lv

Abstract

Ventilation network graph has an important place in the management of a coal mine. In that case, aesthetics plays a major role for generating readable and understandable layouts. Besides, the drawing is required to be oval. The traditional longest path method for drawing ventilation network graph is inefficient and cannot effectively reduce the number of arc crossings because of the geometric intersection method. In this paper, we developed a new approach to draw ventilation network graph, consist of Sugiyama method framework, the longest path method and GA-SA algorithm. The longest path method was employed to rank nodes, and long arcs were removed by solving integer programming problem to minimize the sum length of ventilation network. Then genetic algorithm and simulated annealing algorithm were adopted to optimize the node order on reducing the number of arc crossings. In order to make the drawing be oval, a modified version of the longest path method was made to calculate node coordinates and arc shapes, which is called the longest parallel path method. Finally, computational experiments were carried out on two test ventilation network with our new approach.

Keywords: ventilation network graph, longest path method, integer programming, Sugiyama method, simulated annealing-genetic algorithm

1 Introduction

Mine ventilation system is a complex three-dimensional structure which is composed of crisscross and vertical overlap. So there's much inconvenience if using ventilation system graph in the ventilatory management. The coal mine safety regulation of China stipulates that mine ventilation department must draw the mine ventilation network graph [1]. The ventilation system could be transformed to a ventilation network on the basis of graph theory, which is composed of arcs, nodes and their properties. Ventilation network graph can clearly reflect the structure and flow characteristics of ventilation system, which is the basis of various ventilation calculations.

Li Husheng was the first one to propose the longest path method (LPM) for drawing ventilation network graph [2], searching all the longest paths corresponding to the fans, and lay outting the long paths on both sides and the short paths in the centre. Some minor fixes were contributed to LPM by Wu Bing, et al [3]. Actually, LPM is the most popular method to draw ventilation network graph automatically.

Ventilation network graph is also a kind of digraphs. Most approaches for drawing directed graphs used in practice follow the framework developed by Sugiyama et al. [4], which produces layered layouts. This framework consists of four phases: In the first phase, called Node Rank, the nodes are assigned to horizontal layers. During the second phase, called Node Order (also called Crossing Reducing); an optimal order of the nodes within a layer is

computed such that the number of arc crossings is reduced. The third phase, called Node Coordinate Assignment, calculates an x-coordinate for each node. The fourth phase, called Arc Drawing, calculates the arc shapes of the digraph.

Some minor changes were made to Sugiyama framework for drawing ventilation network graph. In the first phase of the original framework, long arcs between nodes of non-adjacent layers are replaced by chains of dummy nodes and arcs between the corresponding adjacent layers. This work can be achieved by the longest path method. Unfortunately, the longest path method leans to produce long arcs such that more dummy nodes were needed, making an influence on the node order phase. So we turned to use the longest path method in conjunction with integer programming to produce the short arcs. The second phase needs to reduce the number of arc crossings. Garey and Johnson proved that arc crossing reducing is a NP-complete problem [5]. Some heuristics [6] and hybridized genetic algorithm (HGA) [7] were proposed to solve that problem. But the crossover operation in HGA was too complex to effectively minimize arc crossings. Due to many arcs and nodes in the actual mine ventilation system, the number of solutions about node order problem is relatively large, which makes genetic algorithm run into a local optimal solution, not global optimal. So an algorithm mixed with genetic algorithm and simulated annealing algorithm was adopted to optimize the arc crossings problem. Finally node coordinates and arc shapes were calculated by the modified longest parallel path method, ensured that the shape of the drawing is oval.

* *Corresponding author* e-mail: anheihb03dlj@163.com

2 Optimizing on node rank

2.1 THE LONGEST PATH METHOD

For a network with n nodes, the longest path of any the j^{th} node in the network is the longest path length of the i^{th} source of all the j^{th} node's inflow arcs plus one [3]:

$$\begin{cases} l(1) = 0 \\ l(j) = 1 + \max\{l(i_k) \mid k = 1, \dots, m\} , \\ j = 2, \dots, n \end{cases} \quad (1)$$

where $l(j)$ is the longest path length of the j^{th} node, and $l(i_k)$ is the longest path length of i_k^{th} source of all the j^{th} node's inflow arcs.

Let $\lambda(v)$ be the node rank that equal to the longest length of the node v , where $\lambda(v) = l(v)$. Let l_{max} be the max value of the longest path length in the ventilation network, s and t be ventilation network's source and sink node. The nodes of ventilation network are layered according to the vertical direction from bottom to up. The source node is lay out on the lowest layer $\lambda(s) = 0$, while the sink is lay out on the highest layer $\lambda(t) = l_{max}$. Other nodes are lay out on the same layer with the same rank $\lambda(v)$.

The span of an arc (v, w) is defined by $\delta(v, w) = \lambda(w) - \lambda(v)$. If the span δ is larger than 1, $\delta - 1$ dummy nodes are used to split the arc (v, w) into shorter arcs, whose spans are all equal to 1.

2.2 THE OPTIMAL NODE RANK

The node rank based on the longest path method is easy to produce a lot of long arcs, whose span are larger than 1. It will make the arcs and nodes of ventilation network focus on the bottom of the graphic such that the drawing is not beautiful. In order to make the node rank produce short arcs and avoid long arcs, we tried to minimize the sum of the arc lengths on the basis of node rank by the longest path method. The optimization problem was defined as follows:

$$\begin{aligned} \min \sum_{(u,v) \in E} \omega(u,v)(\lambda(v) - \lambda(u)) \\ \text{subject to: } \begin{cases} \lambda(v) - \lambda(u) \geq \delta(u,v) , \\ \lambda(u) \geq 0 \end{cases} \end{aligned} \quad (2)$$

where $\omega(u,v)$ is the weight coefficient of an arc (u,v) , $\delta(u,v)$ is the minimum length of the arc (u,v) . In practical, the default value of $\omega(u,v)$ and $\lambda(u,v)$ is always 1. Obviously, this is a multi-variable integer programming problem, while all the variables must be integers [8].

In fact, the node rank of the longest path method can be used as the initial value of the IP problem to speed up the iterative convergence process.

3 Optimization on node order

The aim of the second phase is to find an optimal node order and reduce arc crossings in ventilation network graph.

3.1 THE PROBLEM OF ARC CROSSINGS MINIMIZATION

After the node rank phase, the nodes of ventilation network $G = (V, E)$ are divided into several layers, then the ventilation network could be transformed into a hierarchical graph $H_G = \{L_1, L_2, \dots, L_h\}$. The span of each arc in H_G is 1, and there are no arcs whose spans are more than 2. The nodes on each layer generate a sequence, called node order. The initial node order is defined by Π_0 .

The number of arc crossings between two adjacent layers is not dependent on the node coordinates x , only related to the nodes order in the layer. We consider the layer L_k and its adjacent layer L_{k+1} . The source and target node of the arc (v, w) and (v', w') is located in L_k and L_{k+1} , the position of the node v in L_k is defined by $\pi_k(v)$. If and only if the following condition is met, there is an arc cross between the arc (v, w) and (v', w') [9]

$$[\pi_k(v) - \pi_k(v')] \times [(\pi_{k+1}(w) - \pi_{k+1}(w'))] < 0 \quad (3)$$

So the problem was transformed into finding an optimal node order, avoiding the complex judgment of geometry intersection.

The optimization process of the drawings was based on an algorithm called GA-SA, which is a mixed algorithm with genetic algorithm and simulated annealing algorithm. We have introduced problem-based chromosome encoding, fitness function and SA algorithm which significantly improve the result of crossings reducing by the enhanced local search ability.

3.2 CHROMOSOME ENCODING

After the node rank phase, the nodes are divided into H layers: L_1, L_2, \dots, L_h . Naturally, the nodes on each layer can be encoded into a permutation. The permutation encoding of the k -th layer is defined by P_k , and the encoding of a node order is defined by Π as follow:

$$\begin{cases} N_k = |L_k| \\ \sum_{k=1}^h N_k = |V| \\ P_k = (\sigma_k(1), \dots, \sigma_k(N_k)) \\ \Pi = (P_1, \dots, P_h) \end{cases}, \quad (4)$$

where L_k is the k^{th} layer, $\sigma_k(i)$ is the node in the i^{th} position on the k^{th} layer, $|L_k|$ is the number of nodes on the k^{th} layer, and $|V|$ is the number of nodes in ventilation network. Therefore, the sum of the nodes on each layer is equal to $|V|$.

When the number of nodes in ventilation network is too large, the encoded string is also too long, which can slow down the convergence rate of genetic algorithm. In this paper, a small adjustment is applied to the encoding in order to reduce the length of the encoded string. A decimal integer can be used to represent the permutation encoding of each layer. Then we detail the new encoding as follow:

$$\begin{cases} k_i \in [1, (N_i)!] \\ \Pi' = (k_1, k_2, \dots, k_h) \end{cases}, \quad (5)$$

where k_i is the integer state of the i^{th} layer corresponding to P_i . The length of new encoding string Π' is h , while the length of Π is equal to $|V|$. In general, it is far less than the number of nodes in the ventilation network, $h \ll |V|$.

Algorithm 1 Pseudo-code of decoding an integer value to the node order of a layer

Where L_k is k -th layer, Π_0 is the original order of nodes on L_k ,

N_k is the number of nodes on L_k and P is an integer value.

function decode_to_order(Π_0, N_k, P) {

$\Pi \leftarrow \Pi_0$;

Pos, Val $\leftarrow 0, P$;

$K \leftarrow N_k - 1$

while $K > 0$ {

$\Pi \leftarrow \Pi_0$;

Cf $\leftarrow \prod_{i=1}^K i$;

$Cp \leftarrow \text{Val}/\text{Cf} + \text{Pos}$;

$\Pi[\text{Pos}] \leftarrow \Pi_0[\text{Cp}]$;

$\Pi[\text{Pos}+1 : \text{Cp}+1] \leftarrow \Pi_0[\text{Pos} : \text{Cp}]$;

$\Pi_0 \leftarrow \Pi$;

Pos, Val $\leftarrow \text{Pos}+1, \text{Val} \% \text{Cf}$;

$K \leftarrow K-1$

}

return Π

}

3.3 OBJECTIVE FUNCTION AND FITNESS FUNCTION

The objective function is defined as the number of arc crossings by $c(\Pi)$, then the fitness function is defined by $f(\Pi) = 2^{-c(\Pi)}$. In order to guarantee the efficiency of genetic algorithm, the fitness function is required as simple as possible. Therefore, the efficiency of the algorithm that counts the number of arc crossings between adjacent layers is critical.

Wilhelm Barth and Petra Mutzel proposed a rapid cross counting algorithm based on radix sort and inverse number counting [9]. For a directed graph $G=(V, E)$, the time complexity of the algorithm is $O(|E| \log |V_{small}|)$, where V_{small} is the smaller part of V that is partitioned into two layers.

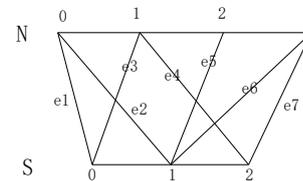


FIGURE 1 Bilayer Graph

3.4 SIMULATED ANNEALING LOCAL SEARCH

Simulated annealing algorithm (SA) is a kind of random optimization algorithm based on the mechanism of metal annealing. The sampling technology based on the probability is adopted to update the optimal solution. Under the temperature of T_k , a new individual Π'_i is got from the disturbed Π_i , which is the i^{th} individual of current population. Then the acceptance probability of Π'_i is determined by Metropolis sampling criterion:

$$p_i = \begin{cases} 1 & \Delta_i < 0 \\ \exp(-\Delta_i / T_k) & \Delta_i \geq 0 \end{cases}, \quad (6)$$

where p_i is the acceptance probability, $\Delta_i = c(\Pi'_i) - c(\Pi_i)$ is the objective function increment. When the new individual Π'_i is a more optimal solution ($\Delta_i < 0$), we completely accept the new individual as the current optimal solution. But when the new individual Π'_i is the poor solution ($\Delta_i \geq 0$), it is accepted by a small probability as the current optimal solution. As the temperature T_k decreased, the acceptance probability of the poor solution would be gradually smaller. Eventually, SA converged to the global optimal solution [10]. The method of geometry cooling is used to control of the temperature as follow:

$$T_k = \lambda T_{k-1} = \lambda^{n-1} T_0, \quad (7)$$

where T_k is the current temperature, T_0 is the initial temperature, λ is the temperature drop coefficient, the value range is $\lambda \in (0,1)$, and k is the iterations of cooling process.

The initial temperature T_0 is calculated as follow:

$$T_0 = c_{\max}(\Pi) / \ln(p_0), \tag{8}$$

where $c_{\max}(\Pi)$ is the number of arc crossings about the worst individual in a population; p_0 is initial acceptance probability.

It is considered that the algorithm is convergent, when the optimal solution does not improve by several cooling iterations. Then the search process of SA will stop.

4 The longest parallel path method

After the node hierarchy through the longest path method, it analyses the hierarchy chart which adds the secondary node, we can find that: "all paths between any two nodes are the longest path, and the longest path length is the same". That is to say that the longest path method has failed at this time.

The longest parallel path method was put forward to calculate node coordinates and arc shapes on the improvement of the original method [3], which made the shape of the drawing seem to be oval.

The node v is laid out on the layer of L_k , all the targets of the node v out-arcs are laid out on the layer of L_{k+1} , the node order of L_{k+1} is $\{w_1, w_2, \dots, w_n\}$. We iterate nodes on the layer of L_{k+1} by the order $\{(v, w_1), \dots, (v, w_n)\}$, called forward search, while backward search is defined by the order $\{(v, w_n), \dots, (v, w_1)\}$.

The longest parallel path method always tries to draw parallel paths. When it searches the longest path between two nodes each time by the depth first search (DFS), it firstly determine whether there is a drawn path that contains two nodes. If present, it will make a backward search try to find a symmetric path, which would be drawn in an arc. If not present, the search direction remains the same as the last search, and the path would be drawn in a straight line.

Forward search and backward search can ensure that the long paths are laid out on both sides, while the short paths are laid out in the middle. Eventually, the shape of ventilation network graph leans to be oval.

```

Algorithm 2 Pseudo-code of the longest parallel path method to
draw nodes and arcs
S, T ← [], [];
Push the source node s into S;
Push the sink node t into T;
While S not empty and T not empty {
    u ← Pop a node from S;
    v ← Pop a node from T;
    If (u,v) in a drawn path P {
        New_P ← DFS search find a symmetric path of P from u
to v without colored arcs;
    }
}

```

```

}
Else {
    New_P ← DFS search find a path from u to v without
colored arcs;
}
Calculate node coordinates and arc shapes in New_P and color
the drawn arcs;
Analyse the path New_P to find more sources and targets;
Update S and T;
}

```

5 Computational Experiments

We developed a procedure to test two ventilation networks, whose size parameters are listed in Table 1. And control parameters of GA-SA algorithm are listed in Tables 2 and 3. The two ventilation network graphs are shown in Figures 2 and 3.

TABLE 1 Ventilation network size parameters

	A simple ventilation network	Ventilation network in Zhao Zhuang coal mine
The number of nodes	26	155
The number of Arcs	35	213

TABLE 2 Control parameters of GA

Control parameters	A simple ventilation network	Ventilation network of Zhaozhuang coal mine
Coding scheme	Decimal integer coding	Decimal integer coding
Population size	15	30
Termination condition	Maximum evolution algebra 20	Maximum evolution algebra 50
Crossover rate	0.75	0.75
Mutation rate	0.05	0.05
Selection Operator	Roulette selection	Roulette selection
Crossover operator	Single-point crossover	Single-point crossover
Mutation Operator	Single-point mutation	Single-point mutation

TABLE 3 Control parameters of SA

Control parameters	A simple ventilation network	Ventilation network of Zhaozhuang coal mine
Termination condition	The best individual has no improvement for 3 generations	The best individual has no improvement for 10 generations
Initial acceptance rate	0.8	0.618
Temperature drop coefficient	0.9	0.99

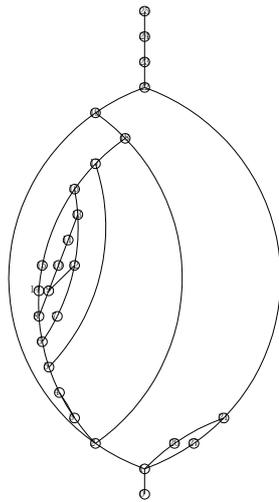


FIGURE 2 A simple ventilation network graph

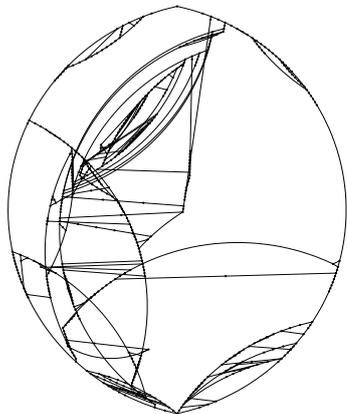


FIGURE 3 Ventilation network graph of Zhao zhuang coal mine

6 Conclusions

We developed a procedure for drawing ventilation network graph based on Sugiyama method, ranking nodes on layers by the longest path method, reducing arc crossings by GA-SA algorithm, and calculating node coordinates and arc shapes by the improved longest parallel path method to make the drawing seem to be oval.

Computational experiments were done with two ventilation networks. They showed that our new approach could find a better layout to reducing arc crossings and make the drawing seem to be oval. Actually, it could be concluded that the drawing is better for the small ventilation network (probably less than 100 arcs).

So there is also another interesting point to study further. We need to simplify the ventilation network, removing part of the arc which does not affect the connectivity and the overall structure of ventilation network. So the ventilation network is splitted into some smaller sub-networks, then this new approach is applied to these small networks to produce a better drawing.

Acknowledgments

Here and now, I would like to extend my sincere thanks to all those who have helped me make this paper possible and better. First, I am deeply grateful to y supervisor, Liu Jian professor, who have check my paper with patience to give me instructive suggestions. Then thanks to the teachers and professors in my colleague who have given a lot of help on my study. Finally, I am very grateful to my lovely friends and classmates who have offered me quiet situation and discussed with me about some knowledge.

References

- [1] Huang L, Liu Y, Li Z, Yang Y 2002 Study on mine ventilation network graph *Journal of Jiaozuo Institute of Technology (NaturalScience)* 2002 **21**(1) 11-4
- [2] Li H 1998 Automatic generation of curved-line network graph from structural ventilation network data *Coal mine safety* (1) 9-12
- [3] Wu B, Lu B, Shui L Automatic Generation of Ventilation Network Graph by Longest Path Algorithm [J] *Coal mine safety* 2006 (6) 1-3
- [4] Sugiyama K, Tagawa S, Toda M 1981 *IEEE Transactions on Systems, Man, and Cybernetics* 1981 **11**(2) 109-25
- [5] Garey M, Johnson D 1983 Crossing number is NP-complete *Algebraic Discrete Methods* **4**(3) 312-6
- [6] Eiglsperger M, Siebenhaller M, Kaufmann M 2005 An Efficient Implementation of Sugiyama's Algorithm for Layered Graph Drawing *Journal of Graph Algorithms and Applications* **9**(3) 305-25
- [7] Kuntz P, Pinaud B, Lehn R 2006 Minimizing crossings in hierarchical digraphs with a hybridized genetic algorithm *Journal of Heuristics* 2006 **1** 1-11
- [8] Du H, Zhao Y 2010 Survey on intelligent optimization algorithms for solving integer programming problems *Application Research of Computers* **02** 408-12
- [9] Barth W, Junger M, Mutzel P 2002 Simple and efficient bilayer cross counting. In *GD'02: Revised Papers from the 10th International Symposium on Graph Drawing* **2528** of LNCS 130-141 Springer Verlag
- [10] Han Y, Shan Y 2012 Using improved simulated annealing genetic algorithm to estimate parameters in groundwater inverse problem *Computer Engineering and Applications* **12** 224-8

Authors



Li-jun Deng, born in October, 1985, Jing Shan, Hubei, China

Current position, grades: PhD Student.
University studies: safety engineering of coal mine.
Scientific interest: the theory and algorithms of mine ventilation system.



Jian Liu, born in December, 1961, Chifeng, Inner Mongolia, China

Current position, grades: professor.
University studies: safety engineering of coal mine.
Scientific interest: mine ventilation and fire prevention.
Publications: 118 papers.
Experience: 108 projects in the national and provincial natural science fund projects.

Multicast routing algorithm based on cloud computing strategy in wireless sensor networks

Zeyu Sun^{1, 2*}, Shouying Li³

¹Department of Computer and Information Engineering, Luoyang Institute of Technology, Luoyang, Henan, 471023

²Electrical and information Engineering, Xi'an JiaoTong University Xi'an 710061, China

³Department of Mathematics and Physics, Luoyang Institute of Technology, Luoyang, Henan, 471023

Received 1 June 2014, www.tsi.lv

Abstract

In the paper, the existing wireless sensor network routing protocol classification and comparative studies, then A targeted selection of typical clustering routing protocol Low Energy Adaptive Clustering Hierarchy (LEACH) protocol for the study, points Insufficient analysis of LEACH, and on this basis to select the optimal number of cluster head LEACH protocol, Selected on the basis of cluster head node, communication between clusters, cluster head node in a distributed fashion to improve "change After the agreement into improving the selection criteria LEACH cluster head, so that the cluster head node distributed more evenly, avoiding Free a single node excessive energy consumption. Finally, the improved LEACH protocol is added by way of randomly selected based on the energy value of closed Residual nodes on MATLAB simulation platform, the simulation results show that adding the residual section Point effectively extending the network life time improves the efficiency of the network.

Keywords: wireless sensor network, routing protocol, network lifetime, sensor nodes

1 Introduction

Wireless sensor network consists of sensors, sensing the object and the observer three basic elements. Wireless sensor networks are between devices, communication between the sensor and the observer. Most or all nodes in wireless sensor network nodes can move their position, which makes the relationship between nodes varies with the mobile node continue to occur. Between nodes in Ad Hoc mode to communicate, each node can act as a router, and each node has a dynamic search, the ability to connect and positioning of recovery. Information collected by observers can actively query or collection of wireless sensor networks can also be passively receiving information transmitted in wireless sensor networks. Information collected by observers for analysis processing, and then take the appropriate measures to reach a certain goal [1]. Wireless sensor networks are self-organizing network, the automatic generation of network topologies. Topology control means to meet the region overlying Under conditions of coverage and network connectivity by selecting the node transmit power control and network key nodes, delete unnecessary links to generate an efficient network topology to improve the efficiency of the whole network, to extend the life of the network cycle [2]. Good network topology can improve efficiency of the link and routing protocol for the upper data fusion, target location, time synchronization basis for many applications, the network load balancing beneficial to extend the lifetime of the entire network [3]. Time

synchronization is directly related to the application of wireless sensor network protocols running and running on it. Network Each node has its own local clock, the nodes within a particular time if the system are synchronized, with the passage of time, there will be some deviation of the clock nodes, the system needs to sleep if many nodes will likely predetermined there may not be time to wake up. Wireless sensor network storage and computing capacity of a single node are weak, the decision of the traditional network protocol is not suitable for wireless sensor networks, the agreement must be to design suitable for wireless sensor networks based on its own characteristics. In wireless sensor networks, routing protocols need to consider not only the energy consumption of a single node, but also need to consider the energy consumption of the entire network.

Wireless sensor networks are self-organizing network (Ad-Hoc), but with self-organizing networks are essentially different [4]. Self-organizing network is composed of hundreds of thousands of nodes in a dynamic wireless multi-hop communication is formed in the form of peer to peer wireless network, which is characteristic for the purpose of data transmission, the node can continue to provide energy, more stable operation of the entire network, the data computing and computing capacity is weak, less the amount of information stored in the data; wireless sensor networks and ad hoc network structures are generally similar, except with the characteristics of self-organization of the network, the main difference is: wireless sensor networks based on data- processing centre,

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

the node is not limited energy supplement, the network has the ability to monitor and control, data processing ability [5,6]. First, the wireless sensor networks in the work process often subject to external factors and the impact on the environment, usually in high-density portion opposed to the node to complete coverage of the target area, tracking and positioning, due to the presence of high-density deployment of nodes, making between nodes frequent phenomenon when transferring the data redundancy, thereby causing disruption and congestion of the network, how to deploy sensor nodes cover the target area of concern and to avoid the occurrence of network redundancy in wireless sensor networks is an important issue. Secondly, the sensor nodes in wireless sensor networks is perceived by the local self-organization in the form of data, the collected data is calculated after processing to the base station, base station forwards the data to the user after the integration. Data acquisition and processing of information in the calculation process involves mechanisms for wireless sensor network coverage, targeting mechanisms, data fusion mechanisms and connectivity mechanisms for wireless sensor networks [7] to achieve this important part of the process is covered by the mechanism. Again, from the perception of the physical world, the wireless sensor networks to handle data collection accuracy and stability and the effective coverage of the target area under certain conditions depends on systematic and complete coverage mechanism [8]. The importance of covering another mechanism also depends on physical performance, is the energy networks of sensor nodes, wireless sensor networks by optimizing coverage, for a reasonable allocation of network resources, improve network quality of service, to suppress the node energy consumption and prolong the network lifetime have played a crucial importance.

2 System architectures

Sensor nodes are usually scattered in a random form of the coverage area, wherein the sensor comprises a combination of nodes in the cluster, the cluster head node, base station or other communication devices, the Internet, from other user groups. Each sensor node After collecting data, through a preliminary analysis and data fusion to pass information to the cluster head node, cluster head node will collect information again computing integration, network link to hop along the way to pass to the base station in data transmission can be collected and processed more sensor nodes, the data reaches the base station or satellite channel in a wired manner calculated developed manner to the user shown in Figure 1.

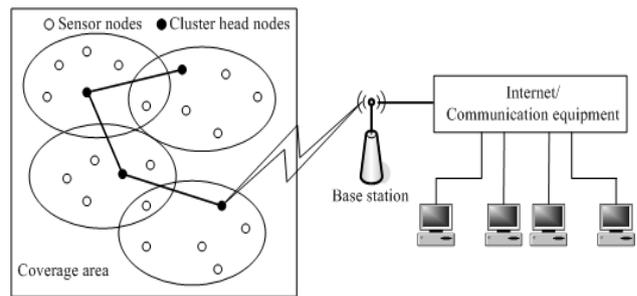


FIGURE1 Wireless Sensor Network Architecture

After randomly distributed in the coverage area of the sensor node cluster, let the sensor node becomes higher energy cluster head node, the sensor nodes collect the relevant data covering the area between the sensor nodes through the self-organization, multichip constitute a communication network system, sensor nodes can collect data cluster head node hop along the network link transmission; through multiple hops after being sent to the base station node, the base station node with a strong computing power, storage capacity, communication ability and control capabilities, and at the same time limiting its energy from outside the base station node can also collected data reprocessing, data processed through a wired network or other communication device applications forwarded to the terminal. For the user, it can make various requests to the wireless sensor network services, such as requiring that access a variety of commands and control information, etc., in order to obtain information and control information related queries.

The sensor nodes of different composition and different practical application, in general, the sensor node by the sensor module, the processor module, a communication module and a power control module is a combination of four parts. The main function of the sensor module is within the coverage area of data collection and data conversion; processor module is mainly responsible for the calculation of the collected data, storage and processing of data transmitted over the other node information; communication module is the main function of the sensor node or a cluster head node for a wireless communication, exchange data and control information; energy control module is mainly responsible for providing energy to the sensor nodes, micro-battery cell is required to, the structure diagram shown in Figure 2.

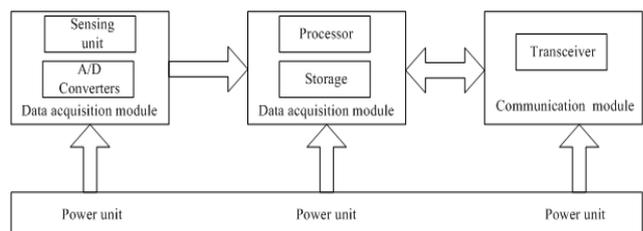


FIGURE 2 Diagram of the structure of the sensor nodes

With the in-depth study of the rapid development of wireless sensor networks as well as staff proposed protocol stack for wireless sensor network nodes. The main

components are divided into: a physical layer, data link layer, network layer, transport layer and application layer. Time synchronization and positioning sub- position of the layer in the protocol stack is rather special, its working principle is dependent on the data transmission channel for collaboration positioning and time synchronization consultation, while also providing a variety of information support for network protocols. As to form the basis of the MAC protocol, geographic routing protocols, etc. These protocols are needed in high-precision positioning and time synchronization on. In addition, the protocol stack also includes three management platforms, namely: energy management platform, mobile management platform and task management platform. The main role of energy management platform is to complete the overall management of the entire network of energy through energy wireless sensor network node scheduling conversion ; mobile management platform, the main role is to move the process when the node forwarding and receiving data in a wireless sensor network, and forwards and processing the received data. Task management platform main role is to coordinate multi- task allocation process, completed between the wireless sensor network nodes work together through a variety of collection and configure the interface to configure and monitor the appropriate mechanisms port also supports multi-task and resource sharing. The main function of the protocol stack layers are: The main function is to handle the physical layer and the data transmission error rate monitor data, to provide transparent data transport stream, while providing a simple radio signal modulation information and technology; data link layer is responsible for the services provided on the basis of the physical layer, communication between entities to establish a data link connection , transfer to a "frame " for the data packet units, and using error control and flow control methods, so that there is an error in the physical line to become error-free data link, monitoring the completion of the frame , media access and error control; main function is the network layer to create logical data transmission link by routing the packet through the communication sub-algorithm to select the most appropriate path to implement congestion control , and , networking and other functions; main function is to transport layer packet processing errors, restructuring packet ordering, data flow control and to ensure the quality of communication services; main function of the application layer is composed of a series of application software to complete monitoring of multi-tasking and control.

3 LEACH protocol and optimization strategies

3.1 LEACH protocol

LEACH protocol by MIT's Heinzelman, who proposed the first WSN clustering hierarchical routing protocol, occupies an important position in the WSN routing protocol, and subsequently clustering based routing

protocols such as TEEN, PEGASIS etc. Most developed by LEACH come the basic idea is to select the protocol LEACH cluster head node randomly cycle through to the energy of the whole network load evenly distributed to each sensor node, which can reduce the energy consumption of the network to improve network lifecycle purpose of this chapter will focus on description and analysis of LEACH shown in Figure 3.

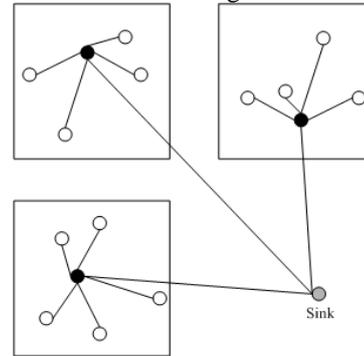


FIGURE 3 Schematic of LEACH

LEACH protocol LEACH-shaped card is put forward on the basis of a centralized routing protocol, which effectively solve the uncertain number of clusters and cluster head node selection optimization problems compared to the LEACH protocol and effectively increase the amount of data received by the base station prolong the survival time of the node. Firstly, the energy of the agreement as a condition of the current node cluster head selection At this point all the nodes are no longer equal probability as a cluster head, but with the current energy is proportional to the current energy more elected as cluster head node probabilities have large number of i-node probability at time t as cluster head for $P_i(t)$, which is calculated as follows equation:

$$p_i(t) = \min \left\{ \frac{E_i(t)}{E_{total}(t)} k, l \right\}, \tag{1}$$

where k is the number of cluster heads in each round, $E_i(t)$ is the energy of the current node i , $E_{total}(t)$, N are N nodes and the energy of the current, so $E_{total}(t) = \sum_{j=1}^n E_j(t)$.

Each one goal is the formation of K clusters. LEACH specific process is as follows: For any node N , node N generates a random number in the range between 0 and 1, if this number is less than a given threshold value $T(n)$, the cluster head node N becomes node, while the node N broadcasts itself as a cluster head information. Threshold $T(n)$ of the expression shown in equation:

$$T(n) = \begin{cases} \frac{P}{1 - p \lceil r \bmod (1/p) \rceil}, & \\ 0 & \end{cases} \tag{2}$$

where, P is the desired cluster head node is the percentage of all, that the percentage of the total number of nodes in the cluster head after every round, its value is generally

from 4% to 5%, for different applications, p values different. r is the number of rounds of the current loop. When $r = 0$, $T(n) = p$, the probability of each cluster head node becomes the same, all for p . The larger the value of r , the value of $T(n)$ is greater, so as not to become too cluster head node, the greater the probability of cluster head. When $r = 1/p - 1$, $T(n) = 1$, then this will become a cluster head node. When $r = 1/p$ and $r = 0$ when $T(n)$ of the same value, $r = 1/p + 1$ and $r = 1$ when $T(n)$ the value of the same after $1/p$ wheels, and all nodes restart a new round of circulation.

3.2 OPTIMIZATION STRATEGIES

In practical applications, it is often encountered in multi-criteria or objectives, design and decision-making problems, "such as securities investment issues, investors in order to get higher returns, you need to select the best stocks to invest in, in general, an outstanding shares have the following characteristics: good performance, low price-earnings ratio, growth higher, but usually these goals are in conflict, such as the current domestic steel industry generally better performance of listed companies, earnings are relatively low, but the steel industry is not sunrise industry, the company's growth is not high; while some small and medium sized companies although growth is high, but the performance is poor, the high price-earnings ratio, and thus to be able to choose a good stock, you need to make investment decisions among these goals a balanced approach that more than a numerical target in a given region of the optimization problem is known as multi-objective optimization.

In order to solve multi-objective optimization problem, we need to create a general mathematical model, we must at first determine its decision variables, the general case, the decision variables n dimensional Euclidean space as a point E_n , namely:

$$x = (x_1, x_2, x_3 \dots x_n) \in E^n. \tag{3}$$

Then is the objective function, in general it can be assumed with P objective functions and decision variables are all about function, namely:

$$f(x) = [f_1(x), f_2(x), \dots, f_p(x)]^T. \tag{4}$$

Finally, its constraints, from a mathematical point of view, there are two constraints: inequality constraints and equality constraints, constraints can be defined as the m inequality constraints and k equality constraints:

$$\begin{cases} g_i(x) \leq 0 & i=1,2,3 \dots m \\ h_j(x) = 0 & j=1,2,3 \dots k \end{cases} \tag{5}$$

If all are the minimization of the objective function value, the multi-objective optimization problem can be described as the following mathematical model:

$$\begin{cases} \min f(x) = [f_1(x), f_2(x), \dots, f_p(x)]^T, \\ x_i^\alpha \leq x_i \leq x_i^\beta \end{cases}, \tag{6}$$

where, x is the decision variable, $f(x)$ is the objective function, X represents the decision vector formed by the decision space x , $g_i(x)$ and $h_i(x)$ constraints x feasible decision variables to determine the range, min represents Vector Minimization, namely vector target $f(x) = [f_1(x), f_2(x), \dots, f_p(x)]^T$ in certain constraints as far as possible the various sub-objective function minimization. It can be seen when the $p = 1$, the mathematical model for a single objective optimization problem mathematical model.

Multi-objective optimization problem is that people in the production or frequently encountered problems in life, in most cases, due to multi-objective optimization problem in all its goals are in conflict, a sub-target improvement may cause the performance of other sub-goals reduced, in order to make optimal multiple targets simultaneously is impossible, and thus in solving multi-objective optimization problem for each sub-goal can only be coordinated and compromise treatment, so that each sub-objective functions are optimal as possible multi-objective optimization problem with a single objective optimization problem is essentially different, in order to properly solve multi-objective optimization problem the optimal solution, we must first multi-objective optimization of the basic concepts of a systematic exposition.

Definition 1: N Viola Space:

$$\begin{cases} x = (x_1, x_2, x_3, \dots, x_n)^T \\ y = (y_1, y_2, y_3, \dots, y_n)^T \\ x = y \quad \text{Iff } x_i = y_i \quad \forall i=1,2,3, \dots, n \\ x > y \quad \text{Iff } x_i > y_i \quad \forall i=1,2,3, \dots, n \end{cases}. \tag{7}$$

Definition 2: Let $X \subseteq R^m$ be a multi-objective optimization model constraint set, $f(x) \in R^p$ is a vector objective function, $x^n \in X$ and x^n the X all the other points are superior, called x^n is the multi-objective minimization model optimal solution.

By definition, multi-objective optimization problem is to make the optimal solution x -vector objective function $f(x)$ for each sub-goal is to achieve the most advantages of the solution, obviously, in most cases; the optimal multi-objective optimization problem solution does not exist.

Definition 3: Pareto optimal solution: Let $X \subseteq R^m$ be a multi-objective optimization model constraint set, $f(x) \in R^p$ is the vector of the objective function. If $\xi \in X$, ξ and there is no more than the superiority of x , then ξ is a minimal model of multi-objective Pareto optimal solution, or non-inferior solution.

Definition 4: No inferior set with the front end: Let $X \subseteq R^m$ be a multi-objective optimization model

constraint set, $f(x) \in R^p$ is a vector objective function. $\lambda \in X$ Is a minimal model of multi- objective Pareto optimal solution set, then λ is called non-inferior set of X , $Y = f(\lambda)$ is called Pareto optimal front.

Seen from the above definition:

a) Multi- objective optimization problem with a single objective optimization problem is essentially different, in general, multi-objective optimization problem Pareto "optimal solution is a collection of the Mu most cases, similar to the single-objective optimization problem in a multi-objective optimal solution optimization problem does not exist, there is only Pareto optimal "multi-objective optimization problem is just a Pareto optimal solution acceptable" not bad "solution, and usually most multi-objective optimization problem with multiple Pareto optimal solution.

b) If a multi-objective optimization problem optimal solution exists, then the optimal solution must be Pareto optimal solution, and the Pareto optimal solution is also the optimal solution by only composed of these, do not contain other solutions, so can be so say, Pareto optimal solution is a multi-objective optimization problem reasonable solution set.

c) For practical application, must be based on the level of understanding of the problem and the decision-makers of personal preference, from a multi-objective optimization problem Pareto optimal solution set of one or more selected solution as a multi-objective optimization problem of optimal solution, so seeking more objective optimization problem the first step is to find all its Pareto optimal.

3.3 ROUTE OPTIMIZATION

Weight coefficient variation method is used to solve multi-objective optimization problem of the earliest methods. The basic idea is: For a multi- objective optimization problem, if for each of its sub-objective function $f_i(x)$ given different weights w_i , where the size of w_i represents the corresponding sub-goals $f_i(x)$ in a multi-objective optimization problem in an important degree, the individual sub-goals weighted linear function can be expressed as:

$$\begin{cases} f(x) = w_1 f_1(x) + w_2 f_2(x) + \dots + w_p f_p(x) = \sum_{i=1}^p w_i f_i(x) \\ w_i = \frac{random_i}{random_1 + random_2 + \dots + random_p} \end{cases} \quad (8)$$

As to the fitness function $\sum_{i=1}^p w_i f_i(x)$ roulette wheel selection can be determined by hybridization and mutation of the individual involved and so on. Thus, this method can provide a lot of random points to a valid interface search direction; this algorithm is used to Flow-shop scheduling problems, and achieved good results.

Here we have normalized the objective function, when the first t . Let $(x_1, x_2, x_3 \dots x_p)$ generation populations, $h_i(t) = \max\{|f_i(x_1)|, |f_i(x_2)|, |f_i(x_3)|, \dots, |f_i(x_p)|\}$ then $f_i(x)$ can be normalized to the new objective function:

$$g_i(x) = \frac{f_i(x)}{h_i(t)} \quad (9)$$

The new fitness function can be redefined as:

$$G(x) = \sum_{i=1}^m w_i g_i(x) \quad (10)$$

Adapted according to the size of the angle $g(x)$ with the roulette wheel selection operator $p_0(t)$ is selected from the initial population of parent and points N for hybridization, parents set point set is P :

$$\begin{cases} x_k^i = r \cdot x^i + (1-r) \cdot x^{i+1} \\ x_k^{i+1} = (1-r) \cdot x^i + r \cdot x^{i+1} \end{cases} \quad (11)$$

Where r is a random number between $[0,1]$, $[x^i, x^{i+1}] \in P$, $[x_k^i, x_k^{i+1}] \in p(t)$, $p(t)$ is set after hybridization offspring.

Let $x = (x_1, x_2, x_3 \dots x_n)$ be a parent population of the body $p(t)$, $j = 1, 2, 3, \dots, n$, x_j is $x \in p(t)$, page j coordinate; $a_j < x_j < b_j$, a_i and b_i , respectively, of the search points in the space coordinate j The lower and upper bounds; $x_j^k \in p(t)$ is the first coordinate j , $p(t)$ is the first step Offspring collection; r is between $[0,1]$ random number between; T is the maximum genetic algebra; t for the current genetic algebra; $r(0,1)$ said that produce 0 or 1 random number. There are:

$$\begin{cases} x_j^k = x_j + (b_j - a_j) \cdot r \cdot \left(1 - \frac{t}{T}\right) & \text{if } r(0,1) = 0 \\ x_j^k = x_j - (b_j - a_j) \cdot r \cdot \left(1 - \frac{t}{T}\right) & \text{if } r(0,1) = 1 \end{cases} \quad (12)$$

Non-uniform mutation is to participate in the weight variation did a random disturbance; this disturbance in the early evolution of a relatively large range, but with the evolution of generation increases, changes in disturbance gradually decreases.

What is part by a description of the problem shows that the mesh size small enough to cover the entire area can be approximated as point coverage, but also to guarantee coverage performance. It proved that when the sensor node communication radius of R_c is greater than or equal to twice the detection radius R_s , the algorithm constructs a cover set is able to guarantee the network connectivity; contrary cannot be guaranteed. To this end, for $R_c < 2R_s$, this paper designed a low time complexity algorithm to maintain network connectivity, as shown in Figure 4:

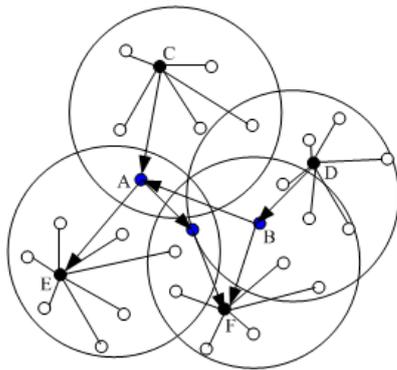


FIGURE 4 Schematic route optimization

Network to obtain the minimum working set of nodes, the central node can take advantage of the depth-first search (depth first search, DFS) or breadth-first search (breadth-first search, BFS) to determine the working node connectivity. If the work can not constitute a connected network nodes, then we put the work could form the set of nodes connected region called the work connected subset of nodes. Therefore, the work of each subset of nodes connected nodes is able to keep working connectivity. Then find work any two connected subset of nodes nearest two working nodes A and B, shown in Figure 4. A wake it working node communication distance range and the nearest node from the Node B. For example the node C is awakened, if the Node B is still in the communication range of the node C, the node C will wake it up and away from the communication range of the node B nearest node until the node A and B can constitute a communicating. , The network connectivity of a further judgment. If there is not a set of nodes connected work, we then calculated above, until the entire network to maintain communications.

4 Evaluation systems

In order to evaluate the feature of the algorithm, this paper MATLAB 6.5 is adopted as a simulation platform in this paper, the sensor nodes are randomly deployed in different network areas, the parameters are included in Table 1.

TABLE1 Simulation parameters

parameter	value	parameter	value
dimension 1	100*100m ²	ϵ_{amp}	20(pJ/b)/m ²
dimension 2	200*200m ²	E_{R-elec}	30nJ/b
dimension 3	400*400m ²	E_{min}	0.05J
Number	80-350	Header	20B
R_s	5m	Initial energy	5J
E_{T-elec}	20nJ/b	broadcast	20B
ϵ_{fs}	10(pJ/b)/m ²	each round	100ms

The wireless communication models for Sensor node transmitting data and receiving data are respectively the following:

$$E_{Tr}(k, d) = E_{T-elec}k + E_{amp}(k, d) = \begin{cases} E_{T-elec}k + \epsilon_{fs}d^2k & d < d_0 \\ E_{T-elec}k + \epsilon_{amp}d^4k & d \geq d_0 \end{cases} \quad (13)$$

In the above formula, E_{T-elec} and E_{R-elec} denote the energy consumption of wireless transmitting module and wireless receiving module; ϵ_{fs} and ϵ_{amp} stand for the energy consumption parameters of spatial model and multiple attenuation models; d_0 is a constant.

Experiment I: The first case is, with the same respective parameters, execute 50 times and get the mean value, then execute for 400 to compare with the LEACH and LEACH-C [9] protocol the quantitative relationship between number of remaining nodes and the number of turns, as shown in Figure 5:

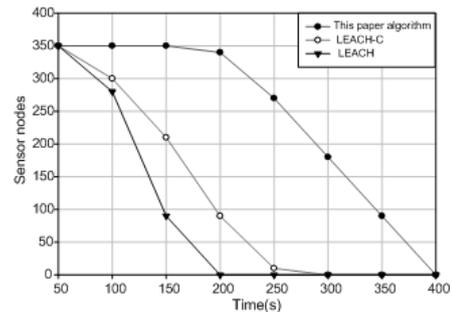


FIGURE 5 Remaining nodes and the round number

As can be seen from Figure 5, with increasing of time, the number of remaining nodes of proposed algorithm is higher than the LEACH protocol, and then the conclusion that with the increasing of time, the energy consumption of the proposed algorithm is lower than that of LEACH protocol, and the network lifetime is extended, also the network resources are optimized.

In order to achieve the scale of network coverage, and thus better evaluate the performance of the model in different sizes, which mainly reflect the minimum number of nodes needs to by deploy in different network coverage, each simulation experiment executed 50 times at average. Curve of node coverage changes is shown in Figure 6:

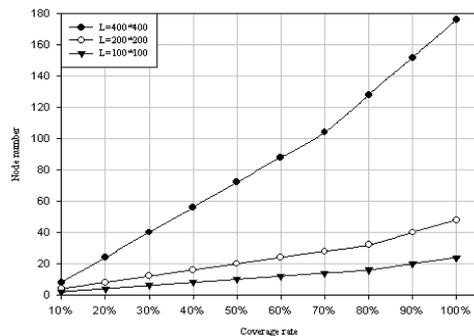


FIGURE 6 Coverage rate for different coverage area

Figure 6 shows the graph of the number of sensor nodes needed to deploy to achieve different node coverage under different network dimensions. The figure shows that, with the expansion of the network, to meet the demand for network coverage, the number of nodes required to be deployed will increase, and the higher the coverage of the network, the number of nodes need to be deployed increases can be obtained from Figure 6 more fast, so that the concern target node can achieve complete coverage.

This algorithm were compared with LEACH protocol, the situation at the same time and the same events occurred mainly as compared to another situation is to choose a random deployment area $400m \times 400m$ sensor nodes placed within $400m$, compare LEACH protocol in the network operation the number of surviving nodes.

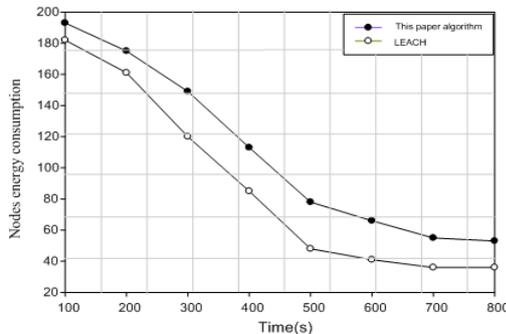


FIGURE 7 NETWORK energy consumption

Figure7 reflects the same number of rounds as having energy consumed by this algorithm is less than LEACH protocol, network energy consumption with time and showing a stable homogeneous state, compared with LEACH protocol is more energy saving, thereby prolonging the network lifetime cycle.

5 Conclusions

Based on in-depth study of the LEACH protocol, for its shortcomings, the improved method, the purpose is to balance node energy consumption and prolong the network life cycle, for the improvement of LEACH protocol is mainly reflected in: select optimal number of clusters heads, helps to reduce energy consumption and improve overall network performance through improved

References

- [1] Zhu R 2011 Intelligent Rate Control for Supporting Real-time Traffic in WLAN Mesh Networks *Journal of Network and Computer Applications* **34**(5) 1449-58
- [2] Wang Y, Jiao Y, Zhang F 2003 Uniform and orthogonal genetic algorithm for multi-objective optimization *Journal of Systems Engineering* **18**(6) 481-6 (in Chinese)
- [3] Jinhua Z, Zhongzhi S, Yong X 2004 A Fast Multi-Objective Genetic Algorithm Based on Clustering, *Journal of Computer Research and Development* **41**(7) 1081-7
- [4] Wang Y, Jiao Y, Zhang F 2003 Uniform and orthogonal genetic algorithm for multi-objective optimization *Journal of Systems Engineering* **18**(6) 481-6 (in Chinese)
- [5] Sun Z, Li Z 2013 Wireless Sensor Network Path Optimization Based on Hybrid Algorithm *Telkommika* **11**(9) 5352-8
- [6] Fang W, Liu F, Yang F 2010 *IEEE Transactions on Consumer Electronics* **2010** **56**(4) 2185-92
- [7] Sun Z, Yu Y 2013 Coverage Control Algorithm Based on Ant Colony Algorithm in Wireless Sensor Networks *International Journal of Applied Mathematics and Statistics* **50**(20) 467-75
- [8] Xiao Y, Chen H, Wu K, Sun B, Zhang Y, Sun X, Liu C 2010 *IEEE Transactions on Computers* **59**(4) 507-21
- [9] Handy M J, Haase M, Timmermann D 2009 LEACH-C: Low Energy Adaptive Clustering Hierarchy with Deterministic Cluster-head Selection 4th *IEEE International Conference on Mobile and Wireless Communication Network* 9 368-72
- [10] Liang L, Xi Z, Huang M 2013 *IEEE Sensors Journal* **14**(3) 3625-36

residual energy more evenly distributed throughout the network, thus avoiding the low energy due as a cluster head node and the phenomenon of excessive energy consumption, better balance energy consumption of the nodes, extending the network life cycle [10]. Adopt multi-hop communication between clusters, the distance between the base station to avoid a free remote cluster head node due to a single hop communication distance and excessive consumption of energy. so evenly distributed in the cluster head node sensor networks, avoiding irrational energy consumption and the impact of network life cycle research work of this paper is not perfect, needs to be improved further exploration, the next step requires in-depth research in the following areas:

1) LEACH algorithm of this paper is based on a single-layer network topology structure, in future studies; we should adopt the improved algorithm, which was applied to a multi-level network topology.

2) The cluster nodes to the cluster head node using single hop communication, this will cause the cluster node energy depletion and end points quickly stop working, resulting in lower system life, may consider the use of a time-based within a cluster single-hop and multi-hop component hybrid routing methods.

Acknowledgements

Science of Technology Research of Foundation Project of Henan Province Education Department under Grant Nos.2014B520099; Natural Science and Technology Research of Foundation Project of Henan Province Department of Science under Grant Nos. 142102210471. The Key Science-Education Project of the Henan province 'Twelfth Five-Year-Plan' under Grant No. 2011-JKGHAB-0051.

Authors



Zeyu Sun, born in 1977, Changchun, Jilin, China

Current position, grades: Lecturer in Luoyang institute of technology of computer and information engineering. Member of China computer society.
University studies: Master of Science at Lanzhou university in 2010. PhD at Xi'an Jiaotong university.
Scientific interest: Wireless sensor networks, parallel computing and Internet.



Shouying Li, born in 1979, Xinxiang, Henan, China

Current position, grades: Lecturer in Luoyang institute of science and technology of mathematics and physics, China.
University studies: Master of logistics engineering Southwest Jiaotong University in 2009.
Scientific interest: optimization algorithm, logistics system optimization.

Computational simulation of the effects of vortex finder diameter on the air core in a hydrocyclone separator

Yuekan Zhang, Peikun Liu*, Linjing Xiao, Xinghua Yang, Junru Yang

College of Mechanical & Electronic Engineering Shandong University of Science and Technology, 266590, Qingdao, China

Received 1 May 2014, www.cmnt.lv

Abstract

Air core is inherent to the solid-liquid hydrocyclone, the air core dimension plays a significant role in the separation efficiency. However, the formation mechanism of the air core in hydrocyclone has not arrived at an agreement. To further understand the flow behaviour of air core in hydrocyclones, in this paper, the volume of fluid (VOF) multiphase model and the Reynolds Stress Model (RSM) were adopted in this study to simulate the flow fields inside a hydrocyclone. The effect of the varying vortex finder diameter on the formation and development of air core was analysed, and the generation and development of the air core were investigated. The results showed that air core could be generated in shorter time and more stable state with larger vortex finder diameter. In addition, the diameter of the air core increased with the vortex finder diameter.

Keywords: hydrocyclone; computational fluid dynamics; VOF multiphase; diameter of vortex finder; air core

1 Introduction

The hydrocyclone is widely used in the mineral, chemical, oil processing industries, because of its high capacity, low maintenance, low operating costs, the design and operational simplicity, and the small physical size of the device. Since 1950, a large number of literatures about the hydrocyclone have been published, including empirical and semi-empirical equations for predicting the equipment performance, mathematical models describing flow distributions [1-5]. After many years of application, the outstanding merits of hydrocyclone have been well acknowledged in increasing application areas. With the development of computational fluid dynamics (CFD), simulation of the flow field inside hydrocyclone has been widely promoted. By using Reynolds stress model (RSM) and volume of fluid (VOF) model, Wang B [6] numerically simulated the gas-liquid two-phase flow inside hydrocyclone. Gupta R et al. [7] studied the development of air core and its influences on the flow field in hydrocyclone. S. M. Mousavian et al. [8] compared the $k-\epsilon$ model, the Reynolds stress model, and the Reynolds stress model with the VOF model for simulating the air core flow fields. Air core is inherent to the solid-liquid hydrocyclone, plays a significant part in the classification efficiency in the hydrocyclone. Till now, the formation mechanism of the air core in hydrocyclone has not arrived at an agreement [9]. To further understand the flow behavior of air core in hydrocyclone, multiphase flow model and RSM model were adopted in this study to predict the flow distribution and formation mechanism of air core inside a hydrocyclone. In this paper, the effect of the varying vortex finder diameter on the formation and development of air core was analysed, and the generation and development of the air core were investigated. The results showed that air core could be generated in shorter time and more stable state

with larger vortex finder diameter. In addition, the diameter of the air core increased with the vortex finder diameter.

2 Numerical model and its boundary conditions

2.1 GEOMETRY AND MESH

The structural parameters and diagram of the geometrical model are as shown in Table 1 and Figure 1. The diameters of tubes for the feed inlet, overflow, underflow and main body of the hydrocyclone are 16.7, (25, 35, 40, 50), 20, 100mm, respectively. Mesh independence analysis were discussed with mesh size 75000, 10000, 125000 and 200000. The results showed that numerical results were independent of the total number of computational cells. In the rest of this paper, we took into account the accuracy of numerical results and the computational time, the total computational domain used hexahedron structure grids, including about 125000 cells for computational domain, and the grids at the centre axis and the surrounding surface were refined (Figure 2).

2.2 BOUNDARY CONDITIONS AND CONTROL PARAMETERS

RSM turbulence model was used in consideration of the turbulence characteristics of high-speed rotation inside the hydrocyclone. Meanwhile, taking the hydrocyclone's inherent multi-phase flow feature into consideration, VOF model was used due to its suitability regarding to gas-liquid two-phase flow. The principal phase was set as water in the VOF model, and the secondary phase was air. The hydrocyclone inlet was set as velocity inlet, and the feeding velocity was assumed to be constant with a magnitude of 6 m/s. The volume fractions of water and air were set as 1 and 0, respectively. The fluid velocity was set perpendicularity to

*Corresponding author e-mail: lpk710128@163.com

the entry section and pointing to internal side. Both the overflow and the underflow outlets were defined based on pressure, and they connect to the atmosphere (zero gauge pressure). The air backflow ratios at both the overflow and underflow ports were set as 1, i.e. air was the only reverse current. No-slip boundary conditions were applied on all walls, and near-wall treatment was applied to the wall using standard wall functions. The computational CFD code FLUENT 6.3 (Fluent Inc., USA) was utilized to solve the continuity equation and Navier-Stokes equation and boundary conditions. This model adopted pressure-based implicit transient three-dimensional solvers. The control parameters were solved by using pressure-velocity coupling SIMPLE method. A pressure-discretization set of simulations was

performed with the pressure staggered option (PRESTO), which works well for high-speed rotation flows. The momentum discretization format was carried out with a higher-order quadratic upwind interpolation (QUICK) spatial scheme. Geometric reconstructive discretization format was used for determining the two-phase volume fractions. First-order upwind format was used for determining the turbulence energy, turbulence dissipation rate and Reynolds stress discretization scheme. Initially, the hydrocyclone was filled with air, i.e. the air volume fraction inside the hydrocyclone was 1. The termination criterion for the computational accuracy was set as maximum relative residual errors of 10^{-5} , unsteady computational method was used, and time increment was defined as 0.0001 s.

TABLE 1 Main structural parameters of the hydrocyclone

Structural parameter of the hydrocyclone	Structural dimensions
Hydrocyclone diameter D (mm)	100
Vortex finder diameter of the hydrocyclone d_0 (mm)	25, 35, 40, 50
Underflow port diameter of the hydrocyclone d_u (mm)	20
Feed inlet dimensions of the hydrocyclone $a \times b$ (mm)	25×12.5
Feed inlet equivalent diameter of the hydrocyclone d_i (mm)	16.7
Overflow pipe insertion depth of the hydrocyclone H4 (mm)	50
Thickness of the overflow pipe of the hydrocyclone S (mm)	7.5
Extension length of the overflow pipe of the hydrocyclone H5 (mm)	30
Cylinder height of the hydrocyclone H1 (mm)	90
Cone section height of the hydrocyclone H2 (mm)	227
Underflow pipe height of the hydrocyclone H3 (mm)	50

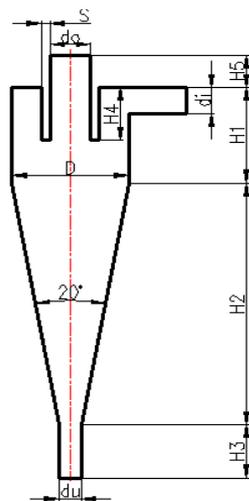


Fig.1. Structural diagram of the hydrocyclone

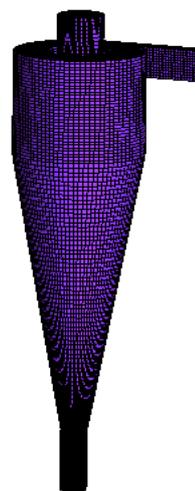


Fig.2. Mesh structure of the hydrocyclone

3 Simulation results and discussion

Simulation completed when all physical parameters in the flow field remained stable. Figure 3 shows the formation and development of the air core in the hydrocyclone during the startup process obtained from CFD simulations. The blue area in Figure 3 represents air, red represents water. In the initial stage, the hydrocyclone was filled with air, i.e. air volume fraction was 1. This value gradually decreased over time, correspondingly, the fluid volume fraction increased. Along with the development of the flow field, negative pressure was generated near the centre axis of the hydrocyclone. Since the outlet of the hydrocyclone was connected to atmo-

sphere, air intake occurred at both the overflow and underflow ports, and air core firstly appeared at the bottom of the overflow outlet. Subsequently, air core at underflow and overflow outlets developed toward each other and finally interconnected.

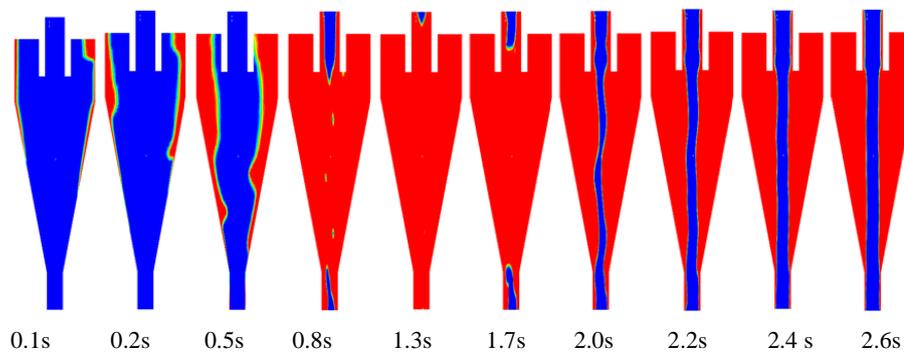
According to Figure 3, the diameter of the overflow pipe affects the air core significantly. Figure 3(a) shows the air core generating process when the diameter of the overflow pipe was set as 25 mm. It is observed that air core structure was formed in about 2.0 s, when it was still unstable at the moment. Gradually developing with its separating process, the air core reached a pseudo-stable state after about 2.6 s of operation and appeared to be in very regular shape with

minimal oscillations. Figure 3(b) shows the transient development of the air core structure in the hydrocyclone when the diameter of the overflow outlet was set as 35 mm. It was observed that the air core structure stabilized after about 2.4 s of simulation, which is obviously shorter in time compared to the 25 mm-diameter-overflow hydrocyclone. Figure 3(c) shows the changing process of the air core structure when the diameter of the overflow outlet was increased to 40 mm. From this figure it can be seen that the air core structure is formed in about 1.4 s, and soon reached a steady state itself. Figure 3(d) shows the changing process of the air core structure when the diameter of the overflow pipe was set as 50 mm. It is obvious that the air core stabilized within the shortest time, which was only 1.7 s.

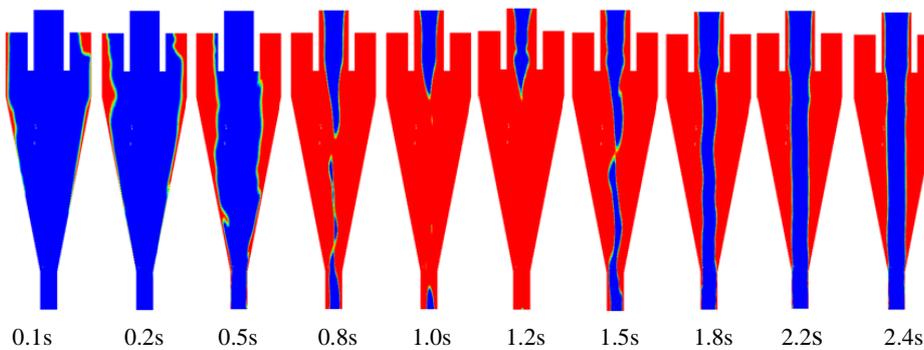
Overall, the diameter of the stabilized air core increased with the diameter of the overflow outlet. The results also showed that the diameter of the overflow outlet affected the air core to some extent, from its generation to stabilization: the larger diameter of the overflow pipe, the shorter time the

air core needed to develop. The air core stabilized itself after about 2.6 s for a 25 mm-diameter-overflow pipe, while for the 50 mm one it stabilized after only about 1.7 s. From the perspective of the air core regularity degree, the smaller diameter of the overflow outlet, the more regular shape the air core had.

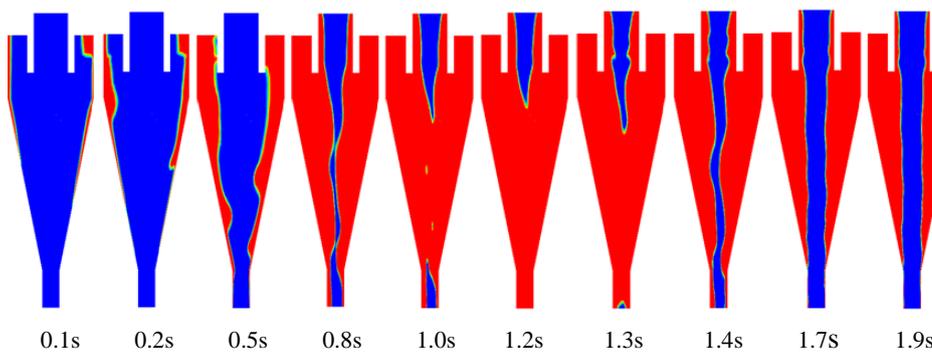
To verify that the generation of the air core is indeed contributed by both the overflow and the underflow outlets, the hydrocyclone with 35 mm-diameter-overflow was set as example as shown in Figure 4. All the parameters and setups remained the same, because the underflow port was blocked by water rather than being connected to atmosphere, and the iteration continued after the air core stabilized. Figure 4 shows that the air core exhibited instability with time, and disconnected after about 2.8s. After that, it gradually disappeared till after about 3.0s, when the air core around the underflow port totally vanished. This phenomenon was due to the disconnection to atmosphere and negative pressure in the hydrocyclone is the prerequisite of forming air core.



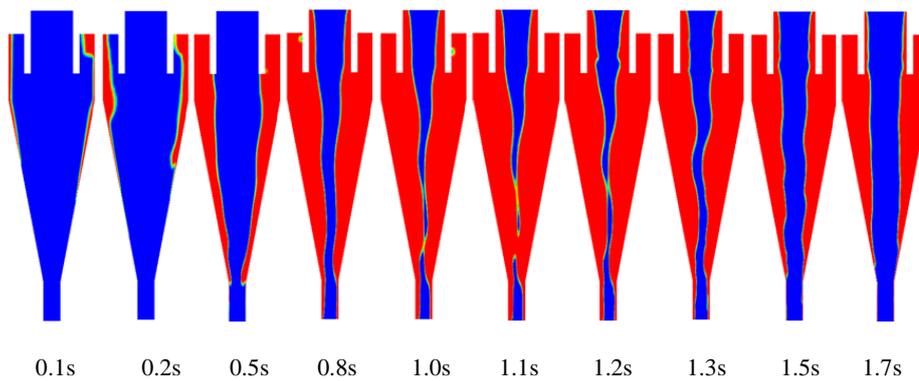
(a) Overflow pipe diameter is 25 mm



(b) Overflow pipe diameter is 35 mm



(c) Overflow pipe diameter is 40 mm



(d) Overflow pipe diameter is 50 mm

Fig.3. Effect s of overflow pipe diameter on air core

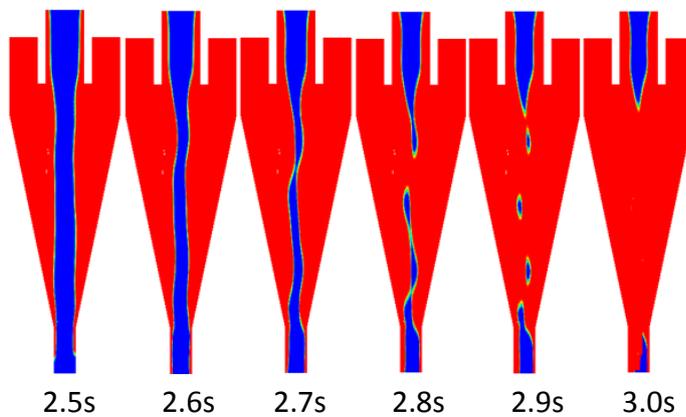


Fig.4. Effects of water-blocked underflow port on air core

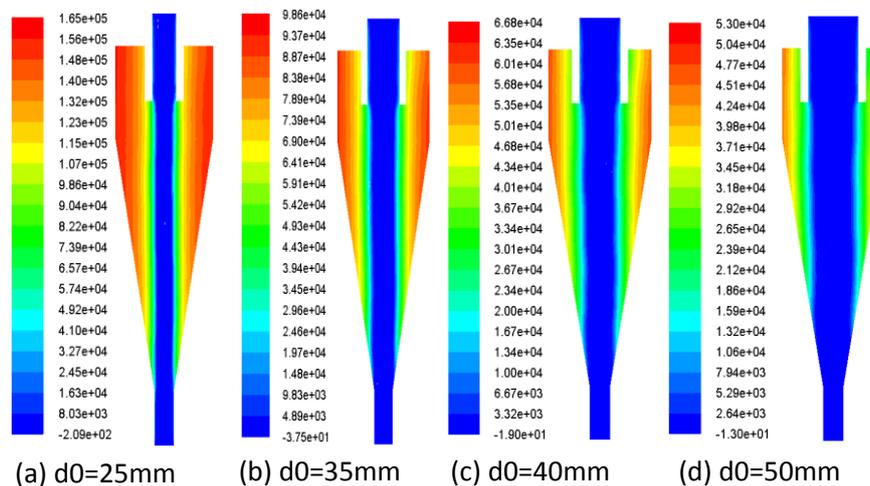


Fig. 5.Static pressure distributions in the hydrocyclone for different diameter of overflow outlet diameters

To better explain the effect of the overflow outlet on the air core, the pressure distributions inside the hydrocyclone with varying overflow outlet diameters were compared as shown in Figure 5. It is shown clearly in Figure 5 that the pressure inside the hydrocyclone was about 165000 Pa when the overflow pipe diameter is 25 mm. The pressure gradually decreased with increasing overflow pipe diameter; for example, the pressure dropped to only 53000 Pa when

the overflow diameter increased to 50 mm. Estimated from these results, the larger overflow pipe diameter, the shorter time during, which negative pressure exists, i.e. the shorter time air core appears. It can also be noticed in Figure 5 that as the overflow diameter increased, larger radial area was featured by negative pressure, further proving the conclusion above: the air column diameter increases with the overflow diameter.

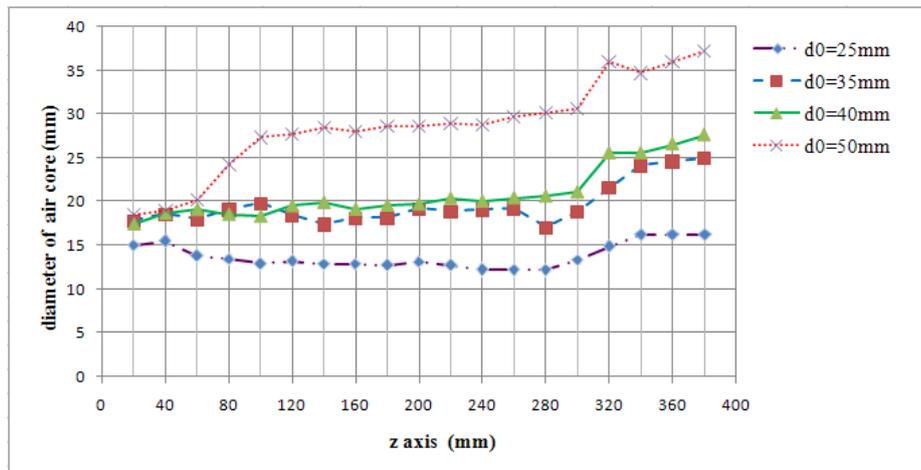


Fig.6 Diameter of air core in the hydrocyclone at z axis

In order to study the effect on the air core with different diameter of vortex finder, Figure 6 showed comparisons of the value of air core diameter with different vortex finder at various axial positions of the hydrocyclone obtained from CFD simulations. It is observed that movement rule of air core inside hydrocyclone could be divided into four parts: the first one is the underflow pipe (from $z=0$ to $z=50$ mm), the air core in this part remain relatively stable, the air core diameter remain basically unchanged at z axis. At vertical position $z=50$ mm, when the vortex finder diameter is $d_0=25, 35, 40, 50$ mm, respectively, the corresponding value of air core diameter in the hydrocyclone is 15.50, 18.57, 18.68, 18.98mm, respectively; The second part is conical section (from $z=50$ mm to $z=277$ mm), the results showed that the diameter of air core there are abrupt changes at the junction of the cylindrical section and conical section of the hydrocyclone. And it is also found that there is a couple of small fluctuation of the air core diameter within conical section. But the change in the rules of air core within this section was not similar. At vertical position $z=227$ mm, when the vortex finder diameter is $d_0=25, 35, 40, 50$ mm, respectively, the corresponding diameter of air core in the hydrocyclone is 12.74, 18.89, 20.38, 28.96mm, respectively; The third part is cylindrical section (from $z=227$ mm to $z=317$ mm), from the Fig.6 above mentioned we can see with the z axis increasing, the diameter of the air core in various vortex finder of the hydrocyclone was increasing trend, and the tendency towards increasing in value was obvious. This phenomenon indicated that the air core movement in this section was not stable. At vertical position $z=317$ mm, when the vortex finder diameter is $d_0=25, 35, 40, 50$ mm, respectively, the corresponding diameter of air core in the hydrocyclone is 14.92, 21.65, 25.55, 35.94mm, respectively; The fourth section is above the vortex finder insertion depth, it was found that with the z axis increasing, the diameter of the air core in various vortex finder of the hydrocyclone was increasing trend, but increased indistinctively with the increasing of z axis. At the top of vortex finder ($z=380$ mm), the diameter of air core reached the maximum value. when the vortex finder

diameter is $d_0=25, 35, 40, 50$ mm, respectively, the corresponding diameter of air core in the hydrocyclone is 16.25, 25.00, 27.61, 37.17mm, respectively

4 Conclusions

In this paper, the computational fluid dynamics (CFD) method was used to simulate the flow fields in a $\Phi 100$ mm hydrocyclone. The flow filed patterns inside the hydrocyclone was revealed and the generation and development of the air core inside the hydrocyclone were discussed. In addition, the simulation helped to improve the understanding of how the vortex finder diameter affects the air core, and the results could be used as a theoretical basis to optimize hydrocyclone design.

(1) The vortex finder diameter significantly affects the dimension and shape of the air core. As the diameter of the vortex finder increased, the diameter of the produced air core increased, and the time for the air core to stabilize became shorter.

(2) As the air core increased in diameter, the pressure-drop and the split ratio inside the hydrocyclone decreased, thus improving the separation performance.

(3) Negative pressure in the hydrocyclone is the prerequisite of air core. The axis of the air core almost superposed the hydrocyclone's. The dimension and shape of the air core changed with the flow field. Even in a stable flow field, the air core still had different dimensions at different positions of the axis. Analysis showed that this could partially because of the unstable flow field, and partially because the structure variation of the hydrocyclone itself.

Acknowledges

This work is supported by the National Natural Science Foundation of China (No.21276145) and guiding program for scientific and technological research of China National Coal Association (No. MTKJ2013-365).

References

- [1] Kelsall D F 1952 A study of the motion of solid particles in a hydraulic cyclone *Trans. Inst. Chem. Eng.* **30** 87–108
- [2] Bradley D 1965 *The Hydrocyclone*. Pergamon, London
- [3] Bloor M I G, Ingham D B 1972 On the efficiency of the industrial cyclone *Trans. Inst. Chem. Eng.* **51** 173–6
- [4] Svarovsky L 1984 *Hydrocyclones* 1st edn. Holt, Rinehart, and Winston, London
- [5] Barrientos A, Sampaio R, Concha F 1993 Effect of the air core on the performance of a hydrocyclone *In: Proceedings XVIII International Mineral Processing Congress* 267–70
- [6] Wang B, Chu K W, Yu A B 2007 Numerical study of particle-fluid flow in a hydrocyclone *Ind. Eng. Chem. Res.* **46**(13) 4695-705
- [7] Gupat R, Kaulaskar M D, Kumar V, et al. 2008 Studies on the understanding mechanism of air core and vortex formation in a hydrocyclone *Chem Eng J.* **144** 153-66
- [8] Mousavian S M, Najafi A F 2009 Numerical simulation of gas-liquid-solid flows in a hydrocyclone separator *Arch Appl Mech* **79** 395-409
- [9] Xu jirun, Luo qian 1988 *Flow fields theory of the hydrocyclone* Bijing (China) Science Press

Authors	
	<p>Yuekan Zhang, born on March 13, 1970, Xinji, Hebei province, China</p> <p>Current position, grades: Instructor of College of Mechanical & Electronic Engineering Shandong University of Science and Technology. University studies: College of Mechanical & Electronic Engineering Shandong University of Science and Technology. Scientific interest: Computational Fluid Dynamics, Solid-liquid separation technology and equipment Publications: 25 papers Experience: I attended the Beijing university of Science and Technology, China from 1988 to 1992. I earned BS degree in mining machinery in 1992, and the M. degree in mechanical design and theory from Shandong University of Science and Technology, China in 2007. Since 2011 I have been working towards the Ph.D. degree in mineral process engineering at the Shandong University of Science and Technology, China. My research interests concentrate on the Computational Fluid Dynamics and solid-liquid separation technology and equipment.</p>
	<p>Peikan Liu, born on January 28, 1971, Wenshang, Shandong province, China</p> <p>Current position, grades: Professor of Shandong University of Science and Technology. University studies: College of Mechanical & Electronic Engineering Shandong University of Science and Technology. Scientific interest: Computational Fluid Dynamics, Solid-liquid separation technology and equipment Publications: 30 papers Experience: I attended the Chengdu university of Science and Technology, China from 1989 to 1993. I earned BS degree in 1993 and M. degree in 1996 in chemical machinery from Sichuan University, and I earned D. degree in mineral processing from Wuhan University of Science and Technology China in 2008. I was with the College of Mechanical & Electronic Engineering Shandong University of Science and Technology, China. My research interests concentrate on the Computational Fluid Dynamics and solid-liquid separation technology and equipment.</p>
	<p>Linjing Xiao, born on January 18, 1966, Yishui, Shandong province, China</p> <p>Current position, grades: Professor of College of Mechanical & Electronic Engineering Shandong University of Science and Technology. University studies: College of Mechanical & Electronic Engineering Shandong University of Science and Technology. Scientific interest: virtual prototype and CAD/CAE/CAE Publications: 55 papers Experience: Linjing Xiao, born in 1966, is currently a professor at Shandong University of Science and Technology, China. He received his PhD degree from Beijing University of Science and Technology, China, in 2001. His research interests include virtual prototype and CAD/CAE/CAE.</p>
	<p>Xinghua Yang, born on October 25, 1978, Luoshan, Henan province, China</p> <p>Current position, grades: Instructor of College of Mechanical & Electronic Engineering Shandong University of Science and Technology. University studies: College of Mechanical & Electronic Engineering Shandong University of Science and Technology. Scientific interest: Computational Fluid Dynamics, Solid-liquid separation technology and equipment Publications: 28 papers Experience: Yang xinghua, born in 1978, is currently Instructor at Shandong University of Science and Technology, China. She received her M. degree from Shandong University, China, in 2001. Her research interests in Computational Fluid Dynamics, Solid-liquid separation technology and equipment</p>
	<p>Junru Yang, born on October 4, 1969, Leting, Hebei province, China</p> <p>Current position, grades: associate Professor of College of Mechanical & Electronic Engineering Shandong University of Science and Technology. University studies: College of Mechanical & Electronic Engineering Shandong University of Science and Technology. Scientific interest: Cladding part design and CAD/CAE Publications: 40 papers. Experience: Yang Junru, born in 1969, is currently an associate professor at Shandong University of Science and Technology, China. She received her PhD degree from Shandong University, China, in 2006. Her research interests include cladding part design and CAD/CAE.</p>

Research on ontology mapping of tourism information resources based on description logic

Jianhou Gan^{1*}, Bin Wen², Jinxu Li²

¹Key Laboratory of Education Informalization for Nationalities, Ministry of Education Yunnan Kunming, China

²School of Information Science and Technology, Yunnan Normal University, Yunnan Kunming, China

Received 12 June, www.tsi.lv

Abstract

In this paper, the Ontology mapping of tourism information resources is discussed. Ontology mapping tries to find the corresponding relationships between two entities, and to achieve interoperability and information sharing in heterogeneous Ontology. We researched the Ontology mapping of tourism information resources and introduced the Description Logics to solve the Ontology mapping. The key relationships and determining relationships among Ontologies in description logic based Ontology mapping is first described. Then a description logic based mapping model is proposed, which can solve the problem of Ontology semantic heterogeneity.

Keywords: tourism information resources, resource integration, ontology isomerism, ontology mapping, description logic

1 Introduction

Information technology contributes greatly to the development of tourism. With the rapid development of global travel e-commerce, as a basic platform for tourism, tourism websites also play an important role in the development of travel e-commerce. However, most of the existing travel information sites have some shortcomings, such as do not update information timely, classifying information illogically, describing resources subjectively, and displaying content incompletely or with limited exhibition ways. Therefore, clients may misunderstand the information or even be misled by the information provided by the websites. The basic reason of these problems is the disorganized information. The current tourism information providers cannot provide interactive and participatory contents for the users; they are also lack of contents, which are systematic and integrative. Problems of the existing tourist information resources are scattered, for example, content sharing is usually poor, low development and utilization level, and inferior quality of services still exist. These problems seriously restrict further development of regional tourism development. In this article, based on Description Logic, we proposed an Ontology mapping model of tourism information resources to overcome the existing problems in tourism information applications [1-3].

An ontology is a "formal, explicit specification of a shared conceptualisation"[4]. An ontology provides a shared vocabulary, which can be used to model a domain, that is, the type of objects and/or concepts that exist, and their properties and relations. Ontologies are the structural frameworks for organizing information and are used in artificial intelligence, the Semantic Web, systems

engineering, software engineering, biomedical informatics, library science, enterprise bookmarking, and information architecture as a form of knowledge representation about the world or some part of it.

In this paper, we consider the tourism information resource ontology building way and built them, as well as put forward the ontology mapping way based on description logics, which better resolved the problem of tourism information resource ontology semantic heterogeneity.

2 Building tourism information resource Ontology

2.1 TARGET OF CONSTRUCTING TOURISM INFORMATION RESOURCE ONTOLOGY

Tourism involves six elements, such as catering, accommodation, sightseeing, transportation, shopping and entertainment. The target of constructing tourism information resource Ontology is to make tourism information integrative and intelligent. This will make online searching become easier for visitors, which can simplify and speed up the process of tourism trade, and also can provide more information about tourism services. The intelligent process can provide wider and more accurate travelling information and personalized services for users.

2.2 PROCESS OF CONSTRUCTING TOURISM INFORMATION RESOURCE ONTOLOGY

To develop Ontology of tourism information resources, several processes are included:

- 1) Determining fields and categories of tourism information resource Ontology;

* *Corresponding author* e-mail: kmganjh@yahoo.com.cn

- 2) Listing important terms, concepts in tourism information resource Ontology;
- 3) Establishing framework of tourism information resource Ontology;
- 4) Defining classes and classes hierarchy system;
- 5) Defining attributes slots and value types in classes;
- 6) Encoding domain Ontology and making it be formal.

amusements, national customs, travel agencies, scenic spot management agencies, transport enterprises, accommodation enterprises, tourism bureau, insurance companies, specialty enterprises, entertainment industries, hydrological landscape, physiographic landscape, cultural landscape, history heritages, national intangible cultural heritages, full lines, pick lines and so on.

2.3 DETERMINING SCOPE AND TERMS IN ONTOLOGY

Before building an Ontology, main contours of the target Ontology is described, which makes the purpose, scope, representation and application of the Ontology clear. The intermediate result is used to describe detailed instruction of Ontology. In this work, related terms and concepts based on determining range of Ontology are listed.

In tourism information resource Ontology, important terms and concepts include: people, organization agencies, scenic spots, travelling routes, transportation, accommodation, traveling, location, specialties,

2.4 DEFINING CLASS AND CLASS HIERARCHY SYSTEM

Class is used to describe abstract entities objects, representing a kind of instances objects, which have common characters. Class has the property of inheritance, formed by hierarchical organization structure. Top class represents the most abstract entity concept, and subclass inherits abstract properties from their parents, which represent entities concepts, which are more specific or narrower than their parent. Adopted top-down approach to define class hierarchy, part of class hierarchy is described in Figure 1.

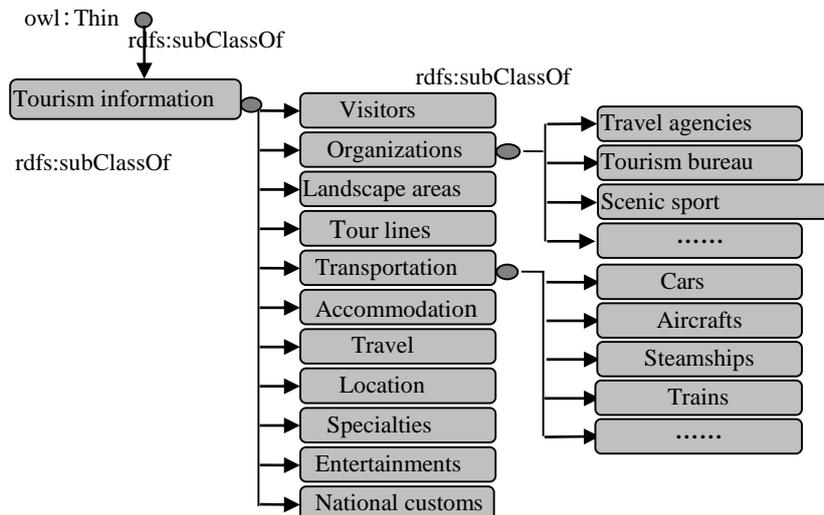


FIGURE 1 Part figure of class hierarchy

2.5 DEFINITION ATTRIBUTES OF CLASS

It is not enough to perform knowledge of field if only having name of each class. Therefore, attributes (properties) for each class are needed. Each class may have a large number of properties. The principle is to define properties of class in corresponding field based on demands. For example, the attributes of scenic spot and tourists can be expressed as:

Tourists (name, gender, ID, age, travel category, hobbies, phone, E-mail).

Scenic spot (name, attractions degree, management agency, category, address, capacity, phone)

2.6 ONTOLOGY ISOMERISM

Nowadays, Ontology has been wide used, and plays a vital role in many applications, such as database integration, P2P system, electronic commerce, Web service and social

networking. As web is the decentralization for cross-cutting areas even in the same field. Therefore, there are always multiple individuals and there are differences produced among different Ontology, which is called heterogeneity [5]. Since multi-Ontology is needed to assist the work, Ontology isomerism will occur through Ontology mapping. Problems of sharing and reusing knowledge, between different Ontology will be solved.

3 Description Logic based on Ontology mapping

3.1 THE RAISE OF ONTOLOGY MAPPING PROBLEMS

Ontology mapping tries to find the corresponding relationships between two entities, and to achieve interoperability and information sharing in heterogeneous Ontology. Ontology mapping is one of the key technologies in Ontology, aims to achieve information

integration based on Ontology, where the most important process is to find semantic associations.

The task of Ontology mapping can be simply described as follows: suppose there are two Ontology, A and B, for every concept in A, we try to find a counterpart concept in B, which has the same or similar properties with the concept in A [6]. Ontology mapping is not a unified expression of Ontology and data, but to achieve conversion between instances according to semantic relation on conceptual level.

Ontology mapping process can be divided into three parts, which are finding mapping, expressing mapping and executing mapping. "finding" is to find related, similar concepts, properties and relationships from two different Ontology through the methods, which can be manual, semi-automatic or automated. "expressing" uses language to express previous mapping relationships; "executing" completes the conversion from source Ontology instances to target Ontology instances, based on mapping relationships.

The types of Ontology mapping are: mapping between Concept-Concept, Property-Property, Property-Concept and so on [7]. Mapping relationship between Concept-Concept, refers to two concepts in different Ontology expressing the same types of information, they are consistent; Mapping relationship between Attribute-Attribute refers to two properties in different concepts indicating the same information; Mapping relationship between Attribute-Concept, refers to the properties of a concept in Ontology express the same information with a concept in another Ontology [8].

Granularity of Ontology mapping have 1:1, 1:n and n:m. Mapping case of 1:1 is often called simple mapping, other cases are called complex mapping.

3.2 FRAMEWORK FOR ONTOLOGY MAPPING MODEL

When we input two isomerism Ontologies, establishing mapping relationship from the source Ontology to the target Ontology, this is the task of Ontology mapping system. Mapping process is an iterative process, as illustrated in Figure 2.

1) Feature extraction: Parsing document, then, extracting vocabularies of Ontology (including concepts, properties, relations and so on).

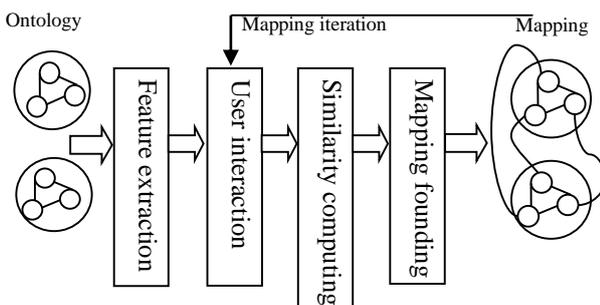


FIGURE 2 Framework for Ontology mapping model

2) User's interaction process (optional process): Ontology mapping system supports an optional user interaction process. Through interaction, users can pre-specify one or more mapping relations, before automatically mapping can also correct error mappings, which are found in mapping system, after automatic discovery of mapping, or create missing mapping relations. User's interaction actions will transmit mapping which impacting other associated elements, with the result, exerting an influence to mapping of the entire Ontology to achieve the purpose of improving the accuracy of mapping.

3) Similarity computing: Consider different methods to calculation according to differences between conceptual similarity and attribute similarity in conceptual similarity calculation. It starts from four aspects: name, attributes, structure and instances. However, attribute similarity calculation only comes from three aspects of matches, which are name, domain and range, to determine the similarity, in calculation process, the value of similarity is limited in [0,1].

4) Mapping discovery: Mapping discovery is based on similarity value after iteration. It aims to select the best mapping relations between elements of Ontology according to some selection strategies, constraints of Ontology and contextual.

5) Mapping: Algorithm outputs mapping table. Each item in the table corresponds to a mapping relation, each item contains four elements set: the first one is element set in source Ontology $O_1 \{e_{i_1}\}$, the second one is element set in target Ontology $O_2 \{e_{i_2}\}$, the third one is corresponding relationship between elements, and the fourth one is similarity values of relationship.

6) Mapping iteration: using the obtained similarity values to conduct iterative operation, to receive comprehensive predictive values of candidate mapping.

3.3 THE CONCEPT RELATIONS OF ONTOLOGY MAPPING

Concept relations in Ontology mapping include: equivalent, generalization, specialization, deviation, and intersection.

1) Equivalent: Assuming two concepts C and D, the interpretation in a certain domain and the corresponding interpretation set of C is equivalent with the corresponding interpretation set of D, so C and D are equal.

2) Generalization (Specialization): Assuming two concepts C and D, the interpretation in a certain domain and the corresponding interpretation set of C is a subset of corresponding interpretation set of D, called D is a generalization of C or C is a specialization of D.

3) Deviation: Assuming two concepts C and D, the interpretation in a certain domain and the corresponding interpretation set of C and the corresponding interpretation set of D has empty intersection, this is called C and D are in deviation.

4) Intersection: Assuming two concepts, the interpretation in a certain domain and the intersection of corresponding interpretation set of C and corresponding interpretation set of D is not empty, called C and D are intersection.

3.4 ONTOLOGY MAPPING BASED ON DESCRIPTION LOGIC

Description Logic is built on concepts and relations, it explains concepts as a collection of objects, and expresses relations as a binary relationship between objects. A Description Logic system contains four basic components: Construction set, which is used to express concepts and relationships. Tbox contains assertions, Abox instance assertions, Reasoning mechanism on Tbox and Abox. Choices and different assumptions on the above elements, decided capacity and reasoning ability of a Description Logic system[9].

A knowledge base $K = \langle T, A \rangle$ consists of two parts: Tbox T and Abox A. Tbox is a finite set which included assertions, has form: $C \sqsubseteq D$, C and D are concepts, usually use $C \equiv D$ as abbreviation of $C \sqsubseteq D$ and $D \sqsubseteq C$, Abox is a finite set of instance assertions, form is $C(a)$, C is a concept, a is the name of an individual, or the form is $P(a, b)$. P is a primitive relation, a and b as names of two individuals [10,11].

Suppose K is a knowledge base. T is Tbox, C and D are concepts:

1) If exist an explanation I make $C^I \neq \emptyset$, then C is satisfied. I shows also known as model C.

2) With each explanation I if have $C^I \subseteq D^I$, then C contained in D, which denotes as $C \sqsubseteq D$;

3) With each explanation I if have $C^I = D^I$, then C and D is equal, denoted as $C \equiv D$. that is $C \sqsubseteq D$ and $D \sqsubseteq C$;

4) With each explanation I if have $C^I \cap D^I = \emptyset$, then, C and D are non-intersect [12].

Elements corresponding relationship in Ontology mapping can express by using Description Logic:

R_1 : concepts C and D equal, if and only if $C \equiv D$ and $D \equiv C$;

R_2 : concepts C and D satisfy generalization, if and only if $D \sqsubseteq C$;

R_3 : concepts C and D satisfy specialization, if and only if $C \sqsubseteq D$;

R_4 : concepts C and D deviate, if and only if $C \sqcap D \sqsubseteq \perp$;

R_5 : concepts C and D deviate, if and only if $C \sqcap D \not\subseteq \perp$.

In the process of Ontology mapping, the above remarks can use following process to judge the relationship between concepts C and D:

Step1: Determine the inclusion relationship between C and D, there are four cases:

a) if $C \sqsubseteq D$ and $D \sqsubseteq C$. according to $\langle R_1 \rangle$. C and D are equivalent;

b) if $D \sqsubseteq C$. and $C \sqsubseteq D$ not established, according to $\langle R_2 \rangle$. C and D satisfy generalization;

c) if $C \sqsubseteq D$. and $D \sqsubseteq C$ not established, according to $\langle R_3 \rangle$. C and D satisfy specialization;

d) if $C \sqsubseteq D$ and $D \sqsubseteq C$ not established. C and D are deviation or C and D satisfy intersection.

Step 2: The case of the previous step appears(4) can be considered;

i) if $C \sqcap D \sqsubseteq \perp$ established, according to $\langle R_4 \rangle$. C and D are in deviation;

ii) if $C \sqcap D \not\subseteq \perp$ not established, C and D are intersection.

3.5 EXAMPLES OF ONTOLOGY MAPPING

In order to build Ontology in tourist information resource field, some concepts are refined.

The main top concept: people, organization agencies, scenic spots, travelling routes, modes of transportation, room and board, travel routes, location, specialties, amusements, national customs and so on.

Organization agencies have sub-concept: travel agencies, scenic spot management agencies, transport enterprises, accommodation enterprises, tourism bureau, insurance companies, specialty enterprises, entertainment industries and so on.

Scenic spots have sub-concept: hydrological landscape, climate biological king, physiographic landscape, cultural landscape, history heritage, other scenic spots and so on.

Travel routes have sub-concept: full lines and pick lines and so on.

Transportation methods have sub-concept: cars, trains, aircrafts, steamships, tour buses, ropeways and so on.

Description Logic expresses some information as defined below.

3.5.1 Describe concept of tourism knowledge

Tour agencies \sqsubseteq organization agencies, scenic spot management agencies \sqsubseteq organization agencies, transport enterprises \sqsubseteq organization agencies, accommodation enterprises \sqsubseteq organization agencies, tourism bureau \sqsubseteq organization agencies, insurance companies \sqsubseteq organization agencies, specialty enterprises \sqsubseteq organization agencies, entertainment industries \sqsubseteq organization agencies.

Hydrological landscape \sqsubseteq scenic spots, climate biological king \sqsubseteq scenic spots, physiographic landscape \sqsubseteq scenic spots, cultural landscape \sqsubseteq scenic spots, history heritage \sqsubseteq scenic spots, other scenic spots \sqsubseteq scenic spots.

Full lines \sqsubseteq travel routes, pick lines \sqsubseteq travel routes.

Cars \sqsubseteq transportation methods, trains \sqsubseteq transportation methods, aircrafts \sqsubseteq transportation methods, steamships \sqsubseteq transportation methods, tour buses \sqsubseteq transportation methods, ropeways \sqsubseteq transportation methods.

3.5.2 Describe relationship

Group (travel agencies, tours) by group (tours, travel agencies), management (scenic spot management agencies, scenic spots) be operating (scenic spots, scenic

spot management agencies), group = by group, management =be operating.

How to determine relationship between organization agency and management agency: scenic spot management agencies \sqsubseteq organization agency is true. However, organization agency \sqsubseteq scenic spot management agencies, is false. According to R_2 , organization agency and scenic spot management agency have generalization relationship. Scenic spot management agency \sqsubseteq organization agency is true. But organization agency \sqsubseteq scenic spot management agencies, is false. According to R_3 , organization agency and scenic spot management agencies have specialization relationship.

How to determine relationship between hydrological landscape and car: hydrological landscape \sqsubseteq car is not true. At the same time, car \sqsubseteq hydrological landscape is also untrue. Hydrological landscape \sqsupseteq car $\sqsubseteq \perp$ is true. According to R_4 , hydrological landscape and car have deviation relationship.

How to determine relationship between hydrological landscape and history heritage: hydrological landscape \sqsubseteq history heritage is not true. And history heritage \sqsubseteq hydrological landscape is also untrue. And hydrological

landscape \sqsupseteq history heritage $\sqsubseteq \perp$ is untrue. Then, hydrological landscape and history heritage have Intersection relationship.

4 Conclusions

Integrating and synchronizing tourism information resources make it become orderly and standardized, which achieves the purposes of services to the tourism industry and promotes economic development. This is the target of information retrieval, knowledge management and knowledge services. In the future, the development of travel information resources services platform will be proceeded, which is based on Description Logic.

Acknowledgements

The work is supported by the National Natural Science Fund Projects (61262071), the General Project of Social Development in Yunnan Province (2011FB044), the humanities and social science research projects of the Ministry of Education (12YJCZH053).

References

[1] Yuan Z 2008 The discussion about integrating resources of regional tourism information *Market Modernization* 21 248-9 (in Chinese)

[2] Lu Y, Kaibo T 2007 Personalized retrieval of tourism information resources based on agent *Researches in Library Science* 3 71-3 (in Chinese)

[3] Shen L 2008 The application of Ontology in tourist information resources sites *Library Information Service* 32(9) 110-2 (in Chinese)

[4] Gruber T R 1993 A translation approach to portable ontology specifications *Knowledge Acquisition* 5(2) 199-220 (in Chinese)

[5] Hu W, Qi Y 2007 Ontology matching: Set up heterogeneous Ontology *Computer World* 45 B12 (in Chinese)

[6] Ehrig M, Sure Y 2004 Ontology Mapping-An Integrated Approach *Proceedings of the 1st European Semantic Web Symposium* 3053 of Lecture Notes in Computer Science Christoph Bussler Greece Springer Verlag 76-91

[7] Maier A, Aguado J, Bernaras A, Laresgoiti I, Pedinaci C, Peña N, Smithers T 2003 Integration with Ontologies *Proceedings of WM2003* Luzern 1-15

[8] Zhang Y 2005 Research of Ontology mapping in data integration *Central South University* 12-6 (in Chinese)

[9] Dong M 2006 A Description Logic with default reasoning *Chinese Journal of Computers* 26(6) 729-36 (in Chinese)

[10] Gan J, Xu T 2007 Semantic web and its application-based on Ontology, Description Logic, Semantic networks *Yunnan Science and Technology Press* 78-85 (in Chinese)

[11] Gan J, Wen B 2008 The research of conceptual modeling based on Description Logic *Journals of Yunnan Normal University* 28(6) 13-6 (in Chinese)

[12] Gao J, Zhu M, Ming H 2004 Foundation of artificial intelligence *Higher Education Press* 231-5 (in Chinese)

Authors	
	<p>Jianhou Gan, born on August 13, 1976, Yunnan, China</p> <p>Current position, grades: Ph.D candidate of Metallurgical Physical Chemistry, professor and supervisor of master student in Yunnan Normal University.</p> <p>University studies: master's degree in Basic Mathematics from the school of information Yunan Normal University in 2004.</p> <p>Scientific interest: education informalization for nationalities, semantic web, database, intelligent information processing.</p> <p>Publications: more than 40 papers.</p> <p>Experience: on education informalization for nationalities, semantic Web, database, intelligent information processing.</p>
	<p>Bin Wen, born on March 15, 1978, Yunnan, China</p> <p>Current position, grades: Ph.D. of Application Technology of Computer, Lecturer in Yunnan Normal University.</p> <p>University studies: Ph.D. degree in Application Technology of Computer from School of Mechanical Electronic & Information Engineering, China University of Mining & Technology, Beijing in 2013. Master's degree in Computer Software and Theory from the school of information, Yunnan Normal University in 2005.</p> <p>Scientific interest: intelligent information processing, emergency management.</p> <p>Publications: more than 20 papers.</p> <p>Experience: a strong experience on intelligent information processing and emergency management.</p>
	<p>Jinxu Li, born on October 11, 1972, Yunnan, China</p> <p>Current position, grades: Master of Computer Software and Theory, lecturer in Yunnan Normal University.</p> <p>University studies: Computer Software and Theory in Yunnan Normal University 2005.</p> <p>Scientific interest: information safety, intelligent information processing</p> <p>Publications: more than 10 papers</p> <p>Experience: a strong experience on information safety and intelligent information processing.</p>

Parametric modelling and simulation on oblique cutting based on MSC.Marc

Jing Sheng*, Hai-Fei Long

Faculty of Mechanical Engineering, Xiamen University of Technology Xiamen, Fujian, P. R. China

Received 1 August 2014, www.tsi.lv

Abstract

The finite element simulation of oblique cutting is a complex and professional process. It is necessary to build a system to construct a model in order to obtain simulation values more conveniently and rapidly. The key techniques of 3-D parametric modelling with MSC.Marc software metal oblique cutting simulation was presented in this investigate. The modelling rule based on the process was carried out. The system, designed using C++ Builder, can access data, which includes the geometrical angles and dimensions of tool, the sizes of workpiece, the relative position between tool and workpiece. Meanwhile their properties and cutting conditions, etc. were stored. The procedure file modelling in the MSC.Marc environment automatically is generated by the program. So the parametric modelling of simulation is completed by calling the procedure file. Further, an example was given and the simulation model was also verified. Therefore, the parametric modelling is a kind of effective way for metal cutting simulation.

Keywords: oblique cutting, numerical simulation, parametric modelling, interface design

1 Introduction

Metal cutting is one of the most common machining methods in manufacture industry. The pursuit of corporations is to obtain high quality, and great efficiency. Studying metal machining mechanism is the main approach to achieve this target [1-4]. However, the machining process of metal cutting is quite complex, relating to physics, mechanics, elastic-plastic theory, metal material, thermology and superficial science (tribology) and so on. There are three ways to study metal cutting: analytics, experiment/analytics and numerical method. Analytics method mainly includes slab stress method, slip-lines field method and upper bound method, which are a part of classical solution in plasticity mechanics; experiment/analytics, i.e. the integration of experiment and analytics include similarity theory and inspection plastic method. The experimental methods in common is used to study the metal cutting process involving the side square deformation observation method, high-speed photography method, scanning electronic microscopy, photo-elastic (photo-plasticity) method and X-ray method. Numerical methods include finite-element method, finite difference method and boundary element method. Traditional analytics are quite difficult when analysing and studying cutting mechanism quantitatively. It takes a long time and is tedious to applying experiment/analytics to obtain cutting forces and temperatures values. With the development of the technology of computer software and hardware and the theories related to metal machining becoming more and more perfect, numerical method is used extensively,

especially finite element method. So far, there is much research on numerical simulation of metal cutting overseas. However, the work on this aspect is limited, and mostly focused on orthogonal cutting [5-7].

It is the trend that finite element numerical simulation will be helpful in knowing about metal machining situation. Whereas modelling is a complex procedure, parametric modelling of the program driven method was considered in this study, and some techniques will be discussed.

2 Finite element modelling

During the process of cutting, for example: oblique cutting its edge is out of the vertical of cutting velocity. Thus, cutters whose inclination angle is not zero are oblique cutting. Figure 1 shows 3-D finite element model of oblique cutting.

The contents of finite element modelling of cutting process include geometric modelling and meshing, boundary conditions, material parameters, geometric parameters, initial conditions, contact conditions, load cases etc.

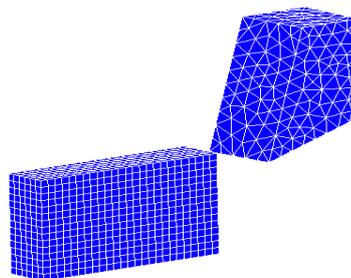


FIGURE 1 The model of oblique cutting

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

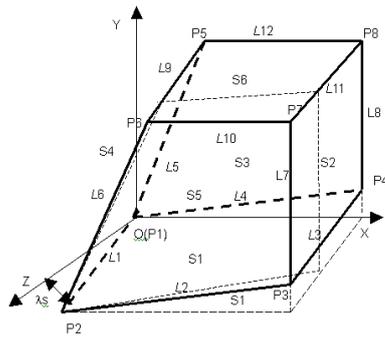


FIGURE 2 Model of cutting tool

2.1 PARAMETRIC MODELING OF TOOL

Based on the physical model and characteristics of finite element analysis software, B-rep (Boundary representation) is employed. The method can express two kinds of information: a geometric one and a topological one. Geometrical data reflects the dimensions and the position of objects. Topological information describes the relative position. The solid model of tool is presented in Figure 2 and Cutting plane and normal sectional plane of the tool are seen in Figure 3. Topological information and geometric information of tool are shown in Figure 4.

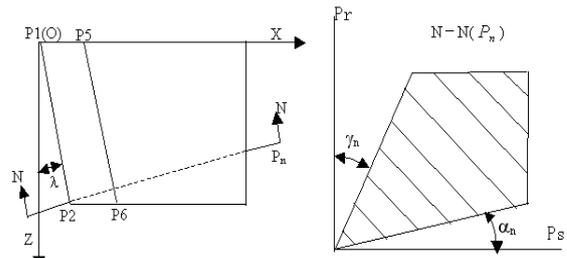


FIGURE 3 Cutting plane and normal sectional plane of t tool

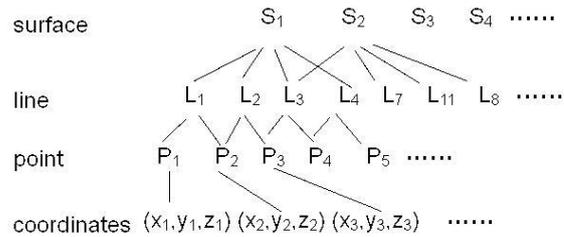


FIGURE 4 Topological and geometric information of tool

By inference, all points' coordinates of cutter are seen in Table 1.

Parametric modelling is a key technique that affects the utility on numeric simulation. Through generating a procedure file, it is possible to do it quickly and conveniently.

By inference, all points' coordinates of tool are seen in Table 1.

TABLE 1 The coordinates of model on tool and workpiece

Point	X-coordinate	Y-coordinate	Z-coordinate
P_1	0	0	0
P_2	$l_1 \times \sin \lambda_s$	0	$l_1 \times \cos \lambda_s$
P_3	$l_4 \times \cos \alpha_0$	$(l_4 \times \cos \alpha_0 - l_1 \times \sin \lambda_s) \times \tan \alpha_0$	$l_1 \times \cos \lambda_s$
P_4	$l_4 \times \cos \alpha_0$	$l_4 \times \sin \alpha_0$	0
P_5	$l_5 \times \sin \gamma_0$	$l_5 \times \cos \gamma_0$	0
P_6	$l_5 \times \sin \gamma_0 + l_1 \times \sin \lambda_s$	$l_5 \times \cos \gamma_0$	$l_1 \times \cos \lambda_s$
P_7	$l_4 \times \cos \alpha_0$	$l_5 \times \cos \gamma_0$	$l_1 \times \cos \lambda_s$
P_8	$l_4 \times \cos \alpha_0$	$l_5 \times \cos \gamma_0$	0
W_1	$-s - l$	$a_p - h$	b
W_2	$-s$	$a_p - h$	b
W_3	$-s$	a_p	b
W_4	$-s - l$	a_p	b

Where l_i is the length of every line, $i=1, 2, \dots, 12$. λ_s is the inclination angle. γ_0 is the rake angle and α_0 is the flank angle. S is the distance between the work and the cutting edge. l is the length of the work along the direction of cutting edge. H is the height of the work. a_p is the depth of cut. b is the distance between tool-tip and the side of the work along the cutting edge direction.

2.2 PARAMETRIC MODELING OF WORK

The work model is constructed in scanning method (Figure 5). Top view and side view of the model of work are seen in Figure 6. Its benchmark is defined by cutting tool. Because the work is a cuboid, the solid work is obtained through its surface that consist of four vertexes (W_1, W_2, W_3 and W_4) in the XOZ sweeping along the direction of Z -

axis. The coordinates of the four points on the work are given in Table 1.

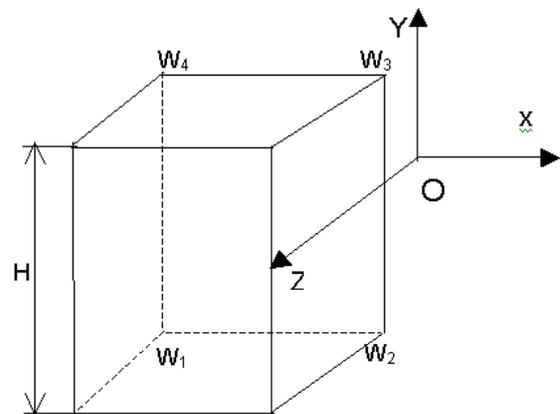


FIGURE 5 Model of work

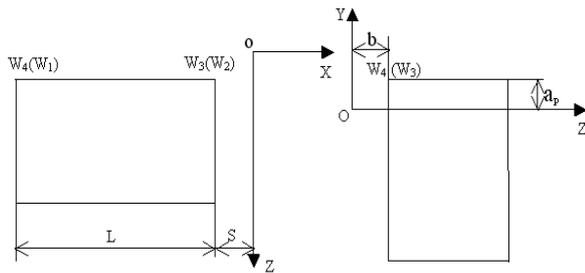


FIGURE 6 Top view and side view of the model of work

A finite element model of a cutter and a work generated and mesh automatically in MSC.Marc by designing interface.

2.3 MATERIAL MODEL

During a practical cutting process, the material of work results in elastic-plastic strain under the condition of high temperature, large strain and large strain ratio. Johnson-Cook model describes material yielding flow properties of the cutting region more accurately. The Johnson-Cook equation is described as follows [1]:

$$\bar{\sigma} = \left[A + B(\bar{\epsilon})^n \right] \left[1 + C \ln \left(\frac{\bar{\epsilon}}{\bar{\epsilon}_0} \right) \right] \left[1 - \left(\frac{T - T_{room}}{T_{melt} - T_{room}} \right)^m \right], \quad (1)$$

where A, B, n, C and m are the parameters determined by a material itself. T_{melt} is melting temperature. T_{room} is the room temperature.

2.4 FRICTION MODEL BETWEEN TOOL AND CHIP

There are two explicit areas on the rake surface: slip region and glue region. On the basis of research, constant coefficient friction is applied in slip region and constant friction stress is used in glue one. The friction stress is written as:

$$f = \begin{cases} \mu \sigma_n & \sigma_f = \mu \sigma_n \\ k & \sigma_f = k \end{cases}, \quad (2)$$

where σ_n is normal stress. μ is friction coefficient and k is shear stress.

2.5 THE CRITERION OF CHIP SEPARATION

During the simulation, there are criterions that make the chip separate from work and rake face. They are divided into geometric criterion and physical criterion. The geometric criterion decides the separation through the changes of geometric dimension of deformable body. The physical one is used to identify whether magnitude of physical quantity causes critical value or not.

In fact, chips are separated by setting a minimum force

or stress of the nodes as threshold.

2.6 EQUATION OF HEAT CONDUCTION

Because the system consists of work, chip and tool generates heat continuously, the first and the second deformation zone of the work go through plastic and elastic deformation. Besides, the rake surface of the tool has severe friction [8].

Equation of the heat conduction in 3-D unsteady-state temperature field (take into account variable thermal conductivity) is defined as follows:

$$\rho c \frac{\partial T}{\partial t} = K \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + \frac{dK}{dT} \left[\left(\frac{\partial T}{\partial x} \right)^2 + \left(\frac{\partial T}{\partial y} \right)^2 + \left(\frac{\partial T}{\partial z} \right)^2 \right] - \rho c \left(w_x \frac{\partial T}{\partial x} + w_y \frac{\partial T}{\partial y} + w_z \frac{\partial T}{\partial z} \right) + q^*, \quad (3)$$

where K represents the thermo conductivity coefficient and T is temperature. ρ is the material density and c is thermal capacity. X, Y and Z are cartesian coordinate system; w_x, w_y and w_z are velocity component of kinematic heat-source in x, y and z -axis. q^* is heat generation rate per unit volume:

$$q^* = W_h \bar{\sigma} \frac{\dot{\bar{\epsilon}}}{J}, \quad (4)$$

where W_h is the ratio that plastic deformation work turn into heat energy. $\bar{\sigma}$ is the equivalence stress. $\dot{\bar{\epsilon}}$ is the equivalence strain ratio. J is coefficient of thermal equivalent of work. As the amount of radiant heat is little, it is ignored.

3 Key techniques

The concept data mode of machining condition is shown Figure 7. The information mentioned above is stored into every table respectively, from which a model of milling process is built by knowledge acquisition machine and explanation machine. The component table includes its geometric and process information based on STEP (standard for the exchange of product model data) APP244 (application protocol). The component material table has physical and chemical properties, and mechanical behaviours. The cutting-tool table describes its dimensions, angles and material). The machine tool table provide with its property parameters and machining parameters. And other tables indicate other modelling data separately.

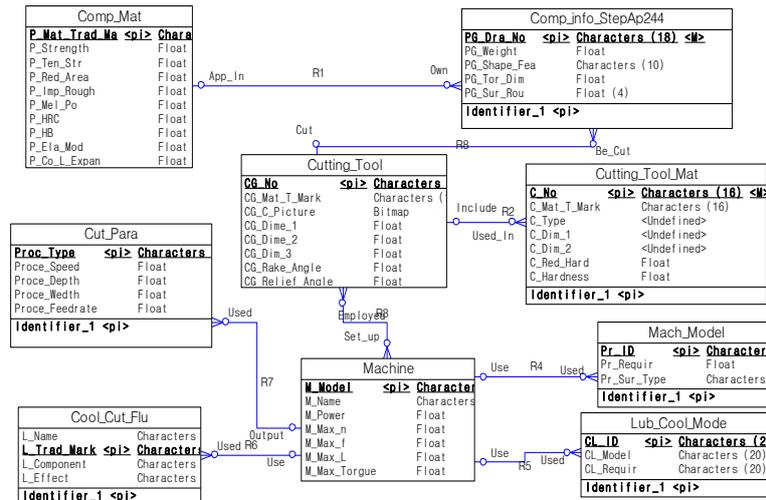


FIGURE 7 The concept data mode of machining condition

3.1 THE INTERFACE DESIGN OF PARAMETRIC MODELING

3.1.1 The interface design between C++Builder and database

Database module and database engine provided by C++ Builder or ADO (Active Data Object) were used to access data. The tables whose type is DBF (Database File) and BDE (Borland Database Engine) were used, while parameters about BDE were set, such as path, type and language drive.

3.1.2 The interface file of parametric modelling in Marc

The system's knowledge based on rule was established according to MSC.Marc software character. And procedure files were written using C++ Builder code according to the rule.

Based on the solid model of cutting tool, and the topology and geometry information of the tool, its geometry information such as points, line and surface were written in program line-by-line, and then meshed automatically. With regard to work with a rectangular, a surface consists of four point first and then swept towards Z-axis to form work model. A work model was also meshed. The modelling procedure mentioned above is written in procedure file. While relative position between tool and work, material model, friction model between tool and chip, properties of cutter and work, cutting conditions, all set about finite element simulation and so on, are written into the program.

The generating procedure file can be operated according to specified format. The modelling process becomes easy. The block diagram of modelling process is shown in Figure 8. The structure of procedure file is seen in Figure 9.

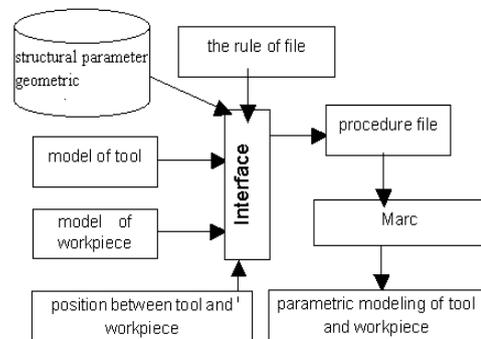


FIGURE 8 Parametric modelling process

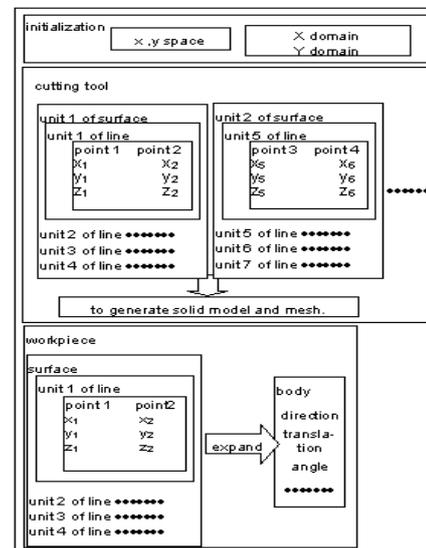


FIGURE 9 Structure of procedure file

3.2 REMESHING TECHNIQUE

With regard to the large deformation, an Updated Lagrange analytical method were adopted. Because excessive deformation of mesh makes successor analysis restrictive, remeshing function was conducted, and original state variable could be mapped.

It is requirement to combine the contact penetration

with increment and re-meshing when simulating cutting process. Then the set of chip separating from work can be performed.

3.3 MESH ADAPTIVE TECHNIQUE

In selected elements, these elements will be re-meshed in incremental step when they disobey given error criterion. The average of strain energy and contact adaptive criterion were employed. The density of mesh can be adjusted automatically, and computation efficiency of the complex problem and precision will be improved.

4 Result of modelling

Figure 10 shows the interface of oblique cutting modelling. Through the interface, operator can finish modelling (Figure 1). The Equivalent Mises stress distribution of simulation is shown in Figure 11.

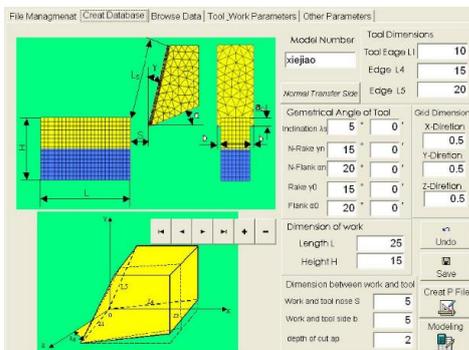


FIGURE 10 The interface of oblique cutting modelling

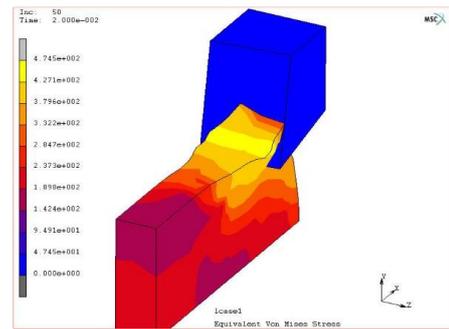


FIGURE 11 Equivalent Mises stress distribution

5 Conclusions

The techniques of parametric modelling, which include database, C++ Builder programming, remeshing and mesh adaptive, were discussed. A model of the oblique cutting process was constructed and employed to simulate for the sake of prediction the change of cutting force and temperature, etc. A running example test and verify the parametric modelling. The intelligent parametric modelling will be further investigated to perform a lot of simulation condition. Parametric modelling method will also be applied to simulate in other aspects such as the quality of machined surface or tool wear during machining process.

6. Acknowledgements

The author is grateful to the Science & Research Foundation of Xiamen University of Technology (grant No. YKJ1100512) for supporting the research.

References

- [1] Maranhão C., Paulo-Davim J 2010 Finite element modelling of machining of AISI 316 steel: numerical simulation and experimental validation *Simulation Modelling Prac-tice and Theory* 18(2) 2010 139-56
- [2] Pedro J A, Tuğrul Ö 2010 Investigations on the effects of friction modeling in finite element simulation of machining *International Journal of Mechanical Sciences* 52(1) 31-42
- [3] Jin X L, Altintas Y 2012 Prediction of micromilling forces with finite element method *Journal of Materials Processing Technology* 212(3) 2012 542-52
- [4] Tang L H, Huang J L, Xie L M 2011 Finite element modeling and simulation in dry hard orthogonal cutting AISI D2 tool steel with CBN cutting tool *International Journal of Advanced Manufacturing technology* 53(9-12) 1167-81
- [5] Zel T Ö, Thepsonthi T, Ulutan D, Kaftanoğlu B 2011 Experiments and finite element simulations on micro-milling of Ti-6Al-4V alloy

- with uncoated and CBN coated micro-tools *Manufacturing Technology* 60(1) 85-8
- [6] Liu F S, Ronaldo I B 2010 Stabilized low-order finite elements for frictional contact with the extended finite element method *Computer Methods in Applied Mechanics and Engineering* 199(37-40) 2456-71
- [7] Stalin-John M R, Shrivastava K, Banerjee N, Madhukar D P, Vinayagam B K 2013 Finite element method based machining simulation for analyzing surface roughness during turning operation with HSS and Carbide Insert Tool *Arabian Journal for Science and Engineering* 38(6) 1615-23
- [8] Filho J M C 2013 Prediction of cutting forces in mill turning through process simulation using a five-axis machining center *International Journal of Advanced Manufacturing Technology* 58(1) 71-80

Authors	
	<p>Jing Sheng</p> <p>Current position, grades: Professor at the Xiamen University of Technology, China. Scientific interest: mechanical manufacture and automation, design and manufacture technology of automobile parts, CAD/CAE and computer integrated manufacturing technology. Publications: 66.</p>
	<p>Li Juan, born on March, 1979, Tangshan City, Hebei Province, China</p> <p>Current position, grades: lector at the Xiamen University of Technology, China Scientific interest: mechanical manufacture and automation, intelligent control Publications: 2.</p>

H_∞ fault-tolerant control for nonlinear singular system via a fault diagnosis observer

Dongwen Zhang^{1*}, Xinguo An²

¹*School of information science & engineering, Hebei University of Science & Technology, Shijiazhuang, China*

²*School of Sciences, Hebei University of Science & Technology, Shijiazhuang, China*

Received 6 June 2014, www.tsi.lv

Abstract

An H_∞ fault-tolerant control scheme based on fault diagnosis observer was developed for a class of nonlinear singular systems with external disturbances and actuator faults. A fault diagnosis observer was designed to estimate the system states and the actuator faults and a sufficient condition for the existence of this observer was presented in the form of feasibility problem of a linear matrix inequality. Based on linear matrix inequality (LMI) technique and the estimates of the states and faults, an H_∞ fault-tolerant control scheme was worked out. The H_∞ fault-tolerant control system via a state feedback controller can be made solvable, impulse free, asymptotically stable, and the effect of external disturbances on the system was attenuated in terms of the prescribed H_∞ performance index. Finally, a simulation example was given to illustrate the procedure of designing the fault diagnosis observer and the state feedback controller, and the simulation result showed the effectiveness of the proposed method.

Keywords: singular systems, H_∞ control, fault-tolerant control, observer, linear matrix inequality (LMI)

1 Introduction

The safety, reliability and maintainability in actual systems and industrial process have motivated researchers to concentrate on the so-called fault-tolerant control (FTC) [1-5]. FTC is primarily meant to ensure safety, i.e., the stability of a system after the occurrence of a fault in the system. There are two approaches to synthesize controllers that are tolerant to system faults. One approach, known as passive FTC, aims at designing a controller which is a priori robust to some given expected faults. Another approach, known as active FTC, relies on the availability of a fault detection and diagnosis (FDD) block that gives, in real-time, information about the nature and intensity of the fault. This information is then used by a control reconfiguration block to adjust online the control effort in such a way to maintain stability and to optimize the performance of the faulty systems. Researches on FDD for systems have long been recognized as one of the important aspects in seeking effective solutions to an improved reliability of practical control systems. Accurate fault estimation can determine the size, location and dynamic behaviour of the fault, which automatically indicates FDD, and has thus attracted interests recently. Many methods have been proposed for FDD, e.g., parity relations approach [6], Kalman filters approach [7], parameter estimation approach [8] as well as observer-based approach [9-11]. Observer-based FDD is one of the most effective methods and has obtained much more attention. So far, various observer-based FDD approaches have been reported in the literatures. Based on Euler approximate discrete model

observer, a fault estimation method was proposed in [9]. A novel augmented fault diagnosis observer design was well addressed in [10], which not only broaden application scopes of adaptive fault diagnosis observer, but also cope with system disturbances. In [11], the observers were designed for both linear and nonlinear systems considering both noise and uncertainties, and the main advantage of these observers is that they can handle both noise and uncertainties simultaneously. Overall, the basic idea behind the use of the observer for FDD is to estimate the state or/and output of the system from the measurement by using some type of observers, and then to construct a residual by a properly weighted the state or/and output error. The residual is then examined for the likelihood of faults by using a fixed or adaptive threshold.

However, only a few efforts were made to investigate FTC for nonlinear singular systems. Nonlinear singular system model characterizes a class of rather complex systems, which not only possesses nonlinearities, but also has singular nature of the algebraic constraints. Therefore, the investigation on this class of systems is more difficult and challenging [12-15]. Several works on FDD and FTC for nonlinear singular systems were reported in [16, 17]. For nonlinear singular systems, [18] designed an observer based on the new parameterization of the generalized Sylvester equations solutions, and the condition for the existence of the observer was given and the sufficient condition for its stability was derived using linear matrix inequality (LMI) formulation. By using the linear matrix inequality (LMI) technique, an interesting descriptor estimator was presented to simultaneously estimate system

* *Corresponding author* e-mail: zdwwtx@hebust.edu.cn

states, output noises and sensor faults for a class of Lipschitz nonlinear descriptor systems [19]. However, some restrictive equivalent transformations were needed for obtaining the state-space observer in the [19]. Obviously, the usages of restrictive equivalent transformations are not desirable from the viewpoint of computation. Moreover, it is not possible to totally decouple the fault effects from the perturbation effects on the system, and the H_∞ control theory has been proved to be an effective tool to tackle the issue. Consequently, this motivates us to investigate the topic of FTC and H_∞ control, which is very important in many practical systems.

Our objective in this paper is to propose an observer-based FTC and H_∞ control method for a class of nonlinear singular systems with input disturbances and actuator faults, a novel design method of fault diagnosis observer is presented, which can overcome the nonlinearity and precisely estimate the values of the states and actuator faults. By linear matrix inequality (LMI) technique and by using the obtained states and faults information, robust fault-tolerant state feedback control scheme is worked out. The solvability, asymptotic stability and H_∞ performance of the closed-loop system are guaranteed after the actuators occur faults.

Throughout this paper, \mathbb{R}, \mathbb{C} denote real number set and complex number set respectively, \mathbb{R}^n denotes the n dimensional Euclidean space and $\mathbb{R}^{n \times n}$ is the set of all $n \times n$ real matrices. I is the identity matrix with appropriate dimensions. $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$ refer to the minimal and maximal eigenvalues of the matrix P respectively. The vector norm $\|x\|$ is defined as $\|x\| = \sqrt{x^T x}$. For a symmetric matrix, $*$ denotes the matrix entries implied by symmetry.

2 Problem statement and preliminaries

Consider the following nonlinear singular system with actuator faults as well as external disturbances:

$$\begin{cases} E\dot{x}(t) = Ax(t) + Bu(t) + g(t, x) + Fu_f(t) + D_1w(t) \\ y(t) = Cx(t) + D_2w(t), \quad x(0) = 0 \end{cases}, \quad (1)$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$ denote respectively the control input and the measurable output vectors, $u_f \in \mathbb{R}^h$ is the unknown actuator fault vector and $E \in \mathbb{R}^n$ is singular with $rank(E) = r < n$. A, B, C, D_1, D_2 and F are known constant real matrices with appropriate dimensions. $g(t, x) \in \mathbb{R}^n$ is a vector-valued time varying nonlinear perturbation with $g(t, 0) = 0$ for all $t \geq 0$ and satisfies the following Lipschitz constraint:

$$\|g(t, x) - g(t, \tilde{x})\| \leq \alpha_0 \|G(x - \tilde{x})\| \leq \alpha \|x - \tilde{x}\| \quad (2)$$

for all $(t, x), (t, \tilde{x}) \in \mathbb{R} \times \mathbb{R}^n$, and G is a known constant real matrix, α and α_0 are both known positive scalars and are called Lipschitz constants. $w(t) \in \mathbb{L}_2$ is the external disturbance on the system, and there exists a position constant f such that $\|u_f(t)\| \leq f$. In the paper, only actuator faults are investigated and it is assumed that, when no fault occurs, $u_f(t) = 0, \forall t \geq 0$.

In this paper, our first goal is to design a fault diagnosis observer to estimate the systems states $x(t)$ and the fault signal $u_f(t)$ simultaneously on the basis of the known input $u(t)$ and the measured output $y(t)$. The second goal is to work out a state feedback controller for FTC and H_∞ control by mean of the estimates of the system states and faults.

Now, one first recalls a lemma that will be used in the next sections.

Lemma 1 (Schur Complement Lemma) Given constant matrices Σ_1, Σ_2 and Σ_3 with appropriate dimensions, where $\Sigma_1^T = \Sigma_1 < 0, \Sigma_2^T = \Sigma_2 < 0$, then $\Sigma_1 - \Sigma_3^T \Sigma_2^{-1} \Sigma_3 < 0$ if and only if:

$$\begin{bmatrix} \Sigma_1 & \Sigma_3^T \\ \Sigma_3 & \Sigma_2 \end{bmatrix} < 0, \text{ or } \begin{bmatrix} \Sigma_2 & \Sigma_3 \\ \Sigma_3^T & \Sigma_1 \end{bmatrix} < 0.$$

3 Fault detection and fault diagnosis

3.1 FAULT DETECTION OBSERVER DESIGN

In this section, a state-space observer will be proposed for FDD, which can provide the information of states and faults. The information is sent to the controller to obtain the control law, which is sent to the actuator.

For the system to admit a feasible FTC solution, the following assumptions are made:

Assumption 1: The row vectors of the matrices E and C in the Equation (1) must be a basis of the n dimensional vector space, that is, $rank \begin{bmatrix} E^T & C^T \end{bmatrix}^T = n$.

Assumption 2: The linear part of the Equation (1) has to be observable, that is

$$rank \begin{bmatrix} (sE - A)^T & C^T \end{bmatrix}^T = n, \quad \forall s \in \mathbb{C}.$$

Under assumptions 1-2, a fault detection observer can be constructed as follows:

$$\begin{cases} \dot{z}(t) = \hat{A}z(t) + \hat{B}u(t) - L_p(\hat{y}(t) - Cx(t)) + \\ \hat{g}(t, \hat{x}) + \hat{F}\hat{u}_f(t) + \hat{D}_1w(t) \\ \hat{x}(t) = z(t) + L_D y(t) \\ \hat{y}(t) = C\hat{x}(t) + \hat{D}_2w(t) \end{cases}, \quad (3)$$

where $\hat{z} \in \mathbb{R}^n$ is the state vector of the detection observer, $\hat{x} \in \mathbb{R}^n$ denotes the observed state vector, $\hat{y} \in \mathbb{R}^p$ is the output vector of the observer, $\hat{u}_f \in \mathbb{R}^h$ is the estimate of the system faults, $\hat{A}, \hat{B}, \hat{D}_1, \hat{D}_2$ and \hat{F} are known parameter matrices with appropriate dimensions.

It's easy to derive that the assumption 1 and the assumption 2 imply that the trio (E, A, C) is completely observable, so the matrix L_D can be selected to make matrix $E + L_D C$ be non-singular. To guarantee the asymptotical stability of the detection Equatuon (3), L_p to be designed and the Lipschtiz constant α should make the following inequality hold [19]:

$$\alpha \leq \frac{\lambda_{\min} \left(\sqrt{1 - \beta^2} C^T C + \eta L_D^T (-L_p)^{-1} \right) - d}{2\lambda_{\max} (C^T L_p \hat{F})}, \quad (4)$$

where d is a pre-specified positive constant. β and η should be chosen properly such that inequality (4) can be satisfied.

Supposing the state error vector $e_m(t) = x(t) - \hat{x}(t)$, the residual signal vector $e_y(t) = y(t) - \hat{y}(t)$, the fault error vector $e_f(t) = u_f(t) - \hat{u}_f(t)$, respectively, then the detection error equation can be written as follows:

$$\begin{aligned} \dot{e}_m(t) &= \hat{A}e_m(t) + (A - \hat{A}E - C)x(t) + \\ &(B - \hat{B})u(t) + g(t, x) - \hat{g}(t, x) - L_p e_y(t) + (\hat{F} - F)e_f(t) \end{aligned} \quad (5)$$

In next subsection, one will give the fault diagnosis rule for the Equation (1).

3.2 FAULT DIAGNOSIS

In this subsection, we will discuss the design method of the fault diagnosis observer. According to the detection error Equation (5), the following fault detection rule is introduced

- 1) If $\|e_y(t)\| = \|Ce_m(t)\| < \lambda$, then no fault occurs at time t ,
- 2) If $\|e_y(t)\| = \|Ce_m(t)\| \geq \lambda$, then faults have occurred at time t ,

where λ is a pre-specified threshold.

According to the above rule, the fault diagnosis observer is presented as follows:

$$\begin{cases} \dot{\tilde{e}}(t) = \hat{A}\tilde{e}(t) + \hat{B}u(t) + \hat{D}_1 w(t) + \\ L_p(\tilde{y}(t) - y) + g(t, \tilde{x}) + \hat{F}\tilde{u}_f(t) \\ \tilde{x}(t) = \tilde{e}(t) + L_D y(t) \\ \tilde{y}(t) = C\tilde{x}(t) + \hat{D}_2 w(t) \end{cases}, \quad (6)$$

where $\tilde{e} \in \mathbb{R}^n$ is the state vector of the diagnosis observer, $\tilde{x} \in \mathbb{R}^n$ denotes the observed state vector, $\tilde{y} \in \mathbb{R}^p$ is the

output vector of the observer. $\tilde{u}_f \in \mathbb{R}^h$ is the estimates of the system faults.

Now, one will provide a sufficient condition for the existence of the fault diagnosis observer.

Theorem 1. For the Equation (1), there exists an asymptotical steady state-space observer in the form of Equation (6) to make the estimated error as small as any desired accuracy, if there exist a positive definite matrix $P \in \mathbb{R}^{n \times n} > 0$ and a matrix $Q \in \mathbb{R}^{p \times n}$ such that

$$\begin{bmatrix} \Psi_{11} & PW \\ W^T P & -I \end{bmatrix} < 0, \quad (7)$$

where $\Psi_{11} = \hat{A}^T W^T P + PW\hat{A} - C^T Q^T - QC + D^T P + PD + \alpha^2 I$, and $W = (E + L_D C)^{-1}$. Specifically, the gain L_D is selected such that $E + L_D C$ is non-singular, and L_p can be computed as $L_p = P^{-1} W^{-1} Q$. Obviously, Equation (7) is a LMI with respect to matrices P, Q .

Proof: For the Equation (1), the detection error Equation (5) and the fault diagnosis observer Equation (6), the error dynamic equation can be characterized as follows:

$$\begin{aligned} \dot{e}(t) &= (E + L_D C)^{-1} (\hat{A} - L_p C) e(t) + \\ &Dw(t) + g(t, \tilde{x}) - g(t, x) \end{aligned}, \quad (8)$$

where $e(t) = \tilde{x}(t) - x(t)$.

Define a Lyapunov function as $V(e(t)) = e^T(t) P e(t)$ with $P > 0$.

Letting $W = (E + L_D C)^{-1}, \Theta = g(t, \tilde{x}) - g(t, x)$ the derivative of $V(e(t))$ along Equation (6) can be obtained as follows:

$$\begin{aligned} \dot{V}(e(t)) &= \\ &e^T ((W(\hat{A} - L_p C) + (D_1 - L_p D_2))^T P + \\ &P(W(\hat{A} - L_p C) + (D_1 - L_p D_2))) e + \Theta^T W^T P e + e P W \Theta \leq \\ &e^T (\hat{A}^T W^T P + PW\hat{A} - (WL_p C)^T P - P(WL_p C) + \\ &D_1^T P + PD_1 - D_2^T L_p^T P - PD_2 L_p) e + \Theta^T \Theta + e^T P W W^T P e \leq \\ &e^T (\hat{A}^T W^T P + PW\hat{A} - (WL_p C)^T P - P(WL_p C) + D_1^T P + \\ &PD_1 - D_2^T L_p^T P - PD_2 L_p + \alpha^2 I + P W W^T P) e \end{aligned}$$

Letting $Q = P W L_p$, thus Equation (7) can be recast to the following inequality by the Schur complement lemma

$$\begin{aligned} &\hat{A}^T W^T P + PW\hat{A} - C^T Q^T - QC + D^T P + \\ &PD - R^T P - PR + \alpha^2 I + P W W^T P < 0 \end{aligned}$$

If Equation (7) holds, one can derive that $\dot{V}(e(t)) < 0$. Furthermore, $e(t)$ converges towards 0 while t converges towards ∞ . The proof is completed.

4 Fault-tolerant control and H_∞ control

Consider the Equation (1) and the fault diagnosis observer Equation (6). One can construct the state -feedback fault tolerant controller as follows:

$$u(t) = K\tilde{x}(t), \tag{9}$$

Where $\tilde{x}(t) \in \mathbb{R}^n$ denotes the observed state vector in (6). In this section, one will discuss how to design the state-feedback gain K .

Applying (9) to the Equation (1), the closed-loop system can be written as follows:

$$\begin{cases} E\dot{\tilde{x}}(t) = A\tilde{x}(t) + BK\tilde{x}(t) + D_1w(t) + \\ g(t, \tilde{x}) + BK e_f(t) \\ \tilde{y}(t) = C\tilde{x}(t) + D_2w(t) \end{cases} \tag{10}$$

Now, one will present the following result.

Theorem 2 For the closed-loop Equation (10) and the given scalar $\gamma < 1$, the closed-loop Equation (10) is solvable, impulse free, asymptotical stable and $\|y(t)\|_\infty \leq \gamma^2 \|w(t)\|_\infty$ if there exist a non-singular matrix $H \in \mathbb{R}^n$ and a controller gain $K \in \mathbb{R}^{m \times n}$ such that the following inequalities holds

$$E^T H = H^T E \geq 0, \tag{11}$$

$$\begin{aligned} (A + BK)^T H + H^T (A + BK) + I + C^T C + \\ D_2^T D_2 + \gamma^{-2} H^T D_1 D_1^T H + \alpha^2 H^T H < 0 \end{aligned} \tag{12}$$

Proof: Choosing a Lyapunov function:

$$V(\tilde{x}(t)) = (E\tilde{x}^T(t))^T H (E\tilde{x}(t)) = (E\tilde{x}(t))^T H^T (E\tilde{x}^T(t))$$

where $E^T H = H^T E \geq 0$ and H is nonsingular.

The derivative of $V(\tilde{x}(t))$ along Equation (10) can be obtained as follows:

$$\begin{aligned} \dot{V}(\tilde{x}(t)) = \tilde{x}^T(t) ((A + BK)^T H + H^T (A + BK) + I + \\ C^T C + H^T \Theta^T \Theta H) \tilde{x}(t) + 2\tilde{x}^T(t) H^T B K e_f(t) + \\ 2\tilde{x}^T(t) H^T D_1 w(t) + 2\tilde{x}^T(t) D_2 D_2^T w(t) \end{aligned}$$

Letting

$$\Omega = (A + BK)^T H + H^T (A + BK) + I + C^T C + \alpha^2 H^T H,$$

so

$$\begin{aligned} \dot{V}(\tilde{x}(t)) \leq \tilde{x}^T(t) \Omega \tilde{x}(t) + 2\tilde{x}^T(t) H^T B K e_f(t) + \\ 2\tilde{x}^T(t) (H^T D_1 + D_2 D_2^T) w(t) \end{aligned} \tag{13}$$

From Equation (2), it is easily derived that

$$\dot{V}(\tilde{x}(t)) \leq -\varepsilon_1 \|\tilde{x}(t)\|^2 + \varepsilon_2 \|e_f(t)\| \|\tilde{x}(t)\| - \varepsilon_3 \|e_f(t)\|^2,$$

where $\varepsilon_1 = \lambda_{\min}(-\Omega)$, $\varepsilon_2 = 2 \|H\| \|BK\|$,

$$\varepsilon_3 = 2 \|H\| \|D_1\| + \|D_2\|^2.$$

Letting $\varepsilon = \min\left(\frac{\varepsilon_1}{2}, \frac{\varepsilon_3}{2\varepsilon_1}\right)$, one has

$$\dot{V}(\tilde{x}(t)) \leq -\varepsilon \left\| \begin{matrix} \tilde{x} \\ e_f \end{matrix} \right\| < 0.$$

By the Schur complement lemma, it is clear that (12) implies

$$\Omega = (A + BK)^T H + H^T (A + BK) + I + C^T C + \alpha^2 H^T H < 0$$

and further indicates

$$(A + BK)^T H + H^T (A + BK) < 0. \tag{14}$$

Moreover, Equations (11) and (14) and $\dot{V}(\tilde{x}(t)) < 0$ indicate that the Equation (10) is solvable, impulse free, asymptotically stable.

Next, one will discuss the H_∞ performance of the Equation (10).

Defining $Z(t) = \dot{V}(\tilde{x}(t)) + y^T y - \gamma^2 w^T w$, using the Equation (10) and inequality (13), one has

$$Z(t) \leq \tilde{x}^T \tilde{\Omega} \tilde{x} + \varepsilon_1 \|\tilde{x}(t)\| \|e_f(t)\| - \varepsilon_3 \|e_f(t)\|^2,$$

$$\text{where } \tilde{\Omega} = \begin{bmatrix} \Omega + C^T C & H^T D_1 + D_2^T \\ D_1 H^T + D_2 & -\gamma^2 I \end{bmatrix}, \tilde{x} = \begin{bmatrix} \tilde{x}^T & w^T \end{bmatrix}^T.$$

Applying the Schur complement lemma to inequality (12), it is clear that $\tilde{\Omega} < 0$.

Letting $\varepsilon_4 = \lambda_{\min}(-\tilde{\Omega})$, it is true that

$$Z(t) \leq -\varepsilon_4 \|\tilde{x}\|^2 + \varepsilon_1 \|\tilde{x}(t)\| \|e_f(t)\| - \varepsilon_3 \|e_f(t)\|^2 \text{ and}$$

$\sqrt{\varepsilon_4 \varepsilon_3} \|\tilde{x}(t)\| \|e_f(t)\| \leq \frac{\varepsilon_4}{2} \|\tilde{x}(t)\|^2 + \frac{\varepsilon_3}{2} \|e_f(t)\|^2$, so one could select $\bar{\varepsilon}$ such that $\bar{\varepsilon} > \varepsilon_1^2 / \bar{\varepsilon} \varepsilon_4$, where $\bar{\varepsilon} = \min(\alpha, \varepsilon_3)$.

Furthermore,

$$Z(t) \leq -\frac{\varepsilon_4}{2} \|\tilde{x}(t)\|^2 - \frac{\alpha \bar{\varepsilon}}{2} \|e_f(t)\|^2.$$

Under zero initial conditions $x(0) = 0$ and the above discussion, it is known that:

$$\int_0^\infty (y^T(t)y(t) - \gamma^2 w^T(t)w(t)) d\tau \leq \int_0^\infty H d\tau \leq 0,$$

is true, that is $\|y(t)\|_\infty \leq \gamma^2 \|w(t)\|_\infty$ the proof is completed.

Remark: Equation (12) with respect to matrices H, K is a nonlinear matrix inequality. One thus has a continuous interest to transform Equation (12) into the LMI form.

Theorem 3. For the closed-loop Equation (10) and the given scalar $\gamma < 1$, the closed-loop Equation (10) is solvable, impulse free, asymptotical stable and $\|y(t)\|_{\infty} \leq \gamma^2 \|w(t)\|_{\infty}$ if there exist a non-singular matrix $W \in \mathbb{R}^n$ and a matrix $T \in \mathbb{R}^{m \times n}$ such that

$$E^T W = W^T E \geq 0, \tag{15}$$

$$\begin{bmatrix} \Xi_{11} & W^T & W^T C^T & W^T D_1 & D_2 D_2 \\ * & -I & 0 & 0 & 0 \\ * & * & -I & 0 & 0 \\ * & * & * & -\gamma^2 I & 0 \\ * & * & * & * & -\gamma^2 I \end{bmatrix} < 0, \tag{16}$$

where $\Xi_{11} = (AW)^T + AW + BT + (BT)^T + \alpha^2 I$.

Moreover, if there exists a feasible solution (T, W) for the above Equations (15) and (16), the state feedback controller gain matrix K can be signed as $K = TW^{-1}$. Obviously, Equations (15) and (16) are LMIs with respect to matrices T, W .

Proof: Pre-multiplying and post-multiplying inequality (12) by $\text{diag}\{H^{-T}, I, I, I\}$ respectively, and letting $H^{-1} = W, KH^{-1} = T$, then using the Schur complement lemma, Equations (15) and (16) can be obtained immediately. This completes the proof.

Now, one will describe the procedure of the fault diagnosis and H_{∞} fault-tolerant control for a class of the Equation (1).

Input: the Equation (1) and the H_{∞} performance index $\gamma < 1$.

Output: the state feedback controller gain K .

Step 1 Choosing scalars β, η , and d such that Equation (4) is satisfied.

Step 2 Choosing suitable matrix L_D such that $E + L_D C$ is non-singular, then solving the Equation (7) by Matlab LMI control toolbox. If there is a feasible solution (P, Q) to the Equation (7), then one can compute $L_p = P^{-1}W^{-1}Q = P^{-1}(E + L_D C)^{-1}Q$.

If there is no feasible solution to the Equation (7), thus the step 2 will be repeated and another matrix L_D is choose until there is a feasible solution to the Equation (7).

Step 3 Solving the Equations (15) and (16) by Matlab LMI control toolbox, if there is a feasible solution (T, W) of the Equations (15) and (16) then one can compute the state feedback controller gain $K = TW^{-1}$.

5 Numerical example

Consider a nonlinear singular system in the form of Equation (1), where

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A = \begin{bmatrix} 0.9 & 3.1 & 1.5 \\ 0 & 0.5 & 0.6 \\ 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0.28 & 1.23 & 0.02 \\ 0 & 0.01 & 0.56 \\ 0.22 & 1.21 & 0.36 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.05 & 0 & 1.29 \\ 0.01 & 0.87 & 0 \\ 0.03 & 0.12 & 0.5 \end{bmatrix}, D_1 = \begin{bmatrix} 4 & 0.6 & 0.4 \\ 1.3 & 1.5 & 0.3 \\ 0.5 & 0.6 & 0.7 \end{bmatrix}, F = \begin{bmatrix} 0.4 & 1.6 \\ 4.2 & 0.2 \\ 1.6 & 9.2 \end{bmatrix}$$

The nonlinear function of Equation (1) is assumed to be

$$g(t, x) = \begin{bmatrix} 0.3 \sin(x_1 + x_2) \\ 0.1 \\ 0.5 \cos x_3 \end{bmatrix}, G = \begin{bmatrix} 0.2 & 0.1 & 0.4 \\ 0.6 & 0 & 0.7 \end{bmatrix}.$$

The Lipschitz constants are $\alpha_0 = 0.7$ and $\alpha = 0.8$. It is clear that the assumption 1 and the assumption 2 are easily satisfied. The parameter matrices of fault diagnosis observer Equation (6) are selected as follows:

$$\hat{A} = \begin{bmatrix} 0.2 & 2.1 & 0.5 \\ 0 & 0.1 & 0.4 \\ 0 & 0 & 1 \end{bmatrix}, \hat{B} = \begin{bmatrix} 0.4 & 0.8 & 0.1 \\ 0.1 & 0 & 0.7 \\ 0.1 & 0.5 & 0.9 \end{bmatrix}, \hat{D}_1 = \begin{bmatrix} 0.7 & 0.2 & 0.7 \\ 0.2 & 0.8 & 0.6 \\ 1.1 & 0.4 & 0.7 \end{bmatrix},$$

$$\hat{D}_2 = \begin{bmatrix} 0.01 & 0 & 0.03 \\ 0.04 & 1 & 0.01 \\ 0.07 & 0.06 & 0.41 \end{bmatrix}, \hat{F} = \begin{bmatrix} 0.5 & 1.1 \\ 0.2 & 1.2 \\ 0.7 & 3.1 \end{bmatrix}$$

The system faults have the following forms:

$$u_{f1} = \begin{cases} 0, & t \in [0, 2.5) \\ 0.002t^2 - 0.3t + 2, & t \in [2.5, 6) \\ 2 \sin(0.3t) + 0.3, & t \in [6, 10) \\ 0.1t + 0.12, & t \in [10, \infty) \end{cases}$$

$$u_{f2} = \begin{cases} 0, & t \in [0, 3) \\ \cos t + 0.02, & t \in [3, 7) \\ \sin 2t - 0.3, & t \in [7, 11) \\ 0.2t + 1.3, & t \in [11, 14) \\ 4.1, & t \in [14, \infty) \end{cases}$$

Using Matlab LMI Control Toolbox to solve Equations (7), (11) and (12), the following results could be obtained:

a) fault diagnosis observer design.

Letting $L_D = \begin{bmatrix} 0 & -3 & 1 \\ 0 & -2 & 0 \end{bmatrix}^T$, $d = 0.35$, $\beta = 0.5$ and

$\eta = 0.2$, one can acquire

$$W = \begin{bmatrix} 1 & 0 & 0 \\ 0.0196 & -1.3514 & -2.2837 \\ -0.0218 & 0 & 0.4376 \end{bmatrix}$$

$$L_p = \begin{bmatrix} 84.08 & -62.45 & -448.98 \\ 71.85 & -104.07 & -669.16 \\ -49.51 & 40.82 & 236.10 \end{bmatrix}.$$

The trajectories of the faults and their estimates are given in Figure 1 and Figure 2. One can see that the tracking performance is desired. In Figure 1, no fault occurs when $0 < t < 2s$, there is no failure false alarm for the designed fault detect observer, The fault f_1 occurs when $2s \leq t < 3.2s$. Similar result can be seen from the Figure 2. The results can be summarized as follows: the method presented in this paper may cope with well the constant value faults, but there is some lag for the fluctuant value faults.

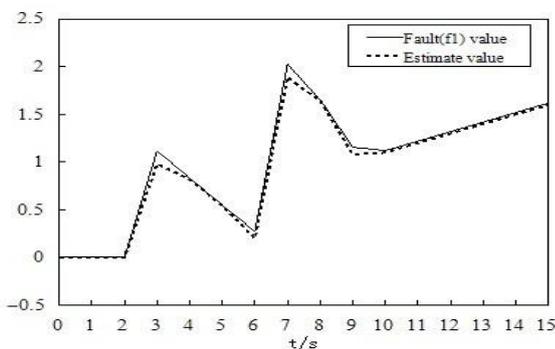


FIGURE 1 The value of fault f_1 and its estimate

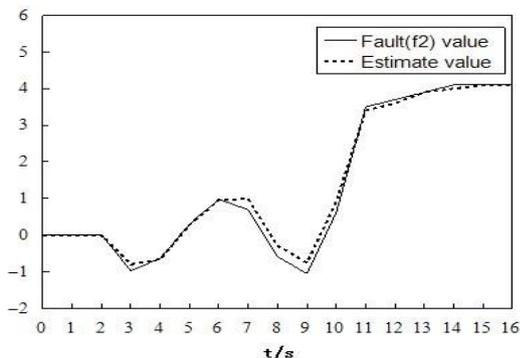


FIGURE 2 The value of fault f_2 and its estimate

b) State-feedback controller design. Letting $\gamma=0.5$, the controller gain K is solved:

$$K = \begin{bmatrix} 1.1178 & 1.6750 & -0.3748 \\ 0.8438 & 1.3940 & -0.7596 \\ -3.7636 & -5.4994 & 1.6743 \end{bmatrix}$$

The fault-tolerant dynamic output responses are

References

[1] Chanlder PR 1984 Self-repairing flight control system reliability and maintain ability program--executive overview *In Proceedings of IEEE National Aerospace and Electronics* 586-90
 [2] Veillette R J 1995 Reliable linear-quadratic state-feedback control *Automatica* 31(1) 137-43

characterized by Figure 3, which show that the closed-loop system is ensured to be stable although the open-loop system is subject to impulsive modes and bounded faults.

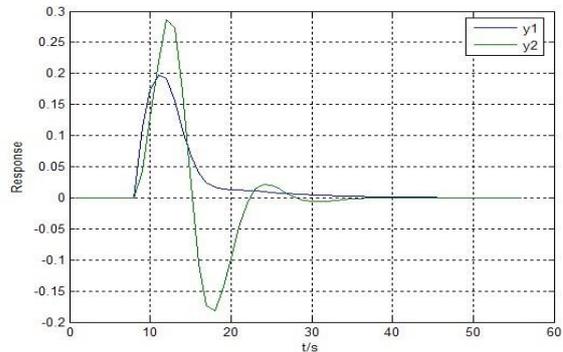


FIGURE 3 Output responses of the system via a fault-tolerant controller

From the above simulation results, it is proven that the fault diagnosis scheme is effective in estimating the states and faults of system, and the designed controller can guarantee the stability and H_∞ performance of the close-loop system when the failures occur.

6 Conclusion

This paper proposed a novel observer-based H_∞ and fault-tolerant control approach for Lipschitz singular systems with bounded perturbations and actuator faults. In terms of Lyapunov theory and linear matrix inequality (LMI) technique, a sufficient condition for the existence of the parameters of fault diagnosis observer was presented. Under assumptions 1-2, a kind of robust full-order observer was developed, which provided the information of both states and faults information. An H_∞ fault-tolerant controller via state feedback was designed. Under this controller, the Lipschitz singular systems could be solvable, impulse free and asymptotically stable, and possess the prescribed H_∞ performance. Moreover, the result obtained in this study is reliable in computation and preferable in application. Furthermore, a numerical example was provided to illustrate the effectiveness of the proposed approach. Future work will focus on the problem of designing the fault detection, the fault diagnosis and fault-tolerant control in an integrated manner for on-line application.

Acknowledgement

We gratefully acknowledge financial support by Natural Science Foundation of Hebei Province under Grant F2014208169.

- [5] Jiang J, Yu X 2012 Fault-tolerant control systems: a comparative study between active and passive approaches *Annual Reviews in Control* **36** 60-72
- [6] Gertler J 1997 Fault detection and isolation using parity relations *Control Engineering Practice* **5**(5) 853-61
- [7] Matía F, Jiménez A, Al-Hadithi BM., Rodríguez-Losada D, Galán R 2006 The fuzzy Kalman filter: state estimation using possibilistic techniques *Fuzzy Sets and Systems* **157**(16) 2145-70
- [8] Zhuang L F, Pan F, Ding F 2012 Parameter and state estimation algorithm for single-input single-output linear systems using the canonical state space models *Applied Mathematical Modelling* **36**(8) 3454-63
- [9] Mao Z H, Jiang B, Shi P 2010 Fault-tolerant control for a class of sampled-data systems via a Euler approximate observer *Automatica* **46** 1852-9
- [10] Zhang K, Jiang B 2010 Fault diagnosis observer-based output feedback fault tolerant control design *Acta Automatica Sinica* **36**(2) 274-81
- [11] Sharifuddin M, Goutam C, Kingshook B. 2010 LMI approach to unknown input observer design for continuous systems with noise and uncertainties *International Journal of Control, Automation, and Systems* **8**(2) 210-9
- [12] Yip E, Sincover R 1981 *IEEE Transactions on Automatic Control* **26**(3) 702-7
- [13] Wu H S, Mizukami K 1994 Stability and stabilization of nonlinear descriptor systems with uncertainties *In Proc.33rd Conf. Decision and Control, Lake Buena Vista FL 2772-7*
- [14] Wang H S, Yung C F, Chang F R 2002 *IEEE Transactions on Automatic Control* **47**(11) 1919-25
- [15] Lin C, Wang Q.G, Lee T H 2006 *IEEE Transactions on Fuzzy Systems* **14**(4) 542-51
- [16] Shields D N 1997 Observer design and detection for nonlinear descriptor systems *International Journal of Control* **67** 153-68
- [17] Lu G P, Ho D W C 2006 *IEEE Transactions on Automatic Control* **51**(5) 818-23
- [18] Wang H J, An X K, Lu R Q 2009 Absolute stability criteria for a class of nonlinear singular systems with time delay *Nonlinear Analysis: Theory, Methods & Applications* **70**(2) 621-30
- [19] Gao Z W, Ho D W C *IEEE Transactions on Signal Processing* **54**(4) 1316-26

Authors



Zhang Dongwen, born in December, 1964, Shijiazhuang County, Hebei Province, China

Current position, grades: the professor of school of information science & engineering, Hebei University of Science & Technology, China.

University studies: M. Eng. from Harbin Institute of Technology in China. D. Eng. from Beijing Institute of Technology in China.

Scientific interest: robust control, fault-tolerant control.

Publications: 50 papers.

Experience: teaching experience of 24 years, 10 scientific research projects.



An Xinguo, born in April, 1988, Shijiazhuang County, Hebei Province, China

Current position, grades: postgraduate student of school of sciences, Hebei University of Science & Technology, China.

University studies: B.Sc. in mathematics and applied mathematics from Hengshui University in China. M.Sc. in applied mathematics from Hebei University of Science and Technology in China.

Scientific interest: robust control, fault-tolerant control.

Publications: 4 papers.

Equilibrium distributions of the queue length in M/M/c queuing system

Jian Xu*

School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing, 100876, China

Received 1 July 2014, www.tsi.lv

Abstract

In this paper, An M/M/c queuing system with multiple working vacations and vacation interruption is considered. All servers work at a lower rate rather than completely stop during a vacation period. Meanwhile, we introduce another vacation policy: vacation interruption. Otherwise using matrix-geometric solution method, we obtain steady-state distribution for queue length.

Keywords: M/M/c; working vacation; matrix-geometric solution

1 Introduction

1.1 LITERATURE REVIEW

During the last two decades, the queuing systems with server vacations or working vacations have been investigated extensively due to their applications in various fields, such as computer systems, communication networks, production managing, etc. General vacation models can be found in Tian and Zhang [1].

About the study of working vacations, in 2002, Servi and Finn [2] first studied an M/M/1 queue with working vacations (Such model is denoted by M/M/1/WV queue), where inter-arrival times, service times during service period, service times during vacation period, and vacation times are all exponentially distributed. They developed the explicit formulae for the mean and variance number of customers in the system. Later in [3], Wu and Takagi extended Servi and Finn's M/M/1/WV queue to an M/G/1/WV queue. They assumed that service times during service period, service times during vacation period as well as vacation times are all generally distributed. Further, they assumed that when a working vacation ends, if there are customers in the system, the server changes to another service rate, where the service times follow a different distribution. In [4], Baba extended Servi and Finn's M/M/1/WV queue to a GI/M/1/WV queue. They not only assumed general independent arrival, they also assumed service times during service period, service times during vacation period as well as vacation times following exponential distribution. Furthermore, Baba derived the steady- state system length distributions at arrival and arbitrary epochs. In [5], Banik et al. studied a finite capacity GI/M/1 queue with multiple working vacations and presented a series of numerical results. For more comprehensive and excellent study on the working

vacation models, the readers may refer to [6-8] and references therein for details.

For the Multi-server vacation models, there are only a limited number of studies due to the complexity of the systems. The M/M/c queue with exponential vacations was first studied by Levy and Yechiali [9]. Chao and Zhao [10] investigated a GI/M/c vacation system and provided an algorithm to compute the performance measures. Tian et al. [11] gave a detailed study of the M/M/c vacation systems in which all servers take multiple vacation policy when the system is empty. Later, Zhang and Xu [12], Zhang and Tian [13] and Ke et al. [14] analysed the M/M/c vacation systems with a "partial server multiple vacation policy".

Existing research works about multi-server vacation models, including those mentioned above, have not related to "the working vacation policy". Besides the lack of research work on this problem, the existing literature about "vacation Interruption" also focuses on a single server model. Li and Tian [15] first introduced and studied an M/M/1 queue. Using the matrix analytic method, Li and Tian [16] generalized their results to the discrete-time GI/Geo/1 queue. The continuous-time GI/M/1queue was analysed by Li et al. [17]. Using the method of a supplementary variable, Zhang and Hou [18] investigated the M/G/1 queue with working vacations and vacation interruption.

1.2 CHARACTERISTICS OF OUR MODEL

The model we consider has some certain implications in practice. In some situations, the number of servers is not one so that a more general model should be used. Therefore, compared with previous studies, our model is more general, and an M/M/1 queue system can be seen as a special case of M/M/c queue system, that is $c = 1$. In the cyclic service queue system, which is always used to

* *Corresponding author* e-mail: xunjian311987@126.com

reconfigure the communication network, we can adopt the working vacation policy to model. Moreover, the analysis of our model can also provide the theory and analysis method to design the optimal lower service rate.

The rest of this paper is organized as follows. In Section 2, we give a brief description of the mathematical model, and a quasi-birth and death (QBD) process is demonstrated. In Section 3, we compute the stationary distributions by using the matrix geometric solution method, and the conditional stochastic decomposition properties are proved in this part. Finally, Section 4 summarizes the investigation and draws the conclusion.

2 Mathematical model and QBD process

We consider the M/M/c/ queue with Working Vacations and vacation Interruption, and the specific application about the “Working Vacations and vacation Interruption” policy in this paper is introduced in 1.2 (Characteristics of our model). It is assumed that customers arrive according to a Poisson process with rate λ . The service times during busy period follow exponential distribution with mean $1/\mu_1$. The service times during vacation period follow another exponential distribution with mean $1/\mu_2$. Vacation times are exponentially distributed with mean $1/\theta$. The service order is assumed to be First Come First Served. In addition, inter-arrival times, service times, and vacation times are mutually independent.

Let $Q_v(t)$ be the number of customers in the system at time t , let $J(t)$ be the indicator variable defined by:

$$J(t) = \begin{cases} 0, & \text{all servers are on a busy period at time } t \\ 1, & \text{all servers are on a working period at time } t \end{cases}$$

Then, the $\{Q_v(t), J(t)\}$ is a quasi-birth-death process (QBD) with the state space:

$$\Omega = \{0, 1\} \cup \{(k, j), k \geq 1, j = 0, 1\} \tag{1}$$

Referring to the state-transition-rate diagram as shown in Figure 1 the infinitesimal generator Q of the QBD describing the M/M/c queuing system with Working Vacations and vacation Interruption is of the form:

$$Q = \begin{matrix} k=0 & \begin{bmatrix} A_0 & C_0 \\ B_1 & A_1 & C_1 \\ B_2 & A_2 & C_2 \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ k=c-1 & \dots & \dots & B_{c-1} & A_{c-1} & C_{c-1} \\ k=c & \dots & \dots & B_c & A_c & C_c \\ k=c+1 & \dots & \dots & B_{c+1} & A_{c+1} & C_{c+1} \\ k=c+2 & \dots & \dots & B_{c+2} & A_{c+2} & C_{c+2} \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} \end{matrix} \tag{2}$$

The entries $B_k (0 \leq k \leq c)$, $A_k (0 \leq k \leq c)$ and $C_k (0 \leq k \leq c)$ are different matrices, define:

$$A_0 = \lambda, C_0 = (0, \lambda), B_1 = (\mu_1, \mu_2)^T$$

Other matrices are 2×2 square matrices, and $C = C_k = \lambda I (1 \leq k \leq c)$, where I is an identity matrix

$$A_k = \begin{bmatrix} -(\lambda + k\mu_1) & \theta \\ \theta & -(\lambda + \theta + k\mu_2) \end{bmatrix}, (1 \leq k \leq c-1),$$

$$A = A_k = \begin{bmatrix} -(\lambda + c\mu_1) & \theta \\ \theta & -(\lambda + \theta + c\mu_2) \end{bmatrix}, (k \geq c),$$

$$B_k = \begin{bmatrix} k\mu_1 & 0 \\ k\mu_2 & 0 \end{bmatrix}, (2 \leq k \leq c-1),$$

$$B = B_k = \begin{bmatrix} c\mu_1 & 0 \\ c\mu_2 & 0 \end{bmatrix}, (k \geq c).$$

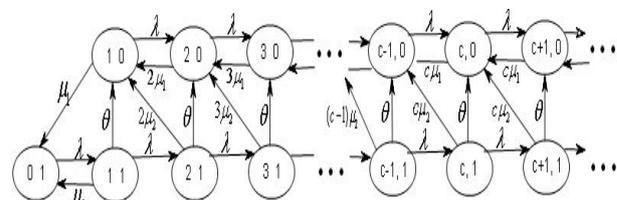


FIGURE 1 State-transition-rate diagram

Note that Q is also viewed as the infinitesimal generator for the QBD process. To analyse this QBD process, a very important matrix in evaluating the performance measures is the matrix R . It is known as the rate matrix, and it is the minimal non-negative solution of the matrix quadratic equation (the readers are referred to Neuts [19]):

$$R^2 B + R A + C = 0. \tag{3}$$

Based on the structures of matrices, A and C , which are represented as the lower triangular matrix, thus the matrix solution R is also the lower triangular matrix.

Doing some arduous algebraic derivations and arrangement, we develop the explicit formula for matrix R in the theorem below. Firstly, let's assume that $\rho = \frac{\lambda}{c\mu_1}$.

Theorem 1. If $\rho < 1$, the matrix Equation (3) has the

minimal nonnegative solution as $R = \begin{bmatrix} \rho & 0 \\ r_{21} & r_{22} \end{bmatrix}$, where:

$$r_{21} = \frac{\lambda(\lambda + \theta)}{c\mu_1(c\mu_2 + \lambda + \theta)}, r_{22} = \frac{\lambda}{c\mu_2 + \lambda + \theta}.$$

Proof: Based on the structures of matrices A, B and C we can assume that R has the same structure as:

$$R = \begin{bmatrix} \rho & 0 \\ r_{21} & r_{22} \end{bmatrix}.$$

Substituting R into the matrix Equation (3), we obtain the following system of equations:

$$\begin{cases} c\mu_1 r_{11}^2 - r_{11}(c\mu_1 + \lambda) + \lambda = 0 & (4.1) \\ -r_{22}(c\mu_2 + \lambda + \theta) + \lambda = 0 & (4.2) \\ c\mu_2 r_{22}^2 + c\mu_1 r_{21}^2 (r_{11} + r_{22}) - r_{21}(c\mu_1 + \lambda) = 0 & (4.3) \end{cases} \quad (4)$$

To obtain the minimal nonnegative solution of Equation (3), in the first equation of the system of Equations (4.1), let $r_{11} = \rho$ (the other root is $r_{11} = 1$). In the second equation of the system of equations (4.2), we can get $r_{22} = \frac{\lambda}{c\mu_2 + \lambda + \theta}$.

Substituting $r_{11} = \rho$ and $r_{22} = \frac{\lambda}{c\mu_2 + \lambda + \theta}$ into the third Equation (4.3), we get:

$$r_{21} = \frac{c\mu_2 r_{22}^2 + r_{22}\theta}{c\mu_1(1-r_{22})} = \frac{\lambda(\lambda + \theta)}{c\mu_1(c\mu_2 + \lambda + \theta)} \text{ and } 0 < r_{21} < 1.$$

In fact: $r_{21} = \frac{\lambda(\lambda + \theta)}{c\mu_1(c\mu_2 + \lambda + \theta)} = \frac{\lambda}{c\mu_1} \frac{\lambda + \theta}{c\mu_2 + \lambda + \theta}$.

When $\rho = \frac{\lambda}{c\mu_1} < 1$, we can get $r_{21} < 1$.

Furthermore, we can verify that all diagonal elements of rate matrix R are less than 1. Therefore, the spectral radius of rate matrix R , $SP(R) = \max\{\rho, r_{22}\}$ is less than 1. Based on the theorem (Neuts [19]), we can prove that the QBD process $\{Q_v(t), J(t)\}$ is positive recurrent if and only if $\rho < 1$.

3 Stationary distributions

$\rho < 1$ is the stability condition of the state process. Using the rate matrix R , we can solve the steady-state probability more efficiently. Let (Q_v, J) be a set of random variables which follows the stationary distribution of the QBD process $\{Q_v(t), J(t)\}$. Define:

$$\Pi = (\pi_0, \pi_1, \pi_2, \dots) \text{ and } \pi_0 = (\pi_{01}), \pi_k = (\pi_{k0}, \pi_{k1}), k \geq 1.$$

$$\pi_{kj} = P\{Q_v = k, J = j\} =$$

$$\lim_{t \rightarrow \infty} P\{Q_v(t) = k, J(t) = j\}, (k, j) \in \Omega$$

So the stationary distribution for this QBD process is given as follows:

Theorem2. If $\rho < 1$, the distribution of (Q_v, J) is:

$$\pi_{m,1} = \begin{cases} \prod_{i=1}^m \frac{\lambda}{i\mu_2 + \lambda + \theta} K, & 1 \leq m \leq c-1 \\ \prod_{i=1}^{c-1} \frac{\lambda}{i\mu_2 + \lambda + \theta} \left(\frac{\lambda}{c\mu_2 + \lambda + \theta}\right)^{m-c+1} K, & m \geq c \end{cases} \quad (5.1), (5)$$

$$\pi_{m,0} = \begin{cases} K \frac{1}{m!} \left(\frac{\lambda}{\mu_1}\right)^m \varphi_m, & 1 \leq m \leq c-1 \\ \pi_{c-1,0} \rho^{m-1+1} + \pi_{c-1,1} r_{21} \sum_{i=0}^{m-c} r_{11}^i r_{22}^{m-c-i}, & m \geq c \end{cases} \quad (6.1), (6.2)$$

$$K = (K_1 + K_2 + K_3)^{-1},$$

where:

$$\varphi_m = 1 - \frac{\mu_2}{\mu_2 + \lambda + \theta} - \alpha - \beta + \sum_{p=1}^{c-1} p! \left(\frac{\mu_1}{\lambda}\right)^p,$$

$$\alpha = \sum_{k=2}^m \sum_{j=1}^k \frac{j(k-1)! \mu_1^{k-1} \mu_2}{\lambda^{k-j} \prod_{i=1}^j (i\mu_2 + \lambda + \theta)},$$

$$\beta = \sum_{k=2}^m \sum_{j=1}^k \frac{\theta(k-1)! \mu_1^{k-1}}{\lambda^{k-j} \prod_{i=1}^j (i\mu_2 + \lambda + \theta)},$$

$$r_{11} = \frac{\lambda}{c\mu_1}, r_{22} = \frac{\lambda}{c\mu_2 + \lambda + \theta}, r_{21} = \frac{\lambda(\lambda + \theta)}{c\mu_1(c\mu_2 + \lambda + \theta)},$$

$$K_1 = \sum_{m=1}^{c-1} \left[\prod_{i=1}^m \frac{\lambda}{i\mu_2 + \lambda + \theta} + \frac{1}{m!} \frac{\lambda}{\mu_1} \varphi_m \right],$$

$$K_2 = \sum_{m=c}^{\infty} \left[\prod_{i=1}^{c-1} \frac{\lambda}{i\mu_2 + \lambda + \theta} + \left(\frac{\lambda}{c\mu_2 + \lambda + \theta}\right)^{m-c+1} \right],$$

$$K_3 = \sum_{m=c}^{\infty} \left[\frac{1}{(c-1)!} \left(\frac{\lambda}{\mu_1}\right)^{c-1} \varphi_{c-1} \rho^{m-c+1} + \right.$$

$$\left. \prod_{i=1}^{c-1} \frac{\lambda}{i\mu_2 + \lambda + \theta} r_{21} \sum_{i=1}^{m-c} r_{11}^i r_{22}^{m-c-i} \right]$$

Proof: Using the matrix geometric solution method(see[19]),we can get:

$$\pi_m = \pi_{c-1} R^{m-c+1}, m \geq c,$$

the equations that $\pi_{01}, \pi_{10}, \pi_{11}, \pi_{20}, \pi_{21}, \dots, \pi_{c-1,0}, \pi_{c-1,1}$ satisfy are:

$$\begin{cases} -\lambda\pi_{01} + \mu_1\pi_{10} + \mu_2\pi_{11} = 0 & (7.1) \\ -(\lambda + \mu_1)\pi_{10} + \theta\pi_{11} + 2\mu_1\pi_{20} + 2\mu_2\pi_{21} = 0 & (7.2) \\ \lambda\pi_{01} - (\lambda + \theta + \mu_2)\pi_{11} & (7.3) \\ \lambda\pi_{10} - (\lambda + 2\mu_1)\pi_{20} + \theta\pi_{21} + 3\mu_1\pi_{30} + 3\mu_2\pi_{31} & (7.4) \\ \lambda\pi_{11} - (\lambda + \theta + 2\mu_2)\pi_{21} & (7.5) \\ \lambda\pi_{20} - (\lambda + 3\mu_1)\pi_{20} + \theta\pi_{31} + 4\mu_1\pi_{40} + 4\mu_2\pi_{41} & (7.6) \\ \lambda\pi_{21} - (\lambda + \theta + 3\mu_2)\pi_{31} & (7.7) \\ \dots & \\ \lambda\pi_{c-3,0} - [\lambda + (c-2)\mu_1]\pi_{c-2,0} + \theta\pi_{c-2,1} + & \\ (c-1)\mu_1\pi_{c-1,0} + (c-1)\mu_2\pi_{c-1,1} = 0 & (7.8) \\ \lambda\pi_{c-3,1} - [\lambda + \theta + (c-2)\mu_2]\pi_{c-2,1} = 0 & (7.9) \\ \lambda\pi_{c-2,1} - [\lambda + \theta + (c-1)\mu_2]\pi_{c-1,1} = 0 & (7.10) \end{cases} \quad (7)$$

Define: $\pi_{01} = K$.

Through a series of calculation, we can get the following result from the Equation (7) {(7.3),(7.5),(7.7),..., (7.10)}.

$$\pi_{11} = \frac{\lambda}{\lambda + \theta + \mu_2} K, \quad (7.3a)$$

$$\pi_{21} = \frac{\lambda}{\lambda + \theta + 2\mu_2} \pi_{11} = \frac{\lambda}{\lambda + \theta + 2\mu_2} \frac{\lambda}{\lambda + \theta + \mu_2} K, \quad (7.5a)$$

$$\pi_{31} = \frac{\lambda}{\lambda + \theta + 3\mu_2} \pi_{21} = \frac{\lambda}{\lambda + \theta + 3\mu_2} \frac{\lambda}{\lambda + \theta + 2\mu_2} \frac{\lambda}{\lambda + \theta + \mu_2} K, \quad (7.7a)$$

...

$$\pi_{c-1,1} = \frac{\lambda}{\lambda + \theta + (c-1)\mu_2} \pi_{c-2,1} = \prod_{i=1}^{c-1} \frac{\lambda}{\lambda + \theta + i\mu_2} K. \quad (7.10a)$$

Clearing up the above formulas, we can obtain Equation (5.1).

Meanwhile, we can get the following result from Equations (7) {(7.2), (7.4), (7.6),..., (7.8)}.

$$2\mu_1\pi_{20} - \lambda\pi_{10} = \mu_1\pi_{10} - \theta\pi_{11} - 2\mu_2\pi_{21} \quad (7.2a)$$

References

[1] Tian N, Zhang Z 2006 Vacation Queuing Models – Theory and Applications Springer-Verlag New York
 [2] Servi L, Finn S 2002 M/M/1 queue with working vacations (M/M/1/WV) *Perform Eval* 50 41–52
 [3] Wu D, Takagi H 2006 M/G/1 queue with multiple working vacations *Perform Eval* 63 654–81

$$3\mu_1\pi_{30} - \lambda\pi_{20} = 2\mu_1\pi_{20} - \lambda\pi_{10} - \theta\pi_{21} - 3\mu_2\pi_{31} \quad (7.4a)$$

$$4\mu_1\pi_{40} - \lambda\pi_{30} = 3\mu_1\pi_{30} - \lambda\pi_{20} - \theta\pi_{31} - 4\mu_2\pi_{41} \quad (7.6a)$$

...

$$(c-1)\mu_1\pi_{c-1,0} - \lambda\pi_{c-2,0} = \dots \quad (7.8a)$$

$$(c-2)\mu_1\pi_{c-2,0} - \lambda\pi_{c-3,0} - \theta\pi_{c-2,1} - (c-1)\mu_2\pi_{c-1,1}$$

The above formulas on both sides of the equal are added, respectively. Through a series of calculation, we can get:

$$\begin{aligned} (c-1)\mu_1\pi_{c-1,0} - \lambda\pi_{c-2,0} &= \mu_1\pi_{10} - \theta\pi_{11} - \\ 2\mu_2\pi_{21} - \theta\pi_{21} - \dots - \theta\pi_{c-2,1} - (c-1)\mu_2\pi_{c-1,1} & \end{aligned} \quad (8)$$

By (7.1), we can get:

$$\mu_1\pi_{10} = \lambda\pi_{01} - \mu_2\pi_{11}. \quad (7.1a)$$

Substituting Equation (7.1a) into Equation (8), we obtain Equation (6.1).

And $R = \begin{bmatrix} \rho & 0 \\ r_{21} & r_{22} \end{bmatrix}$, so we can calculate:

$$R^{m-1+1} = \begin{bmatrix} \rho^{m-c+1} & 0 \\ r_{21} \sum_{i=0}^{m-c} r_{11}^i r_{22}^{m-c-i} & r_{22}^{m-c+1} \end{bmatrix}.$$

Substituting $\pi_{c-1,0}$ (5.1), $\pi_{c-1,1}$ (6.1) and R^{m-c+1} into $(\pi_{m,0}, \pi_{m,1}) = (\pi_{c-1,0}, \pi_{c-1,1}) R^{m-c+1}$, we can obtain Equations (5.2) and (6.2).

Finally by the normalization conditions can be obtained, we can get π_{01} , i.e K .

4 Conclusions

In this paper, we study an M/M/c queuing system with multiple working vacation and vacation interruption using matrix-analytic method. This system is formulated as a QBD process, the necessary and sufficient condition for the stability of the system was deduced. More important, the explicit solution of stable condition and rate matrix of the QBD model are obtained, and then the stationary probability distributions are explicitly developed.

[4] Baba Y 2005 Analysis of a GI/M/1 queue with multiple working vacations *Oper Res Lett* 33 201–9
 [5] Banik A D, Gupta U C, Pathak S S 2007 On the GI/M/1/N queue with multiple vacation-analytic analysis and computation *Appl Math Model* 31 1701–10

- [6] Jain M, Sharma G C, Sharma R 2011 Working vacation queue with service interruption and multi optional repair *Int J Inform Manage Sci* **22** 157–75
- [7] Ke J C, Wu C H, Zhang Z G 2010 Recent developments in vacation queueing models: a short survey *Int J Oper Res* **7** 3–8
- [8] Zhang M, Hou Z 2010 Performance analysis of M/G/1 queue with working vacations and vacation interruption *J Comput Appl Math* **234** 2977–85
- [9] Levy Y, Yechiali U 1976 An M/M/c queue with servers' vacations, *INFOR* **14** 153–63
- [10] Chao X, Zhao Y 1998 Analysis of multiple-server queues with station and server vacations *European Journal of Operational Research* **110** 392–406
- [11] Tian N, Xu X 2001 M/M/c queue with synchronous multiple vacation of partial servers *OR Trans* **5** (3) 85–94
- [12] Xu X, Zhang Z G 2006 Analysis of multiple-server queue with a single vacation (e, d)-policy *Performance Evaluation* **63** 825–38
- [13] Zhang Z G, Tian N 2003 Analysis on queueing systems with synchronous vacations of partial servers *Performance Evaluation* **52** 269–82
- [14] Ke J-C, Lin C H, Yang J Y, Zhang Z G 2009 Optimal (d, c) vacation policy for a finite buffer M/M/c queue with unreliable servers and repairs *Applied Mathematical Modeling* **33** 3949–62
- [15] Li J, Tian N 2007 The M/M/1 queue with working vacations and vacation interruption *J Syst Sci Syst Eng* **16** 121–27
- [16] Li J, Tian N 2007 The discrete-time GI/Geo/1 queue with working vacations and vacation interruption *Appl Math Comput* **185** 1–10
- [17] Li J, Tian N, Ma Z 2008 Performance analysis of GI/M/1 queue with working vacations and vacation interruption *Appl Math Modell* **32** 2715–30
- [18] Zhang M, Hou Z 2010 Performance analysis of M/G/1 queue with working vacations and vacation interruption *J Comput Appl Math* **234** 2977–85
- [19] Neurs M 1981 Matrix-Geometric Solution in Stochastic Models *Johns Hopkins University Press* Baltimore 1981
- [20] Fuhrmann S W, Cooper R B 1985 Stochastic decomposition in the M/G/1 queue with generalized vacations *Oper Res* **33** 1117–29
- [21] Miyazawa N 1994 Decomposition formulas for single server queues with vacations: An unified approach *Stochastic Models* **10** 389–413

Author



Jian Xu, born in 1987, China

University studies: PhD in School of Economics and Management at Beijing University of Posts and Telecommunications in China.

Scientific interest: supply chain management, operations research and optimization, queueing theory, logistics management.

Novel method for quality assessment of computational translation

Juanjuan Suo^{*}, Huimin Dong

Hebei University of Engineering, Handan 056038, China

Received 1 July 2014, www.tsi.lv

Abstract

To overcome the shortcomings that there are few feasible methods and models in the comprehensive assessment on the quality of the computational translation, a novel mathematical tool, the unascertained measure was introduced. After the introduction of the basic knowledge of the Unascertained Sets, the unascertained measure was defined and the comprehensive assessment model was set up. Then the method was introduced to the quality assessment of the machine translation. Engineering practices shows that the method can complete the assessment systematically and scientifically without any assumption.

Keywords: computational translation, quality assessment, unascertained measure, model

1 Introduction

As one of the computational linguistics research field, the emergence of the machine translation drives the development of the information society [1]. In the past years, lots of works have been done on it. And there are many machine translation systems available today [2]. They have the advantages of speed, cost-efficiency, and the ability to deal with sheer volume of translation task. However, there is one thing computer cannot beat human being, at least at the present time and near future, which is the quality of ambiguity. As the key and biggest difficulty of computational linguistics, the ambiguity is the chief bottleneck of computer analysis and understanding. So, the assessment of the computational translation has significance in theory and practice for the development of the computational linguistics and the information society. Many scholars devoted to the related research and have proposed many effective theories and methods [3-10]. But, we still have a long way to go.

According to this situation, a new method, the unascertained set was introduced to solve the unascertained problem of the assessment. The unascertained measure was introduced and the credible identification was set up for the reliability assessment. Application results showed that it could complete the reliability assessment systematically and scientifically.

The rest of the paper was organized as follows. In the introductory part, attention was paid to the basic concepts of the Unascertained Mathematics. In the following part, a reliability assessment model was set up. Then, its application in practice was introduced. Finally, the advantages of the method proposed here were pointed out.

2 Machine translation

Machine translation is also called computer translation and electronic translation. The research of machine translation in China was started in 1956 and the first test was carried out successfully in 1959. With the rapid development of network technique, machine translation becomes more prosperous when more challenges appear. And unsolved problems for many years still exist, many work have to be finished until machine translation technique mature.

2.1 PRINCIPLE AND FUNCTION OF MACHINE TRANSLATION

The process of Machine translation can be divided into five parts, includes original language input, original language analysis, transfer of the original language to target language, target language generation and target language output. Although the research work of machine translation has been underway for many years, its application has great limitations. Currently, the most advanced machine translation systems are only used to replace human translation in a limited range.

The fundamental principle of machine translation is to build machine dictionary, terminology database, data bank for translators by huge storage capacity and rapid retrieval ability of computer. Thus the retrieval time of translators can be saved greatly. Some machine translation system can store, revise and print translated text to help to improve work efficiency apart from providing retrieval function. Machine translation system allows people intervene during analysis, transformation and generation. The problems can be solved easily by people intervene under certain circumstance, such as language construction ambiguity. In order to improve translation efficiency and ensure the

^{*} *Corresponding author* e-mail: suojj@163.com

consistency of translated text, machine translation system usually contains a series of tool components. And two major tool components are translation memory and terminology management.

Translation memory is equivalent database of original and translated text by machine building. Computer can store translated words needed translation in language database when translators are working. During the process of translation, computer will show translators matching sentences translation when the same and similar sentences appear. Most of translation memory software support fuzzy match, users can set minimum matching degree. The translators still can obtain a sentence of fuzzy match by fuzzy match, and then all that is needed is to translate or revise the different parts. Machine memory update constantly and store automatically new translation users providing. Along with constant rich of memory database, the work efficiency of translation will becomes much higher. The work principle of terminology management is to scan one by one original terminology by machine and check in its dictionary base. Terminology management ensures consistency and accuracy of translation terminology and the work efficiency of translators.

2.2 PROBLEM OF COMPUTATIONAL DISTINCTICS AMBIGUITY

The ambiguity problem is one of the main core problems in Computational Linguistics. In the beginning of machine translation, this problem didn't obtain enough attention. So the machine translations soon fell into an unprecedented crisis and directly lead to the appearance of ALPAC report. It made people further realize the importance of ambiguity problem. Language disambiguation is a challenge in machine translation. And the ambiguity phenomenon is the universal phenomenon in nature language. Ambiguity processing is key to improve the translation quality. Ambiguity, according to sources, is divided into vocabulary ambiguity and structural ambiguity. Vocabulary ambiguity is one of parts of speech ambiguity to carry on the syntactic analysis. It easily leads to the extremely syntactic analysis errors. Meaning ambiguity directly leads to the wrong statement. Structural ambiguity is generally caused by the same syntactic structure, and it should be eliminated through the text analysis of the subject and the analysis of sentences by other components. In 1993, Lancaster University Corpus Research Center developed automatic SEMTAG. Through automatic classification of the each word, phrase and sentence, the discourse of the semantic features of general appearance and distribution state, and the calculation formula of the original text can be obtained. This method can solve the exact nature of context translation. The essence of the ambiguity is the shortage of the corresponding relation between the expression of the language form and its meaning. Ambiguity arises when there is a certain concept in language A but there is no such concept in Language B or a concept which is described by one single word in one

language may have several words to express in another language. When words or sentences are translated into other languages, ambiguities may occur because of cultural, grammar or syntactic differences among languages. This is the inherent characteristics of the natural language and it is one of the characteristic of the difference between natural language and artificial language. Human translators can handle this kind of complexity by investigating the cultural differences and conducting research to produce correct translations. However, if translated by machine, it would be impossible. The studies to natural language processing system has guiding significance to researchers, but the complex of the ambiguity phenomenon needs to put forward more perfect and more suitable methods for the ambiguity description and eliminate. There are many factors contributing to the ambiguity of the machine translation translations other than in linguistic perspective, such as computational problems. The studies to natural language processing system has guiding significance to researchers, but the complex of the ambiguity phenomenon needs to put forward more perfect and more suitable methods for the ambiguity description and eliminate. This is the inherent characteristics of the natural language and it is one of the characteristic of the difference between natural language and artificial language.

3 Basic knowledge of unascertained mathematics

The Unascertained Mathematics, proposed by Guangyuan Wang in 1990 [11], is a tool to describe the subjective uncertainty quantitatively. It mainly deals with the unascertained information, which differs from the stochastic information, fuzzy information and grey information. The unascertained information refers to the decision-making-demanded information. The information itself has no uncertainty, but because of situation constrain, the decision-maker cannot grasp the total information of them. The decision-maker himself produces the uncertainty. Since 1990s, Kaidi Liu and other scholars have done a lot of work and the Unascertained Mathematics has been successfully used in many fields [11-13].

The definition of unascertained sets is introduced systematically in [11]. Here we will briefly introduce some key points of the unascertained sets.

3.1 MEMBERSHIP FUNCTION CONSTRUCTION

The membership function of the unascertained set meets the three principles of measure and it is defined in the topology space (F, E) . Yet, the membership function of the Fuzzy Set is a function of a single variable defined in the space of U . The key to the unascertained set is the membership function construction which demands the decision-maker's experience and knowledge background.

3.2 INDEX IDENTIFICATION WEIGHT

The index's weight used in the determining of the composed membership by the single index is and only is the identification weight of the index. In this case, the common methods are used to get the weight value of each assessment index, such as Delphi method, Brainstorming, Analytic Hierarchy Process (AHP) and so on, are helpless. Here, the information entropy is employed to determine the index's identification weight.

Entropy which used to be a thermodynamic concept, it was introduced into information theory in 1948 by C. E. Shannon who put forward the concept of information entropy to measure the level of system chaos or disorder. And Shannon information entropy, which is an objective and applicable method for the determination of weight value, was introduced into the comprehensive assessment. It can calculate weight value of each index more effectively in the comprehensive assessment of marine ecological environment. In the application of Shannon information entropy method, the greater entropy weight indicates greater variation extent of relevant index, much more information and has the greater effect. So, weight value of corresponding index also should be bigger. In contrast, for the smaller entropy weight which has little effect, its weight value should be the smaller [13].

For the discrete stochastic variables, their information entropy is:

$$S = -k \sum_{i=1}^n p_i \ln p_i, \tag{1}$$

where p_i is the probability and $p_i \geq 0, \sum_{i=1}^n p_i = 1$. As u_i

($0 \leq u_i \leq 1, \sum u_i = 1$), suppose $H(\alpha) = -\sum_{k=1}^K u_{ijk} \log u_{ijk}$:

$$u_j^{(i)} = 1 + \frac{1}{\log k} H(\alpha), \tag{2}$$

then $\omega^{(i)} = (\omega_1^{(i)}, \omega_2^{(i)}, \dots, \omega_m^{(i)})$ is the weight of the I_1, I_2, \dots, I_m , where:

$$\omega_j^{(i)} = \frac{u_j^{(i)}}{\sum_{j=1}^m u_j^{(i)}}. \tag{3}$$

For the identification principle, if the ranks are orderly, the principle of maximum degree of membership is not applicable and the credible identification is often used.

4 Quality assessment model of machine translation

4.1 INDEX SYSTEM

The establishment of the assessment index system should use the system engineering theory and it should be closely

associated with the practice. After consulting the specialists, seven independent factors constitute the main assessment indexes. The indexes include readability, formality, convey degree of the related implications, convey degree of the implicit implication, covertly erroneous error, overtly erroneous error and meaning function. The system of assessment indexes is shown in Figure 1.

4.2 MODEL BASED ON UNASCERTAINED MEASURE

Suppose x_1, x_2, \dots, x_n are n translation results, I_1, I_2, \dots, I_m are indexes for the assessment of x_i , $I = \{I_1, I_2, \dots, I_m\}$, x_{ij} is the observed value of x_i under index I_j , and c_k is the k^{th} comment ($1 \leq k \leq K$).

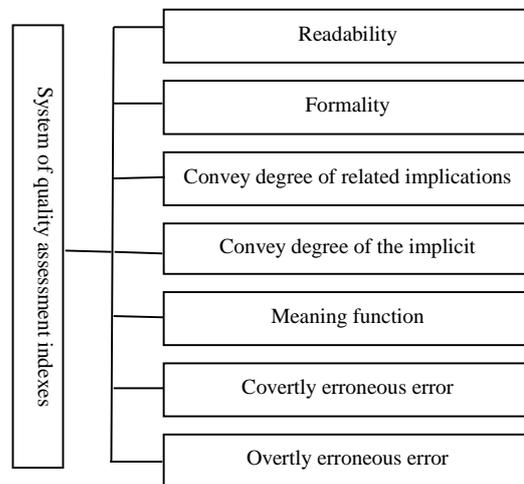


FIGURE 1 The index systems of quality assessment of translation

4.2.1 Single-index unascertained measure.

μ_{ijk} is the degree that the observed value x_{ij} of x_i belongs to the assessment rank c_k .

The Delphi method is employed to get the scores of every factor. The number of the specialists is k . Every specialist should rank the degree that $I_j (1 \leq j \leq m)$ belongs to $c_k (1 \leq k \leq K)$ by using 0-10. If the k^{th} specialist thinks that the degree I_j belongs to c_k is x_{ijk} ,

$\sum_{k=1}^K x_{ijk} = 10$, then $\mu_{ijk} = \frac{x_{ijk}}{10}$ is the unascertained measure.

After obtaining the comprehensive indexes of assessment system, standardized processing of the data should be firstly finished. Suppose the research plan is $x_{ij} (i = 1, \dots, n; j = 1, \dots, m)$. It denotes that there are i samples and j indexes in the research plan. Based on the characteristics of assessment target, the indexes are divided into the positive type and negative type. The

positive type indexes are the indexes whose values are the bigger the better. The negative type indexes refer to the indexes whose values are the smaller the better.

For the positive indexes, the normalization is as follows:

$$x_{ij} = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j} \tag{4}$$

For the negative indexes, the normalization is as follows:

$$x_{ij} = \frac{\max x_j - x_{ij}}{\max x_j - \min x_j} \tag{5}$$

where, $\max x_j$ and $\min x_j$ represent the maximum and minimum value of x_j respectively.

After the standardized processing, the standardized matrix x_{ij} be obtained.

The single-index measure assessment matrix of x_i is:

$$(\mu_{ijk})_{m \times k} = \begin{bmatrix} \mu_{i11} & \mu_{i12} & \dots & \mu_{i1k} \\ \mu_{i21} & \mu_{i22} & \dots & \mu_{i2k} \\ \dots & \dots & \dots & \dots \\ \mu_{im1} & \mu_{im2} & \dots & \mu_{imk} \end{bmatrix}, (i = 1, 2, \dots, n). \tag{6}$$

4.2.2 Identification weight of the index.

$\omega_j^{(x)}$ is the identification weight and:

$$\omega_j^{(x)} = \frac{L_j^{(i)}}{\sum_{i=1}^m L_j^{(i)}} \tag{7}$$

4.2.3 Comprehensive assessment system.

The common unascertained membership functions are as follows and four common measure functions, includes straight line distribution, parabola distribution, exponent distribution and sine distribution are shown in Figures 2-5.

$$\begin{cases} \mu_i(x) = \begin{cases} \frac{-x}{a_{i+1} - a_i} + \frac{a_{i+1}}{a_{i+1} - a_i} & a_i < x \leq a_{i+1} \\ 0 & x > a_{i+1} \end{cases} \\ \mu_{i+1}(x) = \begin{cases} 0 & x \leq a_i \\ \frac{x}{a_{i+1} - a_i} + \frac{a_i}{a_{i+1} - a_i} & a_i < x \leq a_{i+1} \end{cases} \end{cases}, \tag{8}$$

$$\begin{cases} \mu_i(x) = \begin{cases} 1 - \left(\frac{x - a_i}{a_{i+1} - a_i}\right)^2 & a_i < x \leq a_{i+1} \\ 0 & x > a_{i+1} \end{cases} \\ \mu_{i+1}(x) = \begin{cases} 0 & x \leq a_i \\ \left(\frac{x - a_i}{a_{i+1} - a_i}\right)^2 & a_i < x \leq a_{i+1} \end{cases} \end{cases}, \tag{9}$$

$$\begin{cases} \mu_i(x) = \begin{cases} 1 - \frac{1 - e^{x - a_i}}{1 - e^{a_{i+1} - a_i}} & a_i < x \leq a_{i+1} \\ 0 & x > a_{i+1} \end{cases} \\ \mu_{i+1}(x) = \begin{cases} 0 & x < a_i \\ \frac{1 - e^{x - a_i}}{1 - e^{a_{i+1} - a_i}} & a_i < x \leq a_{i+1} \end{cases} \end{cases}, \tag{10}$$

$$\begin{cases} \mu_i(x) = \begin{cases} 1 - \frac{1 - e^{x - a_i}}{1 - e^{a_{i+1} - a_i}} & a_i < x \leq a_{i+1} \\ 0 & x > a_{i+1} \end{cases} \\ \mu_{i+1}(x) = \begin{cases} 0 & x < a_i \\ \frac{1 - e^{x - a_i}}{1 - e^{a_{i+1} - a_i}} & a_i < x \leq a_{i+1} \end{cases} \end{cases}. \tag{11}$$

After the single index matrix and the identification weight are derived, the comprehensive assessment vector μ^i can be derived:

$$\mu^i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{in}) = \omega^i \cdot (\mu_{ijk})_{m \times k}, \tag{12}$$

where μ^i is the unascertained classification, in order to obtain the certainty classification, the identification is needed.

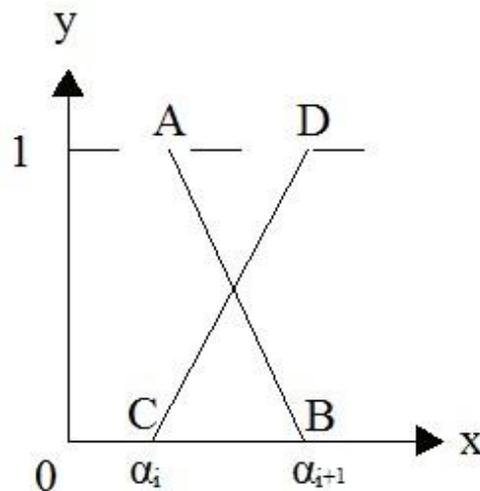


FIGURE 2 Straight line distribution

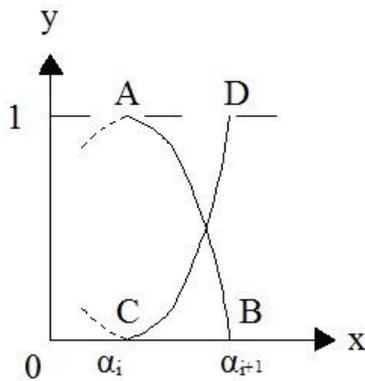


FIGURE 3 Parabola distribution

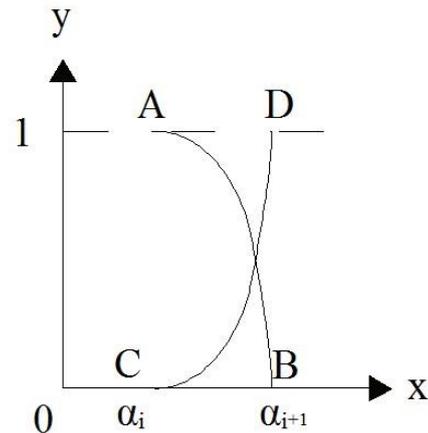


FIGURE 5 Sine distribution

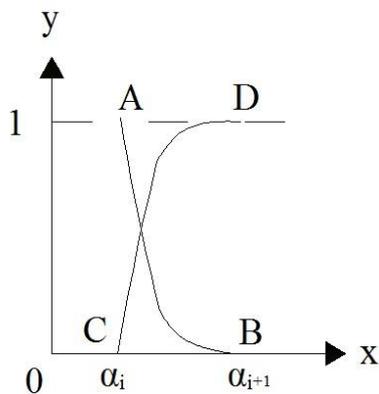


FIGURE 4 Exponent distribution

5 Practice application

Using the method mentioned above, we finished the reliability assessment of a translation result. Five TABLE 1 Scores given by the specialists

Index	Specialist	1	2	3	4	5
Readability		3.7	4.1	3.5	3.7	3.2
Formality		3.5	3.9	3.4	3.6	3.2
Convey degree of related implications		2.9	3.0	2.9	2.8	3.1
Convey degree of the implicit implication		3.1	3.2	3.0	2.9	3.0
Meaning function		3.0	2.8	3.1	3.0	3.1
Covertly erroneous error		3.9	3.8	3.4	4.0	3.8
Overtly erroneous error		3.5	3.4	3.3	3.8	3.5

Thus the policy-making matrix can be obtained:

$$x'_{7 \times 5} = \begin{bmatrix} 3.7 & 4.1 & 3.5 & 3.7 & 3.2 \\ 3.5 & 3.9 & 3.4 & 3.6 & 3.2 \\ 2.9 & 3.0 & 2.9 & 2.8 & 3.1 \\ 3.1 & 3.2 & 3.0 & 2.9 & 3.0 \\ 3.0 & 2.8 & 3.1 & 3.0 & 3.1 \\ 3.9 & 3.8 & 3.4 & 4.0 & 3.8 \\ 3.5 & 3.4 & 3.3 & 3.8 & 3.5 \end{bmatrix}$$

$$(\mu_{ijk})_{m \times k} = \begin{bmatrix} 0.20 & 0.23 & 0.19 & 0.20 & 0.18 \\ 0.20 & 0.22 & 0.19 & 0.20 & 0.19 \\ 0.20 & 0.20 & 0.20 & 0.19 & 0.21 \\ 0.20 & 0.21 & 0.20 & 0.19 & 0.20 \\ 0.20 & 0.19 & 0.21 & 0.20 & 0.20 \\ 0.21 & 0.20 & 0.18 & 0.21 & 0.20 \\ 0.20 & 0.19 & 0.19 & 0.22 & 0.20 \end{bmatrix}$$

The weight can be derived and shown in Table 2.

Then, the single-index unascertained measure is obtained:

4.2.4 Principle of identification.

Because the classification of the comment ranks is orderly, e.g. c_k is “better” than c_{k+1} , the identification principle of “maximum measure” is not available. The credible identification principle is needed. Let the credible identification be λ , it is always 0.6 or 0.7. If:

$$k_0 = \min \left[\left(\sum_{l=0}^k \mu_{jl} \right) \geq \lambda, k = 0, 1, \dots, K-1 \right], \quad (13)$$

then x_i belongs to the rank c_{k_0} .

specialists are invited to give the values of indexes, which are listed in Table 1.

TABLE 2 Weight value of each assessment index

Assessment Index	Weight Value	Assessment Index	Weight Value
Readability	0.16	Meaning function	0.13
Formality	0.15	Covertly erroneous error	0.16
Convey degree of related implications	0.12	Overtly erroneous error	0.15
Convey degree of the implicit implication	0.13		

The assessment result can be obtained:

$$\mu^i = (0.20 \ 0.21 \ 0.19 \ 0.20 \ 0.20).$$

Let $\lambda = 0.7$, the final assessment results can be obtained: the translation belongs to the third rank, which means "normal".

Using the fuzzy comprehensive assessment [14], it belongs to the second rank, "better". Practice demonstrates that the result obtained by using the unascertained measure is more rational. The reasons are the unascertained measure pays more attention to the order of the assessment space and gives the rational rank and credible identification principles. All of those are not possessed by fuzzy comprehensive assessment.

6 Conclusion

The quality assessment of the computational translation can eliminate the possibility of failure and it is the key to ensure the quality of translation. In order to overcome the defects of subjectivity of common methods and evaluate effectively computational translation quality, here the unascertained measure model was established and employed in practice application. The application results show that it can easily realize the assessment without any assumption. And this study has great significance in improvement of machine translation and other fields.

Acknowledgments

The work was supported by the Program of Selection and Cultivating of Disciplinary Talents of Colleges and Universities in Hebei Province (BR2-206).

References

- [1] Hutchins W J, Somers H L 1992 *An Introduction to Machine Translation* Academic Press: London
- [2] Yuyang M, Shuqi S, Junguo Z, Sheng L, Tiejun Z, Xiaoning Z 2011 *Journal of Computer Science and Technology* 26(1) 57-67
- [3] King M, Popescu-Belis A, Hovy E 2003 *Proceedings of MT Summit IX, New Orleans, LA. Sept. 2003* 224-31
- [4] Church K, Hovy E 1993 *Machine Translation* 8 239-258
- [5] Lavie A, Sagae K, Jayaraman S 2004 *Proceedings of AMTA 2004* Washington DC September 2004
- [6] Papineni K, Roukos S, Ward T, Zhu W 2002 *ACL-2002: the 40th Annual meeting of the Association for Computational Linguistics* 311-8
- [7] Turian J, Shen L, Melamed I D 2003 *Proceedings of the MT Summit IX* New Orleans USA 386-93
- [8] Han A L F, Wong D F, Chao L S 2012 *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)* Mumbai India 441-50
- [9] Han A L F, Wong D F, Chao L S, Lu Y, He L, Wang Y, Zhou J 2013 *Proceedings of the Eighth Workshop on Statistical Machine Translation ACL-WMT13* Sofia Bulgaria 414-21
- [10] Wang G 1990 *Journal of Civil Engineering Institute of Haerbin* 23(4) 1-9
- [11] Wu H and Wu H 2000 *Journal of Quantitative & Technical Economics* 2 21-4
- [12] Liu K, Cao Q, Pang Y 2004 *ACTA Automatic Sinica* 30(5) 747-56
- [13] Qiu W 2002 *Management Decision Making and Applied Entropy* China Machine Press: Beijing
- [14] Chen S 1998 *Theory of Project Fuzzy Sets and Its Application* National Defence Industry Press: Beijing

Authors



Juanjuan Suo, born in December, 1977, Hebei Province, P.R. China

Current position, grades: Teacher of Hebei University of Engineering.

University studies: M.D. at Hebei University in 2006.

Scientific interest: computational linguistics.

Publications: more than 10.



Huimin Dong, born in September, 1991, Shandong Province, P.R. China

Current position, grades: Postgraduate at Hebei University of Engineering.

University studies: B.E. at Hebei University in 2012.

Scientific interest: computational linguistics.

An efficient and flexible modelling approach for multi-DSP system

Zheng-Mao Zhou, Shun-Hong Zhong, Ming Cai*

School of Computer Science of Technology, Zhejiang University, Hangzhou of Zhejiang, China

Received 1 July 2014, www.tsi.lv

Abstract

With the development of the information technology, single digital signal processor (DSP) cannot meet the requirements of massive data processing. Multi-DSP parallel processing mode has been commonly used in real-time information processing system. New technology is also making it much easier to integrate multiple DSPs into a single silicon chip. However, designers of a new multi-DSP system and software are confronted with problems such as short product life-time. Meanwhile, product verification is indispensable before launching into the market. In this paper, a multi-DSP simulation platform is developed to solve these problems. The designed multi-DSP platform is based on an ISS-SystemC structure and has three common interconnect interfaces. An AMBA bus-shared memory model is designed for the expansion of the simulation system. A thread-agent method is proposed to optimize the performance of SystemC thread and the experiment results show that the multi-DSP parallel processing mode can improve the processing performance of the system significantly.

Keywords: simulation platform, multi-DSP, ISS-SystemC, SystemC optimization

1 Introduction

With the rapid development of information technology, Digital Signal Processor (DSP), with its unique structure and fast data processing capabilities, has been widely used in mobile communication, radar signal processing, real-time image processing and other fields. However, with the increasing amount of data processing, a single DSP system cannot meet the requirements of large-scale computation. Multi-DSP parallel processing system with characteristics of real-time, high accuracy and large data throughput has already been applied to complex large-scale data processing systems [1]. As designers of new multi-DSP parallel system and software are increasingly faced with short product life-time. The resulting time-to-market constraints are contradicting the continually growing system complexity. Nevertheless an extensive design-space exploration and product verification is indispensable for a successful market launch [2]. In this case, simulation tools are essential both for designers and researchers in computer architecture, due to their ability of studying and validating new designs without the cost of actually building the hardware.

For this purpose, a Hardware/Software co-simulation is very useful for the validation of both hardware and software components in multi-DSP parallel system. Co-simulation can also evaluate the performance of the whole system at an earlier stage before building a prototype. However, there are usually gaps between software components and hardware components in traditional mixed co-simulators. It is not until the emergence of SystemC that the hardware and software are bumped

together, which makes the co-simulations of software and hardware seamlessly. SystemC is one of the most popular system-level modelling languages as it provides a common language for both the hardware and software designers [3]. The single simulation engine (SystemC) ensures the design of the co-simulation to be easier and more efficient.

As is shown in Figure 1, a novel ISS-SystemC framework is proposed for designing multi-DSP systems. An open source ISS (C6Xsim) [4] of DSP is used to abstract the model of the real programmable device where the software should run and SystemC is used for transparent integration of ISSs with other peripherals. The ISS is designed as a c++ class with an ISS-wrapper interface and each DSP core is an instance of this class. As the interconnect between DSPs is very flexible, we designed several peripheral interfaces (EMIF, HPI, McBSP) for the expansion of the platform with SystemC. The ISS-wrapper is used as an intermediate transformation layer for mutual transformation between the read (write) requests from ISS and SystemC transactions, which can be deal with interface module. The SystemC Module in Figure 1 consists of communication mediums (Shared Memory, FIFO), AMBA Bus and Decoding Controller, etc. Each DSP core is connected to this component through a peripheral interface for communicating with each other.

The main contributions of this paper consist of the following two aspects:

- 1) We provide a flexible and scalable multi-DSP model for C62x- series processor at a cycle-accurate level of abstractions, which could meet the needs of debug, functional verification of the multi-DSP system, thus helping shorten the system development cycle, improve

* *Corresponding author* e-mail: zhouprogram@zju.edu.cn

product quality and reliability, and reduce development costs.

2) A serial scheduling mechanism is used in the SystemC-kernel to reduce the design complexity of the system [5], which allows only one `sc_thread` running at the same time in system. We proposed an OS-thread agent approach to avoid the shortcomings of `sc_thread` serial execution. This method makes all the DSP cores to run in parallel when there is no communication between any two DSP cores, which can accelerate the simulation speed greatly, especially in the modern multi-core host machine.

The paper is organized as follows: Section 2 introduces the related work and Section 3 introduces the implementation of multi-DSP simulation platform. Section 4 introduces the system expansion and performance optimization, followed by experiment verification. Section 6 gives the conclusion of the paper.

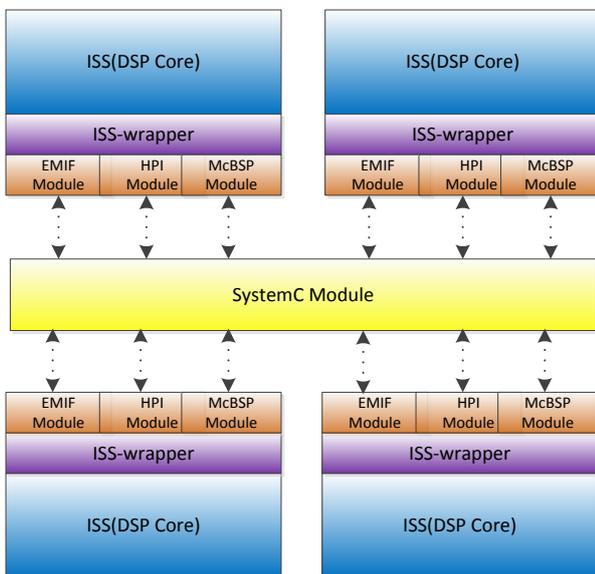


FIGURE 1 ISS-SystemC Framework.

2 Related work

In this section, we begin with a brief introduction of the open source IIS of DSP (C6Xsim). Then we discuss the IIS-SystemC framework. Finally, the flexible interconnect of c62x- series DSP is presented.

2.1 C6XSIM

C6Xsim [4], designed by Vinodh Cuppu in the University of Maryland, is an open source VLIW processor simulation tool. It offers a complete cycle accurate, execution driven simulation environment of Texas Instruments TMS320C62x series of very long instruction words DSP processors. This ISS accurately simulates various stages of the pipeline and gathers statistics to a considerable degree of accuracy of execution. So it can be used for microarchitecture development, performance analysis and application analysis on DSP processors. In

our system, we have added additional functionality to the original C6Xsim to enable the interconnection between the DSPs.

2.2 ISS-SYSTEMC FRAMEWORK

The IIS-SystemC is a popular HW/SW co-simulation framework, which has been used by many academic institutes. [6] is based on qemu-SystemC structure and the experiment results show co-simulation at the cycle-accurate (CA) level is much faster than the conventional ones. [7] developed a complete multi-ARM simulation system based on SWARM-SystemC framework [8] and they found that the effectiveness of a particular system configuration strongly depends on the application domain and generated traffic profiles. [9] proposed a transaction level modelling (TLM) approach for designing an OpenRISC-SystemC co-simulation framework and [10] designed multiple interconnected processors with distributed memory in the SimpleScalar-SystemC framework. [11] proposed an ISS-SystemC co-simulation framework in which HW models can be modified on the fly while keeping the SW parts unchanged. This means that a new ISS can be added to the system with no complex changes.

2.3 INTERCONNECT OF C62X-SERIES DSP

As we know that the cascaded modes of multiple DSPs and high-speed data transfer between them are significant in multi-DSP parallel processing systems. Therefore, the high-speed data or special interfaces are generally used as the cascaded interfaces among multi-DSP systems to meet the requirement of data transfer rate. The Texas Instruments provides three high-speed interfaces for the interconnect of C62X-series DSPs [12]. Table 1 shows the characteristics of these interfaces. The external memory interface (EMIF) is the interface between the external storage and C6X DSP. This interface is a generic mass data transmission channel and the transmission speed can reach 16 Gbit/s in general. The host port interface (HPI) is an interface between master and DSP. The master cannot only access all storage space of the DSP directly, but also the chip memory mapped peripherals. This interface is mainly used to control and configure the slave DSP. The multi-channel buffered serial port (McBSP) is usually used to connect the serial peripheral and the transmission speed is only 0.125Gbit/s.

TABLE 1 Interface of C62x-series DSP

Name	Speed (Gbit/s)	Number of Signal Lines	Typical Interconnect Structure
EMIF	16	36	shared memory
HPI	0.8	40	master-slaver
McBSP	0.125	6	peer-peer

Figure 2 shows the typical interconnect of C62xx-series DSPs with the referred interfaces in our simulation platform. Figure 2a shows the DSP interconnect based on

a shared memory. This interconnect structure is usually used for large amount of data exchanges between DSPs, which could be handled according to the high-speed transmission characteristics of EMIF. Figure 2-b shows a typical master-slave interconnect structure based on HPI and it requires no additional storage medium. As the centre

of the topology, the master-DSP is usually used for the control centre of the whole system and each slave-DSP receives data (control commands and configuration data) from the master-DSP. Figure 2-c shows the peer-to-peer topology structure, which is mainly used for small amount of data communication between the two DSPs.

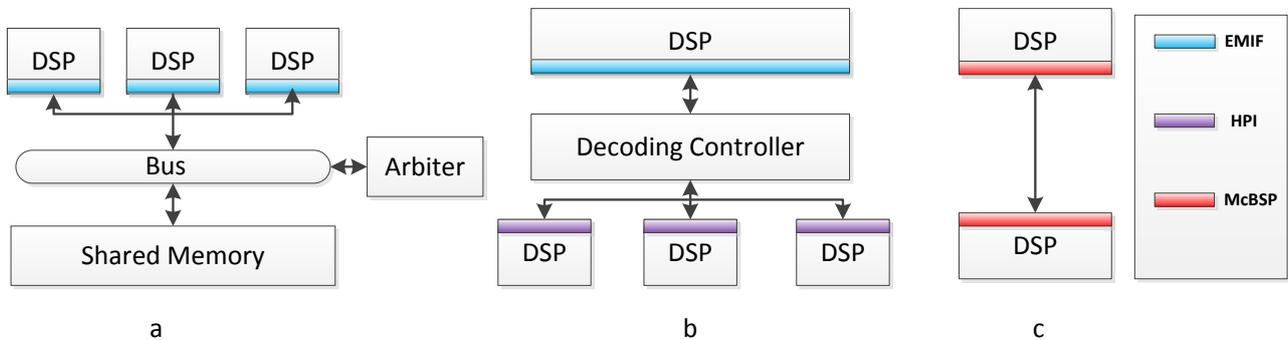


FIGURE 2 Topology of DSP interconnect

3 Multi-DSP simulation platform

Integrating multiple Instruction Set Simulations (ISSs) of DSP into a unified system simulation framework has several non-trivial challenges. In this section, we will introduce each of the major modules of the simulation system. First we add several components to the original open source cycle accurate ISS in order to meet the basic I/O requirements. As SystemC is selected as system model language, a SystemC wrapper is necessary for the conversion between read (write) requests and SystemC transactions. Then three special interfaces are provided for the interconnect of DSPs in accordance with current popular DSP interconnect structure. Finally, an AMBA Bus-Shared Memory Model used in our simulation system is introduced.

3.1 PROCESSING MODULE

As shown in Figure 3, we add three components (Interrupt Controller, Timer Manger and I/O Manager) to the original ISS in order to enable interconnection between the DSPs, which also ensures a full-fledged real-time operating system (RTOS) to run on it. The Interrupt Control component is used for the management of external interrupts, which are mainly from the I/O modules. The clock interrupt is triggered by the Timer, which provides support for a RTOS running on the simulator. The I/O Manager module is used for mapping the I/O address and providing a uniform I/O management interface.

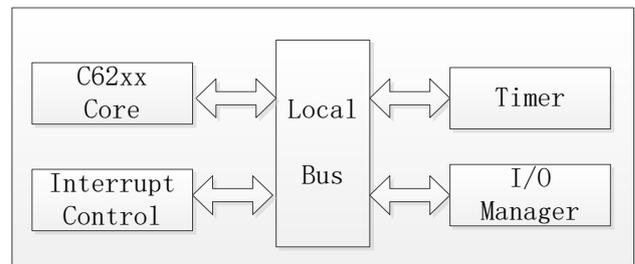


FIGURE 3 Processing module architecture

3.2 SYSTEMC WRAPPER

There are significant differences between the ISS and the interconnect structure, as the ISS is developed with the c program language while the interconnect structure consists of SystemC modules. An approach to make up the difference is to completely embed the ISS within the SystemC module. In other words, we design a class not only with the instruction set simulation function and but also with the function of transforming normal c function call request into the SystemC transaction level request.

As shown in Figure 4, we design two classes, namely the SystemC module *DSP_Wrapper* and the processing module *DSP_Core*. The *DSP_Wrapper* contains an instance of *DSP_Core* and launches the ISS simulation. The two member functions (*tran_acc_in* and *tran_acc_out*) are used for the synchronization with the environment (Bus, peripherals and other DSP Core). The member function, *specail_mem_access* in *DSP_Core* shows that access to specific memory (I/O ports or I/O control registers, etc.) can be detected. The access is suspended not until the *DSP_Wrapper* receives the corresponding response signal and then sets the member variable *s_mem_end* as true.

```

FILE DSP_Wrapper.cpp
#include<systemc.h>
#include "DSP_Core.h"
SC_MODULE(DSP_Wrapper)
{
.....
sc_in_clk clk;
sc_in .....
sc_out .....
sc_inout.....
DSP_Core dsp_core;

/* start instruction set simulation*/
void start_simulation();
/* SystemC transation input*/
void tran_acc_in();
/* SystemC transation output*/
void tran_acc_out();

.....
.....

SC_STOR(DSP_Wrapper)
{
SC_THREAD(start_simulation);
sensitive_pos<<clk;
SC_CTHREAD(tran_acc_in,clk.p
os());
SC_METHOD(tran_acc_out);
sensitive<<dsp_core.s_mem_end;
}
}
.....
.....

FILE DSP_Core.h
Class DSP_Core
{
public:
unit mem_access(.....);
/*special memory access*/
/*(such as I/O,control register)*/
unit special_mem_access(.....);
/*special memory access start*/
sc_signal<bool>s_mem_start;
/*special memory access end*/
sc_signal<bool>s_mem_end;
uint s_add_s;//special address start
uint s_add_e;//special address end
.....
}

FILE DSP_Core.cpp
#include<systemc.h>
#include "DSP_Core.h"
unit DSP_Core::mem_access(uint add)
{
if(add>=s_l_add_s&&add<=s_add_e)
{
return special_memory_access(add);
}
else
.....
}
unit DSP_Core::speacil_mem_access()
{
.....
s_mem_start.write(true);
while(s_mem_end ==false)
wait();
.....
}
.....
}
.....
}

```

FIGURE 4 DSP_wrapper architecture

3.3 INTERFACE DESIGN

3.3.1 External memory interface

External memory interface (EMIF) is the only channel to access external memory in the C62x-series DSP and the transmission speed generally can reach 16 Gbit/s. The main pins and control registers are designed as shown in Tables 2 and 3. A typical access (read) to shared memory with EMIF is as follows:

- Set EM_Add as the corresponding address and EM_Read as *true*;
- Set the EM_Hold as *true*, then wait until EM_HoldA

as *true*;

- Read the EM_Data value after one clock;
- Set the EM_Hold as *false* to release bus when the access finishes.

TABLE 2 Pins of external memory interface

Name	Data Type	Introduction
EM_Enable	sc_in<uint>	enable EMIF
EM_Add	sc_out<uint>	address pins
EM_Data	sc_inout<uint>	data pins
EM_Read	sc_in<bool>	indicate read
EM_Write	sc_in<bool>	indicate write
EM_Ready	sc_inout<bool>	indicate ready
EM_Hold	sc_out<bool>	bus request
EM_HoldA	sc_in<bool>	bus response
EM_Int	sc_out<bool>	interrupt
EM_Clk	sc_inout_clk	clock

TABLE 3 Control register of external memory interface

Name	Data Type	Introduction
GlbCtl	uint	global control register
SdCtl	uint	sdram control register
SdExt	uint	sdram external register

3.3.2 Host port interface

The host port interface (HPI) is a special parallel interface, existing in most of the TI DSP chip. The master-slave interconnect structure with HPI can significantly reduce the complexity of the system with no need for extra chips. The master-slave multi-DSP system is very popular in flight control system in which the reliability is the primary consideration [13]. The main pins and control registers are designed as shown in Table 4 and Table 5 respectively. A typical access to slave memory process is as follows:

- Set HPI_Cntl1 and HPI_Cntl2 as false, false;
- Write special value to the HPIC register to prepare for the access slave;
- Set HPI_Cntl1 and HPI_Cntl2 as false, true and write address to HPIA;
- Set HPI_Cntl1 and HPI_Cntl2 as true, false (or true, true) and then read data from HPID.
- In the previous step, if the values of HPI_Cntl1 and HPI_Cntl2 are set as true and false respectively, the value of HPIA will automatically increase one.

TABLE 4 Pins of host port interface

Name	Data Type	Introduction
HPI_Data	sc_inout<ushort>	data pins
HPI_Cntl1	sc_in<bool>	control pin1
HPI_Cntl2	sc_in<bool>	control pin2
HPI_Hwil	sc_in<bool>	indicate transform
HPI_Ready	sc_out<bool>	indicate ready
HPI_Int	sc_out<bool>	interrupt
HPI_Clk	sc_inout_clk	clock

TABLE 5 Control registers of host port interface

Name	Data Type	Introduction
HPIA	uint	HPI address register
HPIC	uint	HPI control register
HPID	uint	HPI data register

3.3.3 Multi-channel buffered serial port

The Multi-channel buffered serial port is one of the fundamental interfaces in the C62x series DSP. It is usually used for connecting a serial interface peripheral, such as serial AD and serial peripheral interface. It can also be used for the interconnection between the DSPs when there is only a small amount of data exchanged. The simple McBSP-to-McBSP structure can reduce the cost of system design significantly. Figure 6 and Figure 7 show the main pins and control registers of the multi-channel buffered serial port. A typical byte reception process is as follows:

- Set SPCR register as a special value to configure the receiving protocol;
- RSR gets a bit from the MS_DR pin every one clock until one byte transfer is completed;
- Check the received byte according the check code. If the byte is correct, then set the value of DRR as the received byte and set MS_CLR as true to generate an interrupt.

TABLE 6 Pins of multi-channel buffered serial port

Name	Data Type	Introduction
MS_DX	sc_out<bool >	send pin
MS_DR	sc_inout<bool>	receive pin
MS_CLX	sc_inout_clk	send clock
MS_CLR	sc_inout_clk	receive clock
MS_Xint	sc_out<bool>	send interrupt
MS_Rint	sc_out<bool>	receive interrupt

TABLE 7 Control register of multi-channel buffered serial port

Name	Data Type	Introduction
DRR	uchar	receive register
DXR	uchar	send register
RSR	ushort	receive shift register
XSR	ushort	send shift register
SPCR	uint	control register

3.4 AMBA BUS-SHARED MEMORY MODEL

The bus-shared memory model is currently a very popular method of multi-machine interconnect [14]. The masters communicate with each other through the shared memory. In our simulation system, the ISS (DSP core) is the master of the bus and a shared memory is the slave of the bus. In the following parts, the bus model and the shared memory will be introduced in details respectively.

3.4.1 AMBA Bus

The AMBA Bus is applied in our simulation platform as the AMBA is a widely used standard in system on a chip (SoC) [7]. The AMBA Bus contains two standards: an advanced high-performance system bus (AHB) and a peripheral bus (APB) for minimal power consumption and connection with low-performance peripherals. We have developed a SystemC module only for the former one, given the situation of a large number of data exchange between the DSPs. As each ISS (DSP Core) in the simulation system is in peer relationships. Each bus

request from the master is given the same priority. A traditional arbitration strategy (round-robin policy) is implemented in our AMBA Bus model to realize load balancing.

A Bus transaction is triggered by a bus request signal when one master (DSP) wants to access the shared memory. The arbiter receives the request and then determines whether to authorize the request or put the request into the waiting queue according to the current state of the bus. Then the master waits until the bus ownership is granted by the arbiter. At the same time, the address and control lines are driven and the data bus ownership is also granted after one clock cycle. Last, the data transformation starts when a ready signal is asserted by the slave (shared memory), indicating all the preparations have been completed and the single data transformation can be completed after the next rising edge of the clock. Besides this single transfer, specified-length bursts and unspecified-length bursts are also supported in our designed the AHB protocol.

3.4.2 Shared memory

The Shared Memory is connected to the AMBA Bus as a slave. It consists of multiple instantiations of a basic SystemC memory module and each module space is 1MB. It communicates with the masters through the AMBA bus with a typical request-ready asynchronous protocol [7]. The memory read process is as follows: the address lines are assigned and then the memory module checks whether the address is within the scope of the current address space or not. If the address is effective, then a ready signal is asserted by the memory module. After one clock cycle, the value of data address lines is assigned as the content of the corresponding address in memory by the memory module.

4 System expansion and performance optimization

4.1 SYSTEM EXPANSION

As shown in Figure 2, the interconnect structure between the DSPs is very flexible. In many complex scenarios, the interconnect structure is a hybrid structure which consists of two or more interconnect structures in Figure 2 [15]. Each topology involves various parameters, such as the number of DSPs, the number of shared memory, size of each shared memory, any two DSP connection mode, etc. The hybrid topology and configurable system parameters are able to meet the needs of a variety of Multi-DSP systems simulation.

4.1.1 Hybrid topology

Mixed DSP interconnect structure involves multiple external connection interfaces and their communication with the DSP Core. In section 3.3, three external connection interfaces are designed for the DSPs

interconnect. We just need to map the ports of the three interfaces into different memory address space and to ensure the three interface module can simultaneously keep pace with the DSP Core module. Then the hybrid DSP interconnect structure can be supported in our simulation platform.

4.1.2 System configuration

The system configuration describes the DSP interconnection topology. For an ordinary user, figurative descriptions can be understood easily. Therefore, a graphical configuration interface is a very good choice. Figure 4 shows a WYSIWYG DSP interconnection parameter configuration interface.

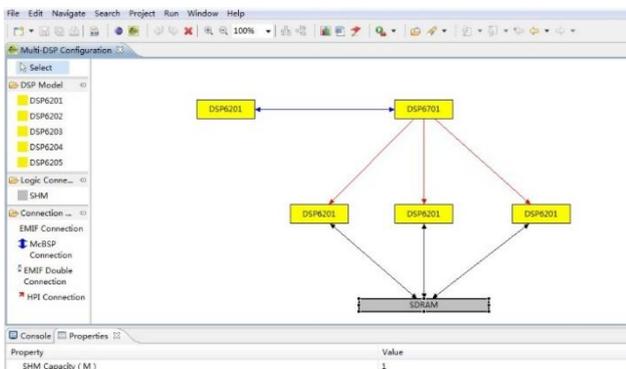


FIGURE 4 Visual configuration of simulation system

4.2 SIMULATION PERFORMANCE OPTIMIZATION

4.2.1 Performance Bottleneck Analysis

Table 5 shows the time cost of quick sort with 20000 random numbers in our simulation system. All the DSP are connected with shared memory as shown in Figure 2-a and a parallel algorithm is used in the multi-DSP system. The host machine is shown in section 5.1.

TABLE 8 Execution time of quick sort

Quick Sort	Time(ms)
single DSP	153824
two DSP	205234
four DSP	285234

As we can see that the execution time of quick sorting rapidly increases with the incremented number of DSPs. This indicates that although the host is a multi-core (eight core) system and supports multiple threads run in parallel, multi-DSP parallel processing does not increase but reduces the system performance instead. As each DSP core is designed independently in a SystemC thread of execution in our system, we suspect that these SystemC threads may not be executed in parallel and [3, 16] confirm our guess. The SystemC kernel adopts fiber mechanism and QuickThread to encapsulate SystemC threads in the Windows and Linux respectively. Therefore, the SystemC

thread is not an OS thread but a lightweight thread. The switches of these lightweight threads do not take place in the OS kernel layer. Therefore, they have very high performance; however they are executed serially as these lightweight threads are equivalent to one OS thread. So multi-DSP parallel system cannot accelerate the sorting process. The synchronization between the DSP cores and peripheral modules can also significantly affect the performance of the multi-DSP system and the synchronization overhead will increase with the number of DSPs in our system. Therefore, the overhead of sorting increases with the number of the DSP in our system, rather than decrease.

4.2.2 Performance optimization implementation

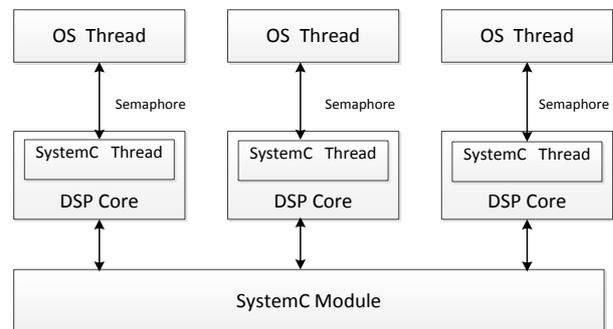


FIGURE 5 SystemC thread parallel optimization

SystemC is one of the most popular system-level modelling languages as it provides a simulation framework, which facilitates design and verification of SoC at different levels. However, the single threaded simulation kernel inherent to SystemC cannot meet requirements of multiple DSP parallel running. We present a method of thread agent to ensure multiple DSPs run in parallel on current popular multi-core machine to speed up the simulation of the whole system.

Figure 5 shows the method of thread agent to avoid SystemC threads' serial execution. Each DSP core creates an OS thread to simulate the instruction execution. As described in section 3, the ISS-SystemC framework is used in our system and the synchronization between DSPs is done through SystemC event mechanism. However, the synchronization is not in every clock cycle but only when the exchange of data between DSPs or access to the peripherals occurs. We create an OS thread in the SystemC thread to simulate the instruction execution. An OS semaphore is used for the synchronization between the OS thread and SystemC thread when the exchange of data between DSPs and access to the peripherals occurs. All OS Threads can run in parallel on the multi-core machine when there is no communication between the DSPs. Therefore, the simulation speed of the entire system could be improved greatly.

5 Experiment verification

5.1 EXPERIMENT ENVIRONMENT

1) Host Machine

- CPU: FX-8350 (4.0GHz), eight cores CPU
- RAM: 8GB (DDR1600)
- Hard Disk: 1TB
- OS: Windows 7 ultimate

2) Development Environment

- Test case compiler: Code Composer Studio 3.3
- Simulation system Compiler: Microsoft Visual Studio 2010
- SystemC version: 2.2.0

5.2 TEST CASE

We make changes on some benchmarks [17] provided by TI, to ensure the benchmarks can run on the multi-DSP parallel system. The modified benchmarks contain Fast Fourier Transformation (FFT), Matrix Multiplication and Quick Sort.

5.2.1 Fast Fourier transformation

Fourier transform, a basic digital signal processing operations, is widely used for presentation and analysis of discrete time-domain signal. Define a discrete finite time sequence $x(n), 0 \leq n \leq N$, and the discrete Fourier transform is:

$$X(k) = \sum_0^{N-1} x(n)w_N^{nk}, k = 0, 1, \dots, N-1, w_N = e^{-j\frac{2\pi}{N}}$$

The algorithm complexity of commonly used Discrete Fourier Transformation (DFT) is $O(n^2)$. In our system, Fast Fourier Transformation (FFT) is used in the multi-DSP simulation system. The DFT sequence is divided into shorter DFT sequences and each shorter DFT sequence is dealt in one DSP. This allows parallel processing of Fourier transform and the algorithm complexity reduces to $O(n \log_m^n)$.

5.2.2 Matrix multiplication

Matrix multiplication is one of the DSP regular processing operations. Define $C = A \times B$, A and B are N -dimensional matrix. Then $C_{i,j} = \sum_{k=1}^M A_{ik} B_{kj} (1 \leq i \leq M, 1 \leq j \leq M)$ and the algorithm complexity is $O(n^3)$. In our system, we have adopted Cannon algorithm of matrix grid division. Assumptions with m^2 DSPs in parallel system, each DSP is assigned to $\frac{n}{m}$ of the data for dealing. The processing

speed of the whole system can be improved obviously and the algorithm complexity is about $o(n^2)/m$.

5.2.3 Quick sort

Quick sort, as a kind of efficient sorting method, is often used to deal with large-scale data sorting. Define that the amount of data to be sorted is n and the number of DSP is m . The algorithm complexity of the sorting data for each DSP is $\frac{n}{m} \log_2^m$ and the algorithm complexity of all the data is $\frac{n}{m} \log_2^m + n \log_2^m$.

5.3 EXPERIMENT RESULTS AND ANALYSIS

Figure 6-8 show relative execution time of three test cases (FFT, Matrix Multiplication and Quick Sort). Each test case runs on simulation systems with three kinds of topology and Table 9 shows the characteristics of these three topologies. SHM and U-SHM indicate the optimized system and the un-optimized system respectively. Finally we discuss the communication overhead and parallel acceleration according to the experiments results.

TABLE 9 Three kinds of topologies

Interconnection Method	Characteristics
SHM	The entire DSPs are connected with one Shared Memory as shown in Figure 2-a
HPI	The entire slave DSPs are connected to Host DSP with HPI as shown in Figure 2-b
Mixed	All the DSPs are connected with mixed interconnection methods as shown in Figure 4

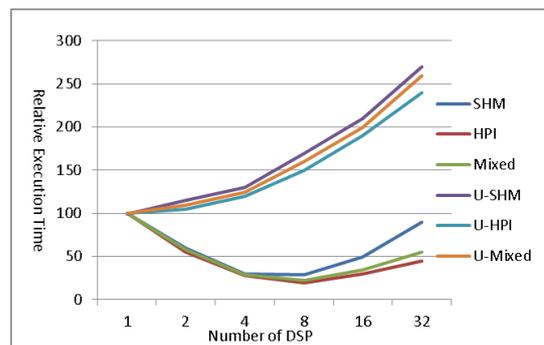


FIGURE 6 Relative execution time of FFT

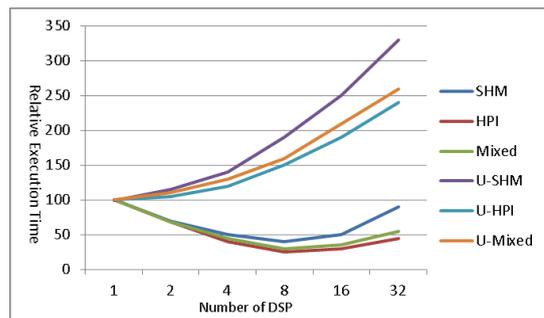


FIGURE 7 Relative execution time of matrix multiplication

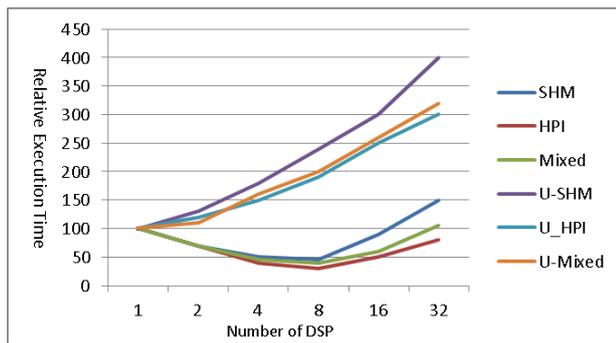


FIGURE 8 Relative execution time of quick sort

5.3.1 Communication overhead analysis

As we can see from section 4.2, the synchronization overhead will increase with the increase of the number of DSP in our system. The results of the three test cases in no performance optimization condition exactly reflect this characteristic. And the communication overhead varies with different interconnect structures. Figure 6-8 show the relative execution time of simulation system with shared memory structure is longer than other simulation systems when the number of DSP reaches four. And the trend is more obvious when the number of DSP exceeds four. This is because the multi-DSP system shared memory structure is restricted to access the only one shared memory. With the increasing of the number of DSP, the bus request time will be longer. The scales of communication are almost the same when the number of DSP is the same in our multi-DSP systems. Therefore, the communication overhead of the system with shared memory structure is greater than other systems.

5.3.2 Multiple DSP parallel acceleration

As we can see from Figure 6-8, the relative execution time of simulation system which performance optimization is used in significantly gets shorter as the number of DSP increases from one to eight. This indicates that multi-DSP simulation system can speed up the system of processing data when the number of DSP is less than the cores of machine which simulation system runs on. This is because each DSP binds to one SystemC thread and the parallel optimization ensures all the SystemC threads take full advantage of threads provided by the operating system to run in parallel. The amount of data processing is fixed, the number of DSP increasing results in an increase in the overhead of communication between DSPs. So the relative execution time does not reduce exponentially with

doubling the number of DSP. When the number of DSP increases to eight or more, the relative execution time does not shorten but extend with the increase of the number of DSP. This is because the host machine has only eight cores, which leads to a maximum of eight OS threads running in parallel on the operating system. Although the number of DSP exceeds eight, there are only eight DSPs parallel running on the operating system at the same time. And additional communication will lead to longer processing time.

The above analysis shows communication overhead varies with different topologies and our multi-DSP simulation system can obviously accelerate the processing of large amounts of data. Creating OS thread in the SystemC thread to complete the time-consuming processing can significantly improve the efficiency of serial running of SystemC threads.

6 Conclusion and future work

We provide a flexible and scalable multi-DSP model for C62x-series DSP at cycle-accurate level of abstractions. An ISS-SystemC framework which provides a hardware and software collaborative development environment is used in our system. Three kinds of interconnection interfaces are designed for a flexible interconnection structure. A kind of optimization for SystemC threads is also used to accelerate the speed of the system simulation and the experiment results verify that the simulation platform is very efficient. Future extensions of our work include two aspects:

- 1) The ISS can be designed to be more perfect and close to the real DSP. For example, direct memory access (DMA) and cache modules need to be added.
- 2) Each SystemC thread creates an OS thread to simulate each DSP's instruction execution, which ensures that multiple DSPs run in parallel on multicore machines in situation of no communication between multiple DSPs. However, the simulation of each DSP's communication is implemented by one SystemC thread, which results serial simulation of all the DSPs' communications. Providing parallelizing SystemC kernel will be an efficient choice to improve the performance of the simulation platform.

Acknowledgments

This work has been supported by China National Natural Science Foundation (No.51175462) and National Defense Research Projects.

References

- [1] Huang F, Qiao C, Wang Y, Wang G 2007 *Computer Engineering* 33(23) 200-33 (in Chinese)
- [2] Braun G, Nohl A, Hoffmann A, Schliebusch O, Leupers R, Meyr H 2004 *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 23(12) 1625-39
- [3] Pasricha S 2002 Transaction level modeling of SoC with SystemC 2.0 *Synopsys Users Group Conference SNUG 2002 India 2002*
- [4] Cuppu V 1999 Cycle Accurate Simulator for TMS320C62x, 8 way VLIW DSP Processor

- <http://www.cs.cmu.edu/afs/cs/academic/class/15745-s07/www/c6xref/c6xsim.pdf>
- [5] Buchmann R, Greiner A 2007 A Fully Static Scheduling Approach for Fast Cycle Accurate SystemC Simulation of MPSoCs *Proc. Int. Conf. Microelectron (ICM 2007)* Cairo Egypt 101-4
- [6] Chao M C, Yeh T C, Tseng G F 2011 *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 30(4) 593-606
- [7] Benini L, Bertozzi D, Bogliolo A, Menichelli F, Olivieri M *Journal of VLSI Signal Process* 41(2) 169-82
- [8] Dales M SWARM-Software arm: <http://www.cl.cam.ac.uk/~mwd24/phd/swarm.html>.
- [9] Boukhechem S, Bourennane E B 2008 TLM Platform Based On SystemC For STARSoC Design Space Exploration *Proc. NASA/ESA Conf AHS 22-5*
- [10] Boyer F, Yang L, Aboulhamid E, Charest L, Nicolescu G 2003 Multiple Simplescalar Processors with Introspection under SystemC *Proceedings of the 46th IEEE International Midwest Symposium on Circuits and Systems* 1400-4
- [11] Fummi F, Perbellini G, Loghi M, Poncino M 2006 ISS-Centric Modular HW/SW Co-Simulation *GLSVLSI'06: Proceedings of the 16th ACM Great Lakes symposium on VLSI* 31-6
- [12] TMS320C6000 Instruction Set Simulator Technical Reference Manual: <http://www.ti.com/lit/ug/spru600i/spru600i.pdf>
- [13] Bao C, Wang L H, Zhang L, Zhang S N 2010 *Science Technology and Engineering* 10(33) 8287-92 (in Chinese)
- [14] Zou Z Q 2006 *Computer Engineering* 32(16) 232-5 (in Chinese)
- [15] Xu T F, Zhang B, Ni G Q 2005 *Transactions of Beijing Institute of Technology* 25(11) 990-2 (in Chinese)
- [16] Ezudheen P, Chandran P, Chandra J, Simon P B, Ravi D 2009 Parallelizing SystemC Kernel for Fast Hardware Simulation on SMP Machines *PADS' 09: Proceedings of the 2009 ACM/IEEE/SCS 23rd Workshop on Principles of Advanced and Distributed Simulation* 80-7 Washington DC USA
- [17] Texas Instruments benchmarks: http://www.ti.com/lstds/ti/dsp/c6000_dsp/c674x/benchmarks.page

Authors



Zheng-Mao Zhou, born in 1988, Hubei, China

Current position, grades: doctoral candidate of Zhejiang University.

University studies: B.S. degree in Computer Science from Huazhong University Science and Technology in 2010.

Scientific interest: digital signal processing, real-time operating system, software reliability.

Publications: 2.



Shun-Hong Zhong, born in 1986, Zhejiang, China

University studies: B.S. degree, M.S. degrees both in Computer Science from Zhejiang University, China, in 2009, 2012, respectively.

Scientific interest: digital signal processing.



Ming Cai, born in 1974, Zhejiang, China

Current position, grades: Associate professor at Zhejiang University.

University studies: M.S. and Ph.D. degrees both in Computer Science at Zhejiang University, China, in 1999 and 2002, respectively.

Scientific interest: digital signal processing, real-time operating system, web service, manufacturing resource discovery.

Publications: 15.

A study on application of judgment matrix intelligent correction method in satisfaction evaluation

Chenxia Suo¹, Yong Yang^{2*}

¹Beijing Institute of Petrochemical Technology, 102617, Beijing, China

²College of Statistics and Mathematics, Zhejiang Gongshang University, 310018, Hangzhou, China

Received 1 August 2014, www.tsi.lv

Abstract

In order to improve rural folk house renovation in the satisfaction evaluation accurately, this paper puts forward a model for rural folk house renovation in the satisfaction evaluation based on intelligent expert judgement matrix adjustment method (AGA-LCAHP). By means of extracting the offset degree resulting in the inconsistency of AHP judgment matrix, this paper puts forward the new method of using accelerating genetic algorithm to locate the element of judgment matrix and calculate the AHP element ranking weight. This algorithm takes offset information as the foundation of correcting judgment matrix to avoid the subjectivity of correction; at the same time, it reserves and extracts the consistency information of judgment matrix, with the consistency index as the orientation of optimization. The case study result shows that the AGA-LCAHP method features high computational accuracy and stable calculation result and also has the popularization and application value in other comprehensive assessment.

Keywords: analytic hierarchy process, judgment matrix, consistency, offset degree, genetic algorithm

1 Introduction

T.L.Saaty et al put forward analytic hierarchy process (AHP) in the 1970's. This method mathematizes the thinking process, quantifies the subjective judgment, quantizes the difference of comparison object, and makes the complex system hierarchical, thus it is a method to make judgment subjective thinking clearer. The key to solve the problem with analytic hierarchy process (AHP) is how to build the judgment matrix. Because there are some unavoidable errors in human judgment, especially in complex system, the given judgment matrix of expert tend to be inconsistent, thus adjusting the given judgment matrix is the usual practice. This turns the test of judgment matrix consistency and how to correct the inconsistent judgment matrix into the key problem of analytic hierarchy process. At present the methods for judgment matrix inconsistency mainly include empirical estimation method, optimal transfer matrix method, vector included angle cosine method, pattern recognition method, induced matrix method. All these are the correction to subjective experience or partial element, which cannot achieve the best correction effect. On the basis of existing accelerating genetic algorithm and correction method of judgment matrix, this paper establishes analytic hierarchy process for locating the correction judgment matrix consistency based on offset information. Genetic algorithm is an algorithm of using coding technology and genetic manipulation to simulate the optimizing search. Compared with non-intelligent optimization method, this algorithm only requires that the problem translated into code can be calculated, and not require whether the solution of problem

is limited by linear, continuous, differentiable, noiseless, etc. Therefore, it is widely used in solving multi-dimensional and nonlinear complex optimization problems. This paper regards the correction of inconsistent matrix as a nonlinear optimization problem, which extracts the information of offset degree by using the inconsistency of judgment matrix, conducts location optimization through self-adaption and global optimization function of accelerating genetic algorithm [8], achieves the correction of matrix consistency, and gives the weight of each element of AHP.

2 Accelerating genetic algorithm for correcting judgment matrix consistency in AHP based on offset information.

The LAGA-CAHP calculation procedure is as following:

Step 1: decompose the system to be evaluated into hierarchical model. According to universality, hierarchical structure is divided into three levels from top to bottom, i.e. *A* – objective level, *B* – criterion level and *C* – scheme level. Level *A* is the general objective of evaluation system consisted of one element. Level *B* consists of *m* criterions for achieving the general evaluation objective, and these criterions weigh the degree that each scheme meets the general objective. Level *C* consists of *n* specific schemes for achieving the general objective. These objectives, criterions and schemes constitute a basic AHP hierarchical model.

Step 2: build the judgment matrix of each level. This includes the judgment matrixes of *B* criterion level and *C* scheme level. Each level of judgment matrix is built with

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

the element of last level as the criterion. The judgment matrix of criterion level is built with the general objective as the criterion; if it is necessary to compare the influence of n criteria B_1, B_2, \dots, B_n , on the general objective A , generally adopt pairwise comparison method to generate a pair comparative matrix. Suppose that a_{ij} is the ratio between the influence of Criterion B_i and Criterion B_j on the general subject A , the matrix $A = (a_{ij})_{n \times n}$ consisted of a_{ij} is called as judgment matrix.

Step 3: test and correct the consistency of each judgment matrix, and calculate the ranking weight. With the weight calculation of Level B as example, suppose that single ranking weight of each element of Level B is w_k ,

$k = 1, 2, \dots, n$, and $w_k > 0$ and $\sum_{k=1}^n w_k = 1$. Based on the definition of judgment matrix, theoretically there should be

$$b_{ij} = \frac{w_i}{w_j}, (i, j = 1, 2, \dots, n). \tag{1}$$

In practical application, determine the single ranking weight $\{w_k | k = 1, 2, \dots, n\}$ of each element through practical judgment matrix $B = \{b_{ij}\}_{n \times n}$. If judgment matrix B meets Equation (1), namely that the judgment matrix is consistent, there is a relation as follows:

$$\sum_{i=1}^n \sum_{j=1}^n |b_{ij} w_j - w_i| = 0. \tag{2}$$

However, because there are some unavoidable errors in human judgment, especially in complex system, the given judgment matrix of expert tend to be inconsistent, namely that the decision maker cannot give an exact comparative result of w_i / w_j . In practical application, most judgment matrixes are inconsistent, thus it is necessary to correct the judgment matrix, till the satisfactory consistency required by AHP is met.

When the judgment matrix is inconsistent, it is necessary to correct original matrix. Based on two hypotheses that most of given judgments of expert are correct and the cognitive ability and judgment basis of experts are roughly the same, this paper puts forward a method of extracting the inconsistency information of judgment matrix based on the logical relation between all information of matrix, and using the inconsistency information to correct.

Suppose that the judgment matrix is $B = (b_{ij})_{n \times n}$, the inconsistency information is extracted according to following steps:

- 1) First conduct consistency test to expert judgment matrix. If the requirement of consistency is met, stop here; otherwise, please go to Step 2.
- 2) The indirect judgment information of relative importance of comparing scheme i and j in expert

judgment matrix is b^k_{ij} . $b^k_{ij} = b_{ik} \times b_{kj}$, of which, $k = 1, 2, \dots, n$ and $k \neq i, j$.

For each element of matrix, there are $(n-2)$ indirect logical judgment information (b^k_{ij}) , from which we can deduce the relative importance of scheme i and j .

3) The offset degree of each element in judgment matrix is defined as β_{ij} :

$$\beta_{ij} = \sum_{\substack{k=1 \\ k \neq i, j}}^n \frac{(b_{ij} - b^k_{ij})^2}{(b_{ij})^2}. \tag{3}$$

4) The overall offset degree of each scheme is defined as β_i . The offset degree of overall scheme is the mean value of all elements included in this scheme in the matrix.

$$\beta_i = \left(\sum_{j=1}^n \beta_{ij} + \sum_{\substack{j=1 \\ j \neq i}}^n \beta_{ij} \right) / (2n+1). \tag{4}$$

5) The overall offset degree of expert judgment matrix is defined as β . The overall offset degree of judgment matrix is the mean value of offset degree of all elements.

$$\beta = \left(\sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \right) / n^2. \tag{5}$$

Through above analysis, this paper breaks up the inconsistency information of judgment matrix into element offset information, scheme offset information and overall offset information.

Suppose that the correction judgment matrix of original judgment matrix $B = \{b_{ij}\}_{n \times n}$ is $X = \{x_{ij}\}_{n \times n}$, of which X is the value of single ranking weight of each element $\{w_k | k = 1, 2, \dots, n\}$, making the minimum matrix X in following formula be the optimal consistency judgment matrix of matrix B .

$$\min CIC(n) = \sum_{i=1}^n \sum_{j=1}^n |x_{ij} - b_{ij}| / n^2 + \sum_{i=1}^n \sum_{j=1}^n |x_{ij} w_j - w_i| / n^2, \tag{6}$$

$$s.t \quad x_{ii} = 1 (i = 1, 2, \dots, n),$$

$$1 / x_{ji} = x_{ij} \in [b_{ij} - db_{ij}, b_{ij} + db_{ij}] \cap [1/9, 9],$$

$$i = 1, 2, \dots, n, j = i + 1, i + 2, \dots, n$$

$$\sum_{i=1}^n w_k = 1, w_k > 0 (k = 1, 2, \dots, n),$$

where, objective function $CIC(n)$ is called as consistency index coefficient; d is non-negative parameter, which can be selected from $[0, 0.5]$ according to experience; this is a nonlinear programming problem, of which, the single ranking weight $w_k (k=1, 2, \dots, n)$ and the element of correction judgment matrix $X = \{x_{ij}\}_{n \times n}$ are optimization variables, and there are $n(n+1)/2$ independent optimization variables in all. The less the value of Equation (6), the higher the consistency of judgment matrix B . When $CIC(n)=0$, $X=B$, namely that Equations (1) and (2) are established, and judgment matrix B has a complete consistency.

This paper uses the global searching function of accelerating genetic algorithm (AGA) to optimize this nonlinear problem. The condition of ending the searching can be that there is satisfactory consistency when $CIC(n)$ value is less than a standard value. When the requirement of satisfactory consistency is not met, we can adjust parameter d and matrix B , till it is satisfactory. The solution of matrix of Level C is the same.

Because it is difficult for the genetic algorithm to adapt to the change of searching space, the computational efficiency is low, and the phenomenon of premature convergence is easy to emerge. This paper uses accelerating genetic algorithm to optimize the consistency index. Suppose that the parameter model to be optimized is:

$$\min f = \sum_{i=1}^m \|F(C, X_i) - Y_i\|^2 \tag{7}$$

$$s.t. a_j \leq c_j \leq b_j, j = 1, 2, \dots, p$$

Of which, $C = \{c_j\}$ is the p optimized parameters, that is each element in judgment matrix; $[c_j - \beta_{ij}, c_j + \beta_{ij}]$ is the initial change range determined by parameter c_j according to the degree of deviation of element, and the higher the degree of deviation, the larger the range of change; X is the N -dimension input variable of model; Y is the N -dimension output variable of model; F is the nonlinear model determined according to consistency index, i.e. $F: R^N \rightarrow R^M$; the value of $\| \cdot \|$ is norm; f is optimization criterion function.

Step 4: overall ranking level and its consistency test. Successively test the judgment matrix and calculate the weight of each level from the highest level A to the lowest level c . The overall ranking weight of Level C is

$$w_i^A = \sum_k w_k w_i^k (i = 1, 2, \dots, n) \text{ and the consistency coefficient of overall ranking is } CIC^A(m) = \sum_{i=1}^n w_k CIC^k(m).$$

It is deemed as that there is a

satisfactory consistency when it is less than a setting standard value.

Step 5: determine the ranking of each decision scheme according to the overall ranking weight $w_i^A (i = 1, 2, \dots, n)$ of each element of Level C .

3 Theoretical analysis of AGA-LCAHP

3.1 JUDGMENT OF THE SATISFACTORY CONSISTENCY OF JUDGMENT MATRIX

If the order of judgment matrix is different, there are different consistency index coefficients of AGA-LCAHP. This paper defines critical random consistency index $LCIC(n)$. This paper constructs 500 order of 3-9 matrixes through stochastic simulation, and these matrixes cannot meet the requirement of consistency, but they meet unit and reciprocity at the same time. On the basis of $LCIC(n)$ analogue data, the 50th data ranked from small to large is taken as the critical value; when $CIC(n) < LCIC(n)$, this judgment matrix is deemed as with satisfactory consistence. $LCIC(n)$ calculated according to analogue data is as shown in following Table 1:

TABLE 1 $LCIC(n)$ Value Calculated According to Analogue Data

Order	3	4	5	6	7	8	9
$LCIC(n)$	0.101	0.132	0.176	0.189	0.192	0.199	0.208

This paper introduces consistency test index coefficient $PCIC(n)$ according to the offset information of judgment matrix, and $PCIC(n)$ is defined as:

$$PCIC(n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (b_{ij} - b_{ij}^k)^2 / 2\sigma^2. \tag{8}$$

Mathematical derivation proves $PCIC \sim \chi^2(n^2)$, therefore the problem of testing whether judgment matrix has a satisfactory consistency is translated into hypothesis testing problem at the right; the original hypothesis considers that the judgment matrix has satisfactory consistence, namely $H_0: \sigma^2 \leq \sigma_0^2$. Construct statistics

$\chi^2_{PCIC} = \sum_{i=1}^n \sum_{j=1}^n (b_{ij} - b_{ij}^k)^2 / 2\sigma^2$, if $\chi^2_{PCIC} > \chi^2_{1-\alpha}(n^2)$, the original hypothesis is refused, or else the judgment matrix is deemed as with satisfactory consistency. This judgment method in this paper is called as consistency χ^2 test method based on offset information $PCIC(n) - \chi^2_{\alpha}(n^2)$.

The coefficient of consistency test index $LCIC(n)$ and $\chi^2_{1-\alpha}(n^2)$ continuously increase with n , indicating that the larger the order n of judgment matrix, the larger the consistency which can be permitted. Compared with the pure use of $CIC(n) < 0.10$ criterion, this method is more flexible. The analogue experiment shows that

$PCIC(n) - \chi^2_\alpha(n^2)$ criterion is more stringent than $LCIC(n)$. The proportion of matrix which meets the requirement of consistency under $PCIC(n) - \chi^2_\alpha(n^2)$ Criterion and $LCIC(n)$ Criterion is as shown in following Table 2.

TABLE 2 The proportion of matrix

Order	3	4	5	6	7	8	9
$PCIC(n)$	9.72	9.78	9.73	9.81	9.82	9.87	9.91
$LCIC(n)$	9.82	9.86	9.91	9.92	9.92	9.89	9.93

3.2 ROBUSTNESS ANALYSIS

With judgment matrix $C = \{c_{ij}\}_{n \times n}$ as example, suppose that relative rate of change is $a \in [0,1]$, any element of matrix is c_{ij} , there are 200 new c_{ij} randomly generated in $[c_j - \beta_{ij}, c_j + \beta_{ij}] \cap [1/9, 9]$, this way 200 random judgment matrixes are obtained. Obtain ranking weight through AGA-LCAHP analysis, and by comparing it with ranking weight of original matrix, the analysis result is as follows:

When relative rate of change is set as 10%, the ratio that 200 random vibration matrixes have satisfactory consistency is 0.01, and the coefficient of satisfactory consistency index $PCIC(n)$ is 0.007, it can be considered that all these matrixes have satisfactory consistency. In the view of ranking weight, when relative rate of change is set as 20% and 50% respectively, the difference between ranking weight of these matrixes and original judgment matrix is not large, indicating that the ranking weight calculated with AGA-LCAHP has certain stability.

3.3 COMPARISON BETWEEN AGA-LCAHP AND OTHER CORRECTION AHP METHODS

The methods for correcting judgment matrix mainly include empirical estimation method, optimal transfer matrix method; vector included angle cosine method, pattern recognition method and induced matrix method. When correcting judgment matrix, mainly include following two aspects of problems, one is that the degree of complex of adjusting the algorithm and the calculated amount are large; another one is that the adjustment of algorithm lacks of the use of information of index or scheme, and the orientation of adjusting lacks of theoretical foundation, sometimes it may go against the subjective intention of expert or affected by the logic error of expert. AGA-LCAHP evaluation model put forward in this paper fully extracts the opinion of expert and corrects the logic error of expert based on expert judgment matrix. Directly proceeded with judgment matrix, this method takes the variable of complete consistency index of judgment matrix as the orientation of optimization, thus it is simple and intuitive; with the global searching ability, AGA-LCAHP evaluation model improves the efficiency of searching through accelerating algorithm; Robustness

analysis of AGA-LCAHP evaluation model indicates that the calculation result of this method is stable; AGA-LCAHP evaluation model achieves the correction function of location and orientation according to the offset degree information of judgment matrix, making its correction amplitude small and that the ranking result and result of most correction method are similar, therefore it has higher universality and adaptation; AGA-LCAHP evaluation model takes full advantage of the information of judgment matrix to achieve intelligent search and optimization through genetic algorithm, which decreases the computational expense and increases the efficiency of correction.

In the view of ranking weight of judgment matrix, characteristic value method is the frequently used method, but the consistency test and weight calculation of judgment matrix in this method is separate, and the weight and consistency are fully determined by judgment matrix. When the consistency of judgment matrix is poor, it is difficult to determine effective characteristic root; row sum normalization method, column sum inversion method and sum product method is just a kind of approximation algorithm, and its accuracy of calculation is not high; AGA-LCAHP method, logarithm regression method, method of least square and minimum deviation method is a kind of initiative method of using all element information of judgment matrix under the condition of meeting the consistency, they obtain the ranking weight through the optimization to the condition of consistency or optimize the value of consistency by means of changing the weight, therefore these methods have many fine natures such as substitution invariance, compatibility, symmetry and complete harmony. However the weight determined through logarithm regression, method of least square and minimum deviation method is small, it is easy to generate large deviation due to that the weight appears on the denominator, thus the robustness of calculated result is poor. AGA-LCAHP method directly deduces the consistency index coefficient according to the definition of judgment matrix, fully extracts the information of judgment matrix, optimizes in the field of degree of deviation and consistency. Therefore, AGA-LCAHP method is a kind of initiative and intuitive method.

4 Comparative analysis of AGA-LCAHP algorithm example

Example 1. Suppose that the judgment matrix is C_1

$$C_1 = \begin{bmatrix} 1 & 1/9 & 2 & 1/5 \\ 9 & 1 & 5 & 2 \\ 1/2 & 1/5 & 1 & 1/2 \\ 5 & 1/2 & 2 & 1 \end{bmatrix},$$

$$C_1^1 = \begin{bmatrix} 1 & 1/9 & 1/9 & 1/5 \\ 9 & 1 & 5 & 2 \\ 9 & 1/5 & 1 & 1/2 \\ 5 & 1/2 & 2 & 1 \end{bmatrix}.$$

The vector of ranking weight of this matrix is (0.1450, 0.5433, 0.0853, 0.2264) obtained by means of sum product method, and the correction judgment matrix obtained by means of included angle cosine method of the vector of each column and eigenvector of normalization judgment matrix is C_1^1 .

The vector of corresponding ranking weight is (0.0427, 0.5210, 0.1860, 0.2513), and the value of consistency index coefficient is $CR = 0.1048$.

The correction matrix obtained by means of induced matrix method is C_1^2 .

$$C_1^2 = \begin{bmatrix} 1 & 1/7 & 2 & 1/5 \\ 7 & 1 & 5 & 2 \\ 1/2 & 1/5 & 1 & 1/2 \\ 5 & 1/2 & 2 & 1 \end{bmatrix},$$

$$C_1^3 = \begin{bmatrix} 1 & 0.1429 & 0.5 & 0.2 \\ 7 & 1 & 5 & 2 \\ 2 & 0.2 & 1 & 0.5 \\ 5 & 0.5 & 2 & 1 \end{bmatrix},$$

$$C_1^4 = \begin{bmatrix} 1 & 0.1111 & 0.5 & 0.2000 \\ 9 & 1 & 3 & 2 \\ 2 & 0.3333 & 1 & 0.5000 \\ 5 & 0.5000 & 2 & 1 \end{bmatrix}.$$

The vector of corresponding ranking weight is (0.1014, 0.5254, 0.0952, 0.2780) and the value of consistency index coefficient is $CR = 0.0933$.

Use AGA-CAHP and AGA-LCAHP to correct, the parameter of rate of change is set as 30%, and the initial change interval of each ranking weight is set as [0,1], use AGA accelerating algorithm to calculate 30 times and respectively obtain correction matrix C_1^3 and C_1^4

The vector of corresponding ranking weight is (0.0643, 0.5345, 0.1237, 0.2776) and (0.0617, 0.5114, 0.1445, 0.2823) the value of consistency index coefficient is $CR = 0.0083$ and $CR = 0.0072$ respectively; the consistency obtained by means of accelerating genetic algorithm is maximum and the correction amplitude is minimum. From the comparison between AGA-CAHP and AGA-LCAHP and other correction methods, we can see that the correction amplitude obtained by means of accelerating genetic algorithm is minimum and the consistency is maximum, and for the vector of ranking weight, AGA-CAHP and

AGA-LCAHP are similar to other correction methods. From the comparison between AGA-CAHP and AGA-LCAHP we can see that AGA-LCAHP can further improve the level of consistency and decrease the amplitude of correction.

Example 2. Suppose that the judgment matrix is C_2 .

$$C_2 = \begin{bmatrix} 1 & 2 & 4 & 1/2 & 2/3 \\ 1/2 & 1 & 3 & 1/3 & 4/9 \\ 1/4 & 1/3 & 1 & 2/9 & 1/9 \\ 2 & 3 & 9/2 & 1 & 1/2 \\ 3/2 & 9/4 & 9 & 2 & 1 \end{bmatrix},$$

$$C_2^1 = \begin{bmatrix} 1 & 1.7896 & 4.0807 & 0.6980 & 0.5087 \\ 0.5588 & 1 & 2.2803 & 0.3900 & 0.2843 \\ 0.2451 & 0.4386 & 1 & 0.1700 & 0.1248 \\ 1.4328 & 2.5642 & 5.8471 & 1 & 0.7290 \\ 1.9656 & 3.5177 & 8.0212 & 1.3718 & 1 \end{bmatrix}.$$

The correction matrix obtained by means of pattern recognition method is C_2^1 . The vector of corresponding ranking weight is (0.1922, 0.1074, 0.0471, 0.2754, 0.3778), and the value of consistency index coefficient is $CR = 0.3114$; use AGA-CAHP and AGA-LCAHP to correct, the parameter of rate of change is set as 30%, and the initial change interval of each ranking weight is set as [0,1], use AGA accelerating algorithm to calculate 20 times and respectively obtain correction matrix C_2^2 and C_2^3 :

$$C_2^2 = \begin{bmatrix} 1 & 2.001 & 4.0011 & 0.4995 & 0.6657 \\ 0.5000 & 1 & 2.9979 & 0.3332 & 0.4449 \\ 0.2499 & 0.3336 & 1 & 0.2221 & 0.1111 \\ 2.0020 & 3.0012 & 4.5025 & 1 & 0.4995 \\ 1.5022 & 2.2477 & 9.0009 & 2.0020 & 1 \end{bmatrix},$$

$$C_2^3 = \begin{bmatrix} 1 & 2.001 & 4.0011 & 0.4995 & 0.6657 \\ 0.5000 & 1 & 2.9979 & 0.3332 & 0.4449 \\ 0.2499 & 0.3336 & 1 & 0.2221 & 0.1111 \\ 2.0020 & 3.0012 & 4.5025 & 1 & 0.4995 \\ 1.5022 & 2.2477 & 9.0009 & 2.0020 & 1 \end{bmatrix}.$$

The vector of corresponding ranking weight is (0.2185, 0.1108, 0.0463, 0.2270, 0.3973) and (0.2185, 0.1108, 0.0463, 0.2270, 0.3973) respectively, and the value of consistency index coefficient is $CR = 0.0407$ and $CR = 0.0407$ respectively. From the comparison between AGA-CAHP, AGA-LCAHP and pattern recognition method we can see that the correction amplitude of AGA-LCAHP method is minimum, and the consistency is high, and for the vector of ranking weight,

the difference between AGA-LCAHP and pattern recognition method is not large.

5 Case studies

Taking the rural folk house renovation in the satisfaction evaluation for example, the paper explores and analyses the relationship between the rural residential building

renovation and the energy use condition and villager satisfaction. The paper focuses on studying on the satisfaction of residents in the residence renovation process and the renovation effect and mainly reflects the satisfaction situation in the housing condition, surrounding environment, energy use and government policies, with the specific index system shown in the Figure 1:

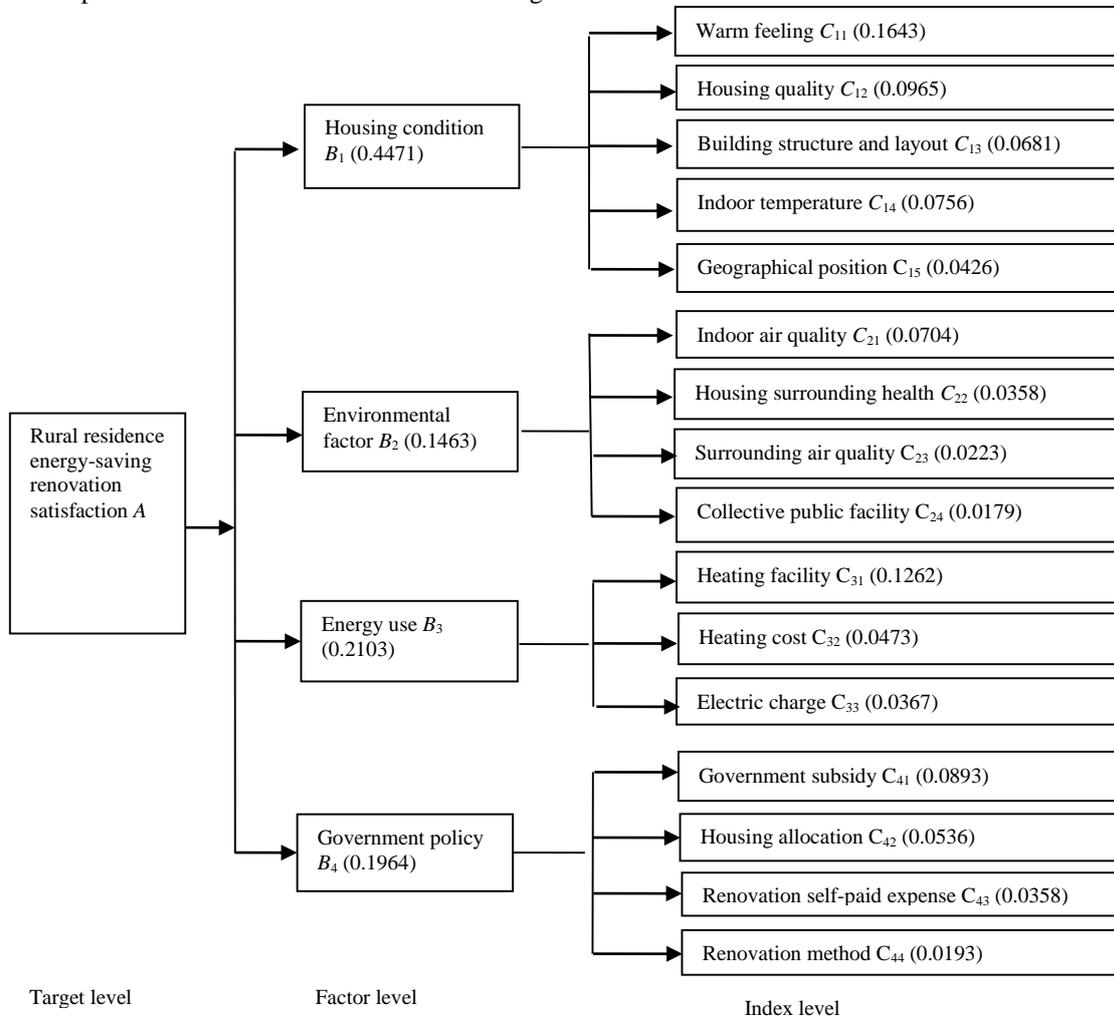


FIGURE 1 Rural residence energy-saving renovation index system and index weight

The judgment matrix in the paper is given out under the general objective of experts, and the direct comparison method is adopted for the importance between indexes to generate the judgment matrix. The paper uses the method of 1-5 ratio scale to quantify the logical judgment of

relative importance. The comparison value can be set as 1, 2, 3, 4 and 5, standing for the importance degrees between two indexes, namely equally important, weakly important, obviously important, very important or extremely important, shown in the following Table 3:

TABLE 3 Pairwise Comparison Saaty Scale

Scale	Significance
1	Comparing two elements, the two are equally important
2	Comparing two elements, the first one is weakly important than the second
3	Comparing two elements, the first one is obviously important than the second
4	Comparing two elements, the first is very important than the second
5	Comparing two elements, the first is extremely important than the second
Reciprocals of numerical values	The above comparative result of the two elements

The judgment matrix is reached through interviewing with experts and inviting them to fill in the consultation

table. The judgment matrix of the factor level and the target level is $A = (a_{ij})_{4 \times 4}$, and the judgment matrixes of the index level and the factor level are $B_1 = (b_{ij})_{5 \times 5}$, $B_2 = (b_{ij})_{4 \times 4}$, $B_3 = (b_{ij})_{3 \times 3}$, $B_4 = (b_{ij})_{4 \times 4}$, shown as follows:

$$A = \begin{bmatrix} 1 & 3 & 1 & 1 \\ \frac{1}{3} & 1 & 1 & \frac{1}{2} \\ 1 & 1 & 1 & 1 \\ \frac{1}{3} & 2 & 1 & 1 \end{bmatrix},$$

$$B_1 = \begin{bmatrix} 1 & 1 & 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & 1 & 1 & \frac{1}{3} \\ 1 & 2 & 1 & 1 & \frac{1}{2} \\ 2 & 1 & 3 & 1 & 2 \\ 4 & 3 & 2 & \frac{1}{2} & 2 \end{bmatrix}.$$

$$B_2 = \begin{bmatrix} 1 & 3 & 4 & 2 \\ \frac{1}{3} & 1 & 2 & 3 \\ \frac{1}{4} & \frac{1}{2} & 1 & 2 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix}$$

$$B_3 = \begin{bmatrix} 1 & 3 & 4 \\ \frac{1}{3} & 1 & 1 \\ \frac{1}{4} & \frac{1}{2} & 2 \end{bmatrix}$$

$$B_4 = \begin{bmatrix} 1 & 2 & 3 & 5 \\ \frac{1}{2} & 1 & 5 & 3 \\ \frac{1}{3} & \frac{1}{5} & 1 & 4 \\ \frac{1}{8} & \frac{1}{3} & \frac{1}{4} & 1 \end{bmatrix}$$

The result of calculating the sorting weight of the above judgment matrixes using AGA-LCAHP is shown in the Table 4:

TABLE 4 The judgment matrix weight comparison of eigenvalue method and AGA-LCAHP method

Method	Judgment matrix	Sorting weight					Consistency index system numerical value
		w ₁	w ₂	w ₃	w ₄	w ₅	
Eigenvalue method	A	0.406	0.148	0.237	0.208	0.0927	0.0655
AGA-LCAHP	A	0.4471	0.1463	0.2103	0.1964	0.0925	0.0569
Eigenvalue method	B ₁	0.3265	0.2348	0.1678	0.1782		0.0591
AGA-LCAHP	B ₁	0.3674	0.2158	0.1524	0.1692		0.0452
Eigenvalue method	B ₂	0.4763	0.2559	0.1522	0.1156		0.0949
AGA-LCAHP	B ₂	0.4812	0.2445	0.1521	0.1222		0.0932
Eigenvalue method	B ₃	0.5937	0.2265	0.1798	0.0909		0.0454
AGA-LCAHP	B ₃	0.6003	0.2251	0.1746	0.0984		0.0432
Eigenvalue method	B ₄	0.4545	0.2727	0.1919			0.1027
AGA-LCAHP	B ₄	0.4546	0.2728	0.1822			0.0748

From Table 4, we can see that the computational accuracy of AGA-LCAHP is higher than the calculation result of the eigenvalue method; the global optimization searching can be conducted based on the sorting value interval, and the calculation result is relatively stable; the consistency coefficient average value of judgment matrix after correction is less than 0.1, with satisfying consistency. The weight coefficient of evaluation indexes and evaluation factors is further reached via accelerating genetic algorithm, shown in Table 4. The calculation result shows that the most influential factor sorting of rural folk house renovation in satisfaction evaluation is housing condition B₁, energy use B₃, environment factor B₂ and government policy B₄. By the computation of index weight, the most influential index sorting is warm feeling C₁₁, heating facility C₃₁, housing quality C₁₂, government subsidy C₄₁, indoor

Temperature C₁₄, indoor air quality C₂₁, building structure and layout C₁₃, house allocation situation C₄₂, heating cost C₃₂, geographical position C₁₅, electric charge C₃₃, renovation self-paid expense C₄₃, housing surrounding health C₂₂, surrounding air quality C₂₃, renovation method C₄₄, and collective public facility C₂₄. The specific weighted value is shown in the figure. Through the above analysis, we can draw a conclusion that the most influential major factors to rural folk house renovation satisfaction are the heat preservation situation after renovation and the renovation situation of heating equipment and housing quality, so the government can focus on strengthening the renovation effort in heating facilities and housing quality in housing energy-saving renovation, so as to improve residents' satisfaction. The Rural residence energy-saving renovation index system and index weight is shown in Figure 1.

6 Conclusions

Firstly, this paper regards the correction of judgment matrix as a nonlinear optimization problem. AGA-LCAHP evaluation model is a new method of extracting the information of offset degree according to the inconsistency of judgment matrix, locating the inconsistency of correction judgment matrix through accelerating genetic algorithm (AGA) under the guide of information of offset degree, and calculating the ranking weight of each element of judgment matrix.

Secondly, AGA-LCAHP evaluation model put forward in this paper fully extracts the opinion of expert and corrects the logic error of expert based on expert judgment matrix. This method directly deduces the consistency index coefficient according to the definition of judgment matrix, fully extracts the information of judgment matrix, and optimizes in the field of degree of deviation and consistency. Therefore, AGA-LCAHP method is initiative and intuitive. The robustness analysis of AGA-LCAHP evaluation model indicates that the calculation result of this method is stable; evaluation model achieves the correction function of location and orientation according to the information of offset degree of judgment matrix, making its correction amplitude small and that the ranking result and result of most correction method are similar, therefore it has higher universality and adaptation; AGA-LCAHP evaluation model takes full advantage of the information of judgment matrix to achieve intelligent

search and optimization through genetic algorithm, which decreases the computational expense and increases the efficiency of correction.

Thirdly, AGA-LCAHP method is a kind of intelligent evaluation method, which calculates the ranking weight of judgment matrix while judging and correcting the consistency of judgment matrix. It provides certain theoretical and practical value for the integration of intelligent method and analytic hierarchy process.

Finally, The difficulty of rural folk house renovation in satisfaction evaluation lies in reasonably confirming the weight of evaluation indexes, so the analytic hierarchy process (AHP) scored by experts is used to confirm the realization process of these weights. The case study result shows that the AGA-LCAHP method features high computational accuracy and stable calculation result and also has the popularization and application value in other comprehensive assessment. The analysis of rural folk house renovation on the satisfaction evaluation result can provide important scientific basis for rural folk house renovation.

Acknowledgments

This paper was funded by these projects: Beijing philosophy and social science planning project (10BeJG338); special funds for the construction of new modern industrial district development research base in Beijing (PXM2013-014222-000051).

References

- [1] Saaty L T 1980 The analytic hierarchy process *New York Mc Graw Hill*
- [2] Liu W, Lei Z 1997 Study on rectification method for the judgment matrix in AHP **17**(6) 30-4 39
- [3] Zhu J J, Wang M G, Liu S X 2007 Research on consistency modification problem of comparison matrix in the analytical hierarchy process *Systems Engineering-Theory & Practice* **5**(1) 18-22
- [4] Holland J H 1992 Genetic algorithm *Scientific American* **9**(7) 44-50
- [5] Xu S B 1988 The principle of analytic hierarchy process *Tianjin University Press (in Chinese)*
- [6] Chen Z M, Liang C Y, Lu W X 2006 A new method of deriving and synthesizing the judgment matrix information in the analytic hierarchy process *Journal of Hefei University of Technology(Natural Science)* **29**(8) 951-4 (in Chinese)
- [7] Jin J L, Yang X H, Ding J 2001 An improved simple genetic algorithm-accelerating genetic algorithm *Systems engineering-theory & practice* **21**(4) 20-4 (in Chinese)
- [8] Cao L X, Feng X Z 2010 A new improved AHP algorithm research and application *Computer Technology and Development* **20**(12) 121-51

Authors	
	<p>Chenxia Suo, born on May 4, 1963, Beijing, China</p> <p>Current position, grades: associate professor of School of Economics and Management, Beijing Institute Of Petrochemical Technology.</p> <p>University studies: TsingHua University, Master.</p> <p>Scientific interest: energy management, engineering economics.</p> <p>Publications : the monograph: "Efficiency and Effectiveness of Energy-efficient Building Technology in Chinese Rural Area".</p> <p>Experience: teaching and research work 19 years.</p>
	<p>Yong Yang, born on December 24, 1983, Xingtai City, Hebei Province, China</p> <p>Current position, grades: PhD.</p> <p>University studies: college of Statistics and Mathematics, Zhejiang Gongshang University.</p> <p>Scientific interest: decision-making; intelligent comprehensive evaluation.</p> <p>Publications: YONG YANG, WEIHUA SU, CHONGHUI ZHANG, SHOUZHEN ZENG 2013 INTUITIONISTIC FUZZY DECISION-MAKING WITH SIMILARITY MEASURES AND OWA OPERATOR <i>International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems</i>. Vol. 21, No. 2 (2013).</p>

Spectral colour calibration for multi-ink printer

Ying Wang*, Zhongmin Wang, Sheping Zhai

School of Computer Science & Technology, Xi'an Univ. of Posts & Telecommunications, Xi'an, China

Received 7 March 2014, www.tsi.lv

Abstract

To implement colour calibration during outputting multispectral images on a multi-ink printer, a new spectral colour calibration method is proposed. Firstly, by uniform sampling in the multi-ink printer colour space, measuring the spectral reflectance of the samples and then transforming the reflectance data to a low dimension spectral space, a forward look-up table is created. Then by sampling in the low dimension spectral space and using a nonlinear optimization to calculate the mapping points of these samples in the printer colour space, a backward look-up table is established. Meanwhile, to improve the optimization accuracy and shorten the computing time, an algorithm is designed to determine the optimization parameters based the samples. Finally, a multi-linear interpolation method is carried out on the forward and backward look-up table to achieve the spectral colour calibration of the multi-ink printer. Experiments show that the new method not only takes advantage of the high calibration precision and less time-consuming of the look-up table, but also solves the problem brought by the high dimension of the spectral data to the look-up table method by utilizing the nonlinear optimization and dimension reduction. Compared with the spectral colour calibration model methods, the new method improves the colorimetric and spectral precision obviously. It also raises the time efficiency of the inverse calibration significantly.

Keywords: spectral colour calibration, multispectral image, multi-ink printing, look-up table

1 Introduction

Multispectral images are those whose pixel values are spectral reflectance of source scenes. They are captured by multi-channel cameras, and mainly used for the accurate and consistent colour reproduction of source scenes under different illuminant. Now they have been widely used in high-end imaging fields such as art archiving [1, 2], medicine [3, 4], military target imaging.

The hard-copy of multispectral images is achieved by printing on a multi-ink printer. During the image printing process, colour calibration is a crucial part. Colour calibration is used to compensate the colour distortion resulted by the nonlinear characteristic of printers and to achieve accurate and consistent colour reproduction of the images on various devices. The data of multispectral images are spectral reflectance and they are narrow sampled in the range of visible light. This leads to the fact that the dimension of image data is high and the amount of the data is large. Moreover, the channel number of the multi-ink printer is more than 6. All these result in that the colour calibration methods designed for the colorimetric images cannot be applied to the reproduction of the multispectral images. Thus, designing new spectral colour calibration method becomes the key in the process of the hard copy of multi-spectral images.

The existing spectral colour calibration methods for the multi-ink printing include Yule-Neilson Spectral Neugebauer (YNSN) model [5, 7], Cellular YNSN model [8, 9], Kubelka-Munk colour mixing model [10], and so on. All these methods can directly calibrate the spectra by

creating the mathematical printing model through the analysis of the physical printing process. The printing accuracy using the model methods is determined by the precision of the model. Since the device is nonlinear and the printing process is changeable, the spectral printing models cannot accurately simulate the actual printing process. This makes the accuracy of spectral colour calibration low. Furthermore, the inverse calibration is implemented by using nonlinear optimization real time. This results in that the consuming time is very long and the efficiency is low. The calibration method commonly used for chrome images is look-up table (LUT). The LUT can be created beforehand. It can avoid the calculating bottleneck in the colour calibration. In addition, its precision is high. Nevertheless, the chrome image is 3-channel image and the printer used for it is 4-ink printer. Thus, the chroma calibration only needs 3-dimension look-up table and tri-linear interpolation. For the multi-ink printing of the multispectral images, the low-dimension LUT cannot meet the requirements obviously. Thus, the high dimension LUT must be established and the multi-linear interpolation must be used. All these become the difficulty when the LUT method is used for spectral colour calibration.

In this paper, a spectral colour calibration using look-up table for multi-ink printing is presented. During the creation of the LUT, a low-dimension spectral space is introduced. The mapping between the low-dimension spectral space and the multi-ink printer colour space is established. It solves the problem that the high dimension spectral data cannot be directly used to create the LUT.

* *Corresponding author* e-mail: wangyingjsj@xupt.edu.cn

During the creation of the inverse LUT, a non-linear optimization is adopted to gain the mapping of the samples from the low-dimension spectral space to the printer colour space. An algorithm to determine the optimization parameters based on the samples is designed to improve the precision and efficiency. This makes the generation of the inverse calibration sample set and the creation of the inverse look-up table become possible. Finally, the multi-linear interpolation is used to achieve the spectral colour calibration. The new method can not only solve the problem brought by the high dimension of the spectral data, but also take the advantage of the high accuracy and efficiency of the look-up table method. It improves the precision and efficiency of the spectral colour calibration effectively.

2 Spectral colour calibration

The colour space of chromatic input/output devices is different and nonlinear. So the colour calibration becomes necessary when images are produced on hard copy devices. Colour calibration is an important step in the Colour Management. When the data in the source device colour space is mapped to the destination device colour space, the Colour Management uses the colour calibration to transform the source data to the device-independent colour space firstly and then transform them to the destination colour space. When images are printed out, it mainly relates to the transformation between the printer colour space and the device-independent colour space. For multi-spectral images, the spectral reflectance space can be utilized as the device-independent space.

Given a point c in the printer colour space, the corresponding spectral reflectance s can be measured by a spectrophotometer. Defining

$$s = \mathcal{F}_{print}(c), \quad c \in \Omega_{print}, \quad s \in \Omega_{spec}, \quad (1)$$

where $\mathcal{F}_{print}(\bullet)$ is the nonlinear mapping from the printer colour space to spectral reflectance space. Ω_{print} is printer colour space and Ω_{spec} is the spectral reflectance space. Correspondingly, for the spectral reflectance that can be printed by the printer, $\mathcal{F}_{print}^{-1}(\bullet)$ is used to transform it to the printer colour space. That is,

$$c = \mathcal{F}_{print}^{-1}(s), \quad s \in G_{print}. \quad (2)$$

G_{print} is the spectral domain of the printer that the spectra in it can be printed but those out of it cannot be printed. It can be denoted as

$$G_{print} = \{s \in \Omega_{spec} \mid \exists c \in \Omega_{print}, \mathcal{F}_{print}(c) = s\}. \quad (3)$$

The spectral colour calibration is to gain the forward mapping $\mathcal{F}_{print}(\bullet)$ and the backward mapping $\mathcal{F}_{print}^{-1}(\bullet)$, thereby, to achieve the calibration transformation between the printer colour space and the spectral space.

3 Spectral colour calibration for multi-ink printing based on LUT

There are two kinds of methods to achieve $\mathcal{F}_{print}(\bullet)$ and $\mathcal{F}_{print}^{-1}(\bullet)$: one is to create the printing model and another is to utilize look-up table. Since the physical printing process is affected by the paper, ink, printing point and many other factors, the printing model cannot simulate the actual printing process accurately. It makes calibration precision low. Moreover, the spectral printing model is usually unable to obtain the analytical inverse model. Using the model to carry out inverse colour calibration needs to utilize the nonlinear optimization in real-time to implement the inverse transformation of the colour space. It leads to that the consuming time of the calibration process is long. The LUT method creates the one-one correspondence of the samples in the source colour space and the destination colour space by sampling in the colour space and measuring the values of the samples. Although this method requires more samples, the actual printing and measuring of the samples makes it better reflect the physical printing process. Thus, the printing precision is high. In addition, the LUT can be established in advance and only needs to utilize the interpolation to execute calibration. Therefore, the calibration can be carried out in real time.

During creating the mapping between the spectral space and the multi-ink printer color space, if the high dimension spectral reflectance data is employed directly, e.g. 31-dimension spectra, the dimension of the LUT will be 31 and the size will be $8 \times 10^{15} GB$ (Given the sampling level per dimension is 6, the printer is 6-ink and the size of the data per channel is 1 byte). Obviously, this makes the LUT cannot be achieved. Therefore, it needs to apply dimensionality reduction to the high dimension spectral reflectance and utilize the low dimension spectral colour space to substitute the spectral reflectance space in the creation of the LUT. Firstly, a low-dimension space Ω_{spec_l} is established,

$$\Omega_{spec_l} = \{r \mid \exists s \in \Omega_{spec}, \mathcal{L}(s) = r\}. \quad (4)$$

$\mathcal{L}(\bullet)$ is the algorithm of the dimensionality reduction.

Correspondingly, $\mathcal{L}^{-1}(\bullet)$ is the inverse transformation from the low dimension space to the spectral reflectance space. The spectral domain of the printer in the new space is

$$G_{print_I} = \{r \in \Omega_{spec_I} \mid \exists c \in \Omega_{print}, \mathcal{F}_{print}(c) = r\} \quad (5)$$

$$c = \mathcal{F}_{print}^{-1}(r), \quad r \in G_{print_I} \quad (7)$$

Then the forward spectral colour calibration can be defined as

$$r = \mathcal{F}_{print}(c), \quad c \in \Omega_{print}, \quad r \in \Omega_{spec_I} \quad (6)$$

The inverse model is

The LUT method is to simulate the forward mapping $\mathcal{F}_{print}(\bullet)$ and backward mapping $\mathcal{F}_{print}^{-1}(\bullet)$ by creating the one-one correspondence of the source and destination space. The process of the spectral colour calibration using LUT is shown in Figure 1 and Figure 2.

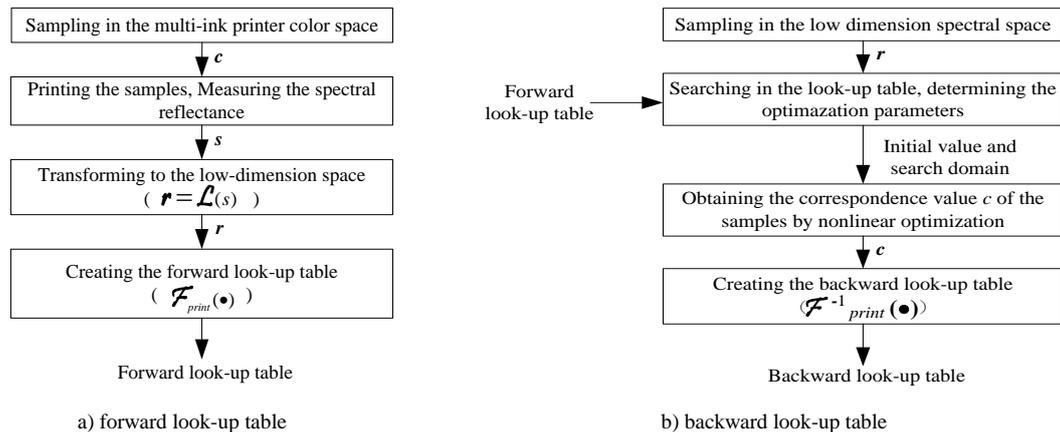


FIGURE 1 Creation of the look-up table

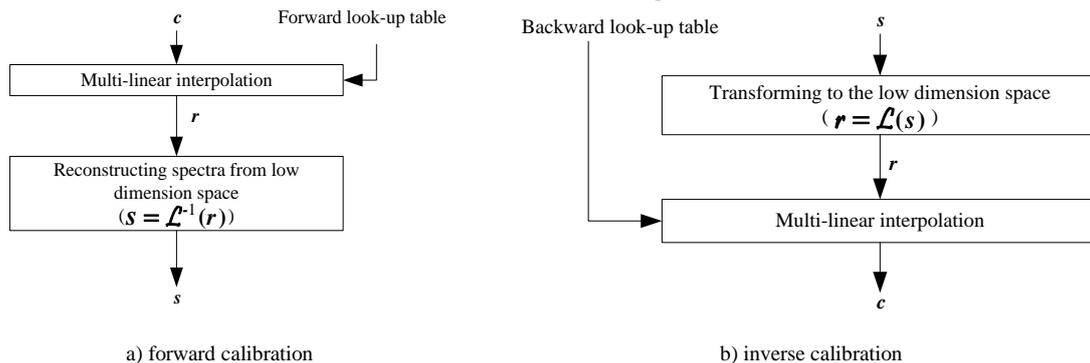


FIGURE 2 Spectral colour calibration for multi-ink printing

During the creation of the forward LUT (Figure 1(a)), the samples in multi-ink printer colour space and its mapping points in the low dimension spectral space are utilized to generate the forward calibration sample set. The forward LUT is established based on this sample set. The creation of the backward LUT (Figure 1(b)) is more complex than the forward LUT. Since the multi-printer colour space and the spectral space are different in nature, the mapping points of the samples sampled uniformly in the printer colour space are non-uniform in the spectral space. Thus, the forward calibration sample set cannot be utilized directly to obtain the mapping points in the printer colour space of the samples in the spectral space. Therefore, we sample uniformly in the low-dimension spectral space and then calculate the mapping points of the samples in the multi-ink printer colour space. Form this, the inverse calibration sample set is generated and the backward LUT is created based on it. Experiments show that the convex volume in the printer colour space is non-convex in the spectral space, so the inverse uniformization

methods used in the chroma colour calibration [11] cannot be applied. In this paper, a nonlinear optimization method is utilized to establish the backward LUT. Its optimization parameters are determined based on the forward LUT.

After creating the forward and backward LUT, the spectral colour calibration is implemented by interpolation. Experiments show that the non-linear interpolation does not have a clear advantage in the interpolation accuracy compared with the linear interpolation in the colour calibration. Therefore, the multi-linear interpolation is applied in this paper to carry out the LUT search.

4 Key technologies of the algorithm

4.1 DETERMINATION OF THE NONLINEAR OPTIMIZATION PARAMETERS

The framework using non-linear interpolation to establish the backward LUT is show in Figure 3. From Figure 3 we

know optimization objective function, initials and the boundary of the optimization variables are the key when using non-linear optimization.

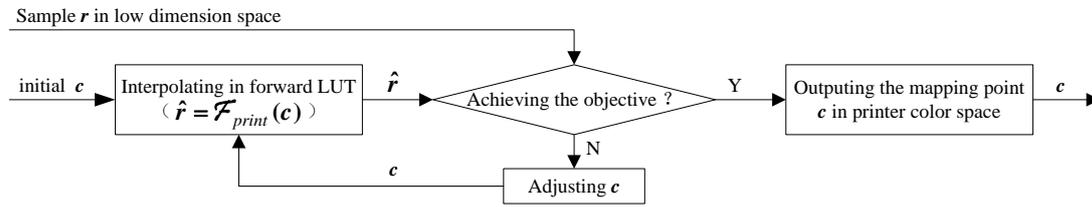


FIGURE 3 Process of generating backward LUT by using non-linear optimization

In order to achieve high quality output of the multispectral images, it is required that the output image and the source image are able to gain a good match in both chrome and spectral reflectance. High chrome precision means that the output image matches the source well under a typical illuminant; high spectral precision means that the output image matches the source well when the illuminant is changed. Therefore, the optimization objective function is set to take into account both the spectral and colorimetric accuracy. We define the function as

$$\begin{aligned} \min f(c) &= \alpha \|r - \mathcal{F}_{print}(c)\|_2^2 \\ &+ \|Col(\mathcal{L}^{-1}(r)) - Col(\mathcal{L}^{-1}(\mathcal{F}_{print}(c)))\|_2^2, \quad (8) \\ \text{s.t. } c &\in sub_BPr \\ sum(c) &\leq c_{limit} \end{aligned}$$

where $Col(\bullet)$ is the transformation from the spectral reflectance space to the CIELAB uniform colour space. Its transforming process is described in Reference [12]. α is a weight. Experiments show when $\alpha = 50$ that $f(c)$ can better reflect the error of both chroma and spectra. sub_BPr is the boundary of the optimization variable. Its definition is shown in Equation (13). c_{limit} is the limitation of the total amount of ink. If the amount of each primary colour ink is too much, the paper will be not able to hang on the ink. The printing quality will drop.

The optimization initial and variable play a crucial role for the convergence speed of optimization function and optimization accuracy. In this paper the algorithm for determining the initial and the boundary of variable is described as follows.

(1) According to the samples in the forward LUT, obtain the hypercube set in the multi-ink printer colour space. It is

$$P_{hypercubes} = \{P_i\} \quad i = 1, \dots, N, \quad (9)$$

Then gain the minimum circumscribed hypercube set of the mapping volume of each above hypercube in the low dimension spectral space. It is

$$R_{hyperbodies} = \{R_i\} \quad i = 1, \dots, N, \quad (10)$$

where R_i and P_i is one-one correspondence. P_i is a hypercube in printer colour space in forward LUT. R_i is the minimum circumscribed hypercube of the mapping volume of hypercube P_i in the low dimension spectral space. N is the number of hypercube in the forward LUT.

During the creation of the forward LUT, the samples are uniformly distributed in the printer colour space. Using the adjacent sample points in the multi-dimension space as the vertices, a hypercube can be formed. Using the mapping points of the above samples in the low dimension spectral space, the mapping volume in the low dimension spectral space of the hypercube in the printer colour space can be achieved.

(2) For any sample r in the low dimension spectral space, obtain the minimum circumscribed hypercube set it locates in,

$$sub_Rr = \{sub_R | sub_R \in R_{hyperbodies}, r \text{ is in } sub_R\}, \quad (11)$$

The hypercube in the printer colour space recorded in the forward LUT is regular convex volume, but its mapping volume in the low dimension spectral space is often non-convex and irregular. This leads to that the hypercube in the multi-ink printer colour space that the mapping point of the sample r located in cannot be positioned directly. The sample r may belong to several minimum circumscribed hypercube.

(3) Find the corresponding hypercube set of sub_Rr in the printer colour space, and calculate the upper and lower bounds on each dimension,

$$sub_Pr = \{sub_P | sub_P \in P_{hypercubes}, sub_R \in sub_Rr, sub_P \text{ and } sub_R \text{ is one-one correspondence}\}, \quad (12)$$

$$sub_BPr = \text{Upper and lower bounds of } sub_Pr \text{ on each dimension} \quad (13)$$

Use sub_BPr as the boundary of optimization variable in Equation (8).

(4) For every vertex c_{vertex} of the hypercube in sub_Pr , utilize the following formula to obtain the closest vertices for optimization objective,

$$c_{ver_closest} = \arg(\min_{c_{vertex}} (\alpha \|r - \mathcal{F}_{print}(c_{vertex})\|_2^2 + \|\text{Col}(\mathcal{L}^{-1}(r)) - \text{Col}(\mathcal{L}^{-1}(\mathcal{F}_{print}(c_{vertex})))\|_2^2)) \quad (14)$$

Use $c_{ver_closest}$ as the optimization initial.

(5) Repeat step (2), (3) and (4) to calculate the optimization initial and the variable boundary of each sample r_j in the low dimension spectral space.

In this paper, we choose Quasi-Newton Method as the non-linear optimization method to establish multi-dimension backward LUT.

4.2 MULTI-LINEAR INTERPOLATION

After creating the forward and backward LUT, the spectral colour calibration can be implemented by multi-linear interpolating in the LUT for every sample in printer colour space or low dimension spectral space. Multi-linear interpolation is the generalization of linear interpolation. The function of linear interpolation is

$$\mathcal{F}_{Fit}^{[m]}(x) = u_0^{[m]}(x) \bullet \mathcal{F}(x^{[m]}) + u_1^{[m]}(x) \bullet \mathcal{F}(x^{[m+1]}), \quad (15)$$

where $m = 0, 1, \dots, M - 1$ is serial number of the samples on x , the sample numbers is M . $\mathcal{F}(x^{[m]})$ is the value of the samples. It can be gained by the forward and backward LUT. $u_0^{[m]}(x)$ and $u_1^{[m]}(x)$ are basic function. They are designed to make the following equation hold,

$$\mathcal{F}_{Fit}^{[m]}(x^{[m]}) = \mathcal{F}(x^{[m]}), \quad \mathcal{F}_{Fit}^{[m]}(x^{[m+1]}) = \mathcal{F}(x^{[m+1]}), \quad (16)$$

Obviously,

$$u_s^{[m]}(x^{[m]}) = 1 - s, \quad u_s^{[m]}(x^{[m+1]}) = s \quad (s \in \{0, 1\}), \quad (17)$$

meets the requirement. Moreover, from Equation (15) we know $u_0^{[m]}(x)$ and $u_1^{[m]}(x)$ is linear function about x . Thus we design

$$u_s^{[m]}(x) = (1 - s) + (2s - 1)(x - x^{[m]}) / (x^{[m+1]} - x^{[m]}), \quad (s \in \{0, 1\}) \quad (18)$$

Multi-linear interpolation is to apply linear interpolation on each dimension. Given a N-linear interpolation, the sampling level on each dimension is M_1, M_2, \dots, M_N , and $m = (m_1, m_2, \dots, m_N)$ is the serial number of samples on each dimension. Then according to Equation (15), the n-linear interpolation function is defined as

$$\mathcal{F}_{Fit}^{[m,n]}(x) = u_0^{[m,n]}(x) \bullet \mathcal{F}_{Fit}^{[m,n-1]}(x_1, x_2, \dots, x_{n-1}, x_n^{[m_n]}, x_{n+1}, \dots, x_N) + u_1^{[m,n]}(x) \bullet \mathcal{F}_{Fit}^{[m,n-1]}(x_1, x_2, \dots, x_{n-1}, x_n^{[m_n+1]}, x_{n+1}, \dots, x_N) \quad (19)$$

where $n = 1, \dots, N$, $x = (x_1, x_2, \dots, x_N)$. Let

$$\mathcal{F}_{Fit}^{[m,0]}(x) = \mathcal{F}(x), \quad (20)$$

from Equation (18) we gain

$$u_s^{[m,n]}(x) = (1 - s) + (2s - 1)(x_n - x_n^{[m_n]}) / (x_n^{[m_n+1]} - x_n^{[m_n]}) \quad (s \in \{0, 1\}) \quad (21)$$

Then

$$\mathcal{F}_{Fit}^{[m,n]}(x) = \sum_{s_1, s_2, \dots, s_n \in \{0, 1\}} \left(\prod_{j \in \{1, \dots, n\}} u_{s_j}^{[m,j]}(x) \right) \bullet \mathcal{F}(x_1^{[m_1+s_1]}, x_2^{[m_2+s_2]}, \dots, x_n^{[m_n+s_n]}, x_{n+1}, \dots, x_N) \quad (22)$$

When $n = N$, we achieve n-linear interpolation function

$$\mathcal{F}_{Fit}^{[m,N]}(x) = \sum_{\substack{s = \{s_1, s_2, \dots, s_N\} \\ s_j \in \{0, 1\}}} u_s^{[m]}(x) \bullet \mathcal{F}(x^{[m+s]}), \quad (23)$$

where

$$u_s^{[m]}(x) = \prod_{j \in \{1, 2, \dots, N\}} u_s^{[m,j]}(x), \quad (24)$$

When using the above interpolation method to implement spectral colour calibration, x is a point in the printer colour space or the low dimension spectral space, and $\mathcal{F}(\bullet)$ is $\mathcal{F}_{print}(\bullet)$ or $\mathcal{F}^{-1}(\bullet)$, which is determined by doing forward or backward spectral colour calibration

5 Experiments

In experiments, the multi-ink printer used is HP designjet 130nr. This printer is a 6 ink printer, including C(cyan), M(magenta), Y(yellow), K(black), c(light cyan), m(light magenta). The spectral measurement device is GretagMacbeth SpectroScan Transmission spectrophotometer. The measured value is spectral reflectance. The spectral range is from 380nm to 730nm and the interval is 10nm. Thus the spectral reflectance is 36 dimension data.

When creating the forward LUT, the calibration sample set used is sampled in the CMYKcm colour space of HP 130nr printer. Each channel of the printer is divided into 6 levels, which is [0 0.2 0.4 0.6 0.8 1]. Then we combine the data of each channel to gain 46656 samples. The low dimension spectral space used is a 6 dimension spectral space LabPQR that is proposed by Mulsell Colour Science Laboratory. This space takes both chroma and spectral characteristics into account. The transformation between it and the spectral reflectance space is simple and the transformation precision is high. Its dimension is adequate. The algorithm about this space is described in [13] and [14]. When creating the backward LUT, we divide each dimension into 6 levels uniformly in this space. Thus, the backward calibration set also includes 46656 samples. Therefore, the LUT in this paper is 6 dimensions LUT.

The test sample set is 1300 samples sampled in the CMYKcm space randomly. By Printing these samples and then measuring their spectral reflectance, we use their colour separation data in the printer colour space and their corresponding spectral reflectance to validate the spectral colour calibration algorithm proposed this paper.

Table 1 shows the experiment result using multi-dimension LUT to implement forward colour calibration for the 1300 samples. Since Yule-Neilsen Spectral Neugebauer (YNSN) model is the most commonly used spectral printing calibration model, Table 1 also shows the calibration result using YNSN model in the same experiment environment. In Table 1, the standard chrome metric ΔE_{ab} of uniform colour space CIELAB is utilized for colour error evaluation. The root-mean-square error Equation E_{RMS} [15] is used for spectral error evaluation. Moreover, the total and mean consuming time using these two methods to execute forward colour calibration for the test sample set is also shown in the table.

From Table 1 we know the accuracy using LUT is much higher than using YNSN model. It is because LUT method uses the sample set measured actually. The sample set in itself embody the various nonlinear facts of the printing process. While the YNSN model applies mathematical method to simulate the printing process, it cannot reflect the nonlinear facts, such as mechanical and

optical dot gain, digital half-tone, and so on. Since the LUT method uses multi-linear interpolation and the YNSN model method uses nonlinear mathematical calculation, the consuming time of these two methods is almost equal. The mean consuming time just has difference on the one part in 10^5 .

The backward calibration sample set cannot achieve by printing and measuring the samples. Thus, the creation method of calibration sample set becomes one of the key factors affecting the calibration accuracy. Table 2 shows the experiment result of the method proposed in this paper by using sample points to determine the optimization parameters. Moreover, according to the requirement that the data in the destination space must be in [0,1] during the backward colour calibration, Table 2 also shows the result that the boundary of the optimization variable is [0,1] and the optimization initial is generated randomly in [0,1].

From Table 2 we know that the method using samples to determine the optimization initial and the boundary of the optimization variable can improve the accuracy of the creation of the backward LUT obviously. At the same time, since the search, range of the optimization variable is reduced and the initial is set as close to the objective as possible, it makes the speed of creating backward LUT increased greatly.

TABLE 1 Comparison of the forward calibration precision using the tow methods

Methods	Chroma error(ΔE_{ab} , D65, 2°observer)				Spectral error(E_{RMS})				Consuming time (second)
	Mean	SD	MIN	MAX	Mean	SD	MIN	MAX	
LUT	2.9524	1.6679	0.2029	11.2935	0.0137	0.0061	0.0029	0.0436	6.11 / 0.0047
YNSN model	4.8695	1.3940	0.4597	9.4172	0.0244	0.0068	0.0037	0.0459	6.11 / 0.0047

TABLE 2 Accuracy of creating backward LUT

Method of determining the optimization parameters	Chroma error(ΔE_{ab} , D65, 2°observer)				Spectral error(E_{RMS})			
	Mean	SD	MIN	MAX	Mean	SD	MIN	MAX
According to the samples	0.1477	0.5078	0	6.1242	0.0054	0.0028	0.0001	0.0245
Using [0,1] interval and random initial	0.2924	2.8606	0	56.1349	0.0057	0.0111	0.0009	0.2238

TABLE 3 Comparison of the backward calibration precision using the tow methods

Method	Chroma error(ΔE_{ab} , D65, 2°observer)				Spectral error(E_{RMS})				Consuming time (second)
	Mean	SD	MIN	MAX	Mean	SD	MIN	MAX	
LUT	2.1966	1.5810	0.8394	4.9732	0.0140	0.0103	0.0060	0.0352	6.11 / 0.0047
YNSN model	4.5213	1.5624	1.0189	8.4628	0.0237	0.0124	0.0063	0.0731	141.031 / 0.1058

Table 3 shows the experiment result of using backward LUT to implement backward colour calibration for the 1300 samples. The result of using YNSN model is also shown in Table 3. Moreover, the total and mean consuming time applying these two methods to execute backward colour calibration for the test sample set is also shown in the Table.

From Table 3 we know during the backward calibration the LUT method not only improves the backward calibration accuracy significantly but also

decreases the calibration time obviously compared with the YNSN model method. This is because the backward calibration only needs multi-linear interpolation in the backward LUT just as the forward calibration. Therefore, its consuming time is almost equal to the forward calibration. However, the YNSN model method utilizes the nonlinear optimization technology real-time to execute inverse transformation on the samples based on the forward model, therefore its consuming time is long.

6 Conclusions

A spectral colour calibration method for multi-ink printing based on LUT is proposed in this paper. The new method makes it possible that using LUT to achieve the spectral colour calibration by introducing a low dimension spectral space and transforming the spectral reflectance to this space. The method utilizes the multi-ink printer colour space and the low dimension space as the calibration space, and takes samples in these spaces to generate the calibration sample set, then uses the sample set to create the forward and backward LUT and finally applies the multi-linear interpolation in the LUT to implement forward and backward colour calibration. During creating the backward LUT, a nonlinear optimization technique is applied. An algorithm is designed to determine the optimization parameters based on the backward calibration

sample set. It improves the accuracy and efficiency of creating the backward LUT obviously. Compared with the printing model methods commonly used, the new method has notable improvement on calibration accuracy and time efficiency.

Acknowledgments

Authors gratefully thank the co-workers for concerning the research project and also acknowledge the support of Science Foundation of Shaanxi province in China (Grant No. 2012JM8044), Research Fund of Shaanxi Provincial Education Department (Grant No. 12JK0733) and Research Fund of Xi'an University of Posts & Telecommunications (Grant NO. 1504).

References

- [1] Sitnik R, Krzeslowski J and Mączkowski G 2012 *Optical Engineering* **51**(2) 021115-1
- [2] Berns R S, Chen T, Wyble D R, et al. 2012 *Proc. of IS&T*, Copenhagen Denmark 162-7
- [3] Bouzid M, Khalfallah A, Bouchot A 2013 *Proc. SPIE, Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues*, San Francisco California USA **8587** 85871J
- [4] Jakovels D, Kuzmina I, Berzina A 2013 *Journal of Biomedical Optics* **18**(12) 126019
- [5] Urban P, Rosen R M, Berns S R 2007 *Journal of Electronic Imaging* **16**(4) 043014-1-11
- [6] Urban P, Grigat R R 2006 *Color Research & Application* **31**(3) 229-38
- [7] Berns S R, Taplin A L, Liang Z T 2004 *Spectral color reproduction with six color output* US Patent No.6698860 (in Chinese)
- [8] Urban P, Rosen R M, Berns S R 2007 *Proc. of IS&T 15th color imaging conf.* Albuquerque New Mexico 178-83
- [9] Chen Y, Berns S R and Taplin A L 2004 *Journal of Imaging Science and Technology* **48**(6) 519-28
- [10] Mohammadi M, Nezamabadi M, Taplin A L 2004 *Technical Report Munsell Color Science Laboratory Rochester New York United States*
- [11] Ding E 2005 Xidian University Xi'an China (in Chinese)
- [12] Tang S 1991 *Colorimetry*. Beijing Beijing Institute of Technology Press (in Chinese)
- [13] Derhak W M and Rosen R M 2006 *Journal of Imaging Science and Technology* **50**(1) 53-63
- [14] Tsutsumi S, Rosen R M, Berns S R 2006 *Proc. of IS&T 14th Color Imaging conf.* Scottsdale Arizona USA 246-51
- [15] Imai H F, Rosen R M, Berns S R 2002 *Proc. of the First European Conference on Color Graphics, Imaging and Vision Poitiers France* 492-6

Authors	
	<p>Ying Wang, born in February, 1977, Xi'an, China</p> <p>Current position, grades: lecturer in Xi'an University of Posts & Telecommunications, China. University studies: Ph.D. School of Computer Science & Technology, Xidian University, China. Scientific interest: colour science and multispectral image input/output technology.</p>
	<p>Zhongmin Wang, born in 1967, Xi'an, China</p> <p>Current position, grades: professor and dean of School of Computer Science & Technology in Xi'an University of Posts & Telecommunications, China. University studies: Ph.D. Beijing Institute of Technology, China. Scientific interest: intelligent information processing, robot technology and computer vision.</p>
	<p>Sheping Zhai, born in June, 1971, Xi'an, China</p> <p>Current position, grades: associate professor and vice dean of School of Computer Science & Technology in Xi'an University of Posts & Telecommunications, China. University studies: Ph.D School of Electronics and Information Engineering, Xi'an Jiaotong University, China. Scientific interest: Intelligent information processing and Semantic Web and Service computing.</p>

A trust-based resource selection algorithm in Cloud Computing

Tong Qin¹, Xinran Liu^{2*}

¹Information Security Center, Beijing University of Posts and Telecommunications, 100876 Beijing, China

²National Computer Network Emergency Response Technical Team/Coordination Center of China (CNCERT/CC), 100029 Beijing, China

Received 1 May 2014, www.tsi.lv

Abstract

From the point of the safety in Cloud Computing and the virtual resource under the Cloud, here proposed a cognitive trust model of Cloud virtual resource, based on Bayesian, and the model supports the effective resource selection as a basis. Furthermore, on the basis of the description of process and question in virtual resource selection, here comes a Trust and Resource Scheduling oriented Cloud resource selection model which takes QoS, trust, resource scheduling and other indexes into account. After applying the improved CHC-TSSM Genetic Algorithm in the model, the simulation experiment confirms the feasibility and efficiency of the algorithm, which can resolve the trust and scheduling problem in resource selection effectively.

Keywords: resource selection, trust, Cloud Computing

1 Introduction

Cloud computing, Internet of things and big data considered as driven forces of the information innovation have become crucial to economic development and social stability [1]. With the emerging market opportunities, many businesses begin to take the advantage of Cloud computing for business upgrade. For example, Amazon develops its own data management centre and virtualizes its resources, providing customers with services as storage, computing, transmission and Internet data. Public Cloud, community Cloud, private Cloud and hybrid Cloud have taken shape [2] to provide services in the form of infrastructure service, platform service and software service. Cloud virtual resource is featured by large quantity and commerciality. It becomes an important issue to select resources that meet the users' demand from vast resources with similar functions [3]. Cloud computing is dynamic, distributed, changeable and in a large scale [4]. Effective resource selection process and method are of great value to meet the users' demand, enhance the selection satisfaction and optimize resources. Service computing has yielded fruitful results and is inspiring Cloud resource selection, which wins attraction from domestic and foreign researchers [5].

2 Related researches

Virtual resource selection under the Cloud mainly rests upon a resource discovery and selection model based on function matching. A new similarity algorithm is proposed based on resource description and relation reasoning. With the algorithm, the matching speed of function and demand and the query rate can be calculated [6]. Cloud resource

selection algorithm constructed on semantic reasoning largely improves the efficiency and efficacy of searching [7]. However, the quality of the Cloud resource is hard to measure and control. Users' increasing attention to QoS (Quality of Service) and individual preference also present challenges to it. Virtual resource evaluation and selection under the Cloud is constructed and fuzzy set theory is introduced to address the dynamic QoS [8, 9].

Menascé et al. [10] and OHSC [11] use heuristic algorithm to address the automatic selection of Cloud resources. D.A. Menascé proposes a restriction of user's QoS and uses genetic algorithm to address this problem [12]. Dillon takes into consideration of the semantic information of resources and QoS level and uses genetic algorithm to address the problem of the optimized resources portfolio [13]. Feng Dengguo and some others introduces trust evaluation into the Cloud computing and enhances the probability of success [14]. Hu Chunhua uses trust spanning tree to construct a trusted set of Cloud resources and trusted evolving mechanism, which addresses the problem of malicious resource node [15]. Xie Xiaolan uses Game Theory to construct a reward matrix for resource nodes. She also proposes a node trust evaluation model and incentive mechanism [16]. Du Ruizhong proposes a trust and user preference oriented Cloud resource selection model and analyses the influence of trust [17].

Based on previous researches, this paper proposes a trust model of Cloud resource based on Bayesian theory. This model combines direct trust and recommendation trust while taking into consideration time-effectiveness and relationship of trust. On the basis of resource node trust computing, virtual resource scheduling is given a place. This paper constructs a virtual Trust and Resource Scheduling oriented Resource Selection Model. After

* Corresponding author e-mail: lxr@cert.org.cn

applying the improved CHC-TSSM Genetic Algorithm in the model, the simulation experiment confirms the feasibility and efficiency of the algorithm, which can solve the trust and scheduling problem in resource selection effectively. Finally, this paper concludes and discusses space for further researches.

3 Trust model of cloud resource based on Bayesian theory

Resource trust refers to the ability of the target resource node to provide reliable services. It is evaluated by historical transaction record and through recommendation from other nodes [18]. Suppose any two Cloud resource nodes x and y have direct and indirect interactions with other nodes, so the probability of successful cooperation is θ . And the successful rate of direct trust degree is expressed by θ_{dt} . For indirect node z , if there is a direct interaction between x and z and y and z , then the successful rate of indirect interaction between x and y is expressed by recommendation trust degree θ_r .

$$f(\lambda_0 \cdot \theta_{dt} + (1 - \lambda_0) \cdot \theta_r), \lambda_0 \in (0, 1). \tag{1}$$

Under non-empty set, for all $\theta_{dt}, \theta_r \in S, \lambda \in (0, 1)$

$$f(\lambda \cdot \theta_{dt} + (1 - \lambda) \cdot \theta_r) \leq \lambda f(\theta_{dt}) + (1 - \lambda) f(\theta_r).$$

So $f(\cdot)$ is the function for aggregated trust degree with the feature of a convex function.

Therefore the aggregated trust function is $\hat{\theta} = \lambda \cdot \theta_{dt} + (1 - \lambda) \cdot \theta_r, \lambda \in (0, 1)$. If the node is prone to trust direct experience, then there is $\lambda > 0.5$.

3.1 ANALYSIS ON TRUST RELATIONSHIP

The relationship of two nodes falls into four categories according to direct trust and (or) recommendation interaction between nodes. When $dt=1$ or 0 , node x and y have (or don't have) direct interaction. When $rt=1$ or 0 , node x and y have indirect recommendation relationship. Four types of relationship can be expressed by $TR(dt, rt)$.

3.2 TIMELINESS ATTENUATING ATTRIBUTE OF TRUST

Trust of resource node dies out with time. Recent transaction does influence the trust. Time is measured by day, reflecting the change of trust degree. Out of convenience for calculation, suppose the interaction order of node is i , the successful interaction times and the failed interaction times are u_i and v_i respectively. The information formula that has considered time attenuating attribute is expressed as:

$$u(n) = \sum_{i=1}^n u_i \cdot \eta^{(n-i)}, v(n) = \sum_{i=1}^n v_i \cdot \eta^{(n-i)}, \tag{2}$$

$0 \leq \eta \leq 1$, $u(n)$ and $v(n)$ refer to successful times and failed times after n interactions. When $\eta = 1$, all interaction records are aggregated. When $\eta = 0$, only the latest service record is used for trust calculation. Recursive method is introduced to address the redundancy of historical records.

$$u(i) = u(i-1) \cdot \eta + u_i, v(i) = v(i-1) \cdot \eta + v_i. \tag{3}$$

3.3 DIRECT TRUST

Suppose x and y are two service nodes in the Cloud virtual resource system. Their interaction can be explained by binomial expression to describe success or failure. When two resource nodes conduct $(n+1)$ transaction and if u times are successful, v times are failure, and then at $(n+1)$ times, the probability of successful cooperation is $\hat{\theta}_{dt}$. The posterior probability of successful cooperation between resource nodes x and y is in accordance with Bayesian distribution. The density function is:

$$Beta(\theta|u, v) = \frac{\Gamma(u+v+2)}{\Gamma(u+1)\Gamma(v+1)}. \tag{4}$$

The direct trust degree of resource node is expressed as:

$$\hat{\theta}_{dt} = E(Beta(\theta|u+1, v+1)) = \frac{u+1}{u+v+2}. \tag{5}$$

At this time $0 < \theta < 1$ and $u, v > 0$.

Without effective evidence it is necessary to evaluate the direct trust degree by intervals. $(\hat{\theta}_{dt} - \varepsilon, \hat{\theta}_{dt} + \varepsilon)$ refers to the confidence interval of node $\hat{\theta}_{dt}$ and it is expressed as:

$$\begin{aligned} \gamma &= P(\hat{\theta}_{dt} - \varepsilon < \theta_{dt} < \hat{\theta}_{dt} + \varepsilon) = \frac{\int_{\hat{\theta}_{dt}-\varepsilon}^{\hat{\theta}_{dt}+\varepsilon} \theta^{u-1} (1-\theta)^{v-1} d\theta}{\int_0^1 \theta^{u-1} (1-\theta)^{v-1} d\theta} \tag{6} \\ &= \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} \int_{\hat{\theta}_{dt}-\varepsilon}^{\hat{\theta}_{dt}+\varepsilon} \theta^{u-1} (1-\theta)^{v-1} d\theta \end{aligned}$$

A balance needs to be addressed between confidence and the accuracy of the interval. When the interaction reaches a certain times, the two cannot be enhanced at the same time. Suppose the cap of the confidence is γ_0 , when the accuracy reaches an acceptable level-that is when $\gamma \geq \gamma_0$, resource trust can be evaluated under such

condition. Suppose there has been n_0 interactions, the relationship between γ_0 and ε is expressed as:

$$n_0 \geq -\frac{1}{2\varepsilon^2} \ln\left(\frac{1-\gamma_0}{2}\right). \quad (7)$$

3.4 RECOMMENDATION TRUST

Recommendation trust consists of several direct interactions among nodes. Recommendation nodes are selected through the calculation of trust degree of other nodes. Suppose resource node x and y , and y and z have direct transaction but are independent from each other. Each pair has n_1 and n_2 times of interaction, u_1 and u_2 times of successful cooperation and v_1 and v_2 times of failure service. Through the recommendation of node z , the recommendation trust of x to y is:

$$\hat{\theta}_n = E\left(\text{Beta}\left(\theta|u_1+u_2+1, v_1+v_2+1\right)\right) = \frac{u_1+u_2+1}{n_1+n_2+2}. \quad (8)$$

When there are several recommendation nodes, the comprehensive recommendation trust according to the above expression and the confidence level is expressed as:

$$\hat{\theta}_n = \frac{\sum_{\gamma \geq \gamma_0} u+1}{\sum_{\gamma \geq \gamma_0} (u+v)+2}. \quad (9)$$

Recommendation of trust of node y to node x equals to the ratio of actual interaction times to demand times:

$$w_{xy} = \begin{cases} \frac{n_{xy}}{n_0}, & \text{if } n_{xy} < n_0 \\ 1, & \text{otherwise} \end{cases}. \quad (10)$$

Given that the overall trust is influenced by active and passive feedback, and the value mapping of $\hat{\theta}_n \in [-1, 1]$, the expression is further adjusted to:

$$\hat{\theta}_n = \frac{\sum w \cdot (u-v)}{\sum w \cdot (u+v)+2}. \quad (11)$$

4 Resource selection algorithm based on trust

A trust-based resource selection method oriented Cloud virtual resource is proposed while considering important factors of resource selection and function matching. It is expected to maximize the Cloud computing performance, realize resource selection as well as the effective implementation of the portfolio.

4.1 DESCRIPTION OF QUESTION

To give a better description of our questions, we introduce some specific definitions and symbols in new resource selection method.

Definition 1. R (resource) refers to units that can implement tasks in the resource pool in cloud computing. It is expressed as:

R = (RID, Group ID, SA, DA, TR, Location).

Namely, resource identifier (RID), Group identifier (Group ID), Static Attribute (SA), Dynamic Attribute (DA), Trust Degree (TD) and Location.

Definition 2. Resource Cluster (RC) refers to a set of Cloud resource nodes that have same or similar functions. It is described as RC = (Group ID, FunSet). All Cloud resource clusters are aggregated to the Resource Cluster Set (RCS) and RCS = {RC₁, RC₂, ..., RC_i, ..., RC_n}.

Definition 3. RFlow (resource flow) is the process in which resource portfolio is motivated to work out. It is expressed as RFlow = (FlowID, FlowFunSet). All these processes form the Resource Flow Set (RFlowS).

Definition 4. RPool (resource pool) refers to a set of resource type and service mode provided by resource cluster or resource flow. It is expressed as:

RPool = (PoolID, PoolName, RCS, RFlowS, PoolOwner)

Service Set or Service Flow carry out semantic search in the resource pool and match the candidate resource nodes that meet the demand of service node to form a candidate service set. Then conduct resource selection on the basis of QoS and trust level and calculate the efficiency and optimal target according to resource scheduling to match every service node with the optimal resource. Finally, the optimal resource portfolio is dawn at sight according to multi-objective optimized model.

4.2 MODEL FACTORS

Factors of resource selection showed in table 1 involves with service quality, trust degree and resource scheduling (location, waiting length, utilization efficiency). After a deep analysis with relevant index calculations, this paper paves the way for constructing a multi-objective programming model based on trust and service scheduling.

(1) Trust degree

Trust degree is crucial to virtual resource selection. It lifts users' satisfaction on resource matching and reduces failures during the implementation of resource services.

Direct trust $\hat{\theta}_{dt}$ and indirect trust $\hat{\theta}_n$ both have a place.

This paper analyses the time-effectiveness of information interaction. Users have preference for current trust level. The overall trust is aggregated as:

$$\hat{\theta} = \lambda \cdot \theta_{dt} + (1-\lambda) \cdot \theta_n, \lambda \in (0, 1).$$

(2) QoS indexes

Users under the Cloud computing is paying more and more attention to service quality. QoS indexes vary from each other with different types and levels of service. If

$Q = (Q^1, Q^2, \dots, Q^K)$ is used to show the preference order of users to indexes, then Q^1 wins the most preference and Q^2 follows it. QoS indexes are available from the resource description and at the registration center.

(3) Indexes of service resource scheduling.

Virtual resource scheduling means to schedule the resources in the pool in an optimized way that plays the performance of the system to the most and yields the most profit. Utilization efficiency R1, load balance R2 and location R3 are given a place. Supervise the parameters of the service state (SS_{ta}) where resource $S_{i,j}$ locates. When the server is open, $X_{i,j}=1$, otherwise $X_{i,j}=0$. R1 is calculated as:

$$R1 = \sum_{i=1}^I \sum_{j=1}^{J_i} (X_{i,j} Y_{i,j} SS_{ta_{ij}}), \quad (12)$$

$$X_{i,j} \in \{0,1\}, i = 1, 2, \dots, I, j = 2, \dots, J_i.$$

TABLE 1 Relevant symbols and description

Symbol	Description
i	the number of resource node
j	Service ID in ResourceSet
$S_{i,j}$	resource node j of resource combinations i
$Y_{i,j}$	$Y_{i,j} \in \{0,1\}$, when resource j is selected by service i , $Y_{i,j}=1$, otherwise $Y_{i,j}=0$
$t_{i,j}$	execution time of resource $S_{i,j}$ instantiation
$wt_{i,j}$	For current users, real waiting time of the selected resource $S_{i,j}$ is decided by waiting time and execution time
Q^K	QoS of selected resource combinations

ServerUtilizRation (SUR) refers to the utilization efficient in virtual server. Resources in the server that has the least utilization efficiency are chosen. The load balance R2 is calculated as:

$$R2 = \sum_{i=1}^I \sum_{j=1}^{J_i} (X_{i,j} Y_{i,j} (1 - SS_{ta_{ij}}) \times SUR_{ij}). \quad (13)$$

In the process of resource selection, those resources in the server that is the closest to users are given priority. $location_{i,j}$ refers to the location of resources. $Uerlocation$ refers to the location of users. R3 is calculated as:

$$R3 = \begin{cases} Distf \left(\sum_{j=1}^{J_i} Y_{i,j} location_{i,j}, Uerlocation \right), I = 1 \\ \frac{1}{I} \sum_{i=1}^I Distf \left(\sum_{j=1}^{J_i} Y_{i,j} location_{i,j}, Uerlocation \right) \\ + \sum_{i=1}^{I-1} Z_{i,j} Distf \left(\sum_{j=1}^{J_i} Y_{i,j} location_{i,j}, \sum_{j=1}^{J_{i+1}} Y_{i+1,j} location_{i+1,j} \right), I > 1 \end{cases} \quad (14)$$

(4) Calculating rules.

Based on previous researches of service flow, resources flow falls into four categories, namely, series type, parallel type, conditional type and circular type. Each

has a different way of calculating QoS and the overall trust. Luckily, there have been successful researches on index calculation. This paper draws merits from these researches to aggregate the indexes. Indexes need to be normalized as they have different types of properties. Normalization serves to the calculation of the multi-objective programming model.

4.3 MULTI-OBJECTIVE PROGRAMMING MODEL OF RESOURCE SELECTION

This paper chooses flow cost, responding time, trust degree and resource scheduling as four objectives of the model together with the ideal plan of users.

(1) The priority of objectives is as follows. First is the flow cost, second is the responding time, third is the trust degree and fourth is the resource scheduling.

(2) The model is mainly restricted by QoS and trust degree.

Based on above objectives and restrictions, TSSM based on trust and resource scheduling is constructed and shown below:

$$lex, \min \{d_c^+ V0, d_t^+ V0, d_r^- V0, d_c^- V0\}, \quad (15)$$

$$ObjFunc = \alpha_1 (d_c^+ V0) + \alpha_2 (d_t^+ V0) + \alpha_3 (d_r^- V0) + \alpha_4 (P1 \times R1V0 + P1 \times R1V0 + P2 \times R2V0 + P3 \times R3V0), \quad (16)$$

$$s.t. d_c^+ = C - C_0, d_t^+ = T - T_0, d_r^- = TR - TR_0, \quad (17)$$

$$\sum_{j=1}^{J_i} c_{i,j} Y_{i,j} \leq C_i, i = 1, 2, 3, \dots, I, \sum_{j=1}^{J_i} t_{i,j} Y_{i,j} \leq T_i, i = 1, 2, 3, \dots, I, \quad (18)$$

$$\sum_{j=1}^{J_i} tr_{i,j} Y_{i,j} \geq TR_i, i = 1, 2, 3, \dots, I, \sum_{j=1}^{J_i} sh_{i,j} Y_{i,j} \leq SH_i, i = 1, 2, 3, \dots, I \quad (19)$$

$$\sum_{j=1}^{J_i} Y_{i,j} = 1, i = 1, 2, 3, \dots, I. \quad (20)$$

Weight of objectives is determined by their priorities. Equation (17) refers to the objective restriction. Equations (18) - (19) are QoS restriction, trust restriction and resource scheduling restriction. Equation (20) is decision variable restriction, which indicates that every node in the process binds with a certain service resource.

4.4 VIRTUAL RESOURCE SELECTION ALGORITHM BASED ON CHC-TSSM

Resource combination is a typical NP problem. The increase of candidate resources means more accessible solutions, which makes the problem even more complicated. The improved CHC-TSSM algorithm makes it possible to obtain the optimal trust service portfolio.

(1) Gene code.

If the resource portfolio is paralleled to chromosomes, then the integer code with fixed length is expressed as:

$$Y_i = [-1, y_1, y_2, \dots, y_1, -1] y_i, \tag{21}$$

y_i accords with number j service ws_{ij} in the candidate service set of the resource portfolio. If the resource doesn't show in the portfolio, the value shall be given 0.

(2) Group initialization.

The initialized group is parallel to the initial virtual resource service set. From the beginning to the end of the resource portfolio, the chromosomes are of the same length. If it is a small-scale group, the limited searching space of the algorithm may result in convergence and knowledge optimal settings.

(3) Adaptive function

Utilization efficiency R1 is expressed as bellow.

$$R1 = \sum_{i=1}^I \sum_{j=1}^{J_i} (X_{i,j} Y_{i,j} SSta_{ij}) \tag{22}$$

and $X_{i,j} \in \{0,1\}, i = 1, 2, \dots, I, j = 2, \dots, J_i$.

According to the objective of virtual resource selection, load balancing R2 is calculated by equation (23).

$$R2 = \sum_{i=1}^I \sum_{j=1}^{J_i} (X_{i,j} Y_{i,j} (1 - SSta_{ij}) \times SUR_{ij}) \tag{23}$$

The adaptive function limited by distance (R3) is expressed as follow:

$$R3 = \begin{cases} Distf \left(\sum_{j=1}^{J_i} Y_{i,j} location_{i,j}, Uerlocation \right), I = 1 \\ \frac{1}{I} \sum_{i=1}^I \left(Distf \left(\sum_{j=1}^{J_i} Y_{i,j} location_{i,j}, Uerlocation \right) + \sum_{i=1}^{I-1} Z_{i,i} Distf \left(\sum_{j=1}^{J_i} Y_{i,j} location_{i,j}, \sum_{j=1}^{J_i} Y_{i,j} location_{i+1,j} \right) \right), I > 1 \end{cases} \tag{24}$$

(4) Genetic operation

Genetic operation involves in selection, crossover, mutation and others. The cross-generational elitist selection strategy is introduced for selection. Calculation of adaption degree is done on the basis of combining the group of previous generation and individual group produced through cross connection. Finally N1 individuals with relatively large adaptation degree will stand out as the group of next generation. The crossover is improved in this paper in the way that if the binary digit number of two different parent generations is M, then individuals at M/2 will be selected. Individuals are picked up randomly. Conduct the multi point cross operation according to cross probability P_c and form N2 new individuals. When the group evolves to T/2 generation, N3 individuals will be selected according to the adaption degree $F(Y_i)$. Select chromosomes according to the mutation probability and empower them with value randomly.

(5) Selection of controlling parameters.

Controlling parameters play an important role to the performance and convergence of the improved CHC genetic algorithm. They include group scale, crossover probability P_c , mutation probability P_m and the maximum evolutionary generations T . Generally speaking, the group scale is from 10 to 200, and generations T from 100 to 1000. P_c is proper from 0.4 to 0.99. If it is too big, the convergence will not be satisfying. The mutation probability P_m is better to take smaller value from 0.1 to 0.6.

(6) Group evolution.

After selection, crossover and mutation, there are (N1+N2+N3) new individuals or chromosomes. Then calculate the adaptive value $F(Y_i)$ of each individual by adaptive function, and select the best N1 as group of the next generation. As a result, some service portfolio with low trust and inefficient use of resources will be left out.

(7) The optimal plan

After repeating the process, the group will evolve to generation T , or the adaptive degree of the optimal plan will become less than the minimum value. The improved CHC genetic algorithm is adapted to objectives and restrictions. Finally, the best individuals and chromosomes will obtain the best virtual service plan after decoding.

5 Experimental evaluations

This paper proposes a trust-based resource selection model. It takes into consideration the optimal resource portfolio by QoS, trust degree of virtual resource and resource scheduling. The simulation experiment is conducted to ensure of the accuracy and the efficiency.

5.1 EXPERIMENTAL ENVIRONMENT AND PARAMETERS SETTING

Experience environment is designed as CPU Intel Core 4.0GHz, memory DDRII4G, operating system Windows7.0 Pro, Java programming language, software environment jdk.5.0-08.

Existing data about virtual resource and Cloud computing is far from enough and is limited by law. Therefore, this paper proposes to produce experiment data randomly according to relevant researches and expertise. Parameters: the maximum generation $G=1000$, group scale 20-100, penetration coefficient is set 3, the crossover probability $P_c=0.6$, the mutation probability $P_m=0.5$, decrement is 0.5. The following two experiments have same parameters.

5.2 RESULT ANALYSIS

All experiments based on above environment and parameters have gone through 50 times of weight average. The trust model and the CHC-TSSM algorithm have been evaluated in terms of effectiveness.

(1) The effectiveness of the trust model

Here designed four types of resource trust relationships together with calculation method to prove the effectiveness of the cognitive trust model. Suppose the direct trust and the indirect trust are both 0.5 and the best sample capacity is 200. Scheduling resources at 6 time (by day), each time (day) represents the trust degree and the evolutionary process of the resource node. Compare the trust weight with trust degree and time-effectiveness. The trust evaluation result is shown in Figure 1:

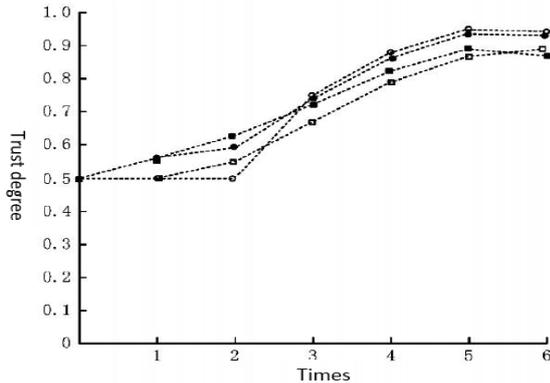


FIGURE 1 Evaluation of trust degree

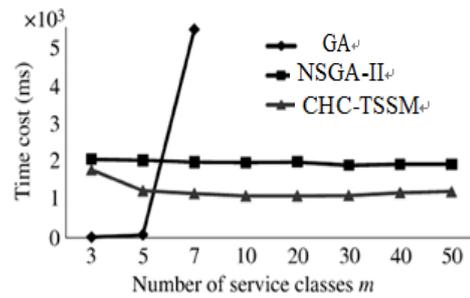
At the beginning, the trust degree of node x to y is 0.5. Only consider direct trust to address, $\lambda_1=1, \lambda_2=0$. At time 2, the interaction fails to reach the optimal 200 times, but the trust degree is still 0.5. At time 5 and 6, the trust degree becomes stable. And the confidence interval is [0.90, 0.95]. Only consider z recommendation trust, $\lambda_2=1, \lambda_1=0$. At time 4, the interaction of z to y reaches the optimal with the capability of recommendation trust. At time 5 and 6, trust degree becomes stable with the recommendation trust of 0.90 and the confidence interval of [0.8255, 0.9325]. The latter two circumstances are based on the previous two with similar calculation. According to Figure 1, the direct trust and the recommendation trust decrease with the time, which proves the hypothesis of time-effectiveness of trust.

(2) Convergence of CHC-TSSM algorithm

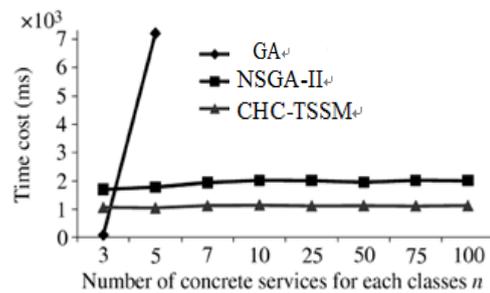
Factors that influence the scale of solutions to the resource portfolio include the number of service class m , the number of concrete service for each cluster and trust degree. Three parameters and the time of CHC-TSSM algorithm are applied to the test. CHC-TSSM algorithm is made comparison with NSGA-II and GA algorithm. The resource cluster m ranges from 5~60 and the candidate resource service of each cluster ranges from 4~100. The test results are shown in Figure 2 (a) and 4 (b).

References

[1] Armbrust M, Fox A, Griffith R 2010 *Communications of the ACM* 53(4) 50-8
 [2] Buyya R, Yeo C.S, Venugopal S 2009 *Future Generation Computer Systems* 25(6) 599-616
 [3] Zissis D, Lekkas D 2012 *Future Generation Computer Systems* 28(3) 583-92
 [4] Kim W 2009 *Journal of Object Technology* 8(1) 65-72



2a)



2b)

FIGURE 2 The relationship between the algorithm time and the number of resource classes well as the number of candidate resource node

6 Conclusions

The trust-based resource selection model considers indexes such as the trust degree of the resource node, QoS and resource scheduling. It offers different index calculation under various service flows constructs a multi-objective programming model and works out relevant restrictions. This paper also proposes an improved CHC-TSSM algorithm with detailed descriptions to address the NP-hard problem. The simulation experiment confirms the feasibility and efficiency of the algorithm featured by an increase of successful rate of resource selection. But there is still space to improve the trust evaluation model or propose new models that may address the increase of QoS indexes or users' preferences.

Acknowledgements

This paper is supported by the National Basis Research Program of China (No.2011CB302605), and the National Natural Science Foundation of China (No. 61121061).

- [8] Yang FangChun, Sen Su, Zhen Li 2008 *Science in China Series F: Information Sciences* **51**(11) 1822-40 (in Chinese)
- [9] Menascé D A, Casalicchio E, Dubey V 2008 *Proceedings of the 7th International Workshop on Software and Performance* ACM 13-24
- [10] Menascé D A, Casalicchio E, Dubey V 2010 *Performance Evaluation* **67**(8) 659-75
- [11] OHSC, Lee D W, Kumara R T S 2008 *IEEE Transactions on Services Computing* (1) 15-32
- [12] Dillon T., Chen W., Chang E 2010 *Proceedings of Advanced Information Networking and Applications (AINA)* 27-33
- [13] Feng D, Zhang M, Zhang Y 2011 *Journal of Software* **22**(1) 71-83 (In Chinese)
- [14] Hu C, Chen X, Wu M 2012 *Science China (Information Science)* **4**(23) 314-32 (in Chinese)
- [15] Xie X, Liu L, Zhao P 2012 *Journal of Electronics & Information Technology* **34**(4) 812-7 (in Chinese)
- [16] Du R, Tian J, Zhang H 2013 *Journal of Zhejiang University (Engineering Science)* **47**(1) 53-61 (In Chinese)
- [17] Jiang L, Xu J, Zhang H 2012 *Expert Systems with applications* **39**(3) 3772-82 (in Chinese)
- [18] Fudzee Md M F, Abawajy J H 2011 *Future Generation Computer Systems* **27**(3) 256-64 (in Chinese)

Authors



Tong Qin, born in February, 1984, Dandong, Liaoning, China

Current position, grades: Ph.D. candidate in Information security center at Beijing University of Posts and Telecommunications.
University studies: M.Sc. degree in software engineering from Beijing University of Posts and Telecommunications (BUPT) in 2007.
Scientific interest: network security and distributed computing.



Xinran Liu, born in February, 1971, Jixi, Heilongjiang, China

Current position, grades: researcher in National computer network emergency response technical team/coordination centre of China (CNCERT/CC), a doctoral supervisor in Information security centre at Beijing University of Posts and Telecommunications.
University studies: Ph.D. in computer architecture from Harbin Institute of Technology in 1998.
Scientific interest: distributed computing and information security.

Optimal routing strategy on weighted networks

Binghua Cheng¹, Fei Shao^{1, 2*}

¹School of Information Technology, Jinling Institute of Technology, Nanjing 211169, China

²Jiangsu Information Analysis Engineering Laboratory, Nanjing 211169, China

Received 1 June 2014, www.tsi.lv

Abstract

It is of great importance to improve the transfer capacity of the weighted networks. In this paper, the traffic dynamics on weighted networks is investigated based on global information. It is shown by simulations that the weighted network transfer capacity depends strongly on the tuneable parameter in three different node delivery capability schemes: constant, proportional to node degree and proportional to node strength. Furthermore simulations on both computer-generated and real world networks show that different tuneable parameter is suitable for different node delivery capability scheme.

Keywords: weighted network, BBV network, routing strategy, transfer capacity

1 Introduction

Due to the constantly growing significance of large communication networks such as the World-Wide-Web and the Internet, the study of network transfer capacity is becoming increasingly important in the past few years. These real world networks can be properly described as complex networks while nodes representing individuals and edges representing the interactions among them. Consequently, the fast-developing theory of complex networks throw light on how to improve the network transfer capacity since the seminal work on the small-world phenomenon [1] and the scale free property [2]. The previous studies have been primarily focused on unweight networks where edges between nodes are either present or not, represented as binary states. However, lots of real world networks present different strength of the edges between nodes such as the mobile communication networks [3], the scientific collaboration networks [4], the world-wide airport network [5] and the Internet [6]. Networks are specified not only by the topology but also by the weight of edges.

Recently, finding optimal routing strategies to control traffic congestion and improve transfer capacity on a large growing communication network is gaining increasing concern. The shortest path routing strategy [7], where the packets are forwarded following the shortest path, is most commonly adapted because packets may reach their destinations quicker than following other paths. However, if all packets follow the shortest path, it will easily lead to the overload of the high-degree nodes and result in traffic congestion. An effective routing strategy [8] is proposed to redistribute traffic load in central nodes to other non-central nodes which can enhance the network capability in processing traffic more than 10 times. Based on the idea, we have put forward a novel routing strategy for BBV

weighted network in which packets are transferred through the path based on the weight of edges with a tuneable parameter α [9]. And different optimal tuneable parameter is suitable to maximize the overall network transfer capacity for different node delivery capability schemes. How to the node characteristics to enhance the transfer capacity deserves special attention.

This paper is organized as follows. In section 2 we describe the BBV weighted network model, the traffic dynamics model and our routing strategy, followed by the experimental evaluations on computer generated networks and real world network in section 3. The conclusions are given in section 4.

2 Models

In those models presented to describe the real world networks [10-12], the BBV model [11] is most widely used. The BBV weighted network can be completely described by an adjacency matrix W , whose elements w_{ij} denote the weight of the edge between node i and j . The definition of the BBV weighted network [11] is based on two coupled mechanisms:

- (i) Growth. Starting from an initial small number of N_0 nodes connected by edges with assigned weight w_0 , a new node is added at every time step. The new added node is connected to m different previously existing nodes with equal weight w_0 for every edge and chooses preferentially nodes with large strength according to the probability $\prod_{n \rightarrow i} = s_i / \sum_l s_l$, where s_i is the node strength described as $s_i = \sum_j w_{ij}$.
- (ii) Weight dynamics. The weight of each new add edge is initially set to a given value w_0 which is often set

* Corresponding author e-mail: shaofei@jit.edu.cn

to 1 for simplicity. But the adding of edge connecting to node i will result in increasing the weight of the other edges linked to node i which is proportional to the edge weights. If the total increase is δ (we will focus on the simplest form: $\delta_i=\delta$), we can get

$$w_{ij} = w_{ij} + \Delta w_{ij} = w_{ij} + \delta \frac{w_{ij}}{s_i} . \tag{1}$$

This will yield the strength increase of node i as:

$$s_i = s_i + \delta + w_0 . \tag{2}$$

The degree distribution of BBV network $P(k) \propto k^{-\gamma_k}$ and the strength distribution $P(s) \propto s^{-\gamma_s}$ yield scale-free properties with the same exponent [11]:

$$\gamma_k = \gamma_s = \frac{4\delta + 3}{2\delta + 1} . \tag{3}$$

The traffic model can be described as follows:

- 1) All nodes can create packets with addresses of destination, receive packets from other nodes, and forward the packets to their destinations.
- 2) At each time step, R packets are created with randomly chosen sources and destinations. Once a packet is created, it is placed at the end of the queue if the node has packets waiting to be forwarded.
- 3) At the same time, the first C_i packets at the head of the queue of each node are forwarded one step to their destinations. If a node has less than C_i packets in its queue, all packets in the queue will be forwarded one step.
- 4) Upon reaching its destination, the packet is removed from the network.

In our model, three node delivery capability schemes are discussed: (i) each node has the same packet delivery capability (we set $C_i=1$ for simplicity); (ii) the node delivery capacity is considered to be proportional to the node degree k_i ($C_i=k_i/\langle k \rangle$); (iii) the node delivery capacity is considered to be proportional to the node strength s_i ($C_i=s_i/\langle s \rangle$). The total node delivery capability of the whole network is equal to the node number n in these three schemes.

Denote $P_{i \rightarrow j}$ as the path between node i and j , which pass through the nodes sequence $x_0(=i), x_1, x_2, \dots, x_{n-1}, x_n(=j)$, we define

$$F(P_{i \rightarrow j}, \alpha) = \sum_{i=1}^n x_i^\alpha . \tag{4}$$

In our routing strategy, we specify the routing path between i and j as the one makes $F(P_{i \rightarrow j}, \alpha)$ minimum under a given tuneable parameter α .

We discussed the situation where x_i is the weight of the edge in the path w_{ij} (WEI stands for this strategy) [9]. In this paper, we explore the other two cases: x_i is the degree of node i ($x_i=d_i$, DEG stands for this strategy), and x_i is the strength of node i ($x_i=s_i$, STR stands for this strategy).

The transfer capacity of networks is most commonly measured by a critical generating rate R_c in those routing strategies [7-9]. A continuous phase transition from free-flow state to congested state occurs at the critical value R_c . In the former state, the numbers of created and delivered packets are balanced, leading to a steady state. The number of accumulated packets increases with time due to the limited delivery capacity or finite queue length of each node in the latter state. We are interested in the critical value R_c which can best reflect the maximum transfer capacity of a network handling its traffic.

The betweenness b_i is often introduced to estimate the possible packet passing through a node i under the given routing strategy, which is defined as

$$b_i = \sum_{s,t} \frac{\sigma(s,i,t)}{\sigma(s,t)} , \tag{5}$$

where $\sigma(s,i,t)$ is the number of paths under the given routing strategy between nodes s and t that pass through node i and $\sigma(s,t)$ is the total number of paths under the given routing strategy between nodes s and t and the sum is over all pairs s, t of all distinct nodes.

A created packet will pass through the node i with the probability $b_i / \sum_{j=1}^n b_j$. Thus, the average number of packets that the node i receives at each time step is $Rb_i / (n(n-1))$. Congestion occurs when the number of incoming packets is larger than the outgoing packets, that is $Rb_i / (n(n-1)) \geq C_i$. So the critical packet generating rate R_c is:

$$R_c = \min(C_i n(n-1) / b_i) . \tag{6}$$

3 Simulations and analysis

To get the optimal value of tuneable parameter α , we obtain the critical packet generating rate R_c versus different parameter α in a BBV network with $n=100$, $\delta=5$, $m=5$ and $\omega_0=1$ with three different routing strategies. The result of the WEI routing strategy is shown in Figure 1, the DEG routing strategy and the STR routing strategy are presented in Figure 2 and 3 correspondingly.

(For every network, 10 instances are generated and for each instance, we run 10 simulations. The results are the average over all the simulations.)

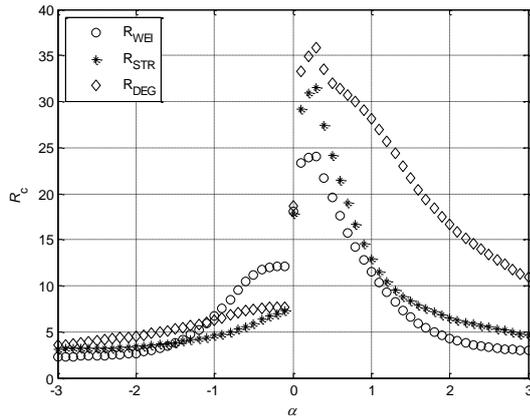


FIGURE 1 R_c VS α . BBV network with $n=100, \delta=5, m=5, \omega_0=1, C_i=1$

Figure 1 exhibits the relationship of the critical packet generating rate and the tuneable parameter where each node has the same packet delivery capability. In all three routing strategies, the critical packet generating rate R_c varies with the tuneable parameter α and R_c reaches the peak when α is 0.3. In this case, when α is 0, it is the same as the traditional shortest path routing strategy [7]. Three routing strategies have almost the same critical packet generating rate at $\alpha=0$. The maximum transfer capacity of the DEG routing strategy ($\alpha=0.3$) is 92.2% greater than the traditional shortest path routing strategy ($\alpha=0$) while the STR routing strategy and the WEI routing strategy is 76.7% and 33.3% correspondingly. And the maximum transfer capacity of the DEG routing strategy is better than the other two strategies (13.9 % higher than STR and 49.1% higher than WEI). The figure demonstrates that when each node has the same packet delivery capability, the network transfer capacity reaches the maximum when packets are forwarded through the path whose sum of the 0.3 power of node degree is the minimum.

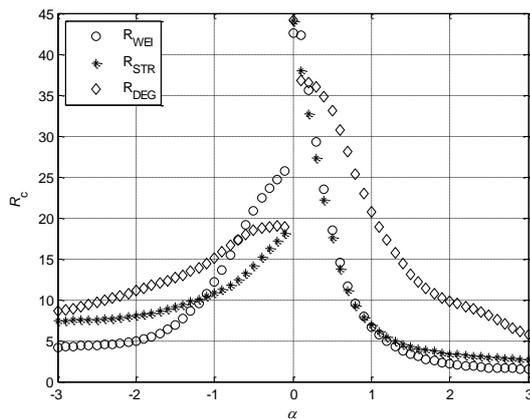


FIGURE 2 R_c VS α . BBV network with $n=100, \delta=5, m=5, \omega_0=1, C_i=k_i/<k>$

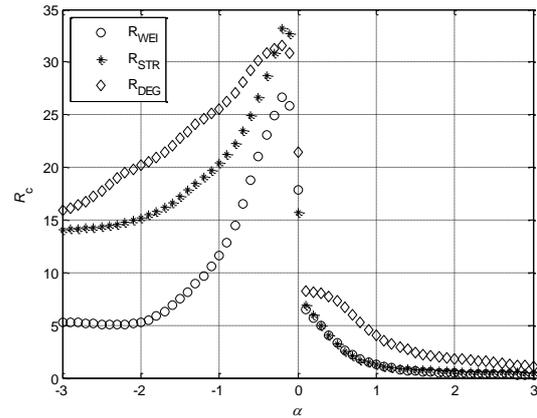


FIGURE 3 R_c VS α . BBV network with $n=100, \delta=5, m=5, \omega_0=1, C_i=s_i/<s>$

The critical packet generating rate versus the tuneable parameter where the node delivery capacity is considered to be proportional to the node degree or the node strength is presented in Figures 2 and 3. The critical packet generating rate R_c reaches the peak when α is 0 while $C_i=k_i/<k>$ and α is -0.2 while $C_i=s_i/<s>$. And while $C_i=k_i/<k>$, all three routing strategies get almost the same maximum transfer capacity at $\alpha=0$. The STR routing strategy works better than the other two routing strategies at $\alpha=-0.2$ when $C_i=s_i/<s>$.

To discuss the implication of the weight on critical packet generating rate, we set $\delta=50$ to get the corresponding simulation results in Figure 4.

Comparing Figures 4(a-c) with Figures 1-3, we can obtain that the critical packet generating rate R_c also reaches the peak when α is 0.3, 0, and -0.2 according to different the node delivery capacity. (We also test the situation where δ is 0.5 to find that the simulation results are followed similar patterns.)

As stated above, in those computer generated BBV networks, the critical packet generating rate R_c varies with the tuneable parameter α in all three routing strategies. In the scheme that each node has the same packet delivery capability, the network achieve maximum transfer capacity with the DEG routing strategy when the tuneable parameter α is 0.3. In the scheme that the node delivery capacity is considered to be proportional to the node degree, the network achieve maximum transfer capacity with all three routing strategies when the tuneable parameter α is 0. And in the scheme that the node delivery capacity is considered to be proportional to the node strength, the network achieve maximum transfer capacity with the STR routing strategy when the tuneable parameter α is -0.2.

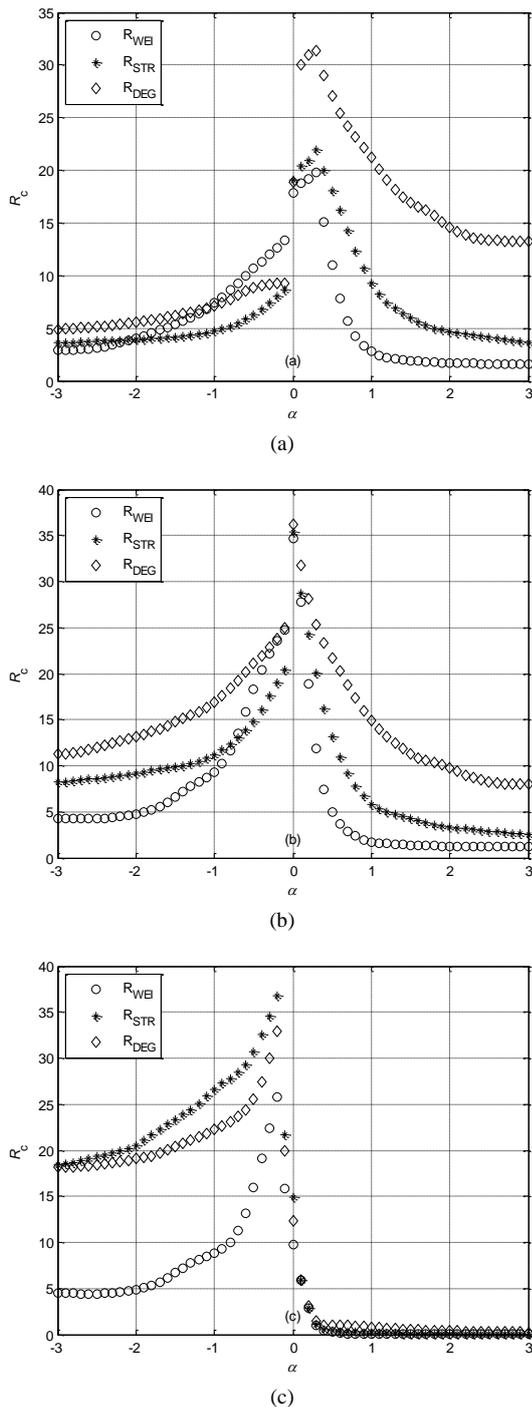


FIGURE 4 R_c VS α . BBV network with $n=100$, $\delta=50$, $m=5$, $\omega_0=1$ (a) $C_i=1$ (b) $C_i=k_i/\langle k \rangle$ (c) $C_i=s_i/\langle s \rangle$

Then we choose the scientific collaboration network [13], which has a giant component of 5835 nodes to test our routing strategy on real world network. Simulation results are shown in Table 1.

As Table 1 shown, the critical packet generating rates reach the peak with the same tuneable parameter as we proposed. It means that our routing strategy is also effective in real world network.

TABLE 1 the critical packet generating rates of the scientific collaboration network

	$C_i=1$			$C_i=k_i/\langle k \rangle$			$C_i=s_i/\langle s \rangle$		
	WEI	DEG	STR	WEI	DEG	STR	WEI	DEG	STR
$\alpha=-0.2$	2.36	2.10	2.10	5.00	4.99	5.01	7.34	10.06	10.06
$\alpha=0$	2.47	2.47	2.47	7.15	7.14	7.16	7.01	6.99	7.01
$\alpha=0.3$	2.49	2.87	2.83	4.15	3.98	3.99	5.81	5.58	5.59

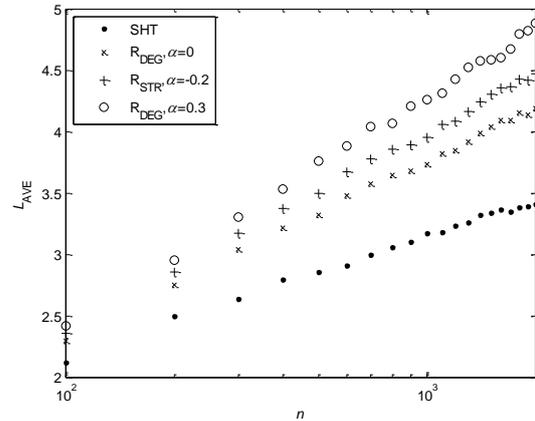


FIGURE 5 L_{AVE} VS n . BBV network with $\delta=4$, $m=4$ and $\omega_0=1$

The average weighted average length [14] L_{AVE} versus the node number n is reported in Figure 5. Although the weighted average length of DEG scheme and the STR scheme are higher than that of the traditional shortest path [7], the small-world phenomenon, i.e. $L_{AVE} \propto \ln n$, is still maintained. The transfer capacity of weighted network is considerably enhanced at the cost of increasing the average weighted average length. Such a sacrifice may be worthwhile when a system requires large transfer capacity.

4 Conclusion

This paper has proposed routing strategies to enhance the transfer capacity of the BBV weighted networks. The transfer capacity varies with the tuneable parameter and reaches the peak at different tuneable parameter according to three different kinds of node delivery capacity. In the scheme that each node has the same packet delivery capability, the DEG routing strategy achieve maximum transfer capacity when the tuneable parameter α is 0.3. When the node delivery capacity is considered to be proportional to the node degree, all three routing strategies achieve maximum transfer capacity when α is 0. And when the node delivery capacity is considered to be proportional to the node strength, the STR routing strategy achieve maximum transfer capacity when α is -0.2. Meanwhile, the small-world phenomenon of the average weighted average length is still maintained. At last, the test on the scientific collaboration network proves that our routing strategies work well in real world network.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61373136), the Natural Science Foundation of Jiangsu Province, China (Grant No. BK2012082), the Research Foundation of

Jinling Institute of Technology (Grant No. JIT-B-201406) and sponsored by Qing Lan Project. The author also gratefully acknowledges the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

- [1] Watts D J, Strogatz S H 1998 *Nature* **393**(6684) 440-2
- [2] Barabási A L, Albert R 1999 *Science* **286**(5439) 509-12
- [3] Onnela J P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A L 2007 *PNAS* **104**(18) 7332-6
- [4] Newman M E J 2001 *Phys. Rev. E* **64**(1) 016132
- [5] Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A 2004 *PNAS* **101**(11) 3747-52
- [6] Pastor-Satorras R and Vespignani A 2007 *Evolution and structure of the Internet: A statistical physics approach Cambridge: Cambridge University Press*
- [7] Zhou T 2008 *Physica A: Statistical Mechanics and its Applications* **387**(12) 3025-32
- [8] Yan G, Zhou T, Hu B, Fu Z Q, Wang B H 2006 *Phys. Rev. E* **73**(4) 046108
- [9] Shao F 2013 *The Scientific World Journal* 378083
- [10] Yook S H, Jeong H, Barabási A L, Tu Y 2001 *Phys. Rev. Lett.* **86**(25) 5835-8
- [11] Barrat A, Barthélemy M, Vespignani A 2004 *Phys. Rev. Lett.* **92**(22) 228701
- [12] Wang W X, Wang B H, Hu B, Yan G, Ou Q 2005 *Phys. Rev. Lett.* **94**(18) 188702
- [13] Newman M E J 2001 *PNAS* **98**(2) 404-9
- [14] Opsahl T, Agneessens F, Skvoretz J 2010 *Social Networks* **32**(3) 245-51

Authors	
	<p>Binghua Cheng, born in December, 1977, Jiangsu China</p> <p>Current position, grades: Master Degree student in Hohai University Since 2009. University studies: Bachelor Degree from Huaihai Institute of Technology in 2000. Scientific interest: information security, complex dynamical networks. Publications: 3 papers.</p>
	<p>Fei Shao, in December, 1978, Jiangsu China</p> <p>Current position, grades: Associate professor at the School of Information Technology, Jinling Institute of Technology and working for Jiangsu Information Analysis Engineering Laboratory since 2010. University studies: Ph. D Degree in the School of Computer Science & Technology at Nanjing University of Posts & Telecommunications. Scientific interest: information security, complex dynamical networks. Publications: 20 papers, 4 of them are indexed by SCI, 10 of them are indexed by EI.</p>

The power supply design software model for mine based on VBA technology

Baoquan Jin^{*}, Xiaohui Hao, Hongjuan Zhang, Yan Gao

Key Lab of Advanced Transducers and Intelligent Control System, Ministry of Education, Taiyuan University of Technology, Taiyuan, China

Received 12 June 2014, www.tsi.lv

Abstract

Coal mine power supply system is increasingly complex, resulting in difficulties for power computing and tuning. A software design program is proposed by using VBA development tools embedded in AutoCAD secondary development, adopt object-oriented program, ActiveX automation technology and visual programming techniques to achieve human-computer interface and database interaction, design one of the mine power supply design software with function mapping, statistical calculations, protection setting, design reports automatically output, data maintenance. The study shows that the software can greatly simplify the power supply design process, improve efficiency, and can effectively improve the coal mine management level.

Keywords: VBA, secondary development, ActiveX automation technology, visual programming

1 Introduction

With the improvement of the level of electrification and automation of coal mining enterprises, the coal mine power supply system become increasing complexity, the traditional supply drawing and calculation methods cannot meet the needs of the coal mine power supply design. Computer-aided design (CAD) has been widely applied in the field of construction, electrical, shipbuilding, mining and other industries. The use of computer-aided design, you can overcome the deficiencies of the traditional methods of power supply design, shorten the development cycle, improve the efficiency and quality of power supply design.

There are a lot of achievements in research for the secondary development of CAD, such, The research studies the object-oriented method and technology of the further development, expecting offer direction of further development methods. An example is illustrated the application of object-oriented technique to the further development of parametric design by means of Pro/Toolkit [1]. ObjectARX files stored in database are shared by users to carry out the parameter-driven design CAD process, Multi-users can run it at multi-points in real time for collaborative design [2]. The parameter design for parts of 3D modelling is mainly introduced, on the basis of CAD (SolidWorks) of second exploitation, illustrates parameter design with correlative characteristics of 3D modelling for vane pump and depicts the correlative requests of parameter design [3]. Along with the extensive use of CAD in the fields of water conservancy and hydropower engineering, secondary exploration of CAD become more and more important. In this paper, combining part drawing of water and electricity station, primary attempt to CAD

secondary exploration of hydraulic engineering is made, including development of special line-type and filling and visual lisp language, etc. [4]. This research introduces a visible theoretic line loss calculation system of distribution network based on two times development of the AutoCAD. This system can draw equipment circuit diagram fast and conveniently based on the geography information. The alternant interface of person and machine raises the accuracy and the efficiency on the equipment and circuit data information to record. At the same time that system according to the network characteristics, with the data of the equipment, circuit and so on, provided two kinds of theoretic line loss calculation methods, really carried out the sketch, data with unify of calculate [5]. This research propose a design scheme of power supply design system of mining area based on AutoCAD. The system uses embedded VBA of AutoCAD plat form to make secondary development, and compiles program to extend functions of AutoCAD to realize integration of drawing, calculation and file management of power supply design by one software. The actual application showed that the system has characteristics of good interaction, convenient operation and high reliability [6]. The research introduces the drawing system of mining equipment power supply system. The drawing system is developed by VBA on the autocad2006 platform, draws quickly and accurately the figure of the power supply system, marks and modifies automatically the attribute of equipment through the menu mode of operation [7]. The process of developing drawing system is introduced in the paper. The developed drawing system has the professional feature on the drawing of mining power supply. It is based on ActiveX technology and developed by using the VBA language embedded the AutoCAD 2004 platform [8].

^{*} *Corresponding author* e-mail: zhengren_wu@163.com

2 Design program

Mine power supply design software is a database at the core, relying embedded in AutoCAD VBA language development, which can run under Windows2000 or later, Office2003 environment. The system is on the basis of CAD drawing function, with the help of CAD blocks technology and data dictionary technology, establish contacts in the CAD graphic device parameters stored in access, run interactively between the statistics of the load in the power supply design, automated operation of the short-circuit current, and check the tuning. After the design is completed, the system uses the word automation technology, automatic output calculation report based on the design of the access process data records. Through the application of the system eliminates the tedious process for power supply design, the original load statistics, short-circuit current calculated and report writing and save a lot of energy for designers , so that designers can put more effort into rational planning. The design technology program is shown in Figure 1.

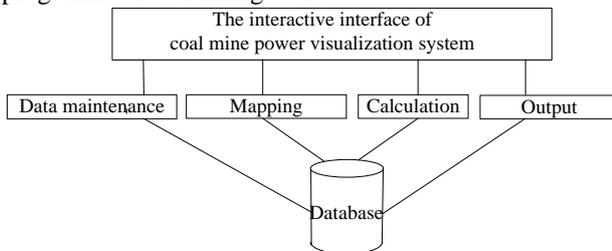


FIGURE 1 The design technology program

From the visualization software design structure, the R&D system is divided into three levels, four functional modules. Upper layer is Coal Mine Power visualization system interactive interface, the middle layer is four functional modules, the bottom layer is the software system's database. The upper man-machine interface are the software system bracket and the main operation of the user interface, it is able to achieve middle-level module functions in the interface layout interface. The underlying database is the foundation of the entire software system. With the support of the database, the function modules can collaborate and work together to achieve the entire functionality of the software system.

The middle layer of the four function module is the core of the development of software systems, including data maintenance module, power supply system mapping module, calculation and statistics module and calculation report output module. Data maintenance module is the auxiliary module system, the use of the module can put the electrical equipment parameters into the database. The mapping module of the power supply system is the basis of the system. Only use function of the module for power graphics in coal mine, which can fulfil power supply computing and power supply design report output. Calculation statistical module's function is to realize the power calculation, the module enables automatic statistics required coefficients, the average power factor calculation, transformer selection and validation, the load center

selection calibration and tuning, the mobile substation selection and validation, high voltage distribution box selection and tuning, high-voltage and low-voltage cables selection and validation, the various protective devices check and tuning. The calculation report output module is to achieve the output of the power supply design report. The power supply design report output with word form, including load tables, electrical equipment selection table, the short-circuit point tables as well as power calculation process.

3 Data information and database design

Access database table object contains seven database objects, Table object, query object, form object, report object, page object, macro objects and module objects. Database table objects and query objects are the basic objects, used to store data and query data, form objects, the report object and page objects is directly user-oriented object for input and output data and control system, the macro object and module code type object, to be completed the complexity of the database management, used to organize the macro operation or programming.

Graphics project file is a system diagram of coal mine power supply calculations. Contains file information, block diagram and calculations information. File information including file name, face type, create and creation time. Cell block information includes basic information and attribute labelling. The cell block basic information represent the device information, electrical node information. The calculation result information including the short-circuit current value and the setting calculation result, the short-circuit current values obtained in the power supply system requires information support, the shorting point located in a loop or branch information as well as the loop involving electrical equipment operation parameters need to know. Tuning calculations need to know the value of the short-circuit current, electrical equipment parameters. Calculating the short-circuit current value and the process of tuning results are very important, linked to a variety of information in the diagram of the power supply system of the underground mine. The data is shown in Figure 2.

The final result of the database structure design is to get a reasonable data model, the model is accurate, that generates the application system will be able to meet the needs of users. When designing the database structure the data need to meet the basic needs of the user, the data have a minimum degree of redundancy, data ensure consistency and completeness, and the data must be very strong independence. According to data information needs of the software system, the coal mine power visualization software system can be divided into five categories: file information class, system class, primitive class of electrical equipment, equipment operator parameters and power supply calculation results class.

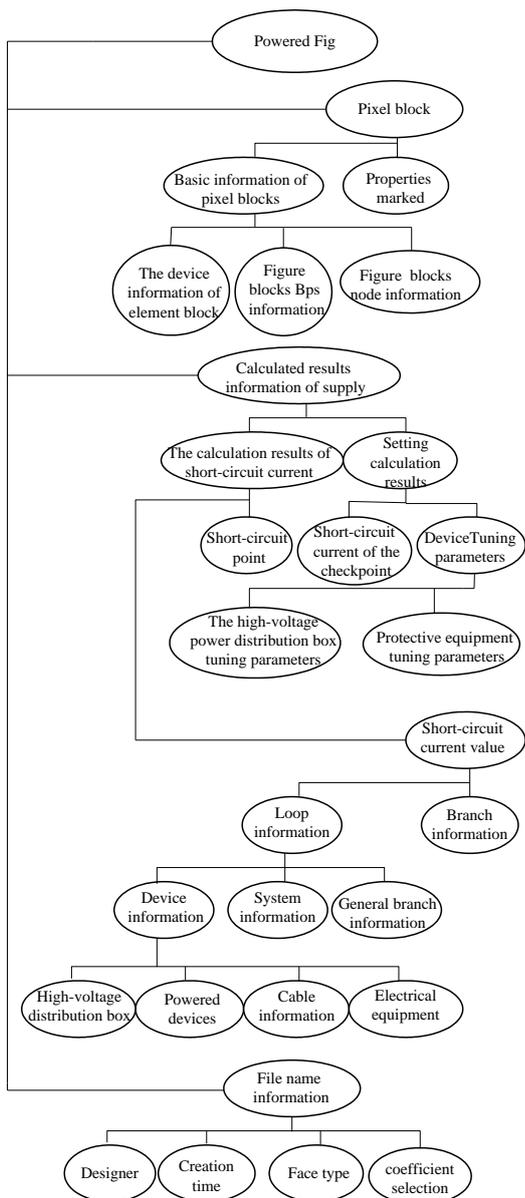


FIGURE 2 Data information

File information class: that reflect the data information what is recorded the information that needed by a power supply system drawing. This information includes: program number, name, the face, power calculation, creator, create time, and required coefficient.

System information class: the system classes data information is recorded information of the logical relationship between the electrical equipment and electrical equipment, cables and electrical equipment, the short-circuit point and the electrical equipment. The information includes: the element number, uniquely identifies of protection device, subjection transformer logo, subjection the high-voltage power distribution box logo, the number of branches, branch Number, uniquely identifies of starting point, the starting point component name, uniquely identifies of the endpoint, the endpoint device name.

Electrical equipment figure element class information: such data is a basic information for device diagram element. The information include: identifies of electrical equipment primitives, basis points, nodes, equipment figure element name, scaling factor, and twiddle factor.

Operation parameter information class information: the data information is a parameter of the equipment needed in the power calculation. the main parameters include the value of the rated voltage, the value of rated current, the value of rated power, the value of starting power factor , the value of start current multiple .

The calculation results class information: such data are the results of the short-circuit current calculation and setting calculation. Short circuit current calculation requires power supply circuit information and loop calculated. Setting calculation reflects the sensitivity of check results, select protection as well as protection setting results.

4 Design of graphics sub-module

The system mapping module use object-oriented technology and with auto cad ActiveX automation technology, the coal mine power supply system equipment components is to be created as device diagram elements. Take advantage of the man-machine interface system’s drawing module, the interface device component primitive button is inserted in equipment components primitive macro. The system equipment components diagram element commonly used in the coal mine power supply is placed in the drawing area, and then take advantage of the system mapping module device component layout button primitive for editing macros, such as moving the macro, delete macros, edit equipment components primitive that placed in the drawing area. Finally, the component parameters are transferred out by auto cad internal events ADO database access technology and human-computer interaction interface database, attribute labeling of equipment components in the drawing area. The design scheme of the system diagram for rendering module shown in Figure 3:

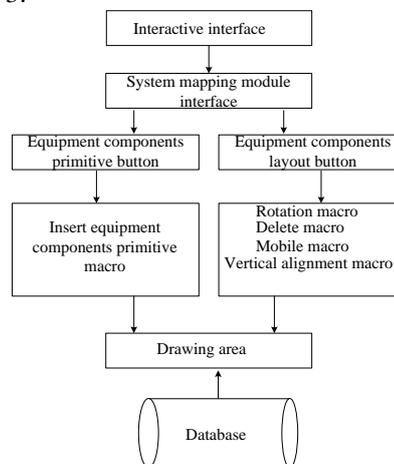


FIGURE 3 The design scheme of the system diagram for rendering module

5 The design for calculation statics and output

Design steps include: database setup and initial configuration, pixel graphics, load parameters configuration (load statistics), mobile substation select, dry-type transformers select, load centers select, high voltage cable and low voltage selection and calibration, short-circuit current calculation, high voltage distribution electrical device selection and calibration, mobile substation and load center tuning, combination switch ,feed switch and starter selection and calibration, calculation report automatically output.

VBA on ActiveX automation technology has the windows interoperability. The technology enables the common use between AutoCAD and other windows applications, and provides a way for the exchange of information between the each other application and control. AutoCAD use ActiveX automation technologies to interact between VBA language and word applications, and its essence is AutoCAD application program quote the word application's object library, then use VBA program to call word application object library.

AutoCAD and word application interact has four step:

1. Complete word application object library reference.
2. Create a word application instance to achieve access.

3. Write code to achieve control of the word application object.

4. Close word application object library reference.

The data interaction design of power supply report output is as follow. There are a lot of description language and a fixed format in calculation report, using typetext objects for text input, it has heavy work load and error. Therefore this work need to use word document template technology, firstly, create a template in a word document, enter the fixed contents in the template, set bookmarks what need to be changes in the content(for example, in the formula contained in the data) , then these templates are saved in the computer, finally, write VBA program in the programming environment, open a word document template, replace the content of the settings tab in the document by the text of the changes in the information or data. Also, this work can be completed by insert the required data or text in the template document using VBA programming. After the power supply system drawing, device parameters configuration and calculation, click system mapping module interface output button, which will pop up a short point output form in the drawing area of the software, select the output short circuit point, and then press the ok button, the power supply system of coal mine design calculation report instructions will automatically output.

References

- [1] Li Y, Du P 2004 Study of the object-oriented method of further development of parameterized CAD *Journal of UEST of china* 33(5) 597-610 (in Chines)
- [2] Zhu R, Zhao Y 2005 Quick cooperative design with customized CAD functions embedded in web pages *Journal of computer-aided design & computer graphics* 17(11) 2570-4 (in Chines)
- [3] Xia T 2011 Parameter Design of Parts for Vane Pump Based on CAD of Second Exploitation *Machinery manufacturing and its automation* 40(1) 102-4 (in Chines)
- [4] Liang G, Zhao L, Wu X 2006 Application of CAD secondary exploration in water conservancy and hydropower engineering drawing *Water Sciences and Engineering Technology* 2(2) 50-1 (in Chines)
- [5] Chen G, Qi X, Shan H 2008 The theoretic line loss calculation of distribution network based on development of the AutoCAD *Microcomputer Information* 24(8) 183-5 (in Chines)
- [6] Wang J, Zhao C 2012 Power supply design system of mining area based on AutoCAD *Industry and Mine Automation* 2(2) 97-9 (in Chines)
- [7] Wang J, Yang J 2011 Development of the drawing system of mining equipment's power supply system based on AutoCAD *Mechanical & Electrical Technology* 2(2) 61-6 (in Chines)
- [8] Hao X, Qu Y, Yao L 2011 Research on drawing methods of mining power supply based on active X technology *Coal Technology* 30(9) 30-1 (in Chines)

Authors	
	<p>Jin Baoquan, born on June 14, 1972, Shanxi, China</p> <p>Current position, grades: Associate professor at Taiyuan university of technology, Taiyuan City, China.</p> <p>University studies: Ph.D. degree in Mechanical and Electronic Engineering from Taiyuan University of Technology, Shanxi, China, 2010.</p> <p>Scientific interest: computer applications, computer software, computer modeling.</p> <p>Publications: 20.</p>
	<p>Hao Xiaohui, born on September 2, 1990, Shanxi, China</p> <p>Current position, grades: Graduate student in Taiyuan University of technology, Taiyuan City, China.</p> <p>University studies: bachelor degree in Measurement control technology and instrument specialty from Taiyuan University of Technology, Shanxi, China, 2013.</p> <p>Scientific interest: computer software, software programming computer application, control technology.</p>
	<p>Zhang Hongjuan, born on September 25, 1974, Shanxi, China</p> <p>Current position, grades: associate professor, Taiyuan University of Technology, Taiyuan City, China.</p> <p>University studies: Ph.D. degree in Mechanical and Electronic Engineering from Taiyuan University of Technology, Shanxi, China, 2011.</p> <p>Scientific interest: computer modeling, system computer modeling, software programming, energy saving control.</p> <p>Publications: 10.</p>
	<p>Gao Yan, born on July 14, 1969, Shanxi, China</p> <p>Current position, grades: associate professor Taiyuan University of technology, Taiyuan City, China.</p> <p>University studies: M.E. in automation and information engineering from Taiyuan University of Technology, Shanxi, China, 1998.</p> <p>Scientific interest: computer application, software programming, control theory, intelligent control.</p> <p>Publications: 10.</p>

Software reliability allocation model of CNC system based on software architecture

**Yan Gu¹, Yiqiang Wang^{2*}, Xiuhua Yuan²,
Xiaoqin Zhou¹, Bangcheng Zhang³**

¹College of Mechanical Science and Engineering, Jilin University, Changchun, Jilin 130022, China

²Ningbo Institute of Technology, Zhejiang University, Ningbo, Zhejiang 315000, China

³School of Electro-mechanical Engineering, Changchun University of Technology, Changchun, Jilin

Received 1 May 2014, www.tsi.lv

Abstract

In order to guarantee the implementation of reliability target, software reliability allocation model of CNC system was established based on software architecture. Software architecture of CNC system was set up, which decomposed CNC system into the functional units, reliability indexes of the system can be distributed into each component from top to bottom. The relative weight of software element in each level of the architecture was determined with analytic hierarchy process (AHP) method. The software reliability allocation model was built by taking the maximum practicability of CNC system as the target function, the reliability and cost function of component as the constraints. The reliability of each component was calculated through culture algorithm (CA). According to the result, the reliability allocation worked out is reasonable and feasible, and during the development of allocation model, the practicability of CNC system was guaranteed and development cost was also saved effectively.

Keywords: software reliability of CNC system, software architecture, reliability allocation, analytic hierarchy process, cultural algorithm

1 Introduction

This CNC system is the core of the CNC machine tool, whose reliability directly impacts the reliability level of the complete CNC machine tool. The software reliability of CNC system is a significant index for measuring the quality of CNC system. As a result, to improve the software reliability of CNC system is a key to improve the overall performance of the CNC machine tool. Therefore, it is imperative to conduct further studies on the software reliability of CNC system. The CNC system is developing towards high-precision, networking, intelligent and complex trend, with more powerful functions, larger calculated quantity. The software scale expands constantly, and it is getting more difficult to guarantee the software reliability [1]. In order to reach certain reliability target, developers must carry out repeatedly tests with unremitting efforts. However, it is extremely complicate in the software, and each component of the software has different influences on the overall software reliability [2-4]. Consequently, in order to avoid unnecessary valuable resources and costs, it must make a difference between each part of the system.

Reliability allocation means that the stipulated system reliability indexes should be regulated properly among the sub-systems forming the whole system, thus to reach the minimum cost of recourses employed by the system development. The software reliability allocation aims to

convert the system reliability index into the reliability index of each sub-system, for guiding the development of sub-system to seek for an optimal design scheme with the resource constraints, thus the system can achieve high reliability [5]. Since the software system is complicated and diversified, each component has distinct influence on the entire software system, and the software developers can only give full considerations during the software development system and assign to each allocation items, software parts till the software units after through comprehension of requirements of the software reliability. So the software developers can be clear about their responsibilities in reliability, carry out reliability analysis and design, and guarantee the implementation of final reliability target.

The hardware reliability allocation technique is mature, but the software reliability allocation is still immature. There have been a series of methods for applying the software reliability allocation already, which mainly includes: Wei Gang has proposed a method of complex software system reliability allocation based on ANP and established reliability allocation model of the complex software system [6]. Kaveh has proposed a dynamic self-adaptive multi-objective particle swarm optimization method to solve binary-state multi-objective reliability redundancy allocation problems [7]. Roberto has corrected the degree of reliability in tandem structure that affects the reliability of the software system components and on this

* Corresponding author e-mail: jluwang@gmail.com

basis, proposed the optimal allocation of time test method [8]. Zhou has proposed embedded real-time multi-tasking software reliability allocation model [9]. Kapur has proposed a mathematical model framework with multiple versions of the software [10].

At present, most reliability allocation is based on the test data, and few consider the architecture of software. Therefore, it is difficult to carry out reliability allocation in the early stage of systematic design. The current reliability distribution fails to consider the relationship between the reliability of software system and the reliability of each component. Meanwhile, there are fewer applications of software architecture study in the software reliability of CNC system, lacking a widely accepted model. The architectural model covers overall systematic information, which makes it convenient for comprehending and analysing the system. The architectural model can guarantee the consistency and completeness of system during the reliability allocation process. In this paper, the software architecture of the CNC system is established, the reliability indexes can be allocated to the components properly to satisfy the reliability requirements of the software system.

The remainder of the paper is organized as follows. Section 2 set up the software architecture of CNC system, software reliability allocation model of CNC system is established on the basis of software architecture. Section 3 presents the AHP method to solve the relative weight of software element in each level of the architecture. Section 4 introduced the theoretical culture algorithm to solve the model. Finally, an application example was given in section 5 to verify the method proposed in this paper.

2 Software reliability allocation model of CNC system

The software architecture is the abstraction of overall result in an abstract level, which provides better methods for developers to comprehend the software and to construct bigger and more complicated software system. As the blueprint of software, the software architecture provides an effective approach for people to grasp the overall software structure, which mainly refers to the organizational structure of the software system, the space connections, constraints, as well as the design and evolution principle and criteria. It is a relatively abstract and macroscopic description of the system composition and system structure, indicating the system division and composition. Gerlan & Shaw model generalized the software architecture as [11].

Software Architecture = {components, connectors, constraints}.

The component is the collection of related objects, which implements certain computational logic after implementation. It is related to structure or logic, and can be embedded into different architecture independently for implementing the reuse of components. The adapting piece is a group of objects, providing high-level communication among the components. It connects different components,

forming a part of the architecture, generally represented as the frame target or transition target. The constraint refers to the rules of component connection, pointing out the posture and condition of component connection. Generally, it consists of semantic constraints and topology constraints.

According to the software architecture, as well as the hierarchical structure of CNC system and functional unit, software architecture of CNC system was established for the corresponding relationship of each level. The software architecture of CNC system is divided into functional level F, program level P, and component level C, as shown in Figure.1.

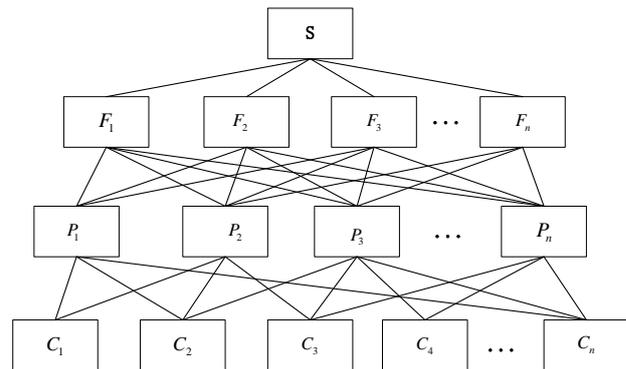


FIGURE 1 Software architecture of CNC system

There is a linear relationship between the software availability and software function, as well as the reliability of these functions [12], which can be defined as:

$$U = \sum_{i=1}^f \omega f_i \cdot r f_i \quad (1)$$

$$U = \sum_{i=1}^p \omega p_i \cdot r p_i \quad (2)$$

where: ωf_i is the global relative right of function i , $r f_i$ is the reliability of function i , ωp_i is the global relative right of program i , $r p_i$ is the reliability of program i .

Suppose all of the components $C\{c1,c2,c3,\dots,cn\}$ in software are mutually independent, and then the reliability of program $r p_i$ is the product of the reliability of all components.

$$r p_i = \prod_{j \in c_i} r c_j \quad (3)$$

According to the software architectural characteristics of the CNC system, the reliability index can be allocated to each functional component with the maximum practicability of CNC system as the target, and the reliability and cost of the component as the constraints. Suppose the system has n functional components,

remarked as $C\{c_1, c_2, c_3, \dots, c_n\}$, the software reliability allocation model of the CNC system is shown as follows:

$$\max[U = \sum_{i=1}^p \omega p_i \cdot \prod_{c_j \in c_i} rc_j], \tag{4}$$

$$\begin{aligned} rc_j &\leq u_i \\ rc_j &\geq l_i \\ \alpha_j + q_j \cdot rc_j &\leq \alpha \cdot v_i, \\ \sum_{j=1}^m \alpha_j + q_j \cdot rc_j &\leq \Omega \end{aligned}$$

where the reliability of component i is rm_i , $0 < \omega p_i < 1$, $\sum_{i=1}^n \omega p_i = 1$, and its upper limit is u_i , while its lower limit is l_i . α_j is the general expense at the time of exerting reliability rc_j constraint on the component i , q_j is the adjustable cost expense, α is the deduction of developers profit from 1, v_i is the cost for accomplishing the component i , and Ω is the development cost budget of the numerical control system software.

3 Reliability allocation model AHP analysis

The significance of each component is determined with AHP method, for it can decrease the artificial factor during the weight determination process, thus to guarantee the practicability and relative objectivity of weight. The method of working out weight collection with AHP is shown as follow: according to the architecture of CNC system, the factors of each level are taken as the criteria successively, and significance of all the elements of the next level related shall be compared, and a value of relative significance can be achieved a_{ij} . A judgment matrix can be obtained by comparing each influencing factors, thus n -order matrix can be formed.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \tag{5}$$

where $a_{ij} > 0$, $a_{ji} = \frac{1}{a_{ij}}$, $a_{ii} = 1$, $i, j = 1, 2, \dots, n$. a_{ij} is selected according to the '1-9' proportion scale improved according to the relative significance [13].

Suppose there are n factors in this level related to a certain index in the previous level, and then the indexes of the previous level above n factors of this level can be the relative significant weight of comparison between the two, namely the weight vector of this level is:

$$W = (w_1, w_2, \dots, w_n)^T \quad 0 \leq w_i \leq 1 (i = 1, 2, \dots, n). \tag{6}$$

According to the judgment matrix constructed, sum and product method is adopted for calculating the normalization processing characteristic value w_i and maximum characteristic root λ_{\max} of each judgment matrix [14], according to the following equation:

$$w_i = \frac{1}{n} \sum_{j=1}^n \frac{a_{ij}}{\sum_{k=1}^n a_{kj}} \quad (i = 1, 2, \dots, n), \tag{7}$$

$$\lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \frac{(Aw)_i}{w_i}, \tag{8}$$

where A is the judgment matrix, $(Aw)_i$ is i element of the vector Aw .

By calculating the maximum characteristic root λ_{\max} of judgment matrix and characteristic vector W , the relative significance weight of a certain factor towards another factor in previous level. AHP measures the consistency degree of the judgment matrix through calculating the consistency ratio CR , when $CR \leq 0.1$, it is in accordance with the consistency, or it is deemed that the judgment is of certain randomness, and it should adjust the element value of the judgment matrix. CR can be calculated according to Equation (9):

$$CR = \frac{CI}{RI}, \tag{9}$$

$$CI = \frac{\lambda_{\max} - n}{n - 1}, \tag{10}$$

where CI is consistency index of the judgment matrix, when the judgment matrix is complete consistent, $CI=0$, RI is average random consistency index, λ_{\max} is maximum characteristic root of judgment matrix, n is orders of the judgment matrix.

When checking the consistency of each judgment matrix, it should calculate the RI value of the judgment matrix gain according to new scale, and in Table 1, new RI value is given:

TABLE 1 Random consistency index RI

Order	1	2	3	4	5
RI	0	0	0.1690	0.2598	0.3287
Order	6	7	8	9	
RI	0.3694	0.4007	0.4167	0.4370	

4 Culture algorithm

Culture algorithm is a bionic intelligent calculation method simulating the cultural evolution. With the double evolutionary mechanism formed by upper reliability level and bottom population space, the algorithm excavates implicit knowledge reflecting the evolutionary degree and advantageous region from the evolutionary individual in

bottom population, and then send feedbacks of the knowledge stored in reliability space to the population evolution process for improving the searching efficiency and improving the performance of algorithm [15]. Culture algorithm simulates the cultural evolution process of human society, with double evolution mechanism on the basis of traditional population-based evolution algorithm [16]. The reliability space is constructed to extract various types of information implied during the evolutionary process, and it is stored in the form of knowledge, which will finally be applied for guiding the evolutionary process. The algorithm flow is shown as follows [17]:

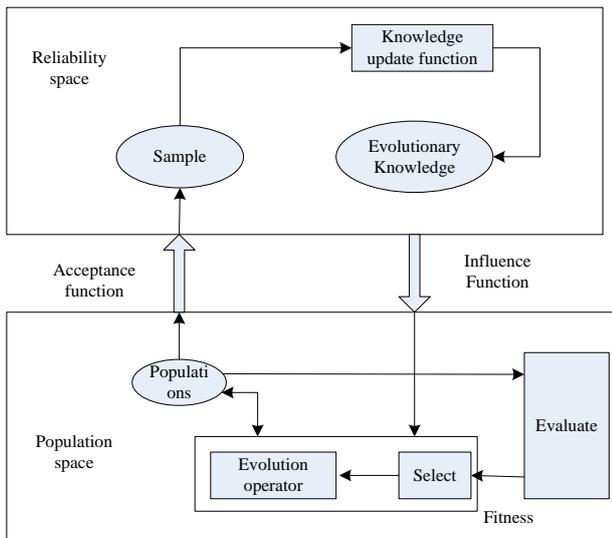


FIGURE 2 Culture algorithm flow

5 Application example

The software architecture of CNC system constructed by taking the maximum practicability of CNC system as the target function is shown as follows:

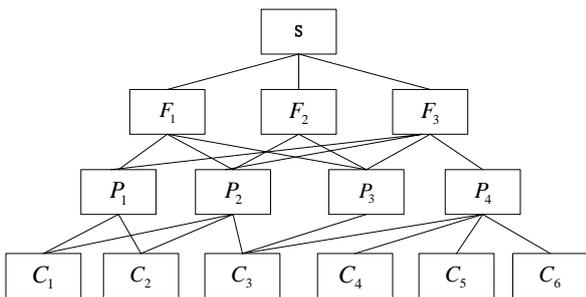


FIGURE 3 Software architecture model of the CNC system

The judgment matrix is constructed by comparing the relative significance of each factor according to the improved 1-9 proportional scale. In the CNC system functional layer, usefulness of the functional importance of F1 F2 F3 CNC system is on scoring pairwise comparison, the judgment matrix U-F as shown in Table 2.

TABLE 2 Evaluation matrix for U-P layer

U	F ₁	F ₂	F ₃
F ₁	1	1.22	1.5
F ₂	1/1.22	1	1.22
F ₃	1/1.5	1/1.22	1

In the CNC system program layer, P1 P2 P3 importance of CNC system functional layer scoring is on scoring pairwise comparison. The judgment matrix F-P as shown in Table 3-5.

TABLE 3 Evaluation matrix for F1-P layer

F ₁	P ₁	P ₂	P ₃
P ₁	1	1.22	1.22
P ₂	1/1.22	1	1
P ₃	1/1.22	1	1

TABLE 4 Evaluation matrix for F2-P layer

F ₂	P ₂	P ₃
P ₂	1	1.5
P ₃	1/1.5	1

TABLE 5 Evaluation matrix for F3-P layer

F ₃	P ₁	P ₂	P ₃	P ₄
P ₁	1	1.22	2.33	1.5
P ₂	1/1.22	1	1.86	1.22
P ₃	1/2.33	1/1.86	1	1/1.5
P ₄	1/1.5	1/1.22	1.5	1

In the CNC system component layer, C1 C2 C3 C4 importance of CNC system program layer scoring is on scoring pairwise comparison. The matrix of P-C judgment as shown in Table 6-8.

TABLE 6 Evaluation matrix for P1-C layer

P ₁	C ₁	C ₂
C ₁	1	1.22
C ₂	1/1.22	1

TABLE 7 Evaluation matrix for P2-C layer

P ₂	C ₁	C ₂	C ₃
C ₁	1	1.22	2.33
C ₂	1/1.22	1	1.86
C ₃	1/2.33	1/1.86	1

P3=1

TABLE 8 Evaluation matrix for P4-C layer

P ₄	C ₁	C ₂	C ₃	C ₄
C ₁	1	1/1.22	1/1.22	1/1.5
C ₂	1.22	1	1	1/1.22
C ₃	1.22	1	1	1/1.22
C ₄	1.5	1.22	1.22	1

The normalization processing characteristic value W_i and maximum characteristic root λ_{max} is calculated according to Equation (7) and Equation (8), the weight vector of each level is obtained, and the single hierarchical arrangement structure is shown in Table 9. The consistency is verified according to Equation (9) and Equation (10) CR is smaller or equal to 0.1, and as a result, the judgment matrix constructed is proper and reliable.

According to the above calculation result, the overall weight vector of calculation procedure level and component level towards the target level is calculated, and the consistency has been verified.

$$WP=(\omega p_1 \ \omega p_2 \ \omega p_3 \ \omega p_4)=(0.245 \ 0.397 \ 0.297 \ 0.061),$$

$$WC=(\omega c_1 \ \omega c_2 \ \omega c_3 \ \omega c_4 \ \omega c_5 \ \omega c_6)=(0.311 \ 0.254 \ 0.386 \ 0.015 \ 0.015 \ 0.019).$$

TABLE 9 Results of single ordering for hierarchy

H	W	λ_{max}	CI	CR
U-F	[0.4023, 0.3289, 0.2688]	3.0	0	0
F ₁ -P	[0.3788, 0.3106, 0.3106]	3.0	0	0
F ₂ -P	[0.6, .04]	2.0	0	0
F ₃ -P	[0.3436, 0.2792, 0.1497, 0.2275]	4.0001	0	0
P ₁ -C	[0.5495, 0.4505]	2	0	0
P ₂ -C	[0.4453, 0.3618, 0.1928]	3.0001	0	0
P ₃ -C	1	1	0	0
P ₄ -C	[0.2025, 0.2475, 0.2475, 0.3025]	4.0	0	0

Suppose the software cost is $V=1000000$, it can be obtained that the component cost vector is $V \times WC$ (311 254 386 15 15 19). Let the profit as 50%, and the component 4-6 has no profit, $\alpha = 1$. The development cost of CNC system software is Ω . Substitute the data to Equation (4), it can be obtained that:

$$\max[U = \sum_{i=1}^p \omega p_i \cdot \prod_{c_j \in c_i} rc_j] = 0.245(rc_1 rc_2) + 0.397(rc_1 rc_2^2 rc_3) + 0.297(rc_3^4) + 0.061(rc_3^2 rc_4^3 rc_5 rc_6)$$

$$0.6 \leq rc_j \leq 1 \quad j = 1, 2, \dots, 6$$

$$95 + 60rc_1 \leq 0.5 \cdot (311)$$

$$70 + 50rc_2 \leq 0.5 \cdot (254)$$

$$85 + 90rc_3 \leq 0.5 \cdot (368)$$

$$5 + 8rc_4 \leq 15$$

$$7 + 6rc_5 \leq 15$$

$$3 + 5rc_6 \leq 19$$

$$265 + 60rc_1 + 50rc_2 + 90rc_3 + 8rc_4 + 6rc_5 + 5rc_6 \leq 490.$$

Run the cultural algorithm by Matlab, the result of the first 8 iterations shows in Table 1. Seen from Table 10, the min adaptation function value is steady at 0.9999.

The reliability of the components is:

$$[rc_1 \ rc_2 \ rc_3 \ rc_4 \ rc_5 \ rc_6] = [0.9758 \ 0.9412 \ 0.6747 \ 0.7026 \ 0.7587 \ 0.8772]$$

References

[1] Richardson I, Casey V, McCaffery F, Burton J, Beecham S 2012 A process framework for global software engineering teams *Information and Software Technology* 1175-91
 [2] Kuo SY, Huang CY, Lyu MR 2001 *IEEE Transactions on Reliability* 50(3) 310-20

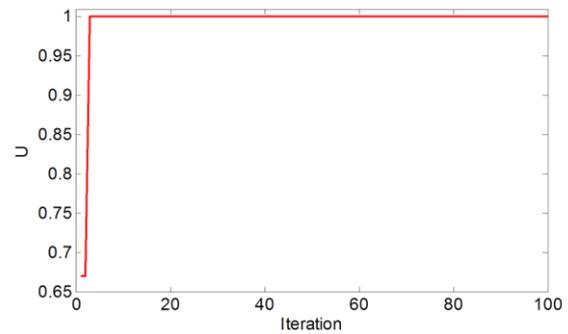


FIGURE 4 Adaptation values and iteration times

TABLE 10 Result of the first 8 iterations

iteration	U
1	0.7490
2	0.7490
3	0.9999
4	0.9999
5	0.9999
6	0.999
7	0.999
8	0.999

Conclusion

The architecture of CNC system software was established in this paper, the software reliability index allocation method of CNC system has been proposed based on software architecture. The reliability index of CNC system has been allocated to the functional units, and the reliability allocation model has been established by taking the system practicability as the target function, and the reliability degree and cost function of functional units as the constraints. The relative weight of each element in the software architecture of CNC system has been calculated with AHP, and the allocation value of each component has been worked out with culture algorithm. The example has shown that this method guarantees that the system reliability index can satisfy the requirements, and meanwhile, it also saves the development cost effectively, improves the validity of allocation method and provide basis for studying the feasibility of its sub-system.

Acknowledgements

This paper is supported by state key science and technology special projects for advanced CNC machine tools and basal manufacturing equipment (2012ZX04011021), and natural science foundation of Zhejiang province (Y1110708), and the National Science Foundation of China (61374138).

[5] Hsieh Y C You P S 2011 An effective immune based two-phase approach for the optimal reliability-redundancy allocation problem *Applied Mathematics and Computation* 1297-307

[6] Wei Gang, Xu Li, Wang Huankun (2012) A Method of Complex Software System Reliability Allocation Based on ANP, 18th ISSAT International Conference on Reliability and Quality in Design, pp:86-90

[7] Damghani K, Abtahi K, Reza A, Madjid T 2013 A new multi-objective particle swarm optimization method for solving reliability redundancy problems 58-75

[8] Pietrantuono R, Russo S, Trivedi S K 2010 *IEEE Transactions on software engineering* 36(3) 323-37

[9] Zhou W, Hao Y, Chen G 2006 Software reliability allocation of warship integrated navigation Journal of Projectiles, Rockets, Missiles and Guidance 4-8

[10] Kapur P K, Pham H, Aggarwal G A, Kaur G 2012 *IEEE Transactions on Reliability* 62(3) 758-68

[11] Garlan D, Shaw M 1994 An Introduction to Software Architecture *Technique Report CMU/SEI-94-TR-21 Carnegie Mellon University* 1-37

[12] Zahedi F, Ashrafi N 1991 *IEEE Transactions on Software Engineering*: 17(4) 345-56

[13] Zhu J, Wang M, Liu S 2007 Research on Consistency Modification Problem of Comparison Matrix in the Analytical Hierarchy Process *Systems Engineering—Theory & Practice* 18-22

[14] Satty T L 1990 How to make a decision: the analytic hierarchy process *European Journal of Operational Research* 9-26

[15] Zhao Z, Yan X, Shi H 2013 Group search optimizer algorithm based on cultural evolution *Journal of East China University of Science and Technology* 95-101 (in Chinese)

[16] Guo Y, Chen M, Wang C, Liu H 2013 Analysis on the convergence of cultural algorithm *Control and Decision* 1631-4

[17] Abs da Cruz V A, Pacheco C M A, Vellasco Marley Hall Barbosa, Barbosa Carlos R. Hall (2005). Cultural operators for a quantum-inspired evolutionary algorithm applied to numerical optimization problems. *Lecture Notes in Computer Science* 1-10

Authors	
	<p>Yan Gu, born in 1980, Jilin, China</p> <p>Current position, grades: doctor at Jilin University, China. University studies: Mechanical Science. Scientific interest: software reliability and manufacturing technology. Publications: 1.</p>
	<p>Yiqiang Wang, born in 1964, Liaoning, China</p> <p>Current position, grades: professor and PhD candidate supervisor at Zhejiang University in China. University studies: PhD in Mechanics (1999) from Jilin University, China. Scientific interest: quality control and digital manufacturing technology. Publications: 2.</p>
	<p>Xiuhua Yuan, born in 1983, Shandong, China</p> <p>Current position, grades: post doctor at Zhejiang University in China. University studies: M.Sc. in Mechanics (2008) and PhD in Mechanics (2011) from Jilin University in China. Scientific interest: software reliability and manufacturing technology. Publications: 3.</p>
	<p>Xiaoqin Zhou, born in 1967, Hubei, China</p> <p>Current position, grades: professor and PhD candidate supervisor at Jilin University, China. University studies: PhD in Mechanics (1998) from Jilin University in China. Scientific interest: advanced optical manufacturing and digital manufacturing technology. Publications: 4.</p>
	<p>Bangcheng Zhang, born in 1972, Jilin, China</p> <p>Current position, grades: professor at Changchun University of Technology, China. University studies: PhD in Mechanics (2011) from Jilin University, China. Scientific interest: digital manufacturing technology. Publications: 5.</p>

A novel unsupervised segmentation for remote sensing image using MRF

Jiajing Wang^{1, 2*}, Shuhong Jiao¹, Zhenyu Sun³

¹Faculty of Information and Communication Engineering, Harbin Engineering University, No.145, Nan Tong Road Harbin, Heilongjiang, China

²No.92677 Unit of PLA, Dalian, Liaoning, China

³No.91550 Unit of PLA, Dalian, Liaoning, China

Received 23 March 2014, www.tsi.lv

Abstract

The image segmentation is the basis of image interpretation in remote sensing applications and plays vital role in image analysis. The Markov Random Field (MRF) approach is widely studied for use in segmentation of remote sensing image, which is an important extraction technique in recognition problems. This paper presents an unsupervised segmentation method for remote sensing image using the MRF. A novel neighbourhood system for the energy function has been proposed, the segmentation of remote sensing image and the optimization process of the parameters are performed simultaneously for the unsupervised segmentation in iterative condition. The experimental results on Synthetic Aperture Radar (SAR) images show that the proposed method performs better than the conventional Bayesian segmentation methods.

Keywords: remote sensing image, MRF, unsupervised segmentation, parameter estimation, SAR

1 Introduction

The rapid development of satellite remote sensing technology has enabled the multi-level, multi-angle, stereoscopic, omni bearing, and all-weather observation of earth. The quality of the observation mainly depends on the interpretation of remote sensing image; therefore, research on the remote sensing image has been a hotspot nowadays [1]. However, segmentation play an important role in the remote sensing technology, it can provide the image structure information, and establish the foundation for automatic target recognition on remote sensing system, it is crucial for understanding and interpreting remote sensing image.

In this paper, a novel method for remote sensing image segmentation is presented, with specific application to Synthetic Aperture Radar (SAR). The SAR is an active remote sensing system that generates and transmits microwave electromagnetic (EM) radiation to the surface of a target region [2]. And the SAR imaging does not get influenced by the weather conditions, geographical location, time of the day, and it can find hidden underground target through the vegetation, therefore, it has been widely applied to military and civil sectors. However, the images of SAR are heavily corrupted by speckle noise due to constructive and destructive EM wave interference during image acquisition [3]. Speckle noise is modelled as a multiplicative noise and gives the images grainy appearance. It can affect automated image segmentation. The MRF model can well suppress the influence of the speckle on segmentation results, so it has been widely

applied to the remote sensing image segmentation applications. Dong [4] used Gaussian Markov Random Field model to complete the segmentation of remote sensing image. Ludwin [5] detected oil spill in SAR images by MRF and image fusion. Recently, Liu et al. [6] presented Triplet Markov Field (TMF) to PolSAR image classification, and D. Elia et al. [7] presented tree structured Markov Random Fields for remote sensing image classification.

A new neighbourhood system of MRF has been developed and presented in this paper, and the image segmentation is performed using the optimized parameters, which acquired through the iterative process for the unsupervised segmentation applications. The rest of the paper is organized as follows: Section 2 introduces the two MRF models including the new neighbourhood system definition and the estimation process of the parameters; section 3 gives a detailed analysis of the unsupervised evaluation method; the results of experiments and comparisons are reported in Section 4; finally, section 5 presents a conclusion and future plans.

2 MRF models

Let $S = \{s = (i, j) | 1 \leq i \leq M, 1 \leq j \leq N\}$ be the set of image sites which specify the location of the pixel, M and N are the width and height of the image respectively. $X = \{X_s, s \in S\}$ and $Y = \{Y_s, s \in S\}$ are two random fields defined in S . X is the hidden field where each X_s represents the class of the site s and takes its value from a

* Corresponding author e-mail: 1656975717@qq.com

finite set $\Psi = \{1, \dots, G\}$. Y is the observed field where each Y_s represents the observed data of site s and takes its value from the gray level value. Both of the two random fields need to be considered in building a uniform model of the whole image.

Assume θ_x and θ_y are the relevant parameters of the field X and Y respectively, and they are distributed uniformly and independent of each other. For image segmentation of MRF model, the task is to recover the hidden field X from the observed field Y . When the observed field data Y are known, seek the estimation class \hat{X} for maximum a posteriori probability, in other words, find the optimal solution $(\hat{X}, \hat{\theta}_x, \hat{\theta}_y)$. The optimal solution belongs to a variable space $\Omega = (X, \theta_x, \theta_y)$:

$$\begin{aligned} \hat{X}_{MAP} &= \arg \max_{\Omega} P(\Omega|Y) = \\ & \arg \max_{\Omega} \frac{P(Y, \Omega)}{P(Y)} = \arg \max_{\Omega} P(Y|\Omega)P(\Omega) = \\ & \arg \max_{(X, \theta_x, \theta_y)} P(Y|X, \theta_x, \theta_y)P(X|\theta_x, \theta_y)P(\theta_x, \theta_y) = \quad (1) \\ & \arg \max_{(X, \theta_x, \theta_y)} P(Y|X, \theta_y)P(X|\theta_x)P(\theta_x)P(\theta_y) = \\ & = \arg \max_{(X, \theta_x, \theta_y)} P(Y|X, \theta_y)P(X|\theta_x). \end{aligned}$$

Hence, finding the optimal solution of the variable space turns into optimizing the Equation (1) under the condition of a given observed field. The models of hidden field X and observed field Y are built separately as follow.

2.1 MODEL OF OBSERVED FIELD

The observed field in the MRF model is based on the original remote sensing image. It can reflect the characteristics of the original image data and fully describe the change of texture details of remote sensing image. The Gaussian Markov Random Fields (GMRF) model is used to represent the conditional distribution of remote sensing image:

$$P(Y|X, \theta_y) = \prod_{s \in S} \left[\frac{1}{\sigma_y^2 (2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} (y_s - \mu_s)^T \sigma_y^{-1} (y_s - \mu_s) \right\} \right]^{d/2}, \quad (2)$$

where μ is the mean value, σ is the standard deviation, and d represents the number of y_s in observed blocks. $\theta_y = \{\theta_y^g; g \in (1, \dots, G)\}$ indicates the relevant parameters of the observed field Y ; $\theta_y^g = \{\mu_g, \sigma_g\}$, $g \in \Psi$ and θ_y can be obtained by Maximum likelihood (ML) [8] estimate:

$$\hat{\mu}_g = \frac{1}{n_g} \sum_{s \in S, x_s \in g} Y_s, \quad (3)$$

$$\hat{\sigma}_g = \frac{1}{n_g} \sum_{s \in S, x_s \in g} (Y_s - \hat{\mu}_g)(Y_s - \hat{\mu}_g)^T. \quad (4)$$

The ML estimate $\hat{\theta}_y$ is obtained using Equations (3) and (4), it has a closed form, and it is iteratively updated with the segmentation calculation.

2.2 MODEL OF HIDDEN FIELD AND A NEW ENERGY SYSTEM

It has been proven by feasibility and rationality in a number of related studies that the hidden field X can be considered as the MRF. Considering the relationship between MRF and Gibbs, the probability distribution of the hidden field X can be described by the neighbourhood system and energy function. Using the MRF-Gibbs equivalence [9], the probability distribution can be directly written as:

$$P(X = x) = Z^{-1} \exp(-\sum_{c \in C} V_c(X)), \quad (5)$$

where Z is a normalization constant, $V_c(X)$ is the potential energy function related to the group c and $U(X) = \sum_{c \in C} V_c(X)$ constitutes the energy function. C is the collection of all the groups. In this paper, the MLL model is used to describe the potential energy function as:

$$V(x_s, x_n) = \begin{cases} \beta_s & (x_s = x_n) \\ -\beta_s & (x_s \neq x_n) \end{cases}. \quad (6)$$

The classic 8 neighbourhood system with the information of location is developed in this study. As shown in Figure 1, x and y are the horizontal and vertical neighbourhoods of the site S , respectively. λ_x and λ_y are unit vectors of x and y . $\lambda_x = \pm 1, \lambda_y = 0$, for s and n are horizontal; $\lambda_x = 0, \lambda_y = \pm 1$, for s and n are vertical; and $\lambda_x = \pm 1, \lambda_y = \pm 1$, for s and n are diagonal.

According to [10] and the theorem of Gravity, the relative position of pixels is closer when the interaction is stronger. The Euclidean distance is used to define the energy parameter related to the group $c = (s, n)$:

$$\beta_s(s, n) = \frac{1}{d^2(s, n) \left[\frac{\lambda_x^2(s, n)}{\beta} + \frac{\lambda_y^2(s, n)}{\beta} \right]}, \quad (7)$$

where $d(s, n) = \sqrt{\lambda_x^2 + \lambda_y^2}$. To achieve the unsupervised segmentation, the optimal values of estimated parameters need to be obtained.

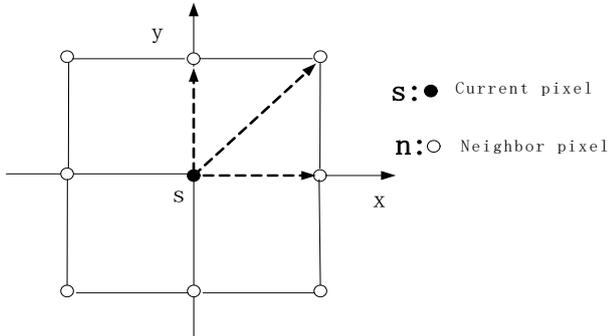


FIGURE 1 New neighbourhood system

3 Unsupervised segmentation of remote sensing image

3.1 PARAMETERS ESTIMATE

Markov Chain Monte Carlo (MCMC) [11] is an easy and effective method for parameter estimation. A number of MCMC algorithms have been reported in related studies; however, the Metropolis-Hasting algorithm is chosen to be used in this study. The posteriori probability $p(\theta_x | X)$ is calculated based on Bayesian theory, where $\theta_x = \{\beta\}$.

$$p(\theta_x | X) = \frac{p(\theta_x)p(X|\theta_x)}{\int p(\theta_x)p(X|\theta_x)d\theta} \propto p(\theta_x)p(X|\theta_x). \quad (8)$$

The Pseudo Likelihood function is introduced to simplify the problem:

$$PL(X|\theta_x) = \ln \prod_{s \in S} p(x_s|\theta_x) \quad (9)$$

and $p(x_s|\theta_x)$ is determined as follows:

$$p(x_s|\theta_x) = \frac{\exp(-U(x_s, \theta_x))}{\sum_{x_s \in G} \exp(-U(x_s, \theta_x))}. \quad (10)$$

The prior probability $p(\theta_x)$ is not considered because the priori information is useless in this application. In the Metropolis-Hasting algorithm, the probability of acceptance $\alpha(\theta_x, \theta'_x)$ can be represented as:

$$\alpha(\theta_x, \theta'_x) = \min \left(1, \frac{\exp(PL(X|\theta_x)q(\theta'_x|\theta_x))}{\exp(PL(X|\theta'_x)q(\theta_x|\theta'_x))} \right). \quad (11)$$

Symmetrical proposal distribution is chosen, Equation (11) turns to:

$$\alpha(\theta_x, \theta'_x) = \min \left(1, \exp(PL(X|\theta_x) - PL(X|\theta'_x)) \right). \quad (12)$$

If the current status is θ'_x , then the new status is obtained using the Equation (12). In summary, find a θ_x as initialization parameter, set the maximum number of iterations N_g for iterative process that operates on $t = 1, \dots, N_g$. Sample the proposal probability $q(\theta'_x|\theta_x)$ to obtain the next state θ'_x , compute the acceptance probability using Equation (12). If θ_x is accepted, then the next state is $\theta'_x = \theta_x$; otherwise $\theta'_x = \theta'_x$; continue the iterative calculation until t reaches the maximum number of iterations. In order to ensure a smooth convergence, the class simulated annealing (SA) algorithm is used before the MCMC estimate. Hence, the Equation (12) can be rewritten as:

$$\alpha(\theta_x, \theta'_x) = \min \left(1, \exp \left(\frac{PL(X|\theta_x) - PL(X|\theta'_x)}{T_t} \right) \right), \quad (13)$$

where T is the coefficient of temperature, $T_t = \gamma T_{t-1}$, $T_0 \gamma^N = 1$, and N is the number of the iteration of SA. The Markov chain converges to the optimal point after a certain number of iterations. Finally, choose an appropriate number M ($M \leq N_g$), and select $\theta_x^1, \theta_x^2, \dots, \theta_x^M$ from the results of MCMC, and approximate them according to:

$$\hat{\theta}_x \approx E(\theta_x | X) = \frac{1}{M} \sum_{t=1}^M \theta_x^t. \quad (14)$$

3.2 ALGORITHM OF SEGMENTATION

Supervised estimation methods are widely used in image segmentation; however, these are not applicable to the automated processing. Therefore, the unsupervised technique that requires simultaneous segmentation and estimation is essential and the trending of the future. In this paper, an unsupervised estimation technique is used to simultaneously optimize the parameters and perform the segmentation. As mentioned earlier, the goal is to find the optimal solution $(\hat{X}, \hat{\theta}_x, \hat{\theta}_y)$ through the rule of the MAP, from Equation (1):

$$(\hat{X}, \hat{\theta}_x, \hat{\theta}_y) = \arg \max_{(X, \theta_x, \theta_y)} P(Y|X, \theta_y)P(X|\theta_x) = \arg \max_{(X, \theta_x, \theta_y)} \ln P(Y|X, \theta_y) + \ln P(X|\theta_x). \quad (15)$$

After taking the logarithm, the optimal solution $(\hat{X}, \hat{\theta}_x, \hat{\theta}_y)$ can be calculated by obtaining maximum of $\ln P(Y|X, \theta_y)$ and $\ln P(X|\theta_x)$. Using the iterative

process described in preceding section $P(Y|X, \theta_y)$ and $P(X|\theta_x)$ can be obtained under different sets of parameters thus get the maximum of them. However, such a direct maximization is complex and computationally expensive. Therefore, a simplification is performed based on the local optimization standard (POS) proposed by Zhang [12].

Specific algorithm is as follows: firstly, use K-means clustering method or Maximum Likelihood estimation for an initial segmentation and parameters $(X', \theta'_x, \theta'_y)$, set 0 to n. Secondly, the initial segmentation result and parameters are substituted into the hidden label field $P(X|\theta_x)$ and observation field $P(Y|X, \theta_y)$. Then, iteratively calculate the new label results and the parameters estimation using the Equations (3), (4) and (14), it is called hidden label field update and parameters update. Finally, according to the rule of POS, compute

$\max(|(\theta'_x, \theta'_y) - (\theta^*_x, \theta^*_y)|) \leq \epsilon$ or $n = n_{\max}$; if either of that set up $(\theta^*_x, \theta^*_y) = (\hat{\theta}_x, \hat{\theta}_y)$ is the optimal solution; otherwise, assign the parameters to the initial value and continue the step 2 to step 4 until it satisfies the termination condition. Here ϵ and n_{\max} are the final parameters of the iterative process, calculate them for different kinds of parameters.

5 Experimental results

In order to describe the quality of segmentation quantitatively, the segmentation is performed on an artificial synthesis of remote sensing images using three different methods: K-means clustering, the classical hidden Markov fields (HMF) method, and the proposed unsupervised segmentation algorithm. Figure 2 summarizes the results of three algorithms and presents a comparison of the segmentation quality of the methods.

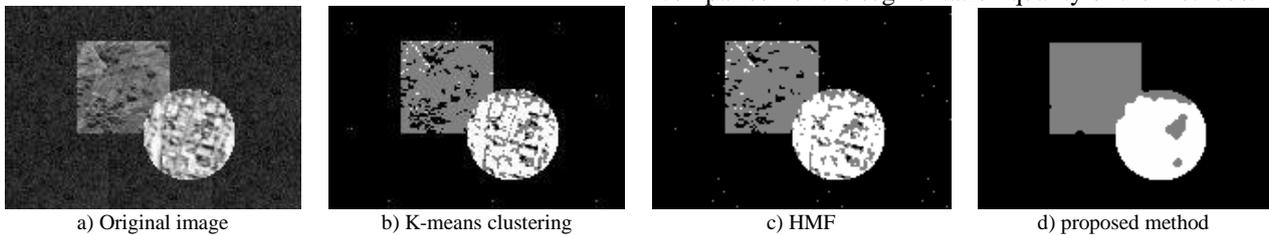


FIGURE 2 Synthesis of remote sensing image and results of segmentation

It can be seen from the diagram that K-means clustering and the HMF segmentation methods are sensitive to noise and the distribution of grey value. In contrast, the method we described has a certain inhibitory effect to the noise since the relationship between the pixels in local space is considered. Additionally, it optimizes the result of K-means clustering by significantly improving the regional consistency. The Kappa coefficient, classification error rate, and running time are calculated and presented in Table 1. The larger value of the Kappa coefficient with the lower of classification error rate shows a greater effect on segmentation. The evaluation results are consistent with the human vision system. However, a longer running time is required for the proposed method due to higher complexity compared to the other two methods.

Test the segmentation effect using real SAR images of different scenarios, as shown in Figure 3, where Figures 3a are original SAR images; Figures 3b are segmentation results of K-means clustering; Figures 3c are segmentation

results of the classical HMF and Figures 3d are the segmentation results of proposed method in this paper.

TABLE 1 Segmentation quality assessment of different method

Method	Criteria of segmentation quality assessment		Running time/(s)
	Segmentation error rate	Kappa coefficient	
K-means clustering	0.0645	0.8272	0.1803
HMF	0.0247	0.8992	1.6508
Proposed method	0.0155	0.9633	55.5314

It can be seen from the segmentation results (Figures 3d) that different objects such as a river flowing across land or the runway of the airport are clearly segmented without the speckle noise effect. However, some false classification points may exist when the distribution of grey value is complicated. For example, further refinement is needed when there is a series of overlapping pixels overlying both the land and river.

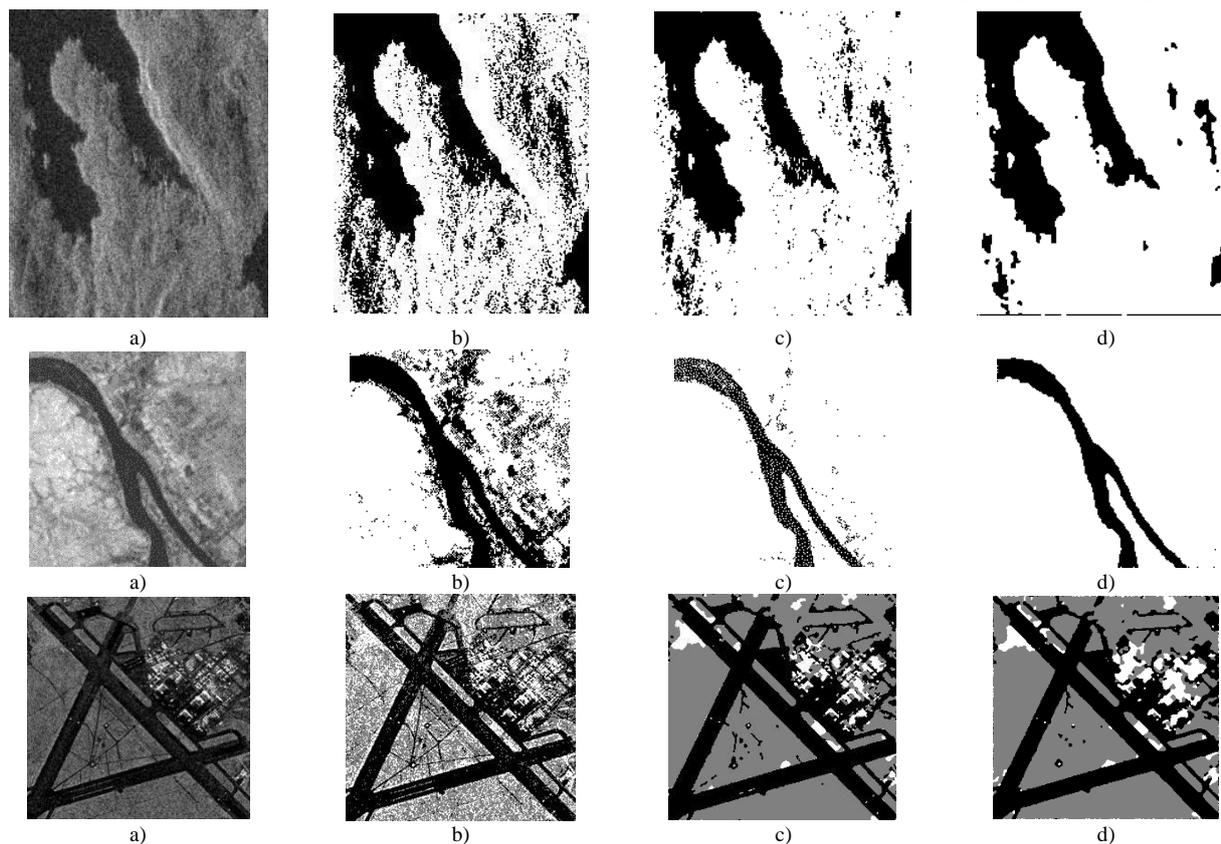


FIGURE 3 Real SAR image and results of segmentation

6 Conclusion

In this paper, a novel neighbourhood system and energy function of MRF model are proposed to improve the accuracy of the image segmentation in remote sensing applications. The proposed method has been applied on SAR image segmentation and the experimental results

show that the proposed method can obtain better segmentation effect with the optimized parameters compared to other methods. However, it has a disadvantage of higher complexity resulting in slightly slow processing. It is planned to reduce the complexity of the proposed method in future studies.

References

- [1] Yuan D B, Cui X M, Xiu W Y, Hong X Q, Yu S W 2014 Research on the application of SIFT algorithm in UAV remote sensing image feature extraction *Journal of Digital Information Management* **12**(2) 67-72
- [2] Curlander J C, McDonough R N 1991 *Synthetic Aperture Radar Systems and Signal Processing* Wiley Interscience New York
- [3] Oliver C, Quegan S 1998 *Understanding Synthetic Aperture Radar Images* Artech House Norwood
- [4] Dong Y, Forster B C 1997 Segmentation of radar imagery using Gaussian Markov random field models and wavelet and transform technique *IEEE Transaction Geoscience and Remote Sensing symposium* **4**(2) 2054-6
- [5] Miguel L L, Parmiggiani M F 2006 Contextual approach for oil spill detection in SAR images using image fusion and Markov random fields *Circuits and Systems IEEE Midwest Symposium on Circuits and Systems* **2**(6) 137-9
- [6] Liu G F, Wu Y, Zhang P, Jia L, Liu H W 2014 PolSAR image classification based on Wishart TMF with specific auxiliary field *IEEE Geoscience and Remote Sensing Letters* **7**(11) 1230-4
- [7] Elia C D, Ruscino S, Abbate M, Aiazzi B, Baronti S, Alparone L 2014 *IEEE Journal of Selected topics in Applied Earth Observations and Remote Sensing* **7**(4) 1116-26
- [8] Tadic V B 2010 *IEEE Transaction on Information Theory* **56**(12) 6406-32
- [9] Geman S, Geman D 1984 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6) 721-41
- [10] Luthon F, CaPlier A, Lievi M 1999 Spatiotemporal MRF approach to video segmentation: application to motion detection and lip segmentation *Signal Processing* **76**(1) 61-80
- [11] Lglesias J E, Sabuncu M R, Leemput K V 2013 Improved inference in Bayesian segmentation using Monte Carlo sampling: application to hippocampal subfield volumetry *Medical Image Analysis* **17**(7) 766-78
- [12] Zhang D 2011 *Spatio-temporal Markov random field based dynamic texture segmentation* Harbin Engineering University Harbin

Authors	
	<p>Jiajing Wang, born in April, 1985, Harbin, Heilongjiang, P.R. China</p> <p>Current position, grades: the Associate Engineer of No.92677 Unit of PLA, China. University studies: B.Sc and M.Sc. in Information and Communication Engineering from Harbin Engineering University in China. Scientific interest: remote sensing image, SAR image processing. Publications: 3 papers. Experience: 2 scientific research projects.</p>
	<p>Shuhong Jiao, born in 1966, Harbin, Heilongjiang, P.R. China</p> <p>Current position, grades: Professor at the College, Information and Communication Engineering at Harbin Engineering University, China. Member of Graphic Image groups and biomedical engineering society of Heilongjiang. University studies: Ph.D in Information and Communication Engineering at Harbin Engineering University in China. Scientific interest: image processing, machine vision and precision guidance. Publications: 40 papers. Experience: teaching experience of 20years.</p>
	<p>Zhenyu Sun, born in August, 1985, Dalian, Liaoning, P.R. China</p> <p>Current position, grades: Associate Engineer of No.91550 Unit of PLA, China. University studies: B.Sc in Electrical Engineering at Harbin Engineering University in China. Scientific interest: remote sensing technology, image processing. Publications: 3 papers. Experience: researching applications of remote sensing technology for 5 years. 2 scientific research projects.</p>

Formal resource request representation for remote environment control system

Shengang Hao^{1*}, Zhang Li¹, Gao Xin²

¹School of computer and information technology, Nanyang Normal University, Nanyang 473000, Henan Province, China

²National center for protein science Shanghai, Institute of biochemistry and cell biology, Shanghai Institutes for Biological Sciences, CAS, Shanghai 201210, China

Received 6 August 2014, www.tsi.lv

Abstract

Many open distributed systems across Internet such as those in grid computing and RECS (remote environment control system) involve the requesting, allocation and maintenance of sorts of resources. The discovery of large amount of resources in different sites is an important issue for the design of these systems. The booming semantic Web technology provides a suitable infrastructure for the publishing, requesting and matchmaking of resources. This paper presents a generic representation for quantified resource requesting with Semantic Web. It allows the representation of complex resource descriptions such as containment hierarchies and disjoint constraints between them. A model-theoretic semantics for matchmaking with countable resources is given for this representation. A constraint-based technique for the matchmaking check with such representation is designed to ensure the correctness for remote environment control system.

Keywords: quantified resource, RECS (remote environment control system), resource matchmaking, service-oriented architecture

1 Introduction

Environment monitoring system is being paid more attention. Distributed environment monitoring system is more and more widely used [1], especially the design and verification of embedded system in environmental monitoring is the guarantee of successful use of environmental monitoring.

A lot of remote environment control system across Internet involve the requesting, allocation and maintenance of many sorts of and large amount of resources in different sites. In e-Commerce, for example, a customer may issue a request to a shop for a quantity of goods. A travel agent may book a number of airline tickets from an airline agent and a number of apartments from a hotel agent. In the field of grid computing, tasks may require for different types of computational resources of certain amounts, such as computers, their memories and disk space, and bandwidth with networks. Most of these Internet applications involve interactions between heterogeneous information sources and agents in open environments, in which the problem of interoperability between the heterogeneous sources is a big issue.

Semantic Web [2] is a booming technology to achieve semantic-level interoperability based on XML. It was motivated to have information sources machine-understandable and agent-sharable by means of annotating their content with common data model and shared ontology. Semantic Web is especially suitable for the task of resource discovery across Internet. First, ontology technology provides a means to conceptualize and manage

different sorts of resources, and to specify resource advertisements and requests. Second, the employment of publicly standardized semantic Web specifications helps to achieve interoperability for the interaction between resource requesters, providers and brokers.

The main concern of this paper is the representation for quantified-resource matchmaking between resource advertisements and resource requests. Quantified resource requesting is mostly investigated in the field of grid computing [3-5], whereas few works is known about quantified-resource matchmaking in the context of e-Commerce although it should have more extensive applications in the area and manifest more complex forms. Our work thus mainly focus on two extensions: one is to allow advertising summarized resource descriptions; another is to allow more expressive queries for quantified resources.

2 Resources, resource advertisements and resource requests

2.1 RESOURCES

The term “resource” is extensively and freely used in information field without a widely-accepted accurate definition. We view resources as anything that is of certain degrees of utility and capacity to some competing processes. In the fields of computer sciences especially grid computing, typical resources include computers, memories, CPU time, disks, printers, network bandwidth, or even programs and data sources.

* Corresponding author e-mail: nythsg@sina.com

In e-commerce, typical resources include various sorts of goods, traffics, energy supplies, human resources, and etc. Resources as a whole can be classified along different dimensions according to features such as if a resource is consumptive, divisible, and sharable. For the context of this paper, we are only concerned with classification based on the ways they are represented, advertised and queried. Since a resource can be either an individual, or a collection of individuals, or an amount of substances or energy, mainly we distinguish resources between resource elements which are individual resource items, and resource portions which may contain other resources. A resource portion is either countable in that it consists of a finite set of resource individuals, or uncountable such as water and fuel in that they are considered to be continuously divisible. Resource portions are main concern of this paper.

2.2 RESOURCE ADVERTISEMENTS

To allow resource discovery across Web, we assume an open architecture in which resource owners advertise their resources in a public resource advertisement base, and resource requesters issue resource requests to the resource advertisement base for availability. It is impractical to register all the resource items in the resource advertisement base when the quantities of resources are so many.

Rather it is reasonable to allow a summarized advertisement for each type of resources. For instance, a resource advertisement base might advertise that there are 50 computers in a LAN rather than list each of them. Furthermore we claim that it is useful to allow multi-view descriptions and hierarchical descriptions in resource advertisements. For an example of multi-view description, it might be advertised that a laboratory has 5 servers and, at the same time, 20 computers installed with Unix. They are multi-view description in that they describe the same resource repository with different capacities.

Hierarchical descriptions involve the representation of inclusive relations between different resource repositories and resources capacities. An example of hierarchical descriptions: "Computing Centre has 2 labs, one lab has 40 PC-486s, the other has 30 PC-586s". It is our objective to extend the existing approach with such multi-view descriptions and hierarchical descriptions.

2.3 RESOURCE REQUESTS

While complex resources are common in e-Commerce, the issue has not been addressed in existing grid-oriented resource request languages [2-3].

Although complex resources could be represented as composition of atomic ones with logic connectives, e.g., using logical conjunction to express two portion of resource as a whole such as "9 PCs and 2 workstations". Such approach may cause confusion when two portions of resources are not disjoint. For instance, "3 professors and

2 female teachers" may denote a set of 3, 4, or 5 teachers depending on the number of female professors in the set. Sometimes such description needs to be clarified with clearer alternatives such as "3 professors plus, in addition, 2 female teachers" or "3 professors including 2 women" which imply respectively the use of exclusive-joining and inclusion between resource portions. Below is a more complex example illustrating the usage of resource exclusive-joining and inclusion:

CS department of Beijing Institute of Technology (BIT) might select a group of senior scholars as the doctoral thesis-defence committee members for A PhD student whose thesis is about the combination of grid and agent. The requirements for the committee members might be specified based on university-policy as follows:

- 1) There must be 7 scholars who are all computer-science professors in Beijing.
- 2) At least 4 of them must be out of BIT.
- 3) At least 3 are experts in grid.
- 4) At least 3 are experts in agent.
- 5) In addition, a secretary for the defence should be selected who must be department teacher with PhD degree in computer science.

This human resource requirement shows how a complex resource request could be composed of simpler ones with joining, exclusive joining, and inclusion. Both (a) and (e) should be included but they should be disjoint. Groups corresponding to (b), (c), and (d) may not be disjoint, and all these 3 groups are included in group corresponding to (a). Later we will show how such requests would be formulated in our representation.

3 Complex resource representation based on semantic web

3.1 RESOURCE ONTOLOGY

In our framework, different forms of resources, including resource repositories, resource portions, and resource items, are uniformly modelled as resource objects. The reason is to gain representational uniformity and simplicity for reasoning with the hierarchical relation.

First, we assume a root class *ResourceObject* for all the resource objects, and its 2 subclasses *ResourceElement* and *ResourcePortion*. In class *ResourcePortion* 2 roles include and disjoint are defined which denote respectively the containment and disjoint relation between two resources. In description-logic style these are written as as:

$$ResourceElement \sqsubset ResourceObject$$

$$ResourcePortion \sqsubset ResourceObject \cap (\forall \text{ include } ResourceObject) \cap (\forall \text{ disjoint } ResourceObject)$$

For our purpose of quantified resource matchmaking, class *QtPortion* are especially defined which inherits *ResourcePortion* and additionally defines 2 roles quantity and *elementClass* which respectively denote how many and what type of resources elements are declared.

$QtPortion \subset ResourcePortion \cap (=1 \text{ quantity Number}) \cap (\geq 1 \text{ elementClass Class})$

Here the value of attribute *elementClass* is in itself a description-logic class constructor which must be a subclass of *ResourceElement*.

QtPortion is divided into two subclasses *DQtPortion* for discrete portions and *CQtPortion* for continuous portions. *ResourceElement* is also divided into two subclasses *DResourceElement* and *CResourceElement*.

In addition to these resource-related concepts, the ontology also includes assertions regarding the properties of these concepts.

For example, "For *QtPortion* *r1* and *QtPortion* *r2*, if the *elementClass* of *r1* *elementClass* of *r2* are disjoint, then $disjoint(r1, r2)$ is true. This might be represented as a *RuleML* rule in the logic layer of semantic Web infrastructure.

3.2 REPRESENTATION OF QUANTIFIED RESOURCE ADVERTISEMENT WITH RDF

In our framework, a resource advertisement base declares a set of resource object instances linked with role include. A resource advertisement base is represented as a set of RDF statements which are subject-predicate-object triples:

- 1) University BIT has 100 classrooms.
- 2) 70 of (1) are multi-media enabled.
- 3) 40 of (1) are large ones that can hold 200 students.
- 4) 50 of (1) are middle ones that can hold 100 students.
- 5) 10 of (1) are small that hold 50 students.
- 6) All large classrooms are multi-media enabled.

For such advertisement, parts of predicate-form RDF statements are as follows:

```
advertise( r0): isa(r0, DQtPortion);elementClass(r0,
Classroom);quantity(r0, 100);
isa(r1, DQtPortion);include(r0, r1);elementClass(r1,
MediaClassroom);quantity(r1, 70);
isa(r2, DQtPortion); include(r0, r2);elementClass(r2,
LargeClassroom); quantity(r2,40);
....
```

where *MediaClassroom* is assumed to be defined in the ontology as the subclasses of *Classroom* and subsumes *LargeClassroom*.

3.3 RESOURCE REQUEST SPECIFICATIONS

While resource advertisements specifies a set of resource instances, a resource request specifies a pattern of resource objects that is to be matched against the declared resource advertisements. As pattern resource request generalizes resource advertisement by introducing pattern variables (prefixed with '?' in below) as well as constraints between them. For example, the request of example (5) in section 2.3 can be formulated as follows:

- 1) *Request* (?X, ?Y, ?Z1, ?Z2, ?Z3);
- 2) *disjoint*(?X, ?Y); *isa*(?X, *DQtPortion*); *quantity*(?X, 7);

3) *elementClass*(?X, *Scholar*[major:computer-science, title:professor, location: Beijing]);

4) *isa*(?Y, *Teacher*);
institute (?Y, bit); 5) *department*(?Y, cs_dept);
degree (?Y, phd_cs);

6) *include* (?X, ?Z1); *isa*(?Z1, *DQtPortion*);
quantity (?Z1, 4);
elementClass(?Z1, *Scholar*[institute ≠ bit]);

7) *include* (?X, ?Z2); *isa*(?Z2, *DQtPortion*);
quantity (?Z2, 3); 8)
elementClass(?Z2, *Scholar*[expertise : grid]);

9) *include* (?X, ?Z3);
isa(?Z3, *DQtPortion*);
quantity(?Z3, 3);
elementClass(?Z3, *Scholar*[expertise: agent]).

In the request specification, RDF-triples are written as binary predicate form, and a frame-like syntax is adopted to denote a specialization of class with role constraints.

4 A semantic model for resource matchmaking

The problem of quantified resource matchmaking with our representation can be formulated as follows: Given a resource advertisement base specified in form as presented in section 3.2, and a resource request specified in form as presented in section 3.3, how can we decide if the request is satisfied with the resource advertisements as a whole, i.e., if the sorts and the amounts of resources specified in a resource request is available in the collection of resources specified in a resource advertisement base? To clearly define the problem, a formal semantics for the representation is necessary.

Definition 1: A resource matchmaking specification is a triple $(O1, O2, A, Q)$ where:

1) *O1* is an ontology, called base ontology, which consists of a hierarchy of first-order classes together with their respective roles;

2) *O2* is an ontology based on *O1* consisting of a hierarchy of second-order classes with root *DQtPortion*, which has roles *elementClass*, *quantity*, *disjoint* and *inclusion* as described in previous section;

3) *A* is an advertisement base formed as *advertise*(*r*): *Tr* which publish resource *r* with a RDF description denoting its hierarchical composition with role inclusion;

4) *Q* is a resource request formed as *request*(*X*): with a finite set *X* of resource variables and a finite set of constraints *C_x* between the variables.

The following question is, given a resource matchmaking specification and an allocation of it, what does mean by "The resource request is certifiable with the resource advertisement". A semantic formalization of our quantified resource representation is thus necessary.

Definition 2. Given a resource matchmaking specification $R = (O1, O2, A, Q)$, an interpretation of *R* is a triple $I = (U, E, [.])$, where *U* is a set of individuals, $E \subset U$ is the set of all individuals of resource items, $[.]I$ is a mapping from any expression in *R* to a set-theoretic construct over *U* such that

- 1) For a class name c in $O1$, $[c] I \in power(U)$, especially $[ResourceElement] I = E$; for an role r in $O1$, $[r] I \in power(U \times U)$;
- 2) For any class c , subclass $c1$ of c , and instance a of c in R , $[c1] I \subset [c] I$; $[a] I \in [c] I$;
- 3) The conventional description logic constructors as well as subsumption relation in $O1$ are the same as those of conventional description logic;
- 4) $[DqtPortion] I = power(E)$; $[quantity] I$ is a function in $power2(E) \rightarrow N$, such that for any $x \in power(E)$, $[quantity] I(x) = |x|$, i.e., the number of elements in x ; $[elementClass] I$ is a function in $power2(E) \times power2(E)$, such that, for any $x, y \in power(E)$, $(x, y) \in [elementClass] I$ iff $x \subset y$; $[include] I \in power2(E) \times power2(E)$, such that for any $x, y \in power(E)$, $(x, y) \in [include] I$ iff $x \supset y$; $[disjoint] I \in power2(E) \times power2(E)$, such that for any $x, y \in power(E)$, $(x, y) \in [disjoint] I$ iff $x \cap y = \emptyset$;
- 5) For $A = advertise(r): Tr$, $[A] I = \cup \{[r] I\}$ such that $[Tr] I$ is true};
- 6) For $Q \equiv ? request(X1, \dots, Xn)$:
 $C, [Q] I \in power2(E)$ and
 $[Q] I = \{ [X1] I, V \cup \dots \cup [Xn] I, V \mid \text{for all valuation } V \text{ of variables } \{X1, \dots, Xn\} \text{ such that } [C] I, V \text{ is true} \}$.

With this interpretation, we can define some semantic properties of a resource matchmaking specification. First, an advertisement must reflect the true containment relation between two portions of resources.

Definition 3. Let $R = (O1, O2, A, Q)$ be a resource matchmaking specification, I be an interpretation of R . I is inadmissible with respect to A iff $[A] I$ is undefined; otherwise I is admissible with respect to A . A is invalid iff all interpretations of R is inadmissible with respect to A ; otherwise A is valid.

An invalid resource advertisement description is illegal because it makes no sense. It is important to be able to check the validness via syntactic inference. An immediate observation is that if A contains an include-clause $DqtPortion[quantity: n1, elementClass: c1]$ include $DqtPortion[quantity: n2, elementClass: c2]$, and $n1 < n2$ or $c1 \cap c2 = \emptyset$ then A is invalid.

Definition 4: Let $R = (O1, O2, A, Q)$ be a valid resource matchmaking specification, I be an admissible interpretation of R . Q is satisfied with A in I iff there exists $x \in [Q] I$ such that $x \subset [A] I$. Q is satisfied with A in I . Q is unsatisfiable with A in I iff for all interpretation I of R , Q is not satisfied with A in I .

We thus established a semantic account for the satisfaction of resource request with resource advertisements.

5 Implementation and application

5.1 RESOURCE MATCHMAKING AS OBJECT CONSTRAINT SATISFACTION

To implement the matchmaking between a complex resource request and a resource advertisement, we take the matchmaking problem as one of object constraint satisfaction (OCS) [13, 14]. The variables of an OCS are resource variables in the resource request, which ranged over instances of *DqtPortion*; the constraints are role constraints in the resource request. The domains of the constraint variables consist of *DqtPortion* instances generated by joining finite number of sub-portions of resource portions in the resource advertisements. For the allocation to be operable, we stipulate that all the sub-portions are from among a set of mutually disjoint resource portions. To make the idea clearer, we give the following definition:

Definition 5: Let $R = (O1, O2, A, Q)$ be a valid resource matchmaking specification. VQ and CQ are respectively the resource variable set and query constraint of Q and quota out of A is a set of pairs $\alpha = \{s1/r1, \dots, sn/rn\}$ here $r1, \dots, rn$ are nodes in A , which satisfied following conditions:

- 1) $s1, \dots, sn$ are respectively sub-portions of $r1, \dots, rn$ in that include (ri, si) holds for each i ;
- 2) $s1, \dots, sn$ are mutually disjoint, i.e., $disjoint(si, sj)$ holds for each i and j ;
- 3) the quantity of si is determined.

For each subset R of $\{s1, \dots, sn\}$, let JR be a new instance of *DqtPortion* by joining all the resource portions of R in following way:

- 1) the quantity value of JR is the sum of those of all the resource portions of R ;
- 2) the element Class value of JR is the DL-union of those of all the resource portions of R ;
- 3) the set of include values of JR is R ;
- 4) the set of disjoint values of JR is the intersection of those of all the resource portions of R .

An assignment of Q with quota α is a mapping λ which maps each resource variable in Q to a subset S of $\{s1, \dots, sn\}$. λ is an allocation of A to Q iff when each resource variable X in Q is replaced in CQ by $J\lambda[X]$, the instantiated constraint is satisfied with A as defined in definition 4.

A resource matchmaking algorithm based on this idea thus need to find one or more mutually disjoint sub-portions of advertised resource portions that satisfied the constraint of the request. The constraint-solving algorithm is currently under development.

5.2 APPLICATION BACKGROUND

The research aims at resource management in an ongoing multi-agent education management system for college. The multi-agent system consists of two set of agents. One is a set of resource agents, such as estate agents, human resource agents, and textbook agents, which provide

services of resource requesting, booking, and allocation. The other set of agents are task agents, such as department clerks, which perform task planning, scheduling, monitoring and execution. The requesting and allocation of resource are important parts in the interaction between the task agents and resource agents. Despite the diversity of various sorts of resources, the behaviours of the resource agents are quite similar [15-16]. Thus, a generic framework for resource modelling is necessary.

6 Conclusion

In this paper, we proposed a representation for remote environment control system of quantified resource matchmaking with a number of novel features. First, it

allows the representation of complex resource requests and advertisements with quantified resource quota, containment hierarchies and disjoint constraints. This enhances the flexibility and expressiveness of the representation. To give an accurate definition of the resource matchmaking with such representation, a semantic theory is established. Second, it is semantic-Web-oriented in that the representation follows conventions of RDF and semantic Web ontology. In addition, the resource-servicing architecture with summarized resource advertisement repository cooperating with resource-requesting agents is in line with the spirit of semantic Web and is suitable for wide range of remote environment control system.

References

- [1] Berners-Lee T, Hendler J, Lassila O 2012 The Semantic Web *Scientific American* **284**(5) 34-43
- [2] Bock C, Gruninger M 2012 PSL: A semantic domain for flow models *Software and Systems Modeling Journal* **186**(10) 35-41
- [3] Liu C, Foster I 2011 A constraint language approach to grid resource selection *University of Chicago Chicago* 60-81
- [4] Liu C, Yang L, Foster I, Angulo D 2012 Design and evaluation of a resource selection framework *In Proceedings of the Eleventh IEEE International Symposium on High-Performance Distributed Computing* Edinburgh Scotland 386-92
- [5] Tangmunarunkit H, Decker S, Kesselman C 2009 Ontology-based resource matching in the grid - the grid meets the semantic web *International Semantic Web Conference* 706-21
- [6] Bechhofer S, Van Harmelen F, Hendler J, Horrocks I, McGuinness D L, Patel-Schneider P F, Stein L A 2009 OWL web ontology language reference Available at <http://www.w3.org/TR/owl-ref/>
- [7] Horrocks I, Patel-Schneider P F, Van Harmelen F 2002 Reviewing the design of DAML+OIL: An ontology language for the semantic web *In Proc. of the 18th Nat. Conf. on Artificial Intelligence (AAAI 2002)* 174-9
- [8] The OWL Services Coalition 2013 OWL-S: Semantic Markup for Web Services Available at <http://www.daml.org/services/owl-s/1.0/owl-s.html>
- [9] Chuanyun X, Yang Z, Dan Y 2011 Ontology based image semantics recognition using description logics *IJACT International Journal of Advancements in Computing Technology* **3**(10) 1-8
- [10] Sycara K, Wido S, Klusch M, Lu J 2002 Larks: Dynamic matchmaking among heterogeneous software agents in cyberspace *Autonomous agents and multi-agent systems Kluwer academia publishers* **5** 173-203
- [11] Fadel F G, Fox M S, Gruninger M M 1994 A generic enterprise resource ontology *Proceedings of the Third Workshop on Enabling Technologies Infrastructures for Collaborative Enterprises* West Virginia University
- [12] Raman R 2010 Matchmaking frameworks for distributed resource management *Computer Science* **96**(2) 56-64
- [13] Palla G, Derenyi I, Vicsek T 2007 The Critical Point of *k*-groups Percolation in the Erdos-Renyi Graph *Journal of Statistical Physics* **128**(1) 219-27
- [14] Palla G, Barabasi A-L, Vicsek T 2007. Community dynamics in social networks *Noise and Stochastics in Complex Systems and Finance* **6601**(3) 273-87
- [15] Chuanyun X, Yang Z, Dan Y 2011 Ontology based image semantics recognition using description logics *IJACT: International Journal of Advancements in Computing Technology* **3**(10) 1-8
- [16] Kim J-M Porter A 2009 A history-based test prioritization technique for regression testing in resource constrained environments *In ICSE '09 Proceedings of the 31th International Conference on Software Engineering* **139**(3) 119-29

Authors	
	<p>Shengang Hao, born in October, 1977, Nanyang County, Henan Province, China</p> <p>Current position, grades: the lecturer in School of Computer ,Nanyang Normal University, China. University studies: M.Sc. in Computer Applications from Wuhan University of Science & Technology in China. Scientific interest: computer modeling, information retrieval. Publications: 8 papers. Experience: teaching experience of 13 years, 5 scientific research projects.</p>
	<p>Li Zhang, born in May, 1978, Nanyang County, Henan Province, China</p> <p>Current position, grades: the associate professor in School of Computer ,Nanyang Normal University, China. University studies: M.Sc. in Computer Applications from Hunan University in China, Ph.D from Beijing Institute of Technology, China. Scientific interest: data mining, data recovery. Publications: 12 papers. Experience: teaching experience of 10 years, 6 scientific research projects.</p>
	<p>Xin Gao, born in June, 1979, Pudong District, Shanghai, China</p> <p>Current position, grades: the engineer, National Center for Protein Science Shanghai, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, CAS. University studies: received her M.Sc. in Computer Applications from Shandong University of Science & Technology in China. Scientific interest: her research interest fields include software engineering, formal modelling. Publications: more than 10 papers published in various journals. Experience: She has teaching experience of 11 years, has completed 4 research projects.</p>

Study on load capacity-based cascading failure model in the computer network

Chang-jun Han*

College of Information Technology Eastern Liaoning University, DanDong City, LiaoNing Province, China, 118003

Received 11 July 2014 www.tsi.lv

Abstract

Studies on the cascading failure process and characteristics in the computer network are beneficial to guiding the system construction and improving the performance. Based on the load initialization capacity of the computer system, routing control strategies and node forwarding rate, this paper constructed one cascading failure model considering service performance in the computer network specific to the influence of cascading failures on the service performance of the computer system network. This model considered multiple influence parameters and effectively measured the variable values of influence parameters of cascading failures on the service performance of the computer system network. Through comprehensive analyses, this model can effectively provide practical guiding significance for the prevention and control of cascading failures in the network.

Keywords: load capacity, computer network, service performance, cascading failure

1 Introduction

In recent years, people have become more and more dependent on the computer network technology and its safe reliability has been paid more and more attention to, along with higher security requirements. Though various remedial measures have been taken, they still cannot completely prevent and control the cascading failure in the network [1] and other problems. Cascading failure refers to that some failure nodes or edges due to natural fault or artificial attack cause the redistribution of 'flow' on nodes or edges by the coupling relationship between nodes or edges and further trigger the problems of other nodes or edges. The chain-reaction finally may lead to the collapse of the network. In aspects of cascading failure in the computer network, scholars at home and abroad have carried out various profound studies and at the same time proposed many security models relevant to cascading failure. For the status in quo of real computer network, computer cascading failure model that utilizes relevant characteristics of load capacity can better conform to the demands. Thus, studies in this aspect also are more than other aspects. However, due to the initial load and weight given by the network nodes, when the network is not smooth or fails, it will redistribute the load capacity according to the scheduled rules. Therefore, studies on computer cascading failure in this aspect are hotspot. However, there are still many defects of those research studies at home and abroad [1-4]. First, the initial load is preset and presents a monotonously linear relationship. However, the results of computer network data obtained by certain routing control strategies showed that there was no functional relationship between the initial load and load

capacity. Second, data packet transmission, forwarding rate of network nodes and routing control strategy may generate various load changes due to cascading failure [2]. Third, current studies on computer network is partial to the network topology while ignores the service performance of the computer in the real network. In the practical situation, users pay more attention to the influences of the service performance of the computer network. Thus, constructing one cascading failure security model in consideration of both service performance of the computer network and load capacity shall provide practical guiding significance for the effective prevention and control of cascading failures.

According to various characteristics of the computer network communication technology, for example, data packet transmission, forwarding rate of network nodes and routing control strategy, this paper constructed one security model of the network data packet transmission [3]. In premise of this, associated with the change factors of load and the non-linear characteristics between the initial load and load capacity, this paper constructed the cascading failure security model in the computer network. Under the condition of guaranteeing the normal computer network topology, this paper designed various parameter factors influencing the network service performance [4] and made a comprehensive analysis and study on the detailed generating process of cascading failures.

2 Cascading failure model in the computer network

The formal description of the definition of the security model is as follows. Simple and undirected graph represents the computer network topology, noted as

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

$G(V,E)$; $V=\{v_1,v_2,\dots,v_n\}$ represents the set of various hosts servers and network connection devices; and $E=\{e_1,e_2,\dots,e_j\}$ represents the physical link (with weighted value) set in the computer network connection and host server. According to the above formal description and in premise of data packet transmission mechanism in the network, one new computer network transmission model is established, DTS model for short.

2.1 CONSTRUCTION OF DTS MODEL

The construction of the security model of data packet transmission in the computer network needs the following assumptions as theoretical foundation, mainly including,

- 1) the general computer network routing selects OSPF computer network protocol;
- 2) most computer networks adopt fiber optic connection, so the physical influence of transmission lines on stream data transmission is not taken into consideration;
- 3) in case of the data packet dropout of the target network node, non-retransmission mechanism is adopted to maintain the unipolarity;
- 4) except the possibility that the computer network interchanger connects with the external interchanger LAN (local area network), only LAN of local interchanger, router or terminal node of each network is used to connect relevant network data information.

On the basis of the above assumption, DTS model is constructed as below. At time point t and data packets occurs at network node v_i and each will randomly select target network node v_j ; and the data packet transmitted between two network modes is noted as U_{ij} . Through the OSPF protocol, the shortest path $P_{ij}(t)$ can be obtained and then forwarded and after it passes the target network node, it will be removed; as is said in the assumptions, if data packet loses in the forwarding process, it fails and cannot reach the target network node. The forwarding capacity is noted as F . F values of different network nodes differ from each other. If the same network node has non-transmitted data packets, these data packets will be left at the end of buffer queue of the network node. Based on the subject theory of information science and under the condition of data packet transmission rules of network nodes, DTS model is defined as below:

$$L_i(t) = \begin{cases} 0, & Q_i(t) \leq F_i \\ Q_i(t) - F_i, & F_i < Q_i(t) < C_i \\ Q_i(t), & Q_i(t) > C_i \end{cases} \quad (1)$$

$$Q_i(t) = L_i(t-1) + S_i(t) + R_i(t), \quad (2)$$

$$F_i = \eta \times d_i, \eta \geq 1, \quad (3)$$

where the above formal description of definitions, the expressive connotation of each symbol is as follows, $L_i(t)$

represents the network load capacity of v_i at t ; $Q_i(t)$ represents the to-be-transmitted data packets of v_i at t ; $L_i(t-1)$ represents the network load capacity of v_i at $t-1$; $S_i(t)$ represents the data packet forwarding rate of v_i at t ; $R_i(t)$ represents the transmitted data packets of v_i received from the neighbouring network nodes at t ; d_i represents the network node degree; η represents F_i coefficient (the coefficient of forwarding capacity) and in a network system of single autonomous area, there is a linear relationship between F_i and d_i ; and C_i represents the maximum forwarding load by network node v_i .

2.2 LOAD CAPACITY-BASED CASCADING FAILURE MODEL IN THE COMPUTER NETWORK

In this model, there are three influence factors of the load changes of network nodes and they may result in abnormality or failure of the computer network, which further may create conditions for cascading failures. These three factors are as follows, first, the changes of the control modes of network routes, which can influence the number of received data packets; second, the forwarding rate of data packet; third, the forwarding capacity of data packets, F . Figure 1 shows the changes of the relationship between the load of the network node and network throughput. When the network node load is small, its forwarding capacity F is enough, the network data packet R is not large, and there is no redundant data packet information in the queue. While when the network node load is large and the forwarding capacity F reaches the maximum, it is easy to generate the phenomenon of saturation and there is much redundant data packet information in the queue. The larger R is, the more prominent the redundancy and congestion of network nodes is. When the network load increases again and exceeds certain index line C , its network throughput will suddenly decrease to void value. At this time, the cascading failure occurs at this network node. The saved capacity of each network node can be described as below:

$$C_i = \alpha \times BC_i, \alpha \geq 1, \quad (4)$$

C_i represents the maximum forwardable load of network node v_i and BC_i represents the mediation number of the shortest path of the network node; represents the coefficient parameter value of C_i , indicating the size of the network node capacity. It can be known from the definition that there is a linear relationship between BC_i and C_i and the coefficient parameter value is taken as the change index. The larger BC_i is, the stronger the forwarding capacity of data packets of network nodes is. To ensure the normal operation of the computer network, C_i also needs synchronous update.

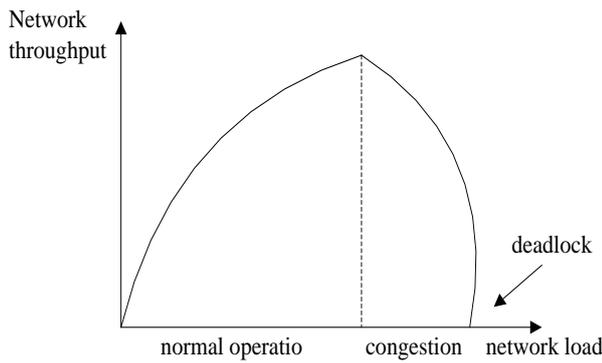


FIGURE 1 Changes of the relationship between the load of network nodes and network throughput

The above mentioned network initial load is based on the previously constructed DTS model, which is different from the traditional network initial load. For the latter, network topology model with uniqueness is the major factor while the decisive factors of the former are various, including network topology, network node load capacity and so on and it has a linear relationship with the load capacity of network nodes. It mainly is used in complex and difficult-to-be-repeatedly-constructed computer network structure model, reaching the two-layer effect results of network topology and data packet forwarding. However, under the condition that the network initial load is irrelevant to C , the actual phenomenon of computer network is more prominent and real [5].

The above DTS model adopts the shortest path control mode to simulate the OSPF network protocol of the computer network [6]. Among them, time interval is an important measure index and the time interval of each network node has a close relationship with the network node load, which can be expressed as below:

$$D_i(t) = \begin{cases} 1, & L_i(t) \leq F_i \\ 1 + \frac{L_i(t-1)}{F_i}, & F_i \leq L_i(t) \leq C_i \\ \infty, & L_i(t) > C_i \end{cases} \quad (5)$$

Equation (5) mainly describes three different conditions of network nodes, including normal operation, network congestion and cascading failure. $D_i(t)$ represents the time interval of network nodes at one moment. It can be known from each definition that the shortest path $P_{ij}(t) = \{v_i, \dots, v_k, \dots, v_j\}$ represents the calculation mode of the shortest path in the OSPF network protocol; U_{ij} forwards network nodes according to this mode and when cascading failure occurs, evade the occurrence network node and adjust $D_i(t)$ to the maximum so as to guarantee the normal operation of other network nodes, reaching the invariability of the structure of the whole computer network. Therefore, when cascading failure occurs, the whole network performance changes dynamically to reach the most effective condition.

It can be known from the above definitions that the whole detailed process of cascading failures can be illustrated in the following three aspects:

- 1) In the initial stage, no flow is transmitted and the load of each computer network is 0; from $t=1$, the network flow gradually increases and each network node starts to have data packet transmitted and the load of network nodes increases simultaneously. Until the preset time, the network is in normal operation stage.
- 2) If the computer network operates abnormally at $t=x$, the network transmission route will change and at this time the route control strategy will adjust the transmission route for redistribution and adjustment.
- 3) When the network routing control strategy changes, the load of network nodes fluctuates at the same time until it exceeds the preset network load capacity. At this time, the network node is in cascading failure and the routing control strategy will again adjust the transmission route for redistribution and adjustment. At this time, the cascading failure into the computer network generates.

3 Application analysis of cascading failure model

3.1 EVALUATION COEFFICIENTS OF THE PERFORMANCE

The evaluation coefficients of cascading failure performance in the computer network play a key role in DTS model and load capacity-based cascading failure model in the computer network [7]. The evaluation coefficients of the performance adopted by current studies mostly are designed and analyzed in aspects of network topology. However, they fail to effectively and comprehensively reflect the cascading failure in the computer network and the changes and relationship of its network service performance [8-10]. In this case, in premise of its service performance, various evaluation coefficients of cascading failure performance are proposed, including service delay time interval, load change rate and network throughput. Below is the detailed analysis of them.

Service delay time interval:

$$DA(t) = \frac{2}{N'(t)(N'(t)-1)} \times \sum_{i \neq j=1}^{N'} \Phi_{ij}(t), \quad (6)$$

$$\Phi_{ij}(t) = \sum_{v_k \in P_{ij}(t)} D_k(t).$$

where, $DA(t)$ represents the service delay time interval; $N'(t)$ represents the number of network nodes in the computer network free from cascading failure at t ; and $\Phi_{ij}(t)$ represents the sum of time interval changes of each network node in $P_{ij}(t)$. $DA(t)$ can express the time interval size of the forwarded data packet between each network node and it shows a proportionally change relationship.

Load change rate:

$$LA(t) = \frac{\sum_{i=1}^N L_i(t)}{N}, \tag{7}$$

where $LA(t)$ represents the load change rate of the computer network and it indicates the number of non-transmitted data packets in network nodes. Usually the data level is at millions. Moreover, N represents the number of network nodes and $L_i(t)$ represents the load of network nodes at time interval t . The change of its LA can demonstrate the relationship of buffer queue change in the computer network and it also presents a proportionally change relationship. At the same time, it influences the network service performance.

Network throughput:

$$TS(t) = \sum_{i=1}^N FR_i(t), \tag{8}$$

$$FR_i(t) = \begin{cases} Q_i(t), & Q_i(t) \leq F_i \\ F_i, & F_i < Q_i(t) \leq C_i \\ 0, & Q_i(t) > C_i \end{cases}$$

where, $TS(t)$ represents the computer network throughput (at time interval t), with unit of Mpps. Under the real condition, the computer network throughput represents the number of the treated and controlled data packets of each network node or port in unit time; N represents the number of computer network nodes; $FR_i(t)$ represents the data packet transmission rate of network nodes at time interval t . It can be known from the description that TS is one key factor and it can comprehensively highlight the change curve of the relationship between network throughput and time t and at the same time it also states the change relationship of network service performance between the cascading failure-free interval and the cascading failure interval, which also presents a proportionally change relationship.

3.2 SIMULATION EXPERIMENT

The theoretical experiment of load capacity-based cascading failure model in the computer network adopts BA scale-free computer network with 100 network nodes. Under this condition, the application analysis of the model is made. The simulation coefficients are as follows, packet sending rate $S=1$ Mpps, simulation time $T=1000s$, the parameter of the network load capacity $\alpha = 2$ and the parameter of the data packet forwarding capacity $\eta=1$. The simulation values adopt the average.

Under the normal operation condition, the network load of the computer network starts from the initial value and then distributes and adjusts. If the network operates normally, each performance coefficient maintains a relatively stable operation state in one time interval.

The cascading failure in the computer network refers to the abnormality of network nodes under the normal network condition, including external attack, artificial faulty operation and others, which may trigger the occurrence of this case. Usually, the computer network maintains stability in the middle time ($t=50s$). When $t=100s$, part of network nodes in reality will be gradually attacked and further result in the computer network cascading failure and influence the network service performance. Figures 2-4 respectively show the changes of three performance evaluation coefficients with time.

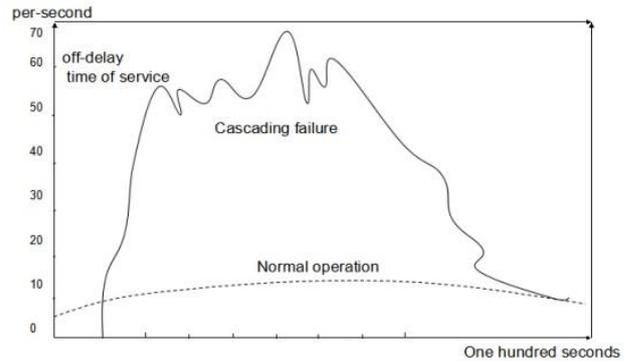


FIGURE 2 Relationship between service delay time interval and t

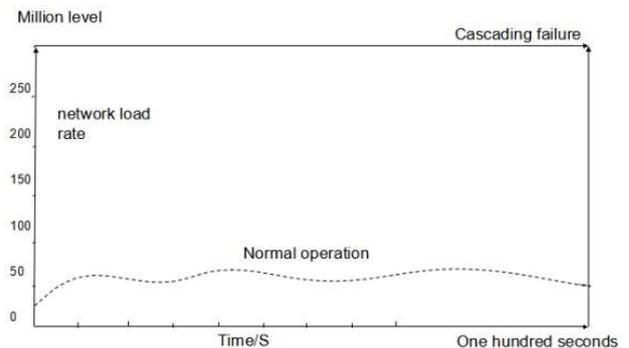


FIGURE 3 Relationship between load change rate and t

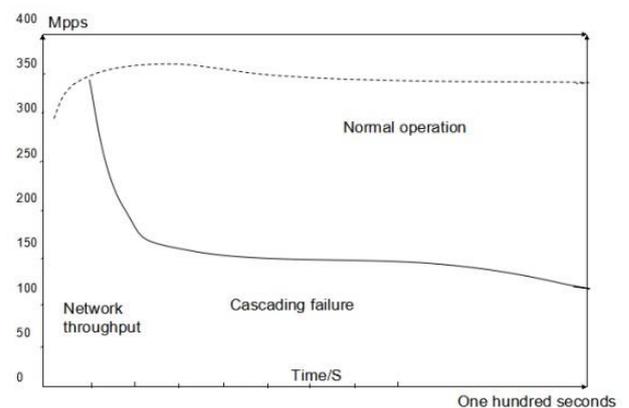


FIGURE 4 Relationship between network throughput and t

From the analysis of the simulation experiment, it can be known that, first, the cascading failure of the computer network makes the service delay time interval gradually reduce after the change; second, the packet loss strategy of network nodes can effectively decrease the network load pressure and effectively reduce its influence on the network service performance; third, the computer network

cascading failure decreases the network throughput and increase the network load pressure, which greatly influence the network service performance.

4 Conclusion

This paper constructed DTS model. In the theoretical premise of its data packet transmission, this paper designed one new cascading failure model considering multiple

factors in the computer network and made a series of application analyses on the model. Through the analysis of the simulation experiment, it is found that the change relationship between the proposed evaluation coefficients of cascading failure performance and time can effectively reflect the influence on the network service performance, which proved the feasibility of the model. In the future, further study and analysis will focus on how to effectively prevent and control the cascading failure on this model.

References

- [1] Motter A E, Lai Y 2002 Cascade-based attacks on complex networks *Physical Review E* **66**(6) 065102
- [2] Kim D H, Motter A E 2008 Fluctuation-driven capacity distribution in complex networks *New Journal of Physics* **10** 053022
- [3] Duan D, Wu J 2013 Adjustable load distribution-based cascading failure model in complex network. *System Engineering Theory and Practice* **33**(1) 123-9
- [4] Barabasi A L, Albert R 1999 Emergence of scaling in random networks *Science* **286**(5439) 509-12
- [5] Ma L, Guo P et al 2013 Cascading failure modeling of computer network considering service performance *Computer Engineering* **39**(12) 83-6
- [6] Sun H J, Zhao H, Wu J J 2008 A robust matching model of capacity to defense cascading failure on complex networks *Physica A* **387** 6431-5
- [7] Kinney R, Crucitti P, Albert R, Latora V 2005 Modeling cascading failures in the North American power grid *European Physical Journal B* **46** 101-7
- [8] Wang B, Zhoutao, He D 2005 Analysis on the latest development trend of statistical physics and complex system Research *China Basic Science* **7**(3) 37-43
- [9] Wang W X, Chen G 2008 Universal robustness characteristic of weighted networks against cascading failure *Physical Review E* **7**(2) 026101
- [10] Deng H, Wu J et al 2008 Analysis of cascading failure model for two-layer small-world networks *Computer Simulation* **10**(28) 167-8
- [11] Li Y, Wu J, Tan Y 2012 Study on the critical invulnerability of cascading failure in logistic support Networks *Journal of System Simulation* **05**(32) 161-3

Author



Chang-jun Han, born on December 25, 1978, DanDong County, LiaoNing Province, China

Current position, grades: the lecturer of College of Information Technology, Eastern Liaoning University, Handan city, Hebei province, China.

Scientific interest: computer network and fibre optic communication technology.

Publications: 4 papers.

Experience: teaching experience of 9 years, completed 5 scientific research projects.

A novel edge detection method based on 2-D Gabor wavelet

Chunya Tong*

School of Electronic and Information Engineering, Ningbo University of Technology, Ningbo, Zhejiang, 315016, China

Received 1 July 2014, www.tsi.lv

Abstract

With the features of substantial data and complex landmark, remote sensing images need a higher requirement for edge detection operator. Due to the limitations of grads operator and Canny operator in edge detection, this paper presents an edge detection method based on 2-D Gabor wavelet real part and the experimental analysis shows this method was better on edge detection.

Keywords: edge detection, 2-D Gabor wavelet, real part, remote sensing images

1 Introduction

Edge detection is the first step of analysing and understanding the image, and also is the part of image pre-processing, which is the key to dealing with many complex issues [1]. Because aerial remote sensing images have access to convenience as well as high resolution, important buildings' edge information can be obtained through edge detection. However, since that remote sensing image has large amounts of data and contains abundant information, it needs a better operator for edge detection.

Now there are some frequently-used nonlinear operators like Grads operator, Laplace operator, LOG operator, Sobel operator, Prewitt operator, Robert operator, Canny operator, direction operator and etc. Besides, there are linear operators like surface fitting and so on [2, 4]. These operators have different characteristics: Grads operators is similar to high-pass filtering, only works in sharpening the edge; Sobel operator is a kind of weighted average operator and it weighted the point near the centre to stress its edge; Robert grads operator is sensitive to noise, so it is rarely used for detecting dense point region edge. Laplace operator is rotation invariant in detecting edge, which is called isotropic. Compared with the edge, it responds stronger to the corner, the endpoint and isolated points; before the LOG operator does differential operator, firstly it smoothest image processing to reduce the noise. Therefore, it relieves the issue that differential operator is sensitive to noise yet influences the result of edge detecting. Canny operator, owing to its strictness with the setting of parameters, is conducive to automated processing of image data. Due to a variety of deficiencies of the algorithms above, it is difficult to be applied to detect the edge of remote sensing images successfully, which contain a large amount of complex landmark. Thus, this paper put forward a novel edge detection method based on 2-D Gabor wavelet.

2 2-D Gabor wavelet transform

2.1 2-D GABOR WAVELET

Being expended from one dimension of Gabor wavelet, the two dimensions Gabor wavelet is a powerful tool for multi-scale image representation and image analysis. Gabor function, as the only way to obtain the spatial and frequency domain of uncertainty relation, is often chosen as wavelet basis function [3]. The basic principle of wavelet transform is to give the representation or approximation of a signal by the convolution of the filter function with a set of signals. Function of the two-dimensional Gabor filter is a complex function, whose real parts and imaginary parts can be expressed as follows:

$$\text{Re}(\psi_j(\vec{x})) = A \times \left[\exp(\vec{k}_j \vec{x}) - \exp\left(-\frac{\delta^2}{2}\right) \right], \quad (1)$$

$$\text{Im}(\psi_j(\vec{x})) = A \times \sin(\vec{k}_j \vec{x}). \quad (2)$$

Two-dimensional Gabor wavelet's transform describes that giving a near area's grey feature of one point \vec{x} on images $I(\vec{x})$, with a convolution is defined as follows:

$$J_j(\vec{x}) = \int I(\vec{x}') \psi_j(\vec{x} - \vec{x}') d^2 \vec{x}'. \quad (3)$$

Gabor filter is a band pass filter in the spatial domain and frequency domain, which has a better ability to distinguish and a good directional selectivity in the spatial domain with good frequency selectivity in the frequency domain. Two-dimensional Gabor wavelet is easy to extract multi-scale and multi-grain direction frequency information.

* Corresponding author e-mail: 77848116@qq.com

2.2 SELECTION OF THE TWO-DIMENSIONAL GABOR FILTER BANK PARAMETERS

Parameters of two-dimensional Gabor wavelet filter bank are mainly reflected in the sampling modes of spatial and frequency domain. 2-D Gabor filter's function determines the scale by scaling and rotating a set of filters to generate its expression of the signals, and the selection of parameters is usually performed in the frequency domain. In order to sample the entire image in frequency, domain parameters of k_v , φ_μ can be chosen to represent the centre frequency and direction to describe images. Different choices of these two parameters reflect the differences in the 2-D Gabor wavelet frequency and direction of spatial sampling approach. Filter bandwidth is available from the Equation, which is expressed as follows:

$$\delta = \sqrt{2 \ln 2} \left(\frac{2^\varphi + 1}{2^\varphi - 1} \right), \tag{4}$$

where φ is half-peak bandwidth (described as octave), the relationship between φ and δ can be expressed in Table 1, where selections of the filter parameters refer to experimental data neurophysiological.

TABLE 1 Correspondence between the values of φ and δ

φ	δ
0.5	2π
1	π
1.5	2.5

The entire frequency space values from 0 to infinity. Because the actual frequency distribution of an image is a limited range, the argument k_v can be within a small range of values. Lades and his group's experiments show that for

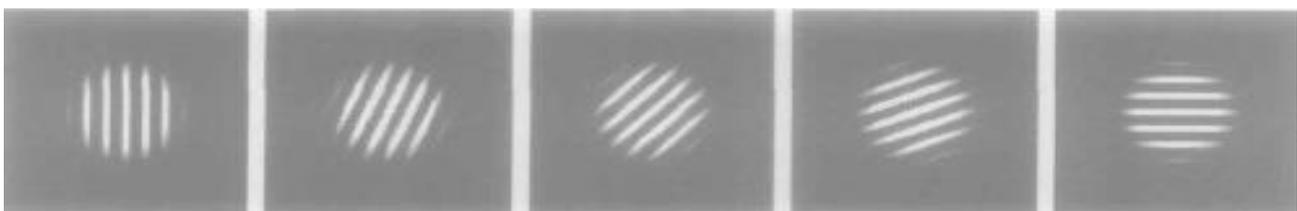


FIGURE 2 Changes in texture direction of 2D Gabor filter function with the real part of the parameter φ_μ

Centre frequency k_v represents λ_v , the wavelength of the 2-D Gabor filters:

$$\lambda_v = \frac{2\pi}{k_v}. \tag{5}$$

When the direction parameter $\varphi_\mu = 0$, the centre frequency equals $\pi/2$, $\sqrt{2}\pi/4$, $\pi/4$, $\sqrt{2}\pi/8$, $\pi/8$, the

the image at the size of $128 * 128$, when the greatest centre frequency of the filter is $\pi/2$ and the filter bandwidth is equal to 0.5 octave, the experiment works best. As the edge of the image is randomly distributed in the range of 0 to 2π , taking into account the symmetry of the Gabor filter, the actual values is in the range of $[0, \pi]$.

For the 40 Gabor filters consists of 5 centre frequency and eight directions, when v is equal to 2, μ is equal to 4, the real and imaginary parts of 2-D Gabor filter can be represented as Figure 1.

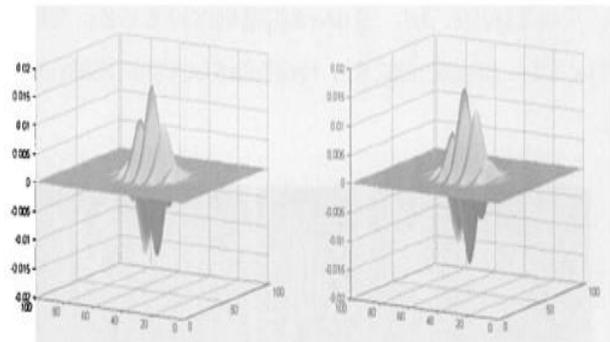


FIGURE 1 The real parts (left) and imaginary parts (right) of 2-D Gabor filter

Filter parameters φ_μ , k_v and δ represents the direction of the filter, respectively, the wavelength and the sizes of the Gaussian window. With the changes of φ_μ in the real and imaginary parts of the 2-D Gabor filter, function performs texture features in different directions.

When the centre frequency of the k is $\pi/4$, the direction parameter is 0, $\pi/8$, $\pi/4$, $3\pi/8$, $\pi/2$, the real parts of the filter function is shown in Figure 2.

Filters in different directions can respond to image texture features in the corresponding direction, when the image texture features are perpendicular to the direction of the filter, it makes response to the optimum.

corresponding wavelengths are: 4, $4\sqrt{2}$, 8, $8\sqrt{2}$, 16. In this case, the real part of the two-dimensional Gabor filter function as shown in Figure 3. With the decrease of the centre frequency, the wavelength of filter increases. And the filter of different wavelength responds to different image features.

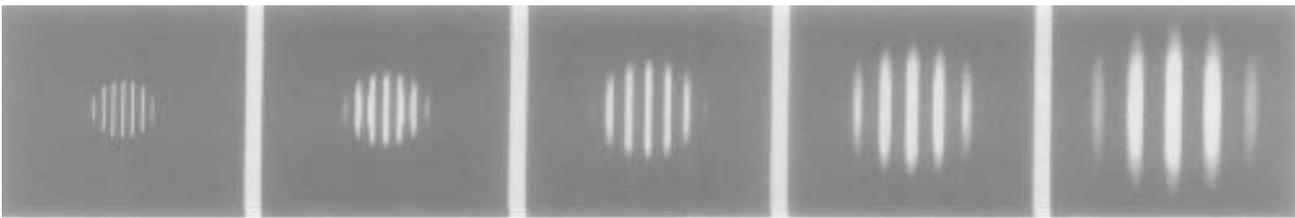


FIGURE 3 The real parts' texture direction of 2D Gabor filter function along with the changes of parameters φ_μ

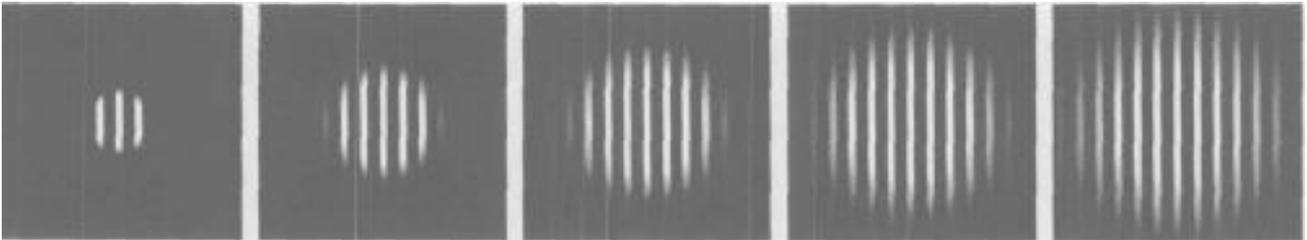


FIGURE 4 The real parts' windows of 2D Gabor filter function along with the changes of the parameter δ

$$r_v = \frac{2\sqrt{2}\delta}{k_v} \tag{6}$$

The effective radius of the Gaussian window r_v (see, Equation (6)), determines equation effective range of the image convolution. When taking centre frequency, the parameters were taken, $\pi, 2\pi, 3\pi, 4\pi, 5\pi$, the 2-D Gabor filter function can be represented in Figure 4. With the increasing Gaussian window, range of local features of a given location on the image of 2-D Gabor wavelet transform increased, but the amount of computation convolution also doubled.

3 Detecting edge method based on 2-D Gabor filters real parts

Because the real parts and imaginary parts of the two-dimensional Gabor filter are similar, for the purpose of reducing the edge detection computation, only the real parts are applied, as shown in Figure 5.

The steps are taken as follows:

- 1) Input the value of centre Frequency k_v and variance δ to determine the size of the convolution template window;
- 2) Input direction parameter φ_μ , φ_μ can be $0, 2\pi/8, \pi/4, 3\pi/8$;
- 3) Use Equation (1) to calculate the value of 2-D Gabor filter, then we have four real template's windows;
- 4) Execute convolution of the template and the image, and generate four convolution images;
- 5) Use the image of the real part to calculate the final margin, then generate four amplitude images;
- 6) Detect four magnitude of the image data in the same position, reserve maximal, then get a maximum value of the images;
- 7) Determine threshold T , and the point greater than the maximum values in the image is retained as the edge point T ;

8) Do some detailed processes for the edge point images.

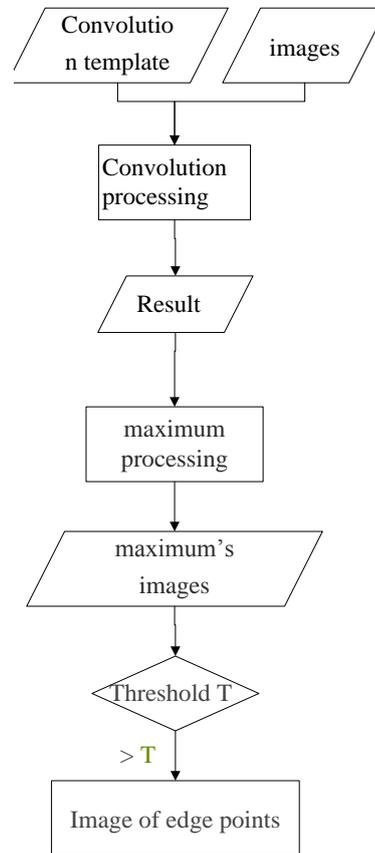


FIGURE 5 Flow chart

4 Experiments and Analysis

4.1 RESULTS BY DIFFERENT EDGE DETECTING METHOD

Edge detection results are shown in Figure 6:

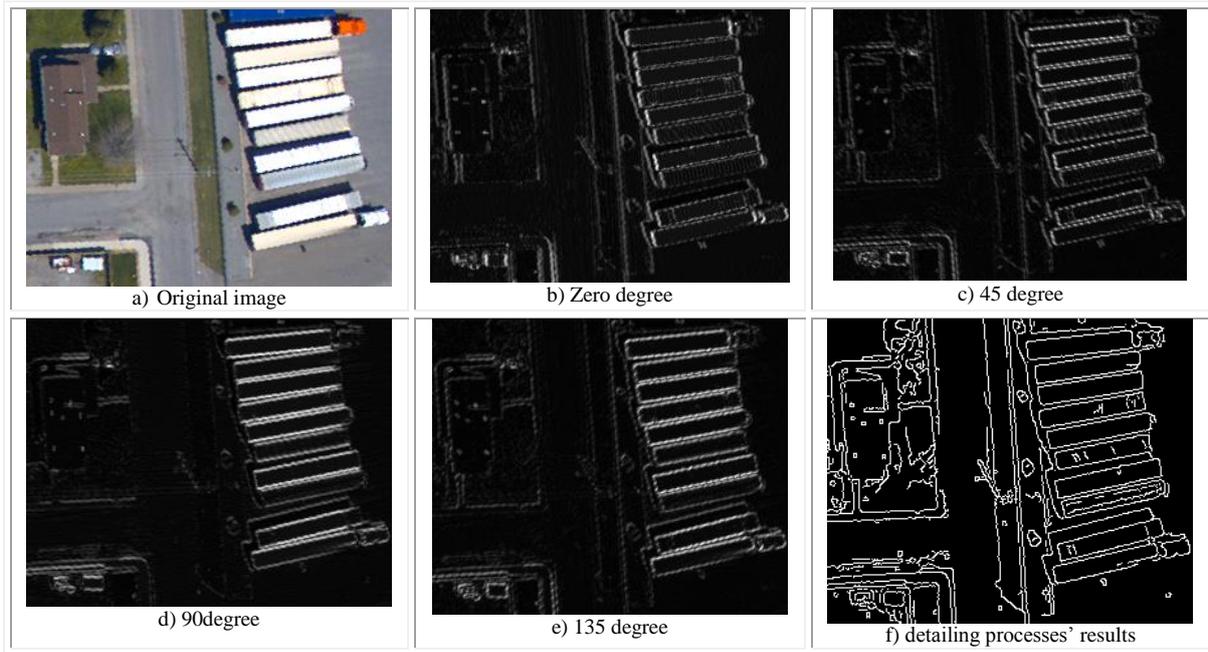


FIGURE 6 2-D Gabor filter used in image edge detecting

4.2 ANALYSIS OF ROC CURVE

ROC curve (receive operating characteristic curve) can be well applied to evaluate different edges' detection operator and detection accuracy.

TABLE 2 Four results of logogram and meaning of edge detecting

logogram	meaning
EE	Edge points correctly detected as edge point
ENE	edge point but erroneously detected as non-edge point
NEE	non-edge point but erroneously detected as edge point
NENE	non-edge point correctly detected as non-edge point

For every pixel in the image, there are four possible kinds of detecting results, respectively named in the Table 2 above.

Therefore, ROC curves used to reflect the relationship between positive edge detection operator class (*TP*, true-positive) and negative category (*FP*, false-positive), *TP* and *FP* can be defined as follows:

$$TP = \frac{n_{EE}}{n_{EE} + n_{ENE}}, \quad FP = \frac{n_{NEE}}{n_{NEE} + n_{NENE}}, \quad (7)$$

where the above equation, n_{EE} , n_{ENE} , n_{NEE} , n_{NENE} is the number of the four results.

Figure 7 shows the ROC curve of Canny operator, Sobel operator and the proposed GW algorithm, known as *FP* when taken from 0 to 0.12, *TP* was significantly higher than the value of the proposed algorithm Canny operator, *FP* is taken from 0 to 0.5, two algorithms proposed by this paper are better than Canny operator. When the *FP* is higher than 0.5, the algorithms mentioned by this paper are either better than Canny operator or equal to it.

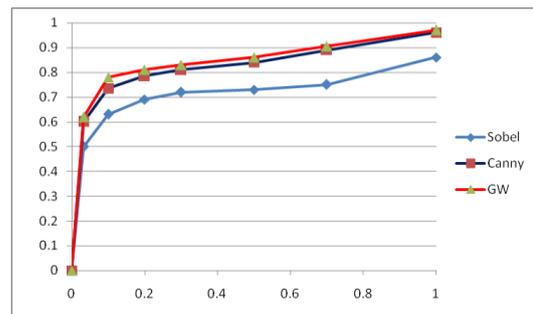


FIGURE 7 GW, Canny and Sobel operators ROC curve

5 Conclusions

With the features of large amount of data and complex landmark, remote sensing images need a higher requirement for edge detection operator. Due to the limitations of grads operator and Canny operator in edge detection, this paper presents an edge detection method based on 2-D Gabor wavelet real part and the experimental analysis shows this method was better on edge detection. However, the method of calculating takes more time to get a result. It will lead to more time-consuming, consequently in further studies, the degree of automation should be enhanced in edge detection.

Acknowledgments

This work was supported by Zhejiang Provincial Natural Science Foundation of China (No.LQ12D01001, No.LQ12F03001), Natural Science Foundation of China (No.61203360), Ningbo City Natural Science Foundation of China (No.2012A610043, No.2012A610009).

References

- [1] Lades M, Vorbruggen J C, Buhmann J, Lange J, von der Malsburg C, Wurtz R P, Konen W 1993 *IEEE Transactions on Computers* **42**(3) 300-11
- [2] Wang X, Chen X, Jin M 2005 Application of line segment extraction to recognize buildings in high-resolution remote sensing images *Journal of Computer Aided Design and Graphics* **17**(5) 928-34
- [3] Lee T S 1996 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(10) 959-71
- [4] Ma W Y, Manjunath B S 1997 Edge flow: a framework of boundary detection and image segmentation *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 1997* San Juan Puerto Rico 744-9
- [5] Nevatia R, Lin C, Huertas A 1997 A System for Building Detection from Aerial Images *In Automatic Extraction of Man-Made Objects from Aerial and Space Images*, edited by Gruen A, Kuebler O, Agouris P 1997
- [6] Guan Y P 2008 Automatic Extraction of lips Based on Multi-scale Wavelet Edge Detection *IET Computer Vision* **2**(1) 23-33
- [7] Fu Z, Tong C, Yan H Fan Z 2010 Parallel Gabor Wavelet Transform for Edge detection *International Conference on Internet Technology and Applications* August 2010 Wuhan China 1-3
- [8] Tomasi C, Manduchi R 1998 Bilateral Filtering for Gray and Color Images *Proceedings of the 1998 IEEE International Conference on Computer Vision* Bombay India 839-46
- [9] Yang Q, Li J 1999 Using the Tag Growth Extracting Line Segments from the Edge Image *Journal of Shanghai Jiaotong University* **33**(4) 466-8

Author



Chunya Tong, born in July, 1980, China

University studies: PhD degree in Photogrammetry and Remote Sensing from Wuhan University, Wuhan, China in 2011.

Scientific interest: remote sensing image processing.

A strategy for fault management in LDC wireless sensor network

Jian Yang^{1, 2, 3*}

¹College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210003, China

²Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing, Jiangsu 210003, China

³Key Lab of Broadband Wireless Communication and Sensor Network Technology (Nanjing University of Posts and Telecommunications), Ministry of Education Jiangsu Province, Nanjing, Jiangsu 210003, China

Received 1 July 2014, www.tsi.lv

Abstract

With rapid development of hardware, wireless sensor networks (WSN) have been applied in a wide range of fields. However, energy cost constrains putting WSN into use. To reduce energy cost, extending life time, WSN in low-duty-cycle (LDC) draws researchers' attention. In general, work time of a node only occupies 0.1%-10% in a cycle. This model certainly reduces the energy for idle listening. On the other hand, it makes the probability of congestion very high due to a node that can only receive packets when waking up. This paper proposes a new LDCWSN model to solve the congestion from duty schedule. With the model, we show a strategy for WSN fault averting, diagnosing and recovery based on congestion in nodes. We include some attributes of LDC WSN in our strategy, i.e. probability of congestion, scheduler, and link quality. By improving the selection of nodes on every level, we get a low rate for network's fault appearance, low E2E delay and long lifetime. The simulation's result shows that our strategy has a better performance in packet loss, energy cost and time delay than proposed WSN fault management.

Keywords: low-duty-cycle, congestion control, congestion recovery, fault management, wireless sensor network

1 Introduction

Wireless sensor network (WSN) is an integrated system consisting of embedded systems, new sensor material, low-power signal processing and wireless networks [1]. In most WSN, battery is the main power source, which makes the whole net have a limited life time. Usually, it is difficult to supply energy with constriction of scale, deploying environment and cost. Thus, many researches focus on energy efficiency in many applications. Researchers found that most cost of energy comes from idle listening [2]. This ennobles the importance of reducing unnecessary communication and sensing duty cycle. For the aim, low-duty-cycle WSN with short active time and long sleep time of nodes are a good choice [3].

Low-duty-cycle WSN use energy management protocol to schedule the active cycle and communicating time [4-6]. This ensures the node will remain dormant in the most time of one cycle and only 0.1%-10% active time [3]. The different schedules of each node cause sleep latency. Once getting a packet from low-level neighbours, a node will store it in a cache and deliver it later.

Therefore, while a node turns into active, neighbouring nodes in low levels will try to send messages to it at the same time, which aggravate the probability of congestion on the whole net.

During the convergence of messages to the sink node, how to find a sequence of nodes in different levels to build

a delivery path in order to improve performance of the whole net (like packet delay, packet loss and life time) is a key issue. This paper based on the specialization of low-duty-cycle WSN, constructs a new research model. According to the model, we designed a strategy for congestion fault management in low-duty-cycle wireless sensor networks. We included the dynamic property, like congestion probability, schedule and link quality, into the strategy. By optimizing the choosing of nodes in different levels, the aim of improving the network is realized. The simulation shows that this strategy has an enhancement in packet loss, energy cost and packet delay.

This article is organized by this: Section 2 introduces the background and related work of low-duty-cycle WSN; Section 3 demonstrates the particularity of congestion in low-duty-cycle WSN; Section 4, we describe our model and strategy; in last section, we give details about our stimulation and future work.

2 Related Work

2.1 LOW-DUTY-CYCLE WIRELESS SENSOR NETWORK

There are only two status of one node in low-duty-cycle WSN: active or dormant [2]. A whole cycle (T) consists of one active time and one dormant time [7, 8]. The node will be active if and only if these two reasons are met:

* Corresponding author e-mail: yangj@njupt.edu.cn

- 1) The schedule wakes up the node to receive packets and do sensing jobs;
- 2) The node has packets to send to its neighbouring nodes.

While the node is dormant, all the hardware is not working except a clock. Nodes can send packets in any time but can only receive packets when it is active. Usually, to reduce redundancy of data, neighbouring nodes often have a different schedule [9, 10]. Because of the different active time of each node, if node A has a packet to node C, it have to wait until node B is active, which will cause a delay called sleep latency.

The congestion in wireless sensor networks can be classified into two types [11]:

One is node congestion. That is to say, the packets that need to be delivered exceed the node's capacity. The cache overflowing causes the packet loss and delay.

Another one is link congestion. Wireless transmission shares a channel. One channel can be used by a node at one time. While many neighbouring nodes compete for the channel, link congestion arises. This also leads to packet loss and delay, and lower throughput of whole net.

Paper [12] proposed a model of multi-channel. In this article, the wireless channel is divided into different paths by rate. The node will use the optimized channel to deliver the packet, avoiding the link congestion. In paper [2], nodes will trim the delivery time based on the link quality. The better the link quality is, the earlier the delivery time is. The two strategies avoid link congestion efficiently, however, in low-duty-cycle wireless sensor networks, the node congestion occurs more frequently due to the short interval of active time and large-amount packets to be received.

In paper [13], every node has multiple paths to choose. Once congestion occurs in one path, the node will try another path to deliver the packet. The frequent update of delivery paths will put a press on network overload, and the retransmit will cost more energy, and aggravate delay as well. In paper [14], sink node plays a role of controlling the nodes' delivery rate to avoid node congestion. In fact, most nodes will be dormant in low-duty-cycle WSN. In this situation the strategy performs poorly. In low-duty-cycle WSN, while taking the sleep latency into consideration, it is also important to lower the probability of congestion. Thus, a strategy that can detect the congestion and recovery from it is necessary.

2.2 CONGESTION IN LOW-DUTY-CYCLE

Congestion leads to the overload of the whole network, higher packet loss and delay. In low-duty-cycle, the results become worse. For example, once a packet is lost, the node has to resend it and consume more energy, which shortens life time. The active-dormant cycle succours the node congestion and effects the whole net. A node has a shot active time in a cycle. Neighbouring nodes will transmit packets at the same time, which cause the link congestion. In addition, the receiver node's parent may be still in dormant. That means receiver node has to store these

packets in a cache and wait until its parent is active. Meanwhile, one delivery failure means that the node has to wait one duty cycle and send the same packet at the next active time. This consoles packet delay and enfeebles the network. Figures 2 and 3 shows congestion with the number of nodes and active time. Figures 4 and 5 show delay with the number of nodes and active time. From Figures 1 and 2, it is easy to conclude that congestion occurs more frequently with more nodes and less active time. Figures 3 and 4 indicates that delay becomes longer due to the congestion. Therefore, we need an algorithm, which can perform well with large-scale and long dormant time.

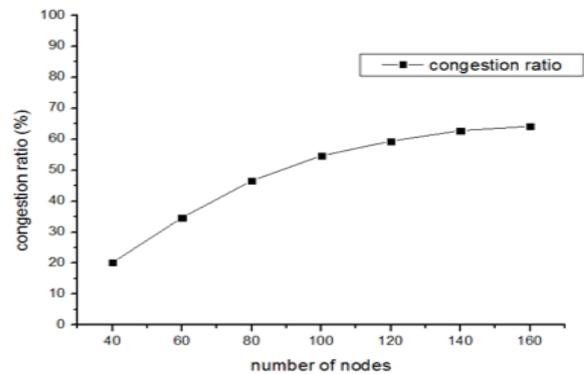


FIGURE 1 Number of nodes VS congestion ratio

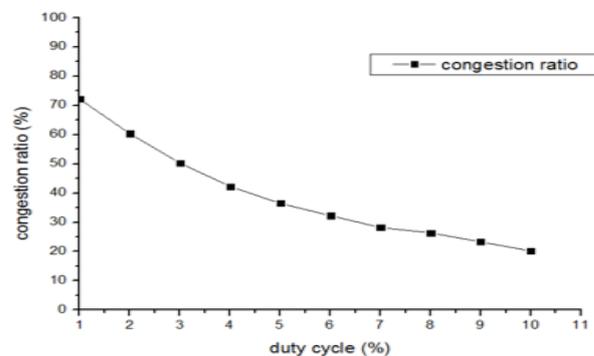


FIGURE 2 Duty cycle VS congestion ratio

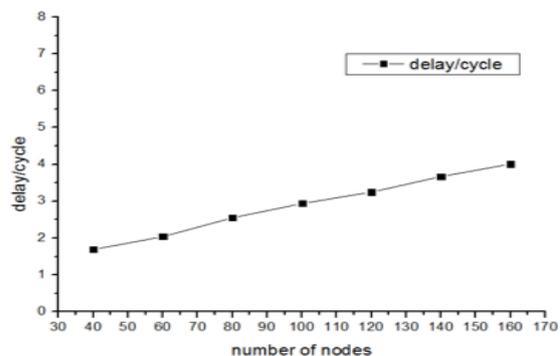


FIGURE 3 Number of nodes VS delay/cycle

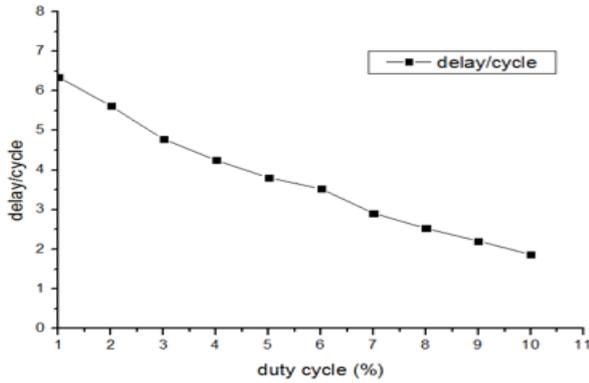


FIGURE 4 Duty cycle VS delay/cycle

3 A strategy for low-duty-cycle WSN

3.1 COMPETITION

Once the network is established, the sink node broadcasts a message. A node decides its level by the smallest hop of the message it received. If a node i receives three messages with 1 hop, 2 hops and 3 hops, then node i belongs to level 1.

The node only communicates to the nodes in different levels. Paper [15] proposed FTSP which can make the whole network simultaneous with 2.24 μ s error (usually, a duty cycle is at least 1ms). The node broadcast its schedule and neighbouring nodes to record just that. In every duty cycle, nodes in lowest level send a test packet to its parent according to the recorded schedule. If the parent received, then it will respond the ACK packet, and store the packet in a buffer, else, the send node marks this packet failure. If the parent received but the buffer is full, the node congestion occurs, and then it will respond the CN packet. While the lowest-level node has all finished, the second-low nodes repeat the process. One level by one level repeat this process until convergence to sink node. Assume this process repeats turns, because of the random link quality and congestion occurrence. The probability of link quality and congestion is described as follows:

Definition 1: represents that for a node a in level K and its neighbouring node A in level $K-1$, the link quality between them is calculated by $P_a^A(LQ) = m/n$, $m \leq n$, where stands for the number of ACK packets received in turns.

Definition 2: represents that for a node a in level K and its neighbouring node A in level $K-1$, the congestion between them is calculated by $P_a^A(NCN) = 1-l/m$, ($l \leq m$) where represents the number of CN packets during successful times.

Definition 3: represents a difference of the wake up time for a node a in level K and its neighbouring node A , which is calculated by:

$$\Delta T_a^A = \begin{cases} t_A - t_a, & t_A \geq t_a \\ t_A - t_a + T, & t_A < t_a \end{cases} \quad (1)$$

where t_a is the wake up time of node a and T is the time of one cycle. We define that $\Delta t_a^A = 1 - \Delta T_a^A / T$. The value of Δt_a^A is meaningless so we just use it to reflect to the latency of two nodes.

3.2 STATIC CONGESTION AVOIDANCE STRATEGY (SCAS)

In the period while, the whole network is established but not in low-duty-cycle yet, according to the value of already known parameters, nodes can optimize convergence path to sink node to reduce the probability of congestion occurrence. We call this Static Congestion Avoidance Strategy (SCAS).

For every node in level K , after n turns, they can calculate their own competition C_a^A in the neighbouring node, which is calculated by $C_a^A = P_a^A(LQ) + P_a^A(NCN) + \alpha \Delta t_a^A$ ($\alpha > 1$), where α is the weight of D -value of wake up time. In our algorithm, the competition reflects the priority of node a in the upper node A . The bigger the competition is, the more possible that node a should be in the receive sequence of the node A . Because there is sleep latency in low-duty-cycle WSN, the most significant value is sleep latency. So we give a weight for Δt_a^A to ensure that the D -value of wake up time will affect the competition most effectively.

Definition 4: $E_a^A(PS)$ represents the exception packet size of node a in its neighbouring node A . We can calculate it by $E_a^A(PS) = P_a^A(LQ) \times PS_a$, where PS_a is the packet size.

When nodes in level K have calculated their competition, they sort the upper nodes by competition to construct a competitive sequence.

Definition 5: sequence $C_a = \{C_a^1, C_a^2, \dots, C_a^n\}$, ($C_a^1 \geq C_a^2 \geq \dots \geq C_a^n$) represents node a in a lower level, sorting all its upper neighbouring nodes by competition.

Nodes in level K then compete with each other for getting parents in the upper level. First, nodes in level K sends a packet, which includes application for joining the receive sequence and competition, to the first node in its competitive sequence. The node in level $K-1$ received these packets, and then sorts their competition. The node in level $K-1$ repeats the process:

- 1) put the node with biggest competition to the receive sequence and delete its packet;
- 2) calculate residual size of buffer $B_A = B_A - E_a^A(PS)$, where B_A is the current buffer size;
- 3) if B_A is larger than the threshold ϕ , then back to step 1.

When the node A in level $K-1$ has finished its receive sequence, it will broadcast the nodes which are in the sequence to all of its neighbours. The nodes in the sequence will update their competition value. For

example, if node a is selected by node A , the next node, which constructs receive sequence is node B and is neighbouring node of a as well. Then, the link quality between node a and node B will be $P_a^B(LQ) = (1 - P_a^A(LQ))P_a^B(LQ)$ and congestion will be $P_a^B(NCN) = (1 - P_a^A(NCN))P_a^B(NCN)$, the competition also updates. By doing the update, for event node a send a packet to node B has become a conditional probability, which can be represented as P (send packet to B | fail to send packet to A). Thus, the competition goes down with probability.

In the broadcast message from node A , there is a message if the buffer size is full. If not, other nodes still have to compete to join A 's receive sequence.

Pseudo code is described as follows:

calculate the competitiveness for every neighbouring node in level $K-1$

sort nodes by the competitiveness

$$C_a = \{C_a^1, C_a^2, \dots, C_a^n\}, (C_a^1 \geq C_a^2 \geq \dots \geq C_a^n)$$

for ($i = 1; i \leq n; i++$)

if isfull (node i) is false

send competitive packet to node i

if (node i) admit

for ($j = i; j < n; j++$)

$$C_a^j = (1 - P_a^i(LQ)) \times P_a^j(LQ) + (1 - P_a^i(NCN)) \times$$

$$P_a^j(NCN) + \alpha \Delta t_a^j$$

else continue

sort nodes by the competitiveness from level K

$$S^A = \{C_1^A, C_2^A, \dots, C_e^A\}, (C_1^A \geq C_2^A \geq \dots \geq C_e^A)$$

for ($i = 1; i \leq e; i++$)

$$B_A = B_A - E_i^A(PS) = B_A - P_i^A(LQ) \times PS_i$$

if ($B_A \leq \phi$) break;

else put node I to receive sequence;

Additionally, according to the receive sequence and the order of lower nodes in the sequence, every lower-level

node calculate its send time TX_i by $TX_i = \frac{t_A}{e} \times e_a + \tau_A$,

where e is the number of nodes in the receive sequence, e_a is the order of node a in the sequence and τ_A is the wake up time of node A .

Level by level, every node constructs its receive sequence for its neighbouring nodes in lower levels. If the nodes exist that are rejected by all of the upper nodes, the node chooses one of the nodes with the largest competition and sends the joining message. The upper ones make it to its receive sequence.

3.3 DYNAMIC CONGESTION AVOIDANCE STRATEGY

When all receive sequences are finished, the sink node broadcasts and the whole network goes into the low-duty-cycle model.

Due to the SCAS, we use exception packet size to judge if the buffer is full. If in one cycle, most packets are sent successfully, the real value must be bigger than the exception. Thus, it is still possible that congestion will occur. Adjustments based on the real situation, we call this Dynamic Congestion Avoidance Strategy (DCAS).

While one node's buffer overflows and a new packet arrives, it will respond the CN packet. The sender node then tries to deliver the packet to the next node in its send sequence. If one node in the lower level gets x CN packets in a continuous cycle (x is a threshold). If the cause is that packet size is too big for upper node's buffer size, the node should surrender the packet and send to sink node. If the buffer size is truly full, the node has to change its send sequence.

To change the send sequence, the node in the lower level sends is full packet. If the response is true, which means the buffer cannot get any packet; the lower level node deletes the upper node from its send sequence. If the response is false, the node will split its packet into small pieces based on the response and send these pieces into different upper nodes. Finally, the sink node will integrate them.

4 Simulation and future work

4.1 SIMULATION

To know the number of retransmissions, packet loss and packet delay in the different algorithms, different duty cycle and different scale; we use NS-3-3.15 to test. The simulation steps are described as follows:

Put n (n from 50 to 600 with the step of 50) nodes in 100×100 square randomly, every node has a level number based on the hop to sink node.

Initialize the character of node. Let the duty cycle be 100s and the node will be active in 1s during 0-99s randomly. Every node has a communication radius and some buffer size. Each node will generate a packet. According to the position of every node, they have their parents with a link quality, which is a random number from 0.4 to 1, a congestion probability which is a random number from 0.1 to 0.6 and a schedule.

Sending packets. We use three different algorithms to construct a different send sequence. Then nodes will deliver packets according to the send sequence. Change the duty cycle and repeat the simulation.

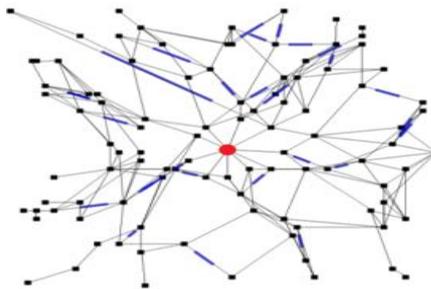


FIGURE 5 Topology of the simulation.

For different network scale and duty-cycle, we compare our algorithm to link quality first and delay first in number of transmissions, send delay and packet loss. The result shows in Figures 6 and 7.

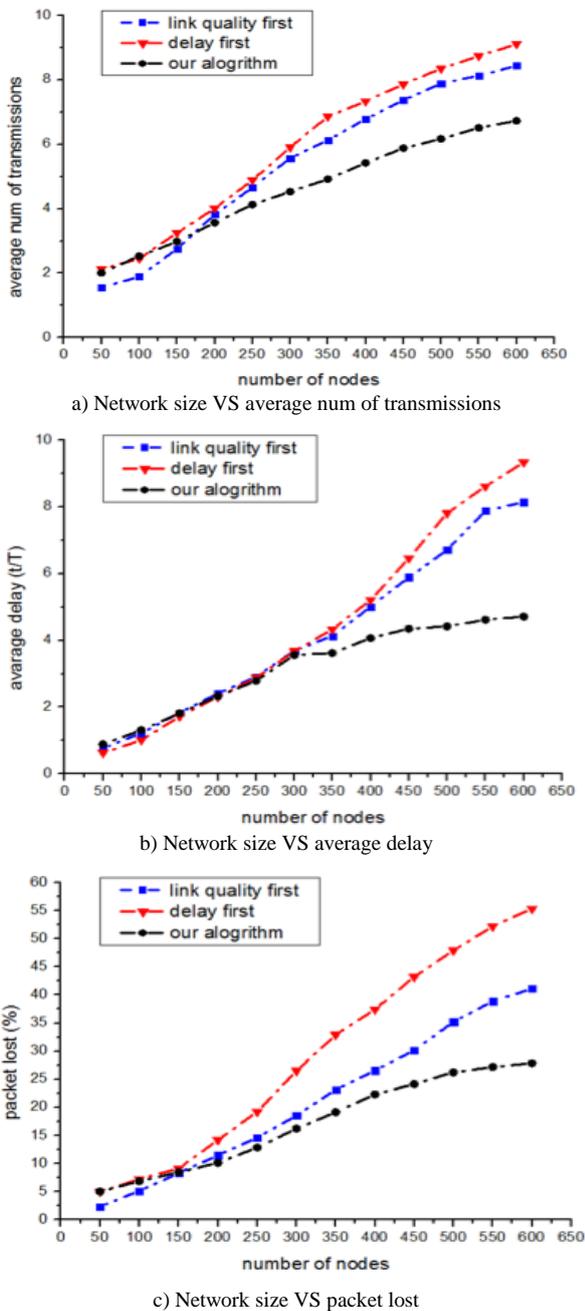


FIGURE 6 Performance comparison under different network scale

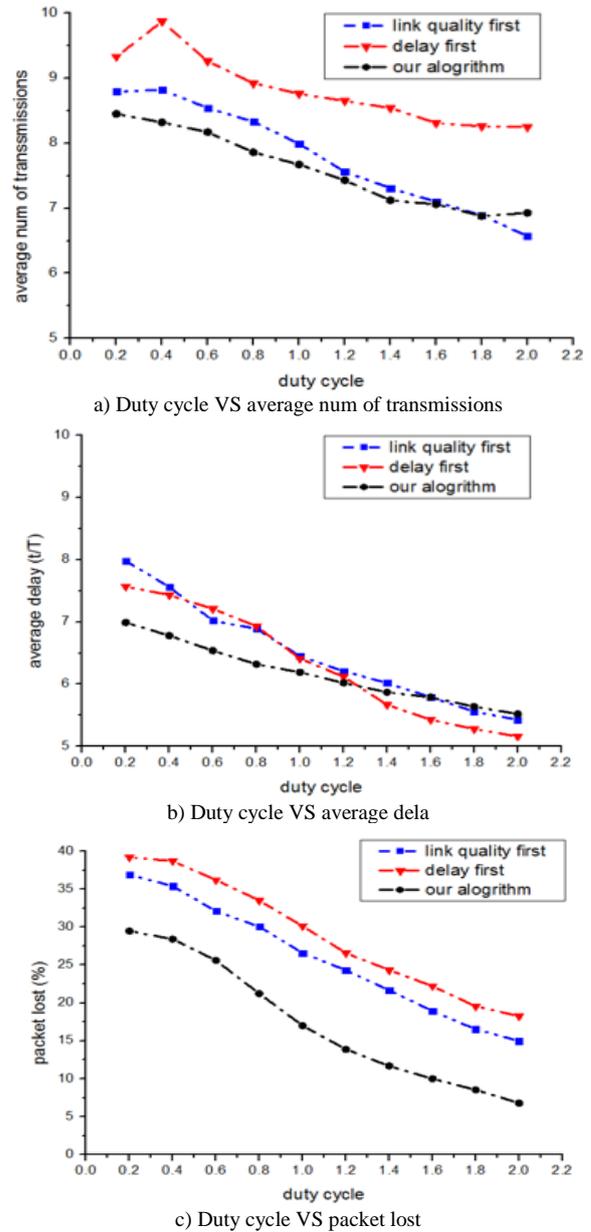


FIGURE 7 Performance comparison under different working duration n

5 Conclusions

This paper proposed the low-duty-cycle congestion strategy. Although the performance is nice, the constriction is ideal, especially the schedule synchronization. In future work, we will discuss how to synchronize the whole network and when the node should adjust its wake up time.

Acknowledgements

Our work supported by A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (yx002001); Graduate student research and innovation program Jiangsu (CXLX12_0480).

References

- [1] Peng Y, Song J, Peng X 2009 Survey of fault management framework in wireless sensor networks *Journal of Electronic Measurement and Instrument* **23**(11) 1-10 (in Chinese)
- [2] Guo S, Gu Y, Jiang B, He T 2009 Opportunistic Flooding in Low-Duty-Cycle Wireless Sensor Networks with Unreliable Links *Proceedings of MobiCom'09* September 2009 133-44
- [3] Pak W, Bahk S 2012 Centralized route recovery based on multi-hop wakeup time estimation for wireless sensor networks with ultra-low duty cycles *Computer Communications* **35**(11) 1355-67
- [4] Jeong G, Gu Y, He T, Du D 2009 VISA: Virtual Scanning Algorithm for Dynamic Protection of Road Networks *Proceedings of IEEE INFOCOM* 927-935
- [5] Wang X, Xing G, Zhang Y, Lu C, Pless R, Gill C 2003 Integrated Coverage and Connectivity Configuration in Wireless Sensor Networks *Proceedings of the First International Conference Embedded Networked Sensor Systems* 28-39
- [6] Kasbekar G S, Bejerano Y, Sarkar S 2011 Lifetime and Coverage Guarantees through Distributed Coordinate-Free Sensor Activation *IEEE/ACM Transactions on Networking* **19**(2) 470-83
- [7] He T, Krishnamurthy S, Luo L, Yan T, Gu L, Stoleru R, Zhou G, Cao Q, Vicaire P, Stankovic J A, Abdelzaher T F, Hui J, Krogh B 2006 VigilNet: An Integrated Sensor Network System for Energy-Efficient Surveillance *ACM Transactions on Sensor Networks* **2**(1) 1-38
- [8] Tolle G, Polastre J, Szewczyk R, Turner N, Tu K, Burgess S, Gay D, Buonadonna P, Hong W, Dawson T, Culler D 2005 A Macroscopic in the Redwoods *Proceedings of SenSys'05* 51-63
- [9] Gupta H, Navda V, Das S, Chowdhary V 2008 Efficient Gathering of Correlated Data in Sensor Networks *ACM Transactions on Sensor Networks* **4**(1) 402-13
- [10] Wang J, Liu Y, Das S 2008 Asynchronous Sampling Benefits Wireless Sensor Networks *Proceedings of the 27th Conference on Computer Communications IEEE* 2207-15
- [11] Li L, Li B, Zhou X 2008 A Survey of Congestion Control Technology for Wireless Sensor Networks *Journal of Computer Research and Development* **45**(1) 63-72 (in Chinese)
- [12] Rathi M K 2010 Data Dissemination in Low Duty-cycle Multi-channel Multi-hop-Wireless-Sensor-Networks *Proceedings of the Computer Engineering and Technology (ICCET) the 2nd International Conference April 2010* 258-62
- [13] Liu Y, Ma Z, Cao Z 2005 A mitigating stagnation based ant colony optimization routing algorithm *Communications and Information Technology Proceedings of IEEE International Symposium on Communications and Information Technology* 36-9
- [14] Akan Ö B, Akyildiz I F 2005 Event-to-Sink Reliable Transport in Wireless Sensor Networks *IEEE/ACM Transactions on Networking* **13**(5) 1003-16
- [15] Maroti G S M, Kusy B, Ledeczi A 2004 The Flooding Time Synchronization Protocol *Proceedings of SenSys'04*

Author



Jian Yang, born in March, 1978, Jiangsu, China

Current position: Associate Professor, PhD candidate in the Nanjing University of Posts and Telecommunications, Department of Computer.

University studies: Nanjing University of Posts and Telecommunications.

Scientific interest: Computer networks, health care services, energy efficiency, reliability, resource allocation, wireless sensor networks

Publications: 4

A performance evaluation model of green supply chain based on fuzzy analysis method of multi-attribute decision-making

Huaping Zhang*

Management & Economics School of North China University of Water Resources and Electric Power, Zhengzhou, Henan, China

Received 1 August 2014, www.tsi.lv

Abstract

This paper proposes a performance evaluation model of green supply chain based on fuzzy analysis method of multi-attribute decision-making. In this model, an evaluation index system is established with economic profit, environment protection, business process and customer service taken into consideration. Fuzzy analysis method of multi-attribute decision-making is introduced to get the fuzzy incidence degree of different performance evaluation indicators. Analysis of performance evaluation of green supply chain is based on the fuzzy incidence degree. Finally, the model and the algorithm are proved to be scientific and feasible through case study.

Keywords: green supply chain, performance evaluation, multi attribute decision making, fuzzy theory, model

1 Introduction

In recent years, with the degradation of environment, people are paying more and more attention to sustainable development. Green supply chain management is born under such circumstance. It is a modern management mode integrated with environment protection, aiming at reducing environment pollution and resources depletion while increasing the benefit of the whole supply chain [1-3]. Performance evaluation of green supply chain helps us to learn about the operation of the supply chain and understand how to improve it. It is significant to increase the competitiveness of products, protect the environment and reach for a sustainable development. Many researches both home and abroad have studies this issue [4-6].

Currently, analysis hierarchy process, grey incidence analysis method and fuzzy evaluation method are major ways to evaluate the performance of green supply chain. However, these methods are more often subjective using index evaluation matrix and indicator weight but overlook the incomparable nature of some indicators or the influence of membership on target evaluation grade [7-10]. Therefore, this paper constructs a performance evaluation index system based on improved fuzzy analysis method of multi-attribute decision-making. It sheds some lights on real practice.

2 The performance evaluation index system of green supply chain

2.1 PRINCIPLES FOR THE CONSTRUCTION OF THE INDEX SYSTEM

The evaluation index system should be objective, fair and accurate that can reflect the sustainable development

capability of the enterprise in an all-round way. Therefore, there are some principles that must be followed in constructing the index system.

(a) Scientific principle: The evaluation index system should be scientific. Data source should be reliable, the indicators should be clear, the evaluation method should be convincing and the evaluation mode should be reasonable.

(b) General principle: The evaluation index system should involve with all features of the sustainable development capability of the enterprise and be able to analyse these features based on their structure, layer and interaction.

(c) Leading principle: Not all factors are significant to the sustainable development capability. Thus, some factors should be given priority with more weight.

(d) Operating principle: The evaluation index system should be practical in real use. Therefore, data should be acquired in a reliable way and those that cannot be available should be kept out of the index system.

(e) Simple principle: Simple and practical indicators are the ones that should be selected. They will reflect the operation state of the supply chain and serve to the calculation and analysis of the performance

2.2 CONSTRUCTING THE PERFORMANCE EVALUATION INDEX SYSTEM OF GREEN SUPPLY CHAIN

The evaluation index system will provide a measurement standard, restriction as well as incentives for green supply chain management. According to abovementioned principles, the evaluation index system falls into three layers: target layer, standard layer and indicator layer. Standard layer consists of economic profit, environment

* *Corresponding author* e-mail: zhanghuaping@ncwu.edu.cn

protection, business process, customer service and indicators. The structure of the evaluation index system is sustainable development and they evolve into 21 shown in Table 1.

TABLE 1 The structure of the evaluation index system of green supply chain

Target layer	Standard layer	Index layer	Class of indicator	Type of indicator
The evaluation index system of green supply chain	Economic profit	Manufacturing and sale rate of products	Quantitative	Positive
		Rate of return on net assets	Quantitative	Positive
		Profit growth rate	Quantitative	Positive
	Environment protection	Investment rate in environment protection	Quantitative	Positive
		Utilization rate of material and energy	Quantitative	Positive
		Return rate of material and energy	Quantitative	Positive
		Impact degree on environment	Qualitative	Adverse
		Energy consumption level	Qualitative	Adverse
		Production capability	Quantitative	Positive
	Business process	Operation efficiency	Qualitative	Positive
		Traffic rate	Quantitative	Positive
		Product quality	Quantitative	Positive
		Transport rate	Quantitative	Adverse
	Customer service	Safe delivery rate	Quantitative	Positive
		Customers satisfaction degree	Qualitative	Positive
		Green identity	Qualitative	Positive
		Market share	Quantitative	Positive
	Sustainable development	Accuracy of market prediction	Quantitative	Positive
		Investment rate in R&D	Quantitative	Positive
		Proportion of design staff	Quantitative	Positive
		Profit rate of new products	Quantitative	Positive

3 The performance evaluation model of green supply chain based on fuzzy analysis method of multi-attribute decision-making

3.1 STANDARDIZATION OF PERFORMANCE EVALUATION INDICATORS

According to Table 1, there are two classes of indicators, one is quantitative and the other is qualitative. Quantitative indicators are available through statistics and calculation analysis. Qualitative indicators need fuzzy description. Detailed descriptions for qualitative indicators are shown in Table 2.

According to Table 1, there are two types of indicators, one is positive indicator and the other is adverse indicator. As indicators have different value and scale, they need to

$$u_{ij}(\Omega) = [u_{ij}^1(\Omega), u_{ij}^2(\Omega)] = \left(v_{ij}^1(\Omega) / \left(v_{ik}^2(\Omega) \mid \max_{1 \leq j \leq m} (v_{ij}^2(\Omega)) \right), v_{ij}^2(\Omega) / \left(v_{ik}^2(\Omega) \mid \max_{1 \leq j \leq m} (v_{ij}^2(\Omega)) \right) \right). \tag{1}$$

If the performance evaluation indicator is an adverse indicator, its value is $v_{ij}(\Omega) = [v_{ij}^1(\Omega), v_{ij}^2(\Omega)]$, and the

$$u_{ij}(\Omega) = [u_{ij}^1(\Omega), u_{ij}^2(\Omega)] = \left(\left(v_{ik}^1(\Omega) \mid \min_{1 \leq j \leq m} (v_{ij}^1(\Omega)) \right) / v_{ij}^2(\Omega), \left(v_{ik}^1(\Omega) \mid \min_{1 \leq j \leq m} (v_{ij}^1(\Omega)) \right) / v_{ij}^1(\Omega) \right). \tag{2}$$

In the expression, $v_{ik}^2(\Omega) \mid \max_{1 \leq j \leq m} (v_{ij}^2(\Omega))$ refers to the maximum value of indicator j under scheme i . $v_{ik}^1(\Omega) \mid \min_{1 \leq j \leq m} (v_{ij}^1(\Omega))$ refers to the minimum value of indicator j under scheme i . After standardization, the value of indicator falls between $[0, 1]$, which means all indicators have unified measurement standard. This will make the analysis more accurate and reliable.

be standardized to ensure the effectiveness and reliability of the performance evaluation.

TABLE 2 Fuzzy descriptions of qualitative indicators

Positive language grade	Negative language grade	Range of score
Very good	Very bad	90-100
Good	Bad	70-90
So-so	So-so	50-70
Poor	A little bad	0-50

If the performance evaluation indicator is a positive indicator, its value is $v_{ij}(\Omega) = [v_{ij}^1(\Omega), v_{ij}^2(\Omega)]$, and the value after standardization is $u_{ij}(\Omega) = [u_{ij}^1(\Omega), u_{ij}^2(\Omega)]$. There is:

value after standardization is $u_{ij}(\Omega) = [u_{ij}^1(\Omega), u_{ij}^2(\Omega)]$. There is:

3.2 FUZZY ANALYSIS METHOD OF GREEN SUPPLY CHAIN OF MULTI-ATTRIBUTE DECISION MAKING

After standardization, construct the positive ideal interval $\bar{u}_{0j}(\Omega)$ for indicators:

$$\bar{u}_{0j}(\Omega) = \left[\bar{u}_{0j}^{-1}(\Omega), \bar{u}_{0j}^{-2}(\Omega) \right] = \left[\max(u_{ij}^1(\Omega) | 1 \leq i \leq m), \max(u_{ij}^2(\Omega) | 1 \leq i \leq m) \right]. \tag{3}$$

The adverse ideal interval $\underline{u}_{0j}(\Omega)$ for the corresponding indicator is described as:

$$\underline{u}_{0j}(\Omega) = \left[\underline{u}_{0j}^{-1}(\Omega), \underline{u}_{0j}^{-2}(\Omega) \right] = \left[\min(u_{ij}^1(\Omega) | 1 \leq i \leq m), \min(u_{ij}^2(\Omega) | 1 \leq i \leq m) \right]. \tag{4}$$

Therefore, the positive ideal scheme S^+ of performance evaluation scheme for green supply chain is:

$$S^+ = \left\{ \left[\underline{u}_{01}^{-1}(\Omega), \underline{u}_{01}^{-2}(\Omega) \right], \left[\underline{u}_{02}^{-1}(\Omega), \underline{u}_{02}^{-2}(\Omega) \right], \dots, \left[\underline{u}_{0n}^{-1}(\Omega), \underline{u}_{0n}^{-2}(\Omega) \right] \right\}. \tag{5}$$

The adverse ideal scheme S^- of performance evaluation scheme for green supply chain is:

$$S^- = \left\{ \left[\underline{u}_{01}^1(\Omega), \underline{u}_{01}^2(\Omega) \right], \left[\underline{u}_{02}^1(\Omega), \underline{u}_{02}^2(\Omega) \right], \dots, \left[\underline{u}_{0n}^1(\Omega), \underline{u}_{0n}^2(\Omega) \right] \right\}. \tag{6}$$

Calculate the distance of fuzzy set by Hamming distance. Construct positive and adverse ideal interval scheme. Establish a fuzzy evaluation model of multi attribute based on fuzzy information. Suppose \tilde{A} and \tilde{B} are two fuzzy sets in the discourse domain Ω . Their membership functions are $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$. The Hamming distance between fuzzy set \tilde{A} and \tilde{B} is:

$$d(\tilde{A}, \tilde{B}) = \int_{\alpha}^{\beta} |\mu_{\tilde{A}}(x) - \mu_{\tilde{B}}(x)| dx. \tag{7}$$

The positive ideal scheme S^+ is featured by membership function $\mu_{S^+}^{\vee}(u_j(\Omega))$:

$$\mu_{S^+}^{\vee}(u_j(\Omega)) = \sup_{\substack{u=u_1 \vee u_2 \vee \dots \vee u_m \\ (u_1, u_2, \dots, u_m) \in R^m}} \min \left\{ \mu_{u_{1j}}(u_1), \mu_{u_{2j}}(u_2), \dots, \mu_{u_{mj}}(u_m) \right\}. \tag{8}$$

The adverse ideal scheme S^- is featured by membership function $\mu_{S^-}^{\wedge}(u_j(\Omega))$:

$$\mu_{S^-}^{\wedge}(u_j(\Omega)) = \sup_{\substack{u=u_1 \wedge u_2 \wedge \dots \wedge u_m \\ (u_1, u_2, \dots, u_m) \in R^m}} \min \left\{ \mu_{u_{1j}}(u_1), \mu_{u_{2j}}(u_2), \dots, \mu_{u_{mj}}(u_m) \right\}. \tag{9}$$

The difference between performance evaluation scheme i and positive ideal scheme S^+ about indicator j is:

$$D_i^+ = \sqrt{\sum_{j=1}^n \left[w_{ij} \times \left(d(u_{ij}^1(\Omega), S_1^+) + d(u_{ij}^2(\Omega), S_2^+) \right) \right]^2}. \tag{10}$$

In the expression, S_1^+ and S_2^+ are left fuzzy maximum set and right fuzzy maximum set of positive ideal scheme S^+ .

The difference between performance evaluation scheme i and positive ideal scheme S^- about indicator j is:

$$D_i^- = \sqrt{\sum_{j=1}^n \left[w_{ij} \times \left(d(u_{ij}^1(\Omega), S_1^-) + d(u_{ij}^2(\Omega), S_2^-) \right) \right]^2}. \tag{11}$$

In the expression, S_1^- and S_2^- are left fuzzy maximum set and right fuzzy maximum set of positive ideal scheme S^- .

So the calculation model μ_i for fuzzy incidence degree of performance evaluation scheme i is:

$$\mu_i = 1 / \left(1 + \left((1 - D_i^-) / (1 - D_i^+) \right)^2 \right). \tag{12}$$

3.3 PERFORMANCE EVALUATION MODEL AND ALGORITHM OF GREEN SUPPLY CHAIN

According to close principle in fuzzy analysis of multi-attribute decision making [11-14], if there is:

$$\mu_0 = \max(\mu_s, \mu_s, \dots, \mu_s) = \mu_s, 1 \leq s \leq m. \tag{13}$$

Then the scheme s has the optimal performance.

As is mentioned above, the algorithm of the performance evaluation model of enterprise green supply chain based on fuzzy analysis method of multi-attribute decision making is described as follows:

Step 1: Construct the performance evaluation index system of green supply chain after survey, statistics and consultation with experts;

Step 2: Under the index system, obtain indicator values of different performance evaluation scheme and standardize them based on Equations (1) and (2);

Step 3: Acquire positive ideal interval and adverse ideal interval according to performance evaluation indicators based on Equations (3) and (4);

Step 4: Acquire positive ideal scheme and adverse ideal scheme according to performance evaluation indicators based on Equations (5) and (6);

Step 5: Get the difference of the performance evaluation scheme from positive ideal scheme and that from adverse ideal scheme;

Step 6: Acquire fuzzy incidence degree of performance evaluation scheme based on Equation (12);

Step 7: Implement the optimal scheme of green supply chain according to fuzzy incidence degree based on Equation (13).

4 Case study and model test

This paper tests the model and algorithm by analysing the performance of green supply chain of three brand enterprises in a certain industry. After survey, data collection and consultation with experts, management team and relevant technicians, the performance evaluation indicators are available and shown in Table 3.

TABLE 3 Information of performance evaluation indicator of green supply chain

Standard layer	Weight	Indicator layer	Weight	Indicator information		
				Scheme 1	Scheme 2	Scheme 3
Economic profit	0.305	Manufacturing and sale rate of products	0.25	0.88-0.93	0.90-0.95	0.88-0.93
		Rate of return on net assets	0.25	0.35	0.35	0.30
		Profit growth rate	0.35	0.20	0.15	0.22
		Investment rate in environment protection	0.15	0.16	0.20	0.15
Environment protection	0.120	Utilization rate of material and energy	0.30	0.92-0.95	0.88-0.93	0.92-0.95
		Return rate of material and energy	0.25	0.18	0.26	0.23
		Impact degree on environment	0.20	50-60	60-70	50-60
		Energy consumption level	0.25	40-50	40-50	50-60
		Production capability	0.25	95	92	85
Business process	0.285	Operation efficiency	0.22	85	90	90
		Traffic rate	0.15	0.85-0.90	0.80-0.85	0.85-0.90
		Product quality	0.25	985	962	895
		Transport rate	0.13	24	48	24
		Safe delivery rate	0.20	0.95	0.90	0.90
Customer service	0.165	Customers satisfaction degree	0.25	85-90	80-85	85-90
		Green identity	0.20	80-85	80-85	85-90
		Market share	0.35	0.18	0.22	0.13
Sustainable development	0.125	Accuracy of market prediction	0.22	0.35-0.40	0.60-0.65	0.45-0.50
		Investment rate in R&D	0.21	0.35	0.25	0.25
		Proportion of design staff	0.19	0.55	0.45	0.55
		Profit rate of new products	0.38	0.36	0.23	0.30

According to Section 3.1, standardize indicators of different classes and types and get the standardized indicator values, as is shown in Table 4.

TABLE 4 Standardization of performance evaluation indicators of green supply chain

Indicator layer	Indicator information		
	Scheme 1	Scheme 2	Scheme 3
Manufacturing and sale rate of products	0.926-0.979	0.947-1.000	0.926-0.979
Rate of return on net assets	1.000	1.000	0.857
Profit growth rate	0.909	0.682	1.000
Investment rate in environment protection	0.800	1.000	0.750
Utilization rate of material and energy	0.968-0.95	0.926-0.979	0.968-1.000
Return rate of material and energy	0.692	1.000	0.885
Impact degree on environment	0.833-1.000	0.714-0.833	0.833-1.000
Energy consumption level	0.800-1.000	0.800-1.000	0.667-0.800
Production capability	1.000	0.968	0.924
Operation efficiency	0.944	1.000	1.000
Traffic rate	0.944-1.000	0.889-0.944	0.944-1.000
Product quality	1.000	0.977	0.909
Transport rate	1.000	0.500	1.000
Safe delivery rate	1.000	0.947	0.947
Customers satisfaction degree	0.944-1.000	0.889-0.944	0.944-1.000
Green identity	0.889-0.944	0.889-0.944	0.944-1.000
Market share	0.818	1.000	0.591
Accuracy of market prediction	0.538-0.615	0.923-1.000	0.692-0.769
Investment rate in R&D	1.000	0.714	0.714
Proportion of design staff	1.000	0.818	1.000
Profit rate of new products	1.000	0.639	0.833

According to Section 3.2, acquire the Hamming distance of performance evaluation indicators by

constructing positive ideal interval and adverse ideal interval, as is shown in Table 5.

TABLE 5 Hamming distance of performance evaluation indicators of green supply chain

Indicator layer	Scheme 1		Scheme 2		Scheme 3	
	Positive	Adverse	Positive	Adverse	Positive	Adverse
Manufacturing and sale rate of products	0.021	0.000	0.000	0.021	0.021	0.000
Rate of return on net assets	0.000	0.143	0.000	0.143	0.143	0.000
Profit growth rate	0.091	0.227	0.318	0.000	0.000	0.318
Investment rate in environment protection	0.200	0.050	0.000	0.250	0.250	0.000
Utilization rate of material and energy	0.000	0.032	0.032	0.000	0.000	0.032
Return rate of material and energy	0.308	0.000	0.000	0.308	0.115	0.193
Impact degree on environment	0.000	0.143	0.143	0.000	0.000	0.143
Energy consumption level	0.000	0.167	0.000	0.167	0.167	0.000
Production capability	0.000	0.076	0.032	0.044	0.076	0.000
Operation efficiency	0.056	0.000	0.000	0.056	0.000	0.056
Traffic rate	0.000	0.056	0.056	0.000	0.000	0.056
Product quality	0.000	0.091	0.023	0.068	0.091	0.000
Transport rate	0.000	0.500	0.500	0.000	0.000	0.500
Safe delivery rate	0.000	0.053	0.053	0.000	0.053	0.000
Customers satisfaction degree	0.000	0.056	0.056	0.000	0.056	0.000
Green identity	0.056	0.000	0.056	0.000	0.000	0.056
Market share	0.182	0.227	0.000	0.409	0.409	0.000
Accuracy of market prediction	0.385	0.000	0.000	0.385	0.231	0.154
Investment rate in R&D	0.000	0.286	0.286	0.000	0.286	0.000
Proportion of design staff	0.000	0.182	0.182	0.000	0.000	0.182
Profit rate of new products	0.000	0.361	0.361	0.000	0.167	0.194

Therefore, the differences of performance evaluation indicators from positive ideal scheme and those from adverse ideal scheme are shown in Table 6.

TABLE 6 Differences of performance evaluation indicators from positive ideal scheme and those from adverse ideal scheme

Difference	Scheme 1	Scheme 2	Scheme 3
Positive ideal scheme	0.053	0.092	0.099
Adverse ideal scheme	0.118	0.088	0.082

Thus, the fuzzy incidence degree of different performance evaluation schemes is. Therefore, enterprise 1 has the highest performance level of green supply chain.

5 Conclusions

This paper constructs a performance evaluation index system of enterprise green supply chain. Indicators of different classes and types are standardized according to their features and a performance evaluation model of fuzzy analysis method of multi attribute decision making is

References

- [1] Xiong W 2008 Research on the performance evaluation of green supply chain *China Logistics and Purchasing* **12** 74-5
- [2] Bala Subrahmanya M H 2005 Pattern of technological innovations in small enterprises: comparative perspective of Bangalore (India) and Northeast England (UK) *Technovation* **25**(3) 269-80
- [3] Ji J, Liu Q, Guo Z 2013 Performance Appraisal of Green Supply Chain Management of Steel Enterprises *Science and Technology Management Research* **16** 53-7
- [4] Liu W 2010 The research on performance evaluation of green supply chain management literature reviews and trends *Logistics Sci-Tech* **6** 103-5
- [5] Huang H 2010 Analysis of successful factors of green supply chain implementation *Productivity Research* **10** 215-6

constructed. This provides a quantitative analysis method for performance evaluation of enterprise green supply chain management.

Fuzzy analysis method of multi attribute decision making is suitable for evaluating green supply chain in real situation. It is simple and clear. More importantly, it can realize the uncertain evaluation by qualitative description and evaluate the performance of green supply chain in a quantitative way. It serves as guidance to scientific management and decision making and helps to increase the competitiveness of the enterprise.

Acknowledgments

This paper is supported by Henan Province Soft Science Research Foundation (No. 122400450107), Scientific Research Foundation of Henan Province Education Department (No. 13A630717) and Human-Social Sciences Foundation of Henan Province Education Department (No. 2013-ZD-130).

- [6] Bai S, Li Y 2013 Research on method for green supply chain management of enterprises *Logistics Engineering and Management* **3** 117-9
- [7] Li H, Dan B, Rao K, Zhao H 2012 Study on the cooperation decision making model in interrelated supply chains for wastes reusing considerations *Advanced Materials Research* **2012** 1397-401
- [8] Lv L, Liang Y, Peng C 2008 Green degree evaluation of green supply chain of enterprises based on Fuzzy-AHP model *Science Technology and Industry* **3** 1-5
- [9] Liu X, Wang Z 2011 The performance evaluation of green supply chain of enterprise 2011 International Symposium on Applied Economics *Business and Development* **8** 304-9
- [10] Per Erik Eriksson 2010 Improving construction supply chain collaboration and performance: a lean construction pilot project *Supply Chain Management: An International Journal* **15**(5) 394-403

- [11] Wang T, Yang A, Zhong S 2014 Multi-Attribute Extension Fuzzy Optimized Decision Making Model of Scheme Design *Tehnički vjesnik/Technical Gazette* 21(2) 239-47
- [12] Li S, Hu F 2006 Selecting and estimating green partner based on fuzzy evaluation model *Journal of Hunan University (Natural Science)* 33(3) 137-40
- [13] Zhang M, Wu H 2005 Study of green supply chain and its performance measurement system *Science and Management* 3 23-5
- [14] Rao C, Xiao X, Peng J 2007 Novel combinatorial algorithm for the problems of fuzzy grey multi-attribute group decision making *Journal of Systems Engineering and Electronics* 18(4) 774-80

Authors



Huaping Zhang, born in May, 1980, Dongying, Shandong, China

Current position, grades: Associate professor, business management in Management & Economics School of North China University of Water Resources and Electric Power.

University studies: Bachelor's degree in Trade Economics (2002), Master's degree in Management Science and Engineering (2008) at North China University of Water Resources and Electric Power and Doctor's degree in National Economics (2011) at Henan University.

Scientific interest: business strategic management, cost management accounting, technological economic management.

A performance evaluation model of supply Chain based on extension correlation function

Jiaojin Ci*

Institute of Economy and Management, Normal College of Nanyang, Nanyang, Henan, China

Received 1 September 2014, www.tsi.lv

Abstract

Performance evaluation of supply chain is complex and of uncertainty, and is influenced by factors of multiple levels. Directing at the features mentioned above, this paper studied the performance evaluation of supply chain of multiple attributes, and put forward a performance evaluation model and algorithm of supply chain based on extension correlation function. Via analysis of relative factors in the process of performance evaluation of supply chain, the model gave out an index system of performance evaluation of supply chain and by standardization of different evaluation indexes, it built an improved extension correlation function between evaluation indexes of enterprise supply chain performance and the ideal range of supply chain performance. Thus, the comprehensive weighted extension goodness between enterprise supply chain performance and the ideal range of supply chain performance could be obtained. According to the value of comprehensive weighted extension goodness, the supply chain with the optimal implementation effect could be selected, so as to offer effective support for the follow-up implementation of supply chain. Finally, in order to offer a scientific method for improving the supply chain and the enterprise competitiveness, the model and algorithm was tested an actual case.

Keywords: supply chain, performance evaluation, extension correlation function, model

1 Introduction

Supply chain management is an operation model forming and developing in fierce competition with the aim to improve the core competitiveness. Along with the rapid development of computer science and technology and IT science and technology, in today's commercial environment that tends to be disordered and dynamic, traditional supply chain is evolving to a dynamic supply chain of higher agility. It will make the supply chain structure more diversified, and the relationship among enterprises as supply chain members more complex. Thus, the performance evaluation of supply chain will be more complex with multiple levels of influence factors [1-3]. To evaluate and measure the performance of supply chains and management effects, relative indexes need to be chosen. However, if the organization performance is only evaluated from return of investment (ROI), the static measurement of organization performance based on financial accounting and management can no longer effectively evaluate the performance of supply chain. By far, researches directing at supply chain performance mainly concentrate on logistics, information flow, capital flow, and systematically analysis on supply chain performance is few. What's more, most of performance evaluation models of supply chain only considered single performance index. The indexes considered includes inventory level, cost, date of delivery, etc. And generally, those models only provide ideal decision-making and operation method, and do not provide discussion on measurement of concrete supply chain performance. And

aspects including choices of indexes and construction of evaluation methods call for further study [4-6]. By far, there have been many decision-making and evaluation methods of complex system. In different system, different decision-making and evaluation methods focus on different aspects in solving problems. Some studies on those methods have made corresponding achievements [7-10]. For performance evaluation system of supply chain, there are multiple and complex factors that influence the performance value of supply chain and some indexes can only be described with qualitative fuzzy description rather than accurate quantitative value. Thus, the process of performance evaluation of supply chain is a process of decision-making analysis of complex system with incomplete information. Thus, this paper studied the performance evaluation of chain supply based on extension theory [11-14]. A normative evaluation index system of supply chain performance was established. It can analyse the supply chain performance of enterprise comprehensively based on extension correlation function, which made the modelling and analysing of supply chain more comprehensive, accurate and effective. Overall, with this model, the management of supply chain can be more targeted, which can provide a strong support for enterprise development.

* *Corresponding author* e-mail: 15938469500@163.com

2 Construction of evaluation index system of supply chain

2.1 BASIS CHARACTERISTICS OF EVALUATION INDEXES OF SUPPLY CHAIN PERFORMANCE

Different scholars have different consideration in determination of evaluation index of supply chain. For example, at present, the typical ROF model analyses supply chain performance from the aspects of resources, output, flexibility, etc.; SCOR model evaluates the supply chain performance based on four management process, namely material supply, manufacturing, transportation and planning; Lummus R R model analyses performance evaluation of supply chain form supply, transformation, transportation and requirement management. It can be observed that these models improve and supplement traditional performance evaluation of supply chain from different perspectives. However, by far there are still some limitations in performance evaluation of supply chain. For example, there are few studies focusing on the performance evaluation of the whole supply chain, and most of the studies only focused on the chosen of indexes but did not evaluate and analyse the whole supply chain, and performance evaluation of supply chain based on the operation flow is not reflected. Thus, choice of evaluation indexes of performance evaluation of supply chain in the new circumstance needs to be focused on the connection

between enterprises in the chain and the interior and exterior environment of supply chains. What’s more, the relationship and integrally of members in the whole chain and the consistency of operation of members should be emphasized, and meanwhile, the potential relationship among business process, pecuniary condition and market condition should be reflected. Thus, choice of evaluation indexed of supply chain performance should follow some basic rules including comprehensiveness that the evaluation indexes of supply chain reflect the operating condition of the whole supply chain. Systematises that the evaluation indexes of supply chain reflect the integration and coordination of the operation of the whole supply chain, importance that the evaluation indexes chosen should be key performance indicator in performance evaluation of supply chain.

2.2 EVALUATION INDEX SYSTEM OF SUPPLY CHAIN PERFORMANCE

Directing at some problems in choice of evaluation indexes for existing performance evaluation of chain supply, by combining relative principles in choice of evaluation indexes of supply chain performance, this study provided an improved evaluation index system of supply chain performance from the perspectives of market, finance and operation flow. Its structure is presented in Figure1.

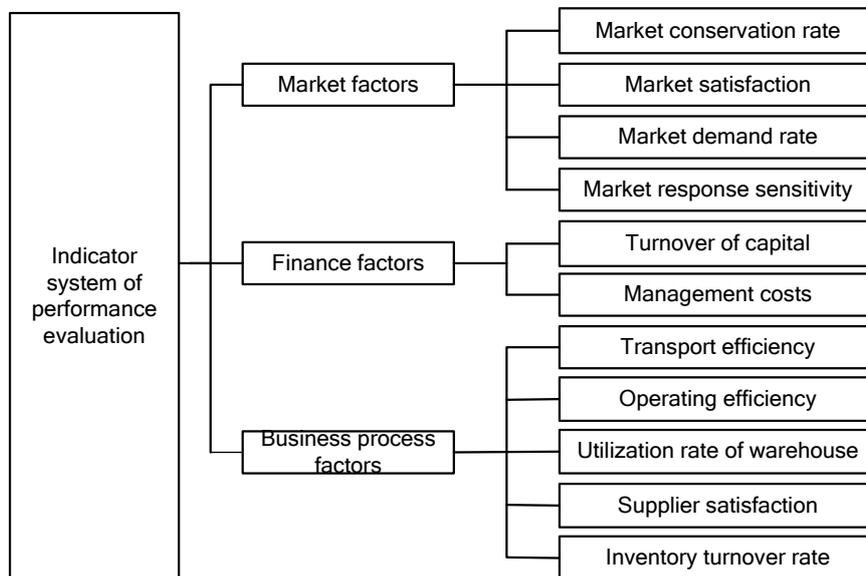


FIGURE 1 Evaluation index system of supply chain performance

3 Performance evaluation model and algorithm of supply chain based on extension correlation function

3.1 STANDARDIZATION OF EVALUATION INDEXES OF SUPPLY CHAIN PERFORMANCE

In the process of choosing evaluation indexes of supply chain performance, some indexes are not a certain value, so these indexes are fuzzy; some indexes are positive

indexes while some are negative indexes. Thus, these indexes of multiple levels, attributes and categories need to be standardized. Thus, directing at qualitative and quantitative measurement, this paper puts forward the process of standardization presented as below.

If the evaluation index of supply chain performance p_i is a positive index, its value $v_i(p) = [v_i^a(p), v_i^b(p)]$. The corresponding standardized index is:

$$v_i^*(p) = [v_i^{*a}(p), v_i^{*b}(p)] = \left[v_i^a(p) / \max_{1 \leq j \leq m} (v_{ij}^b(p)), v_i^b(p) / \max_{1 \leq j \leq m} (v_{ij}^b(p)) \right], \tag{1}$$

where: $\left[v_i^a(p) / \max_{1 \leq j \leq m} (v_{ij}^b(p)), v_i^b(p) / \max_{1 \leq j \leq m} (v_{ij}^b(p)) \right]$ is the maximum value in the interval of evaluation index of supply performance p_i .

If the evaluation index of supply chain performance p_i is a negative index, its value $v_i(p) = [v_i^a(p), v_i^b(p)]$, the corresponding standardized index is:

$$v_i^*(p) = [v_i^{*a}(p), v_i^{*b}(p)] = \left[\min_{1 \leq j \leq m} (v_{ij}^b(p)) / v_i^a(p), \min_{1 \leq j \leq m} (v_{ij}^b(p)) / v_i^b(p) \right], \tag{2}$$

where: $\left[\min_{1 \leq j \leq m} (v_{ij}^b(p)) / v_i^a(p), \min_{1 \leq j \leq m} (v_{ij}^b(p)) / v_i^b(p) \right]$ is the maximum value in the interval of evaluation index of supply performance p_i .

If the evaluation index of supply chain performance p_i is a fuzzy quantitative description, fuzzy evaluation and corresponding fuzzy membership should be applied. The values and indexes are presented in Table 1.

TABLE 1 The membership degree of fuzzy evaluation of evaluation indexes of supply chain performance

Degree of fuzzy membership	State level of index
0.1	Vary bad
0.3	Bad
0.5	Fair
0.7	Good
0.9	Very good
0.2,0.4,0.6,0.8	Between the neighbouring state levels mentioned above

By standardization of evaluation indexed of supply chain mentioned above, the discrepancy between different indexes was removed and the measuring standard was unified. In this way, the performance evaluation of supply chain can be more accurate.

3.2 CALCULATION MODEL OF EXTENSION CORRELATION OF SUPPLY CHAIN PERFORMANCE

Extenics is a subject of intelligent design that studies the rule and method of contradictory issues among or within things with formalized model. This subject, with characteristics of formalization, logicalization and mathematization, has corresponding achievements of

$$\rho_{ij}^o = \frac{\left(\left| v_{ij}^a(p) - \frac{v_i^{o* a}(p) + v_i^{o* b}(p)}{2} \right| + \left| v_{ij}^b(p) - \frac{v_i^{o* a}(p) + v_i^{o* b}(p)}{2} \right| + v_i^{o* a}(p) - v_i^{o* b}(p) \right)}{2}. \tag{7}$$

engineering application in many fields and has formed extension engineering extension direction. This paper provides an extension correlation calculation model of supply chain performance by improving the classical extension distance.

If the corresponding positive ideal domain $v_i^{o*}(p)$ of standardized evaluation index of supply chain performance p_i is:

$$v_i^{o*}(p) = [v_i^{o* a}(p), v_i^{o* b}(p)] = \left[\max_{1 \leq j \leq m} (v_{ij}^{* a}(p)), \max_{1 \leq j \leq m} (v_{ij}^{* b}(p)) \right]. \tag{3}$$

Then the corresponding negative ideal domain $v_i^{N*}(p)$ of standardized evaluation index of supply chain performance p_i is:

$$v_i^{N*}(p) = [v_i^{N* a}(p), v_i^{N* b}(p)] = \left[\min_{1 \leq j \leq m} (v_{ij}^{* a}(p)), \min_{1 \leq j \leq m} (v_{ij}^{* b}(p)) \right]. \tag{4}$$

When the standardized evaluation index of supply chain performance p_i is an accurate value, namely $v_i(p) = v_i^a(p) = v_i^b(p)$, the extension distance ρ_{ij}^o of the evaluation phase of supply chain performance j between evaluation index of supply performance p_i and positive ideal domain $v_i^{o*}(p)$ is:

$$\rho_{ij}^o = \left| v_i(p) - \frac{v_i^{o* a}(p) + v_i^{o* b}(p)}{2} \right| - \frac{v_i^{o* b}(p) - v_i^{o* a}(p)}{2}. \tag{5}$$

The extension distance ρ_{ij}^N of the evaluation phase of supply chain performance j between evaluation index of supply performance p_i and negative ideal domain $v_i^{N*}(p)$ is:

$$\rho_{ij}^N = \left| v_{ij}(p) - \frac{v_i^{N* a}(p) + v_i^{N* b}(p)}{2} \right| - \frac{v_i^{N* b}(p) - v_i^{N* a}(p)}{2}. \tag{6}$$

If the standardized evaluation index of supply chain performance p_i is fuzzy value, namely $v_i(p) = [v_i^a(p), v_i^b(p)]$, the extension distance ρ_{ij}^o of the evaluation phase of supply chain performance j between evaluation index of supply performance p_i and positive ideal domain $v_i^{o*}(p)$ is:

The extension distance ρ_{ij}^N of the evaluation phase of supply chain performance j between evaluation index of supply performance p_i and negative ideal domain $v_i^{o*}(p)$ is:

$$\rho_{ij}^N = \frac{\left(\left| v_{ij}^a(p) - \frac{v_i^{N*a}(p) + v_i^{N*b}(p)}{2} \right| + \left| v_{ij}^b(p) - \frac{v_i^{N*a}(p) + v_i^{N*b}(p)}{2} \right| + v_i^{N*a}(p) - v_i^{N*b}(p) \right)}{2} \tag{8}$$

When all the extension distances are obtained, the extension correlation function K_j^O between evaluation phase of supply chain performance j and ideal positive domain $v_i^{o*}(p)$ and the extension correlation K_j^N between evaluation phase of supply chain performance j and negative ideal domain $v_i^{o*}(p)$ are respectively:

$$K_j^O = \begin{cases} \rho_{ij}^O / \max_{0 \leq j \leq 1} |\rho_{ij}^O| & \max_{0 \leq j \leq 1} |\rho_{ij}^O| \neq 0 \\ 0 & \max_{0 \leq j \leq 1} |\rho_{ij}^O| = 0 \end{cases}, \tag{9}$$

$$K_j^N = \begin{cases} \rho_{ij}^N / \max_{0 \leq j \leq 1} |\rho_{ij}^N| & \max_{0 \leq j \leq 1} |\rho_{ij}^N| \neq 0 \\ 0 & \max_{0 \leq j \leq 1} |\rho_{ij}^N| = 0 \end{cases}. \tag{10}$$

Thus, the extension correlation ψ_j^O between evaluation phase of supply chain performance j and positive ideal domain $v_i^{o*}(p)$, and the extension correlation ψ_j^N between evaluation phase of supply chain performance j and negative ideal domain $v_i^{N*}(p)$ are respectively:

$$\begin{cases} \psi_j^O = \sum_{i=1}^n (w_i * K_j^O) \\ \psi_j^N = \sum_{i=1}^n (w_i * K_j^N) \end{cases}. \tag{11}$$

3.3 REALIZATION OF PERFORMANCE EVALUATION AND ALGORITHM OF SUPPLY CHAIN BASED ON EXTENSION CORRELATION FUNCTION

If the extension correlation of the evaluation phrase of supply chain performance j subordinate to the positive ideal domain of evaluation performance of supply chain $v_i^{o*}(p)$ is ψ_j ($0 \leq \psi_j \leq 1$), the extension correlation of the evaluation phrase of supply chain performance j subordinate to the negative ideal domain of evaluation performance of supply chain $v_i^{N*}(p)$ is $1 - \psi_j$. In order to determine the extension correlation ψ_j , the objective function was established according to the advantages and disadvantages of positioning implementation plan:

$$F(\psi_j) = \min \left\{ (\psi_j * \psi_j^N)^2 + ((1 - \psi_j) * \psi_j^O)^2 \right\}. \tag{12}$$

The extension correlation ψ_j can be obtained according to extremum principle:

$$\psi_j = 1 / \left(1 + (\psi_j^N / \psi_j^O)^2 \right). \tag{13}$$

According to what mentioned above, the implementation steps of evaluation model and algorithm of supply chain based on extension correlation function is as below:

Step 1: According to the principles of choosing evaluation index of supply chain performance, establish an evaluation index system of supply chain performance of multiple levels and attributes by consulting relative design specialist.

Step 2: Based on interrelated thesis about standardization of evaluation indexes of supply chain performance mentioned in Section 2.1, standardize positive evaluation indexes according to Equation (1), negative evaluation indexed according to Equation (2) and fuzzy evaluation indexes according to Figure 1.

Step 3: Build positive ideal domain and negative ideal domain of different types of evaluation indexes under the evaluation index system of supply chain performance by Equations (3) and (4)

Step 4: Obtain the extension distance between different evaluation indexes of supply chain performance and the positive ideal domain according to Equations (5) and (7).

Step 5: Obtain the extension distance between different evaluation indexes of supply chain performance and the negative ideal domain according to Equations (5) and (7).

Step 6: Obtain the extension correlation function between performance evaluation of supply chain and positive ideal domain of evaluation index of supply chain performance according to Equation (9), and the extension correlation function between performance evaluation of supply chain and negative ideal domain of evaluation index of supply chain performance according to Equation (10).

Step 7: Considering the weight of different evaluation index of supply chain performance, obtain the extension correlation between performance evaluation of supply chain and positive ideal domain of evaluation index of supply chain performance, and the extension correlation between performance evaluation of supply chain and negative ideal domain of evaluation index of supply chain performance based on Equation (11).

Step 8: Obtain the comprehensive correlation of performance evaluation of supply chain of all the evaluation indexed according to Equation (12).

Step 9: Analyse the performance evaluation of supply chain according to the value of comprehensive extension correlation.

4 Case studies

In this paper, a supply chain in which a brand clothing enterprise is the core was chosen as the major research subject to explain and analyse the model and algorithm in this paper. This clothing enterprise is a collective operation

enterprise with production-supply-marketing integration. It has a wholesome system of production, selling and designing, and also relevant raw material suppliers, distribution alliance business and third party logistic service provider. The market operating area of the enterprise is wide, and the business cooperation with other sections is stable. Thus, the supply chain with this enterprise as its core is formed. Relative data of this enterprise from 2011 to 2012 was collected. Moreover, by the evaluation index system of supply chain performance and corresponding calculation model and algorithm, different evaluation indexes of supply chain performance was standardized. Results are presented in Table 2.

TABLE 2 evaluation indexes of supply chain performance and standardized data

Evaluation index	Index type	2011		2012	
		Original data	Standardized data	Original data	Standardized data
Market retention rate	Positive index	0.78	1.000	0.72	0.923
Market satisfaction	Positive index	0.85-0.88	0.966-1.000	0.78-0.80	0.886-0.909
Market demand rate	Positive index	0.12-0.15	0.750-0.938	0.13-0.16	0.813-1.000
Market reaction sensitivity	Positive index	0.80	1.000	0.80	1.000
Rate of capital turnover	Positive index	0.61	0.938	0.65	1.000
Management cost	Negative index	560	1.000	620	0.903
Transport efficiency	Positive index	0.68-0.71	0.958-1.000	0.63-0.65	0.887-0.915
Operating efficiency	Positive index	0.57-0.59	0.934-0.967	0.58-0.61	0.951-1.000
Utilization rate of warehouse	Positive index	0.78	0.951	0.82	1.000
Supplier cooperation satisfaction	Positive index	0.66	1.000	0.54	0.818
Inventory turnover rate	Positive index	0.85	1.000	0.83	0.976

According to the Table of extension distance of evaluation index of supply chain performance, the extension distance between evaluation index of supply chain performance and positive ideal domain and the

extension distance between evaluation index of supply chain performance and negative ideal domain can be obtained respectively. Results are presented in Table 3.

TABLE 3 Extension domain of evaluation index of supply chain performance

Evaluation index	2011		2012	
	Extension distance of positive ideal domain	Extension distance of negative ideal domain	Extension distance of positive ideal domain	Extension distance of negative ideal domain
Market retention rate	0	0.077	0.077	0
Market satisfaction	0	0.074	0.069	0
Market demand rate	0.094	0.063	0.063	0.094
Market reaction sensitivity	0	0	0	0
Rate of capital turnover	0.062	0	0	0.062
Management cost	0	0.097	0.097	0
Transport efficiency	0	0.054	0.047	0
Operating efficiency	0.017	0.015	0.008	0.040
Utilization rate of warehouse	0.049	0	0	0.049
Supplier cooperation satisfaction	0	0.182	0.182	0
Inventory turnover rate	0	0.024	0.024	0

According to the extension correlation function of evaluation index of supply chain performance, the extension correlation function between evaluation index of supply chain performance and positive ideal domain and the extension correlation function between evaluation index of supply chain performance and negative ideal domain can be obtained. The results are presented in Table 4.

Thus, the comprehensive extension correlation of performance evaluation of supply chain of this enterprise

is 0.485 in 2011 and 0.651 in 2012. According to the evaluation analysis above, the supply chain capacity of the supply chain with brand clothing enterprise presented a tendency of rapid decline in 2011-2012. It shows that this enterprise should put more resources into the evaluation indexes that presented a decline tendency in the evaluation index system of supply chain performance to stop the decline.

TABLE 4 Extension correlation function of evaluation index of supply chain performance

Evaluation index	2011		2012	
	Extension correlation function of positive ideal domain	Extension correlation function of negative ideal domain	Extension correlation function of positive ideal domain	Extension correlation function of negative ideal domain
Market retention rate	0	1.000	1.000	0
Market satisfaction	0	1.000	0.932	0
Market demand rate	1.000	0.670	0.670	1.000
Market reaction sensitivity	0	0	0	0
Rate of capital turnover	1.000	0	0	1.000
Management cost	0	1.000	1.000	0
Transport efficiency	0	1.000	0.870	0
Operating efficiency	1.000	0.375	0.471	1.000
Utilization rate of warehouse	1.000	0	0	1.000
Supplier cooperation satisfaction	0	1.000	1.000	0
Inventory turnover rate	0	1.000	1.000	0

5 Conclusions

This paper studied the performance evaluation of supply chain of complexity, multiple attributes and multiple levels of influence factors and relative problems in decision-making. Based on it, this study provided an improved performance evaluation model and algorithm, which is based on extension correlation function. The paper firstly analysed the relative constraints in the implementation process of performance evaluation of supply chain. Based on it, corresponding evaluation index system of supply chain performance was established. Then the ideal

domains of different types of evaluation indexes under this evaluation index system was constructed, and the extension distance between the improved evaluation index of supply chain performance and the ideal domain of each evaluation index of supply chain performance. And based on it, the comprehensive extension correlation of supply chain performance was constructed. In this way, supply chain performance can be evaluated effectively. This model is simple in calculation with high resolving capacity and its performance and operation is quite simple. Thus, it can provide support for computer-assisted performance evaluation of supply chain.

References

- [1] Freed and Reed 2000 How to evaluate supply chain partner-ship *Journal of Business Logistics* **21**(2) 127-45
- [2] Lagace, Dahlstorm, Gassnheimer 1997 The impact of supply chain integration on Cooperating performance *Logistics Information Management* **6**(4) 10-9
- [3] Ma S, Lin Y, Chen Z 2005 Supply Chain Management *Beijing Mechanics Press*
- [4] Liu L 2003 Survey on evolution of SCM theory and methods *Journal of Management Sciences in China* **6**(2) 81-8
- [5] Persson F, Olhager J 2002 Performance simulation of supply chain designs *International Journal of Production Economics* **77** 231-45
- [6] Huo J, Sui M, Liu Z 2002 Construction of integrated supply chain performance measurement system *Journal of Tongji University* **30**(4) 495-9
- [7] Liu Sifeng, Cai Hua, Cao Ying, Yang Y 2011 Advance in grey incidence analysis modelling *2011 IEEE International Conference on Systems, Man and Cybernetics (SMC)* October 2011 1886-90
- [8] Qi X W, Liang C Y, Zhang E Q, Ding Y 2011 Approach to interval-valued intuitionistic fuzzy multiple attributes group decision making based on maximum entropy *Systems Engineering - Theory & Practice* **31**(10) 1940-8 (in Chinese)
- [9] Blaszczynski J, Greco S, Slowinski R, Szelag M 2009 Monotonic Variable consistency rough set approaches *International Journal of Approximate Reasoning* **50**(7) 979-9
- [10] Wang T, Yang A, Bu L 2013 Mechanism scheme design based on multi-attribute extension gray relevant optimized decision-making model *Systems Engineering - Theory & Practice* **33**(9) 2321-9 (in Chinese)
- [11] Yang C Y, Cai W 2007 Extension Engineering (Ke Tuo Gong Cheng) *Beijing: Science Press*
- [12] Zhao Y W Study of Extension Conceptual Design of Mechanical Product *Chinese Engineering Science* 2001 **3**(5) 67-71
- [13] Gong J Z, Qiu J, Li G X Product Configuration Design Based On Extension Theory *Computer Integration Manufacturing System* 2007,13(9):1700-9 (in Chinese)
- [14] Yang C Y, Cai W 2012 Research Progress of Extension Concentration Correlation Function *Journal of Guangdong University of Technology* **29** 7-14

Authors



Jiaojin Ci, born in November, 1982, Anging, Anhui, China

Current position, grades: Lecturer at the Institute of Economy and Management, Normal College of Nanyang. Director of the Training Center.

Scientific interest: Logistic management and Electronic Commerce.

Sourcing and pricing strategy research of competition supply chain under supply disruption

Ling Hou^{1*}, Meng Li², Dongyan Chen¹

¹Department of Applied Mathematics, Harbin University of Science and Technology, 150080, Harbin, China

²School of Management, Harbin Institute of Technology, 150001, Harbin, China

Received 1 August 2014, www.tsi.lv

Abstract

Under the environment of supply disruption, it is significant to study decision-making, because sourcing strategies of retailers impact the profit of the supply chain while the pricing strategies of suppliers affect all aspects of the supply chain. In this paper, the demand distribution function of each supply chain is obtained, which is based on the total demand of two supply chains with given distribution function, and the sourcing and pricing problems are obtained in supply chain network under the environment of supply disruption. In order to decompose the total demand with the given distribution function, customer choice theory is adopted to acquire the demand of each supply chain. By game theory and optimization theory, we obtain the sourcing strategies of two retailers and the pricing strategies of two suppliers in this system. Finally based on the assumption of a uniform demand distribution, the outcomes of the proposed models are demonstrated with a numerical example. The results show that when disruption probability or delivery cost are high, retailers will only order from the spot market although the spot market wholesale prices are a little high; but when the disruption probability is moderate or low, the retailer would rather place orders from suppliers. Specific purchasing method depends on the competition ability between suppliers.

Keywords: supply chain network, supply disruption, sourcing strategies, pricing strategies

1 Introduction

With the deepening of globalization and integration of supply chain, competition has become a hot topic in the area of supply chain management. Supply chain management, which was treated as a method for the enterprise accessing to the core competitiveness, has turned to a more complex system, namely the supply chain network management. At present, research on supply chain network mainly focuses on some simple network structures, such as one-to-many and many-to-one. Wu explored the equilibrium structure for two competing supply chains. Each chain has one manufacturer with two exclusive retailers, that is, the supply chain network structure is of 1-2 type [1]. Ha investigated the contracting problem using a two-stage game in two competing supply chains with information sharing, each consisting of one manufacturer and one retailer, that is, the supply chain network structure is of 1-1 type [2]. For more similar study, see literatures [3-5] for details. Note that above literatures do not involve competitive decision-making problems of many-to-many network. Therefore, in this paper we just discuss the case of supply network of 2-2 type which is a generalization of the above 1-2 and 1-1 type structures. It proves that this kind of network structure is fit for the actual operation state of supply chain. For example, products of Haier and ChangHong are both in Gome and

Suning's supply chain and there is dramatic competition for market share between Gome and Suning.

The research of this paper is closely related to supply chain sourcing management and supply disruptions management. Early literatures on sourcing management often assume that cost, quantity and distribution ability are three important factors that wholesalers need to consider before making a decision (Dickson [6]; Verma and Pullman [7]; Weber [8] et al.). For example, Weber [8] concluded that the order quantity was the most important criteria for retailers to develop a sourcing strategy, while the cost and picking (delivery ability) followed. Talluri and Narasimhan extended the work of Weber by treating prices as an input variable and the amount of revenue and delivery ability as system outputs [9].

In the above literature, the authors did not consider the impact of the competition of several companies on the sourcing strategy, where a single supplier model easily leads to duopoly suppliers. In recent years, some scholars began to study the sourcing strategy for competitive enterprises. Parlar and Perry consider a firm that faces constant demand and sources from two identical-cost capacitated suppliers which are subject to production failure. Inter failure and repair times are exponentially distributed for both suppliers. The authors propose a suboptimal ordering policy that is solved numerically [10]. Gurler and Parlar extended the work of Parlar and Perry by considering the case of Erlang inter failure times and

* Corresponding author e-mail: houling0729@163.com

general repair times. They propose that, for the same order costs and order scale without restrictions, downstream of supply chain can reduce order quantity and/or diverted to other route; Otherwise, the supply chain would use other interference management strategy, such as double sourcing and emergency purchase [11]. Xiao et al. studied the vendor selection problem which says that a supplier procures major raw materials from which raw materials suppliers [12]. Although the above two articles explore the impact of competition on the sourcing strategy and wholesale pricing strategies of supplier, they are in a stable market supply environment and do not involve the supply uncertain environment, which more reflects scenario in reality.

Supply disruptions management has become an increasing concern of the business and academia. Many companies began to realize that the supply disruption has seriously affected the ability to successfully manage their supply chain. Literature on supply disruptions management has a huge body. However, most of these studies assume a single supplier and alternative energy is not available in the system of a single supplier. While a large number of studies have shown that the effective method for weakening supply disruptions is the multi-sourcing strategy. Therefore, some scholars began to study the multi-sourcing strategy for supply disruptions. Tomlin and Wang developed a single period dual-sourcing model by two suppliers with yield uncertainty. But the information between two suppliers is not completely symmetrical: one unreliable supplier and one reliable (and thus more expensive) supplier. They focus on inventory and sourcing mitigation. They concluded that retailers can reduce interrupt risk by an appropriate purchasing strategy although the two-supplier information is not symmetrical [13]. Chopra considered the mitigation-disruption strategy when the unreliable supplier is subject to both recurrent and disruption uncertainty [14]. Further, Federgruen and Kleindorfer considered Type I service-level-related constraints in their yield management models [15, 16], while Yang et al. propose an interesting analytical approach on the multiple-sourcing random yield problem [17]. Here Yang and Dada consider the problem of a newsvendor that is served by multiple suppliers, where a supplier is defined to be either perfectly reliable or unreliable. They showed that in the optimal solution a supplier will be selected only if all less-expensive suppliers are selected, regardless of the supplier's reliability [17, 18]. Literature on supply disruptions mentioned above only investigated the retailer's strategy and assumed that the supplier's strategy is exogenously given. However, the supplier's response, for example their pricing strategies, also influences the decisions of the supply chain members. The wholesale price setting caused many scholars' attention, representative literatures give the optimal pricing strategies of suppliers in different situations, including Lariviere and Porteus [19], Wang and Gerchak [20], Bernstein and DeCroix [21], Serel [22], He [23], Cho and Tang [24], Surti C and Hassini [25] etc.

However, the above literatures only discuss the many-to-one supply chain structure, without considering the many-to-many supply chain network pricing problem under supply disruptions.

This paper differs from the existing studies in the following aspects. First, the structure of the supply chain is an extension of the above supply chain structure; secondly, supply disruption is introduced into the sourcing strategy; Third, we consider how to deal with unmet demand. The above-mentioned literature assumes that every retailer only has one ordering opportunity in the entire sales process and is not allowed emergency replenishment and that unmet demand will be lost. In our paper, we assume that retailers may procure from the spot market, rather than lost, for unmet demand after observing the demand. Finally, we investigate the impacts of supply disruption on the retailer's sourcing strategy by both theoretical and computational analyses.

The remainder of this paper is as follows. In the next section, we give problem statement and model of demand function. Then, we obtain the sourcing strategies of the retailer and the supplier in Section 3 and Section 4. In Section 5, we give numerical examples to verify the validity of the results. The last section summarizes the research findings and future research directions.

2 The problem statement and model of demand function

2.1 THE PROBLEM STATEMENT

This paper analyses the competition between two supply chains with only two echelons. That is, both of two chains consist of one manufacturer and one retailer. Because members in the one supply chain often belong to the other supply chain in reality, we consider the competition between supply chains with a crossover structure in order to actually reflect competitive scenario in reality. Namely, there are cooperative relations between the two suppliers and two retailers, where two suppliers can supply to two retailers. That is 2-type 2 network structure which is shown in Figure 1.

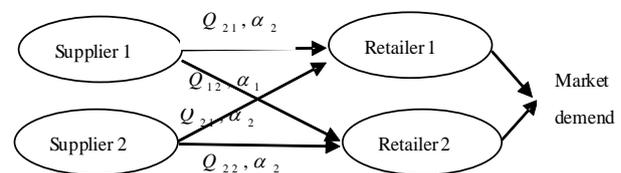


FIGURE 1 Supply chain network structure

There is lots of practical background about this model. For example, Haier and ChangHong are both in Gome and Suning's supply chain, and there is dramatic competition for market share between Suning and Gome. Similarly, two large supermarkets in a city sell the same products from two manufacturers.-This model in Figure 1 can be applied to analysis the competition decision between these enterprises. The assumptions for the 2-2 network model are listed below:

1) All parties are risk neutral. The two supply chains, suppliers and retailers are indexed by i and j , where $i, j \in \{1, 2\}$. There exists the possibility of disruption for the two suppliers. We use α_i to denote the reliability level of supplier i . With probability α_i , supplier i is “up” and fully fulfils retailer i 's order; with probability $1 - \alpha_i$, supplier i is “down” and retailer i receives no inventory. We say that “reliability is high” or “disruption risk is low” if α_i is high. We consider the general case where two suppliers have different reliability levels. When an interrupt occurs, we assume that the retailer's order cannot be met, which means that the retailer is completely disrupted. At the same time, we assume that there is a reliable spot market as an emergency supply points for the two retailers.

2) c_i is the unit delivery cost of the product of supplier i and ω_{ij} is the unit wholesale price of the product for retailer i offered by supplier j .

3) p_i is the fixed unit selling price of the product for retailer i . Q_{ij} is the order quantity of retailer i placed with supplier j . Q_{i3} is the inventory level after making an emergency order from the spot market. ω_{i3} is the fixed unit wholesale price of the product offered by the spot market. The unit goodwill cost of unmet demand is denoted by b_i . The surplus stock that remains unsold at the end of the period can be sold to a secondary market at a unit salvage value v_i . We always assume that $b_i < \omega_{ij}$.

(4) Marginal cost $\gamma_i c_i$ is incurred under supply disruption, where $0 < \gamma_i < 1$ denotes the total proportion of the marginal delivery cost of supply chain i in the event of a failure. We suppose that this part of costs is shared by the failing supplier and the retailer. The proportion of the cost incurred by the supplier i is η_i where $0 < \eta_i < 1$. This cost structure is different from that used in most of the literatures in which only the retailer assumes the cost in the event of a failure. However, this is not always true. In fact, before supply failures occur, both the retailer and suppliers usually have incurred some costs, which may include fixed set-up costs and variable costs. To simplify the analysis, we assume that all the setup costs are zero and all the variable costs in the event of a supply failure are proportional to the delivery cost and to the order quantity

Among the above variables, Q_{ij} , Q_{i3} and ω_{ij} is decision variables and the others exogenous variables, respectively, which are known to all the members of the supply chain. In this paper we focus on the revenues of two suppliers and two retailers. The revenue of the spot market and its delivery cost are not considered. The spot market is not a decision- maker in our paper. In addition, the retailer sells the product at a fixed price in the market in the selling season. Any unmet demand will incur a goodwill cost to the retailer. After the selling season, the residual product will be salvaged. We assume that $0 < v_i < c_i < \omega_{ij} < \omega_{i3} < p_i$.

In this article, the Nash game is obeyed by the two supply chains, while the Stackelberg game is subjected by the internal of the two supply chains. The sequence of

events is as follows:

1) The two suppliers decide their individual wholesale prices simultaneously (stage 0);

2) The two retailers decide their individual order quantities with suppliers 1 and 2 simultaneously in anticipation of supply disruption and demand (stage 1);

3) The retailer i makes an emergency order from the spot market after a supply disruption but before demand occurs (stage 2).

2.2 MODEL OF DEMAND FUNCTION

Total demand D_i is assumed to be a positive stochastic random variable with probability density function $f(x)$ and the differentiable and strictly increasing cumulative distribution function $F(x)$, where $F_p(0) = 0$, $F_p(\infty) = 1$ and its reverse function is F_p^{-1} , $p = (p_1, p_2)$ denotes price vectors for two products, which means price influences on the demand. Obviously, F_p^{-1} is still strictly monotone increasing and second differentiable.

Due to different quality of these two kinds of products, consumer choice is influenced by their own salary level, product prices and the effect of brand. Hence, this paper always assumes that consumer demand for product i is influenced by product price p_i and quality level S_i . Denote consumer preference coefficient for product characteristics S_i by a . Then the consumer utility function for product i may be described as:

$$U_i = U_0 + aS_i - p_i, \tag{1}$$

where U_0 is the fixed utility of two kinds of product. Obviously, if $p_1 < p_2$ then $S_1 < S_2$, this shows that higher quality means higher price. Figure 2 illustrates the utility functions of two products, where a_0 is the intersection point of the utility function for two products, $a_0 \in (0, +\infty)$.

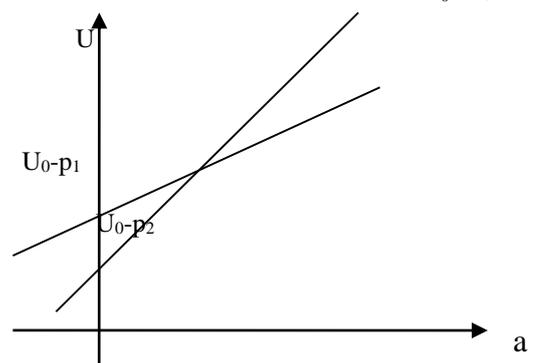


FIGURE 2 Utility function of the two products

Since different consumer has different preference coefficient in the market, consumer preference is considered to be a positive stochastic random variable with probability density function $h(a)$ and cumulative distribution function $H(a)$. From Figure 2, a_0 satisfies the following equality:

$$U_0 + a_0 S_1 - p_1 = U_0 + a_0 S_2 - p_2.$$

Hence:

$$a_0 = (p_1 - p_2) / (S_1 - S_2).$$

Generally, consumers will choose the product with greater utility in the market. Thus two products market share λ_i can be obtained respectively:

$$\lambda_1 = \int_0^{a_0} h(a) da = H\left(\frac{p_1 - p_2}{S_1 - S_2}\right) = H(a_0),$$

$$\lambda_2 = \int_{a_0}^{+\infty} h(a) da = 1 - H(a_0) = 1 - H\left(\frac{p_1 - p_2}{S_1 - S_2}\right).$$

It is not difficult to see that if the quality standard S_i of product i increases, then the market share λ_i for the product i will increase, and therefore $\frac{d\lambda_i}{dS_i} > 0$. Similarly, if the quality standard S_j for competitor increases, then the market share λ_i for the product i will decrease, and therefore $\frac{d\lambda_i}{dS_j} < 0$. Furthermore, if the price p_i of the product i increases, then the market share λ_i for the product i will decrease, and $\frac{d\lambda_i}{dp_i} < 0$; if the price p_j for competitor increases, then the market share λ_i for the product i will decrease, which follows that $\frac{d\lambda_i}{dp_j} > 0$.

The market share of the two products will be fixed if product preferences are fixed. It is assumed that the market

shares for the two products are λ_1 and λ_2 respectively. Note that $\lambda_i + \lambda_j = 1$. Hence, the demand function for product i and j can be expressed as

$$F_{ip}(x) = P\{\lambda_i D \leq x\} = F_p\left(\frac{x}{\lambda_i}\right), \tag{2}$$

$$f_{ip}(x) = \frac{1}{\lambda_i} f_p\left(\frac{x}{\lambda_i}\right). \tag{3}$$

It is easy to see that the demand is dependent on quality level, prices and consumer preferences of the two products. Through designing a reasonable price or improving the quality level, we can obtain a higher market share and the competitiveness for the supply chain in the market can be enhanced.

3 The sourcing strategy of the retailer

3.1 THE SOURCING STRATEGY OF THE RETAILER OF THE RETAILER i IN STAGE 2

Denote z_i as the inventory level of the supply chain before the emergency order is placed. Let $\pi_{R_i}^{II}(Q_{i3} | z_i)$ be random profit of the retailer i in stage 2, where the superscript "II" represents stage 2. We have:

$$\pi_{R_i}^{II}(Q_{i3} | z_i) = p_i E[\min(Q_{i3}, d_i)] + v_i E(Q_{i3} - d_i)^+ - b_i (d_i - Q_{i3})^+ - \omega_{i3} (Q_{i3} - z_i)^+$$

By a simple calculation, we can deduce the retailer i 's expected profit in stage 2, denoted as $\Pi_{R_i}^{II}(Q_{i3} | z_i)$, which is given by:

$$\Pi_{R_i}^{II}(Q_{i3} | z_i) = \begin{cases} (p_i + b_i - \omega_{i3})Q_{i3} - (p_i + b_i - v_i) \times \int_0^{Q_{i3}} F_p\left(\frac{x}{\lambda_i}\right) dx + \omega_{i3} z_i - b_i E(D), & Q_{i3} \geq z_i \\ (p_i + b_i)z_i - (p_i + b_i - v_i) \times \int_0^{z_i} F_p\left(\frac{x}{\lambda_i}\right) dx - b_i E(D), & Q_{i3} \leq z_i \end{cases}. \tag{4}$$

As a classical newsvendor problem, the retailer i 's sourcing problem in stage 2 is to choose the emergency order quantity to maximize its expected profit for any given initial inventory level z_i . We obtain that the order-up-to-level (OUL) policy is optimal for the retailer by using the first and second-order optimality conditions. The threshold value of the inventory level is:

$$\hat{Q}_{i3} = \lambda_i F^{-1}\left(\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i}\right)$$

Therefore, the optimal inventory level after the retailer placing an emergency order is as follows:

$$Q_{i3}^* = \max\{\hat{Q}_{i3}, z_i\}. \tag{5}$$

Then the maximum expected profit of retailer i in stage 2 for any given initial inventory level z_i is shown as follows:

$$\pi_{R_i}^{II*}(z_i) = \begin{cases} (p_i + b_i - \omega_{i3})\hat{Q}_{i3} - (p_i + b_i - v_i) \times \int_0^{\hat{Q}_{i3}} F_p\left(\frac{x}{\lambda_i}\right) dx + \omega_{i3}z_i - b_i E(D), & \hat{Q}_{i3} \geq z_i \\ (p_i + b_i)z_i - (p_i + b_i - v_i) \times \int_0^{z_i} F_p\left(\frac{x}{\lambda_i}\right) dx - b_i E(D), & \hat{Q}_{i3} \leq z_i \end{cases} \quad (6)$$

3.2 THE SOURCING STRATEGY OF THE RETAILER OF THE RETAILER *i* IN STAGE 1

The retailer *i*'s sourcing problem in stage 1 is to choose the order quantities Q_{i1} and Q_{i2} from supplier *i* and supplier *j* to maximize its expected profit for any given wholesale price.

Initially, when disruptions occur at the same time to both suppliers (with probability $(1-\alpha_i)(1-\alpha_j)$ for any given wholesale price the expected profit $G_0(Q_{i1}, Q_{i2})$ of the retailers is given by:

$$G_0(Q_{i1}, Q_{i2}) = (1-\alpha_i)(1-\alpha_j)[\pi_{R_i}^{II*}(0) - (1-\eta)\gamma_i c_i Q_{i1} - (1-\eta)\gamma_j c_j Q_{i2}]$$

When disruption occurs solely to one supplier, the expected profit $G_1(Q_{i1}, Q_{i2})$ of the retailers is given by:

$$G_1(Q_{i1}, Q_{i2}) = \alpha_i(1-\alpha_j)[\pi_{R_i}^{II*}(Q_{i1}) - \omega_{i1}Q_{i1} - (1-\eta)\gamma_i c_i Q_{i1}] + (1-\alpha_i)\alpha_j \times [\pi_{R_i}^{II*}(Q_{i2}) - \omega_{i2}Q_{i2} - (1-\eta)\gamma_j c_j Q_{i2}]$$

When the two supply chains do not face a disruption (with probability $\alpha_i\alpha_j$), the expected profit $G_2(Q_{i1}, Q_{i2})$ of the retailers is given by:

$$G_2(Q_{i1}, Q_{i2}) = \alpha_i\alpha_j[\pi_{R_i}^{II*}(Q_{i1} + Q_{i2}) - \omega_{i1}Q_{i1} - \omega_{i2}Q_{i2}]$$

Let $\pi_{R_i}^I(Q_{i1}, Q_{i2})$ be the retailer *i*'s expected profit in stage 1, where the superscript "I" represents stage 1. Then we have:

$$\pi_{R_i}^I(Q_{i1}, Q_{i2}) = G_0(Q_{i1}, Q_{i2}) + G_1(Q_{i1}, Q_{i2}) + G_2(Q_{i1}, Q_{i2}). \quad (7)$$

The optimization model, which represents the maximum of the total weighted expected profit (considering all possible combinations of disruption events on none, on one, or on both supply chains) is:

$$(P): (Q_{i1}, Q_{i2}) \in \arg \max \pi_{R_i}^I(Q_{i1}, Q_{i2}), \text{ s.t. } Q_{i1} \geq 0, Q_{i2} \geq 0.$$

For the optimization problem (P), we have the following conclusions about the optimal sourcing strategy of the retailer *i*.

Theorem 1. After supply disruption has occurred, equilibrium sourcing quantity of the retailer *i* from the spot market satisfies the OUL policy and the threshold value of the inventory level is:

$$\hat{Q}_{i3} = \lambda_i F^{-1}[(p_i + b_i - \omega_{i3}) / (p_i + b_i - v_i)].$$

The equilibrium sourcing strategies from suppliers 1 and 2 are as follows:

a) If $A < 0$ and $B < 0$, then two suppliers are placed with zero order quantity, where:

$$A = \alpha_i(\omega_{i3} - \omega_{i1}) - (1-\alpha_i)(1-\eta_i)\gamma_i c_i,$$

$$B = \alpha_i(\omega_{i3} - \omega_{i1}) - (1-\alpha_i)(1-\eta_i)\gamma_i c_i.$$

Thus, the retailer only sources from the spot market and the emergency order quantity is \hat{Q}_{i3} .

b) If $\alpha_i\alpha_j(\omega_{i3} - v_j) > B \geq 0 > A$, then supplier 1 is placed with zero order quantity. The retailer only sources from supplier 2 and the equilibrium sources quantity from supplier 2 is \hat{Q}_{i3} .

c) If $\alpha_i\alpha_j(\omega_{i3} - v_j) > A \geq 0 > B$, then supplier 2 is placed with zero order quantity. The retailer only sources from supplier 1 and the equilibrium quantity ordered from supplier 1 is \hat{Q}_{i3} .

d) When $Q_{i2} \leq \hat{Q}_{i3} \leq Q_{i1}$, if:

$$A \geq B \geq \alpha_j A \geq 0, \quad (8)$$

then two suppliers are selected to be placed with positive orders, which are given by:

$$Q_{i1}^* = \lambda_i F_p^{-1}\left[\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i} + \frac{A - B}{\alpha_i(1-\alpha_j)(p_i + b_i - v_i)}\right], \quad (9)$$

$$Q_{i2}^* = \lambda_j \{F_p^{-1}\left[\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i} + \frac{B}{\alpha_i\alpha_j(p_i + b_i - v_i)}\right] - F_p^{-1}\left[\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i} + \frac{A - B}{\alpha_i(1-\alpha_j)(p_i + b_i - v_i)}\right]\}, \quad (10)$$

e) When $Q_{i1} \leq \hat{Q}_{i3} \leq Q_{i2}$, if:

$$B \geq A \geq \alpha_i B \geq 0, \quad (11)$$

then two suppliers are selected to be placed with positive orders, which are given by:

$$Q_{i2}^* = \lambda_j F_p^{-1}\left[\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i} + \frac{B - A}{\alpha_j(1-\alpha_i)(p_i + b_i - v_i)}\right], \quad (12)$$

$$Q_{i1}^* = \lambda_i \{F_p^{-1}\left[\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i} + \frac{A}{\alpha_i\alpha_j(p_i + b_i - v_i)}\right] - F_p^{-1}\left[\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i} + \frac{B - A}{\alpha_j(1-\alpha_i)(p_i + b_i - v_i)}\right]\}. \quad (13)$$

Proof: We discuss the equilibrium sources strategies Q_{i1} and Q_{i2} of the retailers based on the following different cases.

Case1: $\hat{Q}_{i3} \geq Q_{i1} + Q_{i2}$.

In the profit function of retailer, G_0 , G_1 and G_2 are respectively given by:

$$G_0(Q_{i1}, Q_{i2}) = (1 - \alpha_i)(1 - \alpha_j)[(p_i + b_i - \omega_{i2})\hat{Q}_{i3} - (p_i + b_i - v_i) \int_0^{\hat{Q}_{i3}} F_p\left(\frac{x}{\lambda_i}\right) dx - b_i E(D) - \gamma_i(1 - \eta_i)c_i Q_{i1} - \gamma_j(1 - \eta_j)c_j Q_{i2}],$$

$$G_1(Q_{i1}, Q_{i2}) = \alpha_i(1 - \alpha_j)[(p_i + b_i - \omega_{i2})\hat{Q}_{i3} - (p_i + b_i - v_i) \times \int_0^{\hat{Q}_{i3}} F_p\left(\frac{x}{\lambda_i}\right) dx + \omega_{i3}Q_{i1} - b_i E(D) - \omega_{i1}Q_{i1} - \gamma_j(1 - \eta_j)c_j Q_{i2}] + (1 - \alpha_i)\alpha_j[(p_i + b_i - \omega_{i2}) \times \hat{Q}_{i3} - (p_i + b_i - v_i) \int_0^{\hat{Q}_{i3}} F_p\left(\frac{x}{\lambda_i}\right) dx + \omega_{i3}Q_{i2} - b_i E(D) - \omega_{i2}Q_{i2} - \gamma_i(1 - \eta_i)c_i Q_{i1}]$$

$$G_2(Q_{i1}, Q_{i2}) = \alpha_i\alpha_j[(p_i + b_i - \omega_{i2})\hat{Q}_{i3} - (p_i + b_i - v_i) \times \int_0^{\hat{Q}_{i3}} F_p\left(\frac{x}{\lambda_i}\right) dx - b_i E(D) - \omega_{i1}Q_{i1} - \omega_{i2}Q_{i2}]$$

It is easy to find that G_0 , G_1 and G_2 are linear functions of the order quantities Q_{i1} and Q_{i2} in this case. So retailer i 's expected profit $\pi_{R_i}^I(Q_{i1}, Q_{i2})$ is also a linear function of the order quantities Q_{i1} and Q_{i2} . From the equalities:

$$\frac{\partial \pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i1}} = A, \quad \frac{\partial \pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i2}} = B,$$

the following conclusions follow:

1) If $A \geq 0$ and $B \geq 0$ then the expected profit $\pi_{R_i}^I(Q_{i1}, Q_{i2})$ will increase as the order quantities Q_{i1} and Q_{i2} increase. Hence $Q_{i1}^* + Q_{i2}^* \geq \hat{Q}_{i3}$. On the other hand, we have $Q_{i1}^* + Q_{i2}^* \leq \hat{Q}_{i3}$ from the assumption. Therefore, we have $Q_{i1}^* + Q_{i2}^* = \hat{Q}_{i3}$.

2) If $A < 0$ and $B < 0$, then the expected profit $\pi_{R_i}^I(Q_{i1}, Q_{i2})$ will increase as the order quantities Q_{i1} and Q_{i2} decrease. Hence $Q_{i1}^* = Q_{i2}^* = 0$.

3) If $A < 0$ and $B \geq 0$, then the expected profit $\pi_{R_i}^I(Q_{i1}, Q_{i2})$ will increase as Q_{i1} decreases or as Q_{i2} increases. Hence we have $Q_{i2}^* \geq \hat{Q}_{i3}$ and $Q_{i1}^* = 0$.

4) If $A < 0$ and $B \geq 0$ then the expected profit $\pi_{R_i}^I(Q_{i1}, Q_{i2})$ will increase as Q_{i2} decreases or as Q_{i1} increases. Hence $Q_{i1}^* \geq \hat{Q}_{i3}$ and $Q_{i2}^* = 0$.

Case 2: $\max\{Q_{i1}, Q_{i2}\} \leq \hat{Q}_{i3} \leq Q_{i1} + Q_{i2}$.

Similar to Case 1, we can obtain expressions of G_0 , G_1 and G_2 in this case. The first-order partial derivatives of $\pi_{R_i}^I(Q_{i1}, Q_{i2})$ with respect to Q_{i1} and Q_{i2} are given by

$$\frac{\partial \pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i1}} = \alpha_i\alpha_j[(p_i + b_i - \omega_{i3}) - (p_i + b_i - v_i)F\left(\frac{Q_{i1} + Q_{i2}}{\lambda_i}\right)] + A,$$

$$\frac{\partial \pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i2}} = \alpha_i\alpha_j[(p_i + b_i - \omega_{i3}) - (p_i + b_i - v_i)F\left(\frac{Q_{i1} + Q_{i2}}{\lambda_i}\right)] + B.$$

But we cannot determine whether the Hessian matrix of $\pi_{R_i}^I(Q_{i1}, Q_{i2})$ is negative definite or not. So the equilibrium order quantities cannot be deduced by the first-order optimality condition.

From the assumption $\max\{Q_{i1}, Q_{i2}\} \leq \hat{Q}_{i3} \leq Q_{i1} + Q_{i2}$ and the analysis in Case 1, it is straightforward to deduce that $A \geq 0$. So we have:

$$\left. \frac{\partial \pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i1}} \right|_{Q_{i1} + Q_{i2} = \hat{Q}_{i3}} = A \geq 0,$$

$$\left. \frac{\partial \pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i1}} \right|_{Q_{i1} + Q_{i2} = +\infty} = \alpha_i\alpha_j(v_i - \omega_{i3}) + A.$$

Note that when $\alpha_i\alpha_j(\omega_{i3} - v_i) + A < 0$, $\frac{\pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i1}}$ has a unique zero point as follows:

$$(Q_{i1} + Q_{i2})^* = \lambda_i F^{-1}\left(\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i} + \frac{A}{\alpha_i\alpha_j(p_i + b_i - v_i)}\right).$$

Furthermore, we deduce that

$$\frac{\partial \pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i2}} = B - A.$$

Thus, we have the following conclusions:

1) If $B - A < 0$. i.e., $\frac{\pi_{R_i}^I(Q_{i1}, Q_{i2})}{\partial Q_{i1}} < 0$, then the expected profit will increase as Q_{i2} . Hence $Q_{i1}^* = (Q_{i1} + Q_{i2})^* \geq \hat{Q}_{i3}$ and $Q_{i2}^* = 0$. But from the assumption $\max\{Q_{i1}, Q_{i2}\} \leq \hat{Q}_{i3} \leq Q_{i1} + Q_{i2}$, we have $A = 0$, $Q_{i1}^* = \hat{Q}_{i3}$.

2) If $B - A \geq 0$, i.e. $\frac{\partial \pi_{R_i}^l(Q_{i1}, Q_{i2})}{\partial Q_{i2}} < 0$, then the expected profit will decrease as Q_{i2} . Hence $Q_{i1}^* + Q_{i2}^* = (Q_{i1} + Q_{i2})^*$ and $Q_{i2}^* \geq \hat{Q}_{i3}$.

Case 3: $Q_{i2} \leq \hat{Q}_{i3} \leq Q_{i1}$ and $Q_{i1} \leq \hat{Q}_{i3} \leq Q_{i2}$.

It is straightforward to verify that the Hessian matrix of $\pi_{R_i}^l(Q_{i1}, Q_{i2})$ is negative definite. Hence, $\pi_{R_i}^l(Q_{i1}, Q_{i2})$ is jointly concave to Q_{i1} and Q_{i2} . The equilibrium order quantity can be uniquely deduced by the first-order optimality condition.

$$\frac{\partial \pi_{R_i}^l(Q_{i1}, Q_{i2})}{\partial Q_{i1}} = \alpha_i(p_i + b_i - \omega_{i3}) - \alpha_i \alpha_j (p_i + b_i - v_i) \times F_p \left(\frac{Q_{i1} + Q_{i2}}{\lambda_i} \right) - \alpha_i (1 - \alpha_j) \times (p_i + b_i - v_i) F_p \left(\frac{Q_{i1}}{\lambda_i} \right) + A = 0, \tag{14}$$

$$\frac{\partial \pi_{R_i}^l(Q_{i1}, Q_{i2})}{\partial Q_{i2}} = \alpha_i \alpha_j (p_i + b_i - \omega_{i3}) - \alpha_i \alpha_j (p_i + b_i - v_i) \times F_p \left(\frac{Q_{i1} + Q_{i2}}{\lambda_i} \right) + B = 0. \tag{15}$$

From Equation (12) the unique optimal total order quantity is deduced as follows:

$$(Q_{i1} + Q_{i2})^* = \lambda_i F_p^{-1} \left[\frac{p_i + b_i - \omega_{i3}}{p_i + b_i - v_i} + \frac{B}{\alpha_i \alpha_j (p_i + b_i - v_i)} \right], \tag{16}$$

Substituting Equation (16) into Equation (14), it is easy to deduce Equation (9).

Moreover, since $Q_{i2}^* \geq 0 \Leftrightarrow Q_{i1}^* \leq (Q_{i1}^* + Q_{i2}^*)$, the equilibrium order quantity can be uniquely deduced via Equations (9) and (10) when Equation (8) holds. In the same way, we can obtain sourcing strategy of the retailer i when $Q_{i2} \leq \hat{Q}_{i3} \leq Q_{i1}$, i.e., conclusion (e) holds.

From above analysis for the three different cases, we reach the conclusions about the equilibrium sourcing strategy of the retailer.

From **Theorem 1**, we can easily obtain the conditions for both suppliers being placed with positive order quantities as follows.

Corollary 1. After supply disruption has occurred, both suppliers are placed with positive order quantities if and only if Equation (8) or Equation (9) holds.

From Theorem 1, it can be observed that:

1) Retailer i places with at most one supplier when $A < 0$ and $B < 0$. The reason is that the supplier's supply reliability is very low or its delivery cost is very high.

2) If $Q_{i1}^* + Q_{i2}^* > 0$, then $Q_{i1}^* + Q_{i2}^* > \hat{Q}_{i3}^*$. This means that the total order quantity is not less than the threshold value

\hat{Q}_{i3}^* , when any supplier is placed with a positive order quantity. This indicates that it prefers the supplier(s) to the spot market once the retailer selects one supplier or two suppliers.

(3) The sourcing strategy of the retailer is affected mainly by two key factors. They are

$$\alpha_i (\omega_{i3} - \omega_{i1}) - (1 - \alpha_i)(1 - \eta_i) \gamma_i c_i$$

and

$$\alpha_j (\omega_{j3} - \omega_{j2}) - (1 - \alpha_j)(1 - \eta_j) \gamma_j c_j.$$

The larger the value of a factor is, the more powerful the corresponding supplier is. So we regard these two factors as the competitiveness of the two suppliers. Furthermore, the other factors that affect supplier competitiveness consist of the unit delivery cost of the product for the supplier, the fixed wholesale price of the spot market, the total proportion of the marginal delivery cost, and the probability of delivering orders on time. The supplier can improve his competitiveness by improving his probability of delivering orders on time or decreasing his delivery cost. However, the marginal delivery cost usually increases when delivery is stable. Thus, a trade off exists between the marginal cost of delivery and the probability of on-time delivery.

In the following, we will indicate the impact of system parameters on order quantity in the case that both suppliers are placed with positive order quantities.

Corollary 2. If both suppliers are placed with positive order quantities, then the change trend of order quantity with system parameters is as follows:

- 1) $Q_{i1}^*(Q_{i2}^*)$ will increase as wholesale price $\omega_{i1}(\omega_{i2})$ decreases or as wholesale price $\omega_{i2}(\omega_{i1})$ increases.
- 2) $Q_{i1}^*(Q_{i2}^*)$ will increase as disruption probability $\alpha_i(\alpha_j)$ increase or as disruption probability $\alpha_j(\alpha_i)$ decreases.
- 3) $Q_{i1}^*(Q_{i2}^*)$ will increase as delivery cost $c_i(c_j)$ decreases or as delivery cost $c_j(c_i)$ increases.
- 4) $Q_{i1}^*(Q_{i2}^*)$ will increase as disruption probability $\gamma_i(\gamma_j)$ increase or as disruption probability $\gamma_j(\gamma_i)$ decreases.

4 Pricing strategy of the supplier

In this section, two suppliers set their individual wholesale prices simultaneously to maximize their respective expected profits before the retailer places its orders and the suppliers do not collude. This is a static non-cooperative game between two suppliers. When the order quantity cannot be met, setting of the wholesale prices will lose its meaning. So we only derive a sufficient condition for the existence of an equilibrium price strategy in the case that both suppliers are placed with positive order quantities in this section. The expected revenue function of supplier 1

in stage 0 in the case that both suppliers are placed with positive order quantities is given by:

$$\pi_{s_i}(\omega_{i1}, \omega_{j1}) = [\alpha_i(\omega_{i1} - c_i) - (1 - \alpha_i)\eta_i\gamma_i c_i]Q_{i1}^* + [\alpha_i(\omega_{j1} - c_i) - (1 - \alpha_i)\eta_i\gamma_i c_i]Q_{j1}^*,$$

$$\pi_{s_j}(\omega_{i2}, \omega_{j2}) = [\alpha_j(\omega_{i2} - c_j) - (1 - \alpha_j)\eta_j\gamma_j c_j]Q_{i2}^* + [\alpha_j(\omega_{j2} - c_j) - (1 - \alpha_j)\eta_j\gamma_j c_j]Q_{j2}^*.$$

Therefore, the optimization model that two-supplier will set their individual wholesale prices simultaneously to maximize their respective expected profits after anticipating the order quantity of retailer is:

$$(M): \begin{cases} (\omega_{i1}^*, \omega_{j1}^*) \in \arg \max \pi_{s_1}(\omega_{i1}, \omega_{j1}) \\ (\omega_{i2}^*, \omega_{j2}^*) \in \arg \max \pi_{s_2}(\omega_{i2}, \omega_{j2}) \end{cases} \cdot s.t \ \omega_{i1} \geq 0, \omega_{j1} \geq 0, \omega_{i2} \geq 0, \omega_{j2} \geq 0$$

From above equations, we can derive the profit function of $\omega_{i1}(\omega_{j1})$ and $\omega_{i2}(\omega_{j2})$, but we cannot provide a close-form solution of $\omega_{i1}(\omega_{j1})$ and $\omega_{i2}(\omega_{j2})$. A sufficient condition for the existence of an equilibrium price strategy is as follows.

Theorem 2. In the environment of supply disruptions, if the equilibrium order quantity is a decreasing concave function of the wholesale price of its supply chain in the case that both suppliers are placed with positive order

quantities, then the unique equilibrium prices satisfy the following conditions:

$$\alpha_i Q_{i1}^* + [\alpha_i(\omega_{i1}^* - c_i) - (1 - \alpha_i)\eta_i\gamma_i c_i] \frac{\partial Q_{i1}^*}{\partial \omega_{i1}^*} = 0,$$

$$\alpha_j Q_{j1}^* + [\alpha_j(\omega_{j1}^* - c_j) - (1 - \alpha_j)\eta_j\gamma_j c_j] \frac{\partial Q_{j1}^*}{\partial \omega_{j1}^*} = 0,$$

$$\alpha_j Q_{i2}^* + [\alpha_j(\omega_{i2}^* - c_j) - (1 - \alpha_j)\eta_j\gamma_j c_j] \frac{\partial Q_{i2}^*}{\partial \omega_{i2}^*} = 0,$$

$$\alpha_i Q_{j2}^* + [\alpha_i(\omega_{j2}^* - c_i) - (1 - \alpha_i)\eta_i\gamma_i c_i] \frac{\partial Q_{j2}^*}{\partial \omega_{j2}^*} = 0.$$

From **Theorem 2**, it can be observed that equilibrium price strategy exists and is unique if and only if the equilibrium order quantity is a decreasing concave function of the wholesale price of its supply chain. Many popular distributions in reality have this feature, such as uniform distribution and normal distribution.

5 Numerical examples

In this section we will use numerical example to verify the effectiveness of the conclusion. Suppose that the total demand follows the uniform distribution in the interval [300,400]. The basic parameter values are given as:

TABLE 1 System parameter values

Parameter	$\omega_{i3}(i=1,2)$	c_1	c_2	p_1	p_2	b_1	b_2	v_1	v_2	α_1	α_2	η_i	γ_i
values	16	10.5	12	18	16	5	4	3	4	0.3	0.9	0.2	0.4

Then distribution function of supplier i is given by:

$$F_{ip}(x) = F_p\left(\frac{5x}{2}\right) = \begin{cases} 0 & x < 100 \\ \frac{x-100}{160} & x \in [100,160), \\ 1 & x \geq 160 \end{cases}$$

we can obtain equilibrium order quantity and equilibrium pricing by **Theorems 1** and **2**. They are $Q_{i1}^* = 201.60$, $Q_{i2}^* = 127.36$, $\omega_{i1}^* = 8$, $\omega_{i2}^* = 10$. This has fully demonstrated the effectiveness of our theorems in actual situation. Next we verify the validity of Corollary 2 in Figure 3-7.

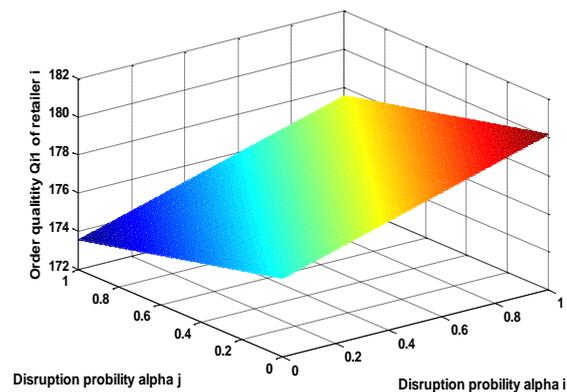


FIGURE 3 The change trend of retailer's order as α_i and α_j

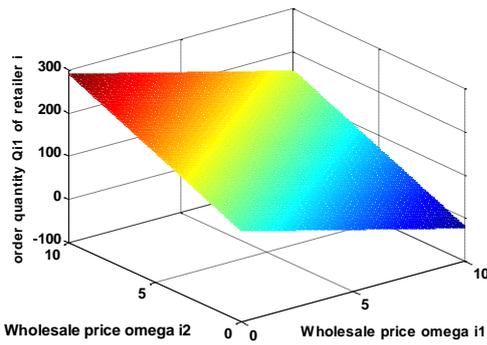


FIGURE 4 The change trend of retailer's order as ω_{i1} and ω_{i2}

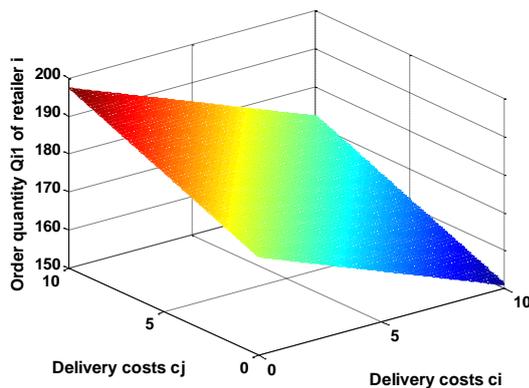


FIGURE 5 The change trend of retailer's order as c_i and c_j

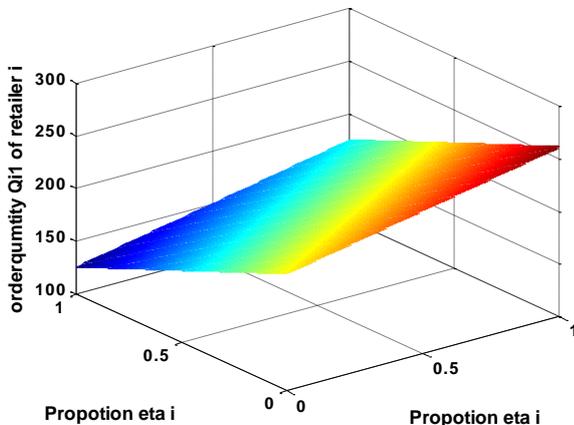


FIGURE 6 The change trend of retailer's order as η_i and η_j

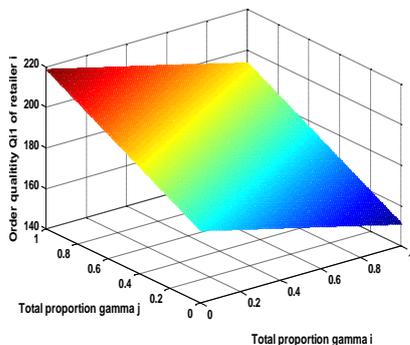


FIGURE 7 The change trend of retailer's order as γ_i and γ_j

From Figure 3-7, we can verify the validity of **Corollary 2**. In addition, the following conclusions can be obtained:

1) Supply disruption probability α_i on the order Q_{i1} of retailers i has more influence than competitive supply chain disruption probability α_j by comparing in Figure 3.

2) The wholesale price ω_{i1} on the order Q_{i1} of retailers i has more influence than competitive supply chain wholesale price ω_{i2} by comparing in Figure 4.

3) The delivery cost c_i on the order Q_{i1} of retailers i has the same influence with competitive supply chain delivery cost c_j by comparing in Figures 5.

4) On the whole, sourcing strategies Q_{i1} of the retailer i are affected mainly by disruptions probability, delivery costs and wholesale, but parameters $\eta_i(\eta_j)$ and $\gamma_i(\gamma_j)$ has less influence.

6 Conclusion

An effective sourcing strategies enhancing supply chain resilience is a necessary component of a firm's overall hedging strategy. This paper investigates sourcing strategies of the two retailers and the pricing strategies of the two suppliers in a 2-2 supply chain network under an environment of supply disruption. We obtain a sufficient condition for existence of an equilibrium sourcing and pricing strategies. The results show that equilibrium sourcing and pricing strategies are affected mainly by disruptions probability and delivery costs. Therefore, the appropriate parameters should be designed to obtain the optimal strategy in the actual operation of the market. These finding can guide suppliers to find a trade-off between the wholesale price and order quantity and a trade-off between the probability of on-time delivery and the marginal delivery cost.

We believe that several promising avenues exist for further research in this field. How to devise a mechanism to coordinate the whole channel is a potential topic for the research in the future. Furthermore, extensions can be made which includes multi-period problem or risk-averse participants.

Acknowledgements

This study is supported by the Science and Technology Foundation of the Ministry of Education of Heilongjiang province (12531138), the Fundamental Research Funds for the Central Universities (Grant No. HIT.HSS.201120), the China Postdoctoral Science Foundation under Grant No. 2013M541351 and the National Natural Science Foundation of China (11271103). The authors are also grateful for the valuable comments and suggestions of the editor and the reviewers, which have improved the presentation and the quality of this article.

References

- [1] Wu O Q, Chen H 2003 Chain-chain competition under demand uncertainty *Vancouver: The University of British Columbia* 1-10
- [2] Ha A Y, Tong S L 2008 Contracting and information sharing under supply chain competition *Management Science* 54(4) 701-15
- [3] Wu D S, Baron O, Berman O 2009 Bargaining in competing supply chains with uncertainty *European Journal of Operational Research* 197(2) 548-56
- [4] Boyaci T, Gallego G 2004 Supply chain coordination in a market with customer service competition *Production and Operation Management* 13(1) 3-22
- [5] Shou B Y, Huang J W, Li Z L 2009 Managing supply uncertainty under chain-to-chain competition *Working Paper* City University of Hong Kong
- [6] Dickson G 1966 An analysis of vendor selection systems and decisions *Journal of Purchasing* 2(1) 28-41
- [7] Verma R, Pullman M E 1998 An analysis of the supplier selection process *Omega* 26(6) 8739-50
- [8] Weber C A, Current J R, Benton W C 1991 Vender selection criteria and methods *European Journal of Operational Research* 50(1) 2-18
- [9] Talluri S, Narasimhan R 2003 Vendor evaluation with performance variability: A max-min approach *European Journal of Operational Research* 146(3) 543-52
- [10] Parlar M, Perry D 1996 Inventory models of future supply uncertainty with single and multiple suppliers *Naval Research Logistics* 43(2) 191-210
- [11] Gürlér U, Parlar M 1997 An inventory problem with two randomly available suppliers *Operation Research* 45(6) 904-918
- [12] Xiao T J, Xia Y S, Zhang G P 2007 Strategic outsourcing decisions for manufacturers that produce partially substitutable products in a quantity-setting duopoly situation *Decision Sciences* 38(1) 81-106
- [13] Tomlin B, Wang Y 2005 On the value of mix flexibility and fual sourcing in unreliable newsvendor networks *Manufacturing & Service Operations Management* 7(1) 37-57
- [14] Chopra S, Reinhardt G, Mohan U 2007 The importance of decoupling recurrent and disruption risks in a supply chain *Naval Research Logistics* 54(5) 544-55
- [15] Edergruen A, Yang N 2009 Optimal supply diversification under general supply risks *Operations Research* 57(6) 1451-68
- [16] Kleindorfer P R, Saad G H 2005 Managing disruption Risks in supply chains *Production and Operations Management* 14(1) 53-68
- [17] Yang S, Yang J, Abdel M L 2007 Sourcing with random yields and stochastic demand: a newsvendor approach *Computers & Operations Research* 34(12) 3682-90
- [18] Dada M, Petrucci N C, Schwarz L B 2007 A newsvendor's procurement problem when suppliers are unreliable *Manufacturing & Service Operations Management* 9(1) 9-32
- [19] Lariviere M A, Porteus E L 2001 Selling to a newsvendor: an analysis of price-only contracts *Manufacturing & Service Operations Management* 3(4) 293-305
- [20] Wang Y, Gerchak Y 2001 Capacity games in assembly systems with uncertain demand *Manufacturing & Service Operations Management* 5(3) 252-267
- [21] Bernstein F, DeCroix G A 2004 Decentralized pricing and capacity decisions in a multitier system with modular assembly *Management Science* 50(9) 1293-1308
- [22] Serel D A 2008 Inventory and Pricing Decisions in a single-period problem involving risky supply *International Journal of Production Economics* 116(1) 115-28
- [23] He Y, Zhao X 2012 Coordination in multi-echelon supply Chain under supply and demand uncertainty *International Journal of Production Economics* 139(1) 106-15
- [24] Cho S H, Tang C S 2013 Advance selling in a supply chain under uncertain supply and demand *Manufacture & Service Operations Management* 15(2) 305-19
- [25] Surti C, Hassini E, Abad P 2013 Pricing and inventory decisions with uncertain supply and stochastic demand *Asia-Pacific Journal of Operational Research* 30(6) 1350030-1-1350030-25

Authors



Ling Hou, born in June, 1978, Suihua, Heilongjiang province, China

Current position, grades: Lecturer in Department of Applied Mathematics, Harbin University of Science and Technology, China.
University studies: PhD from School of Management Harbin University of Science and Technology, China.
Scientific interest: Supply chain management, inventory, marketing science, and risk management.
Publications: 6



Meng Li, born in July, 1978, Harbin, Heilongjiang province, China

Current position, grades: Lecturer in Department of Applied Mathematics, Harbin University of Science and Technology, China.
University studies: PhD from University of Tsukuba, Japan.
Scientific interest: Supply chain management, mathematical finance and risk management.
Publications: 10 included in SCI or EI.



Dongyan Chen, born in February, 1964, Ganzhou, Jiangsu province, China

Current position, grades: Professor and doctoral tutor in Applied Science College, Harbin University of Science and Technology, China.
University studies: PhD in Harbin institute of technology, China.
Scientific interest: Supply chain management, optimization methods, robust control and time-delay system control.
Publications: more than 100

Lane changing intent identification based on logistic regression model

Jinshuan Peng*, Lei Xu

Chongqing Key Lab of Traffic System & Safety in Mountain Cities, Chongqing Jiaotong University, Chongqing 400074, China

Received 1 may 2014, www.tsi.lv

Abstract

To reduce the risk of the lane changing behaviors, based on integrated collection platform, the research group conducts experiments under real road environment for the purpose of studying drivers' lane changing intent identification. On the basis of the drivers' fixation characteristics of the rearview mirrors before changing lanes, the length of lane changing intent time window is determined. Based upon differential analysis of visual characteristics between lane keeping and lane changing intent stages, saccade numbers, visual search extent, saccade amplitude, standard deviation of head rotation angles in the horizontal direction are selected as the characteristic indexes of the identification. The logistic model is built according to feature extraction of the leaning samples, then applied to the identification process after the validity test. Results show that the identification success rate may reach 90.42%, thus verifying the feasibility and effectiveness of the logistic model to identify drivers' lane changing intent.

Keywords: lane change, intent identification, logistic model, index system

1 Introduction

Lane change is a common driving behavior, due to the restricts of vehicle condition, road environment, drivers' decision-making level and so on, lane changing behaviors have complexity and uncertainty attributes. Once coupling disorders occur, it may cause a traffic accident, which will bring about huge economic losses and casualties. In order to reduce the possibilities of lane change accidents, various of assistance systems become available, their general working principles are: Monitor the vehicle observed in rear view with radars or cameras, once there exists conflict vehicles within the given distance threshold, warning signals will be sent to the drivers [1]. The default rule of the lane changing assistance system is regarding the turn signals as the main basis to identify drivers' lane changing intent. According to the previous statistical results of the researching group, by the initial time of the lane changes, the opening rate of the turn signals is about 48.4% [2]. During practical applications, due to the irregular operational behaviors of the drivers, high false alarm rate always trouble the lane changing assistance system, requiring more reliable methods to identify drivers' lane changing intent.

Many scholars spent their efforts on lane changing behavior researching, which may provide new ideas to lane changing intent identification. Salvucci D D designed simulated test, results showed that by the initial time of the lane changes, the opening rate of the turn signals was about 50%, and during the lane changing process, drivers' fixation points always shifted from current lane to the target lane [3, 4]. Tijerina L conducted experiments under

real road environment, results indicated that before lane change occurred, drivers would pay more attention than lane keeping behaviors [5]. Liu A proposed that visual characteristics were the essential tool to identify drivers' operative intention, and eye movements could provide key information for design and development of the intelligent vehicle [6]. Doshi A affirmed that besides eye movements information, head movements characteristics also could be used to effectively identify drivers' operative intention [7]. Lethaus F insisted that drivers would pay more attention to rearview mirrors than to inside mirrors when executing left lane changes [8]. Henning M J asserted that both turn signals and visual characteristics could identify drivers' operative intention, and the latter had the identification time series advantage [9]. Olsen E C B divided interest regions of drivers' lane changing process, based on analysis of fixation parameters, drew the conclusion that when executing left lane changes, drivers would double their fixation duration at rearview mirrors in lane changing intent stage compared with lane keeping stage [10]. Based upon above researching results, this paper tries to propose a new method to identify drivers' lane changing intent, so as to send warning signals before potential dangers take place, guarantee the safety during lane changing process.

2 Experiments

The research group built an integrated collection platform around driving behavior characteristics (Figure 1), the platform consisted of several data collecting equipment and sensors, such as FaceLAB 5, millimeter wave radar, VBOX inertia sensor, lane identification system and so on.

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

The platform could synchronously collect multi-parameter from different sensors, which contained drivers' visual characteristics, vehicles' motion states, driving conditions (distance, relative speed between target vehicle and own vehicle), etc.



FIGURE 1 Driving behaviour collection platform

The researching group conducted experiment under real road environment. 16 professional drivers (2 women and 14 men, between the ages of 28 and 50, Mean = 41.1, standard deviation = 5.85) were recruited to take part in the naturalistic driving test. All the subjects passed the strict physical examination, without visual or hearing disorders, uncorrected visual acuity all over 1.3, and had 5 years or more driving experience.

The research group selected G25 Changxing-Huzhou (Zhejiang, China) expressway as the experimental routes, a total length of 25 km. The routes were bidirectional 4 lanes, separated by green belts, and the speed limit was 110 km/h.

Disposal plans for different emergencies were established before experiment could be run. After a general introduction to the experiment, the drivers had 10 minutes to be familiar with the testing vehicle. Meanwhile, experimenters recorded drivers' name, gender, age, driving experience and so on. Subjects were told to perform the driving missions completely according to his or her driving habits and expectation. Furthermore, experimenter should reduce the interference with the subjects, so as to get the parameters which could characterize subjects' real driving behaviors.

3 Lane changing intent time window

Precious research results have revealed that drivers always show specific visual search law in a time window before changing lanes. Lee S E proposed that drivers may show typical visual characteristics within the 3-second before changing lanes [11]. Guo Y S observed the experiment video with fixation point, drew the conclusion that lane changing intent time window was about 6-second [12]. In order to reduce negative effects of subjective assumption

of the time window, this paper proposes a new method to determine the lane changing intent time window according to drivers' fixation characteristics of rear-view mirrors.

Suppose that drivers' fixation times of rearview mirror regions (inside rearview mirror, side mirror) before a lane change is a natural number N ($N=1,2,3,\dots$), the interval between drivers' first fixation at rearview mirror and the initial time of the lane change is regarded as the time window length. The initial time of the lane change is determined by changing trends of steering wheel angle during lane changing process. According to the method mentioned above, distribution laws of the lane changing intent time window for the 16 subjects are shown in Figure 2. Time window length difference of the drivers is measured by single factor variance analysis, $F=0.923 < F_{0.05}(15, 213)=1.67$, which indicates that there is no significant difference in the time window length of the subjects. This could provide objective basis for determining an agreed time window length. The median reflected overall level of the data, we can see that median time window lengths of the subjects are distributed between 1.5-second and 4-second. The dotted line in Figure 2 represents 5-second time window length, and most of the subjects' third quartiles are below it, so ultimately we determine 5-second as the lane changing time window length.

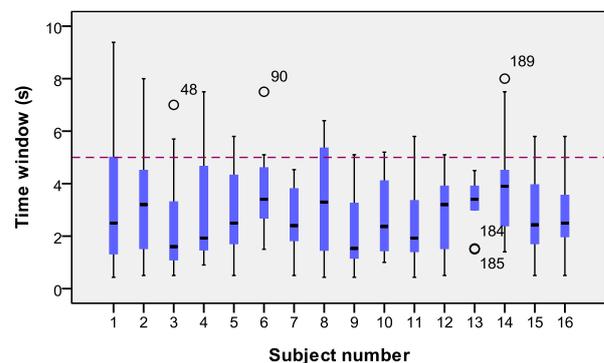


FIGURE 2 Time window length distribution of different subjects

In order to achieve the goal of lane changing intent identification, firstly, the characterization index should be determined. This paper tries to establish index system based on the differential analysis between lane changing intent and lane keeping stage. Given the 5-second time window, lane changing intent and lane keeping samples are selected by offline mode. More specifically, by moving forward 5-second from initial time of the lane changes, intent samples could be achieved. Correspondingly, car following behavior, as well as free driving, are cut down to 5 seconds to be lane keeping samples. According to the method mentioned above, ultimately the research group selects 401 lane keeping samples, 406 lane changing intent samples. Among them, the amount of learning samples of lane changing intent and lane keeping are both 200, others are the samples to be identified.

4 Characterization index

4.1 NUMBER OF SACCADDES

Generally speaking, when driver shifts his or her attention from one target to another, saccade behaviors may be needed to accomplish the shift process. To some extent, saccade behaviors could reflect complexity of the driving mission, there is a fine linear correlation between them.

Given the 5-second time window length, Figure 3 shows distribution difference of saccade numbers between lane changing intent and lane keeping stages. It seems that there exists significantly more saccade process in lane changing intent than in lane keeping stages, the difference of the two groups is analyzed using independent sample t-test ($p < 0.05$), and the diversity is remarkable. The reason may be that compared with lane keeping behaviors, drivers should pay more attention to the surrounding target objects in lane changing intent stages, which leads to the increase of fixation shift process, as well as the saccade numbers.

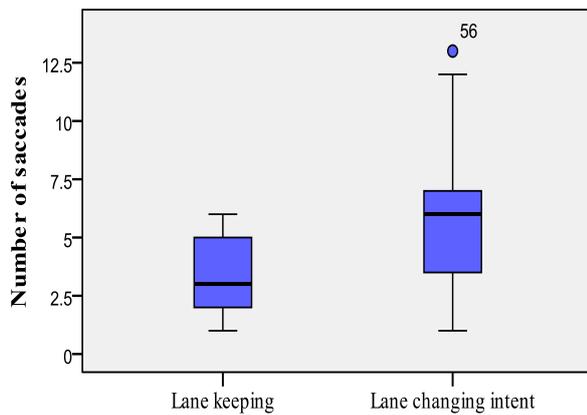


FIGURE 3 Distribution of saccade numbers

4.2 VISUAL SEARCH EXTENT

Underwood G raised that rotation degree standard deviation of the eyes in the horizontal direction could be used to evaluate drivers' visual search extent [13]. The bigger the value is, which may indicate that the more information drivers obtain from the surrounding environment. Figure 4 shows the schematic of rotation degree of the eyes. Where E is the position of the eyeballs, EF is the visual line, α is the rotation degree in the horizontal direction, and β is the rotation degree in the vertical direction.

Figure 5 depicts the distribution differences of visual search extent between lane keeping and lane changing intent stages. It seems that each quartile in lane changing intent stage is remarkably bigger than that of lane keeping stage. Independent sample t-test results ($p < 0.05$) shows that the difference between the two groups is remarkable. Further statistics indicates that the average visual search

extent of the lane keeping stage is about 4.6° , which is significantly smaller than its lane changing intent counterpart's 12.6° .

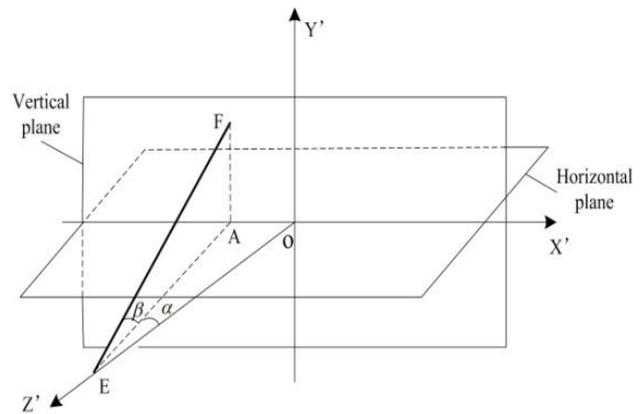


FIGURE 4 Rotation degree of the eyes

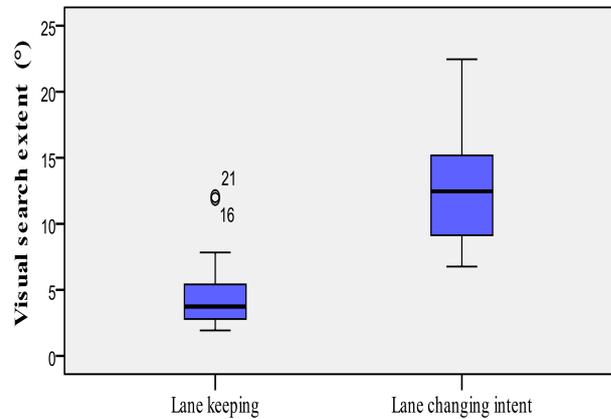


FIGURE 5 Visual search extent under different driving stages

4.3 SACCADDE AMPLITUDE

The angle between adjacent two fixation points is defined as saccade amplitude (Figure 6). The bigger the value is, the farther between the two fixation points. Suppose that there exists several saccade processes in the given 5-second time window length, mean of the values is defined as average saccade amplitude of the learning sample.

As shown in Figure 7, each quartile of saccade amplitude in lane changing intent stage is remarkably bigger than lane keeping stage. Further data processing results indicates that the average of all the lane keeping learning samples is about 12.5° , which is obviously smaller than lane changing intent stages (about 25.6°). One reason may be that there are more fixation shift routes in lane changing intent stages, for instance, the shift route from forward view to side mirrors, which is accompanied by large saccade amplitude, thereby confirming the safety before lane changing operations.

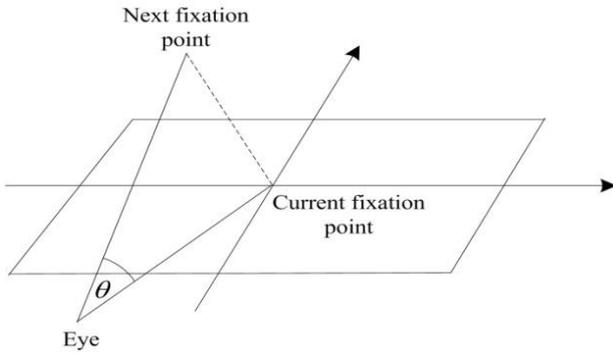


FIGURE 6 Schematic of saccade amplitude

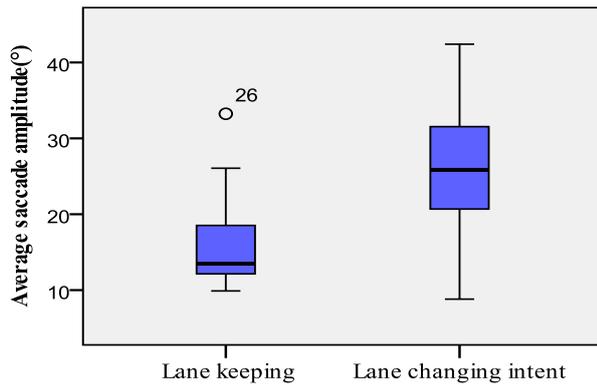


FIGURE 7 Distribution of average saccade amplitude

4.4 HEAD ROTATION DEGREE

Besides eye movements information, faceLAB 5 can also track drivers' head movement status. Head rotation degree in the horizontal direction could reflect drivers' operational intention to some extent [9]. This paper uses Std (standard deviation) of HRD (head rotation degree) in the horizontal direction to depict the discrete degree of the head movements.

Figure 8 depicts Std of HRD distribution difference in the horizontal direction, showing that each quartile in lane changing intent stage is remarkably bigger than that of lane keeping stage ($p < 0.05$). The average of HRD Std in lane keeping stage is 1.5° , which is significantly smaller than counterpart's 8.8° in lane changing intent stage. The reason for the difference may be that drivers need head movements to compensate the eye movements in the lane changing intent stage, so as to accomplish the fixation shift process, which is accompanied by large saccade amplitude.

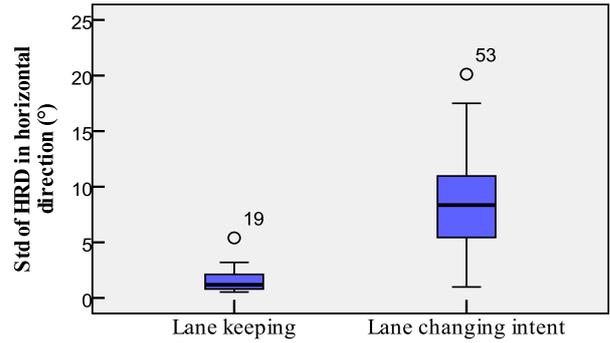


FIGURE 8 Standard deviation of head rotation degree

5 Lane changing intent identification

5.1 BINARY LOGISTIC MODEL

The logistic model is quite different from traditional regression analysis, it could be better suited to solve the regression case with discrete dependent variable than the latter one. The logistic model needs fewer restrictions of the independent variable's distribution characteristics. By means of nonlinear transformation, linear combination of the independent variable could be transformed to the probability value of the dependent variable [14]. For lane changing intent identification (define it as event Y), the identification results may be a dualistic problem, that is, "lane changing intent" ($Y=1$) or "lane keeping" ($Y=0$), so binary logistic model could be used to solve lane changing intent identification problem. The logistic model could be defined as follows:

$$\log it(p) = \ln \frac{p}{1-p} = b_0 + b_1x_1 + \dots + b_nx_n, \tag{1}$$

$$p = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}, \tag{2}$$

$$Z = b_0 + b_1x_1 + \dots + b_nx_n, \tag{3}$$

where, p is the probability of drivers having lane changing intent, $p \in [0,1]$, x is the independent variable related to the event, b is the regression coefficient of the independent variables, and e is the natural constant. When $p \geq 0.5$, we consider that drivers intend to execute the lane changing behavior ($Y=1$). Otherwise, the judgement results would be lane keeping behavior ($Y=0$).

5.2 INTENT IDENTIFICATION

Based on the analysis of characteristic parameters in the intent time window, characteristics indexes for lane changing intent identification are determined as number of saccades (x_1), visual search extent (x_2), saccade amplitude (x_3), and Std of head rotation degree in the horizontal direction (x_4). The basic idea of the intent identification is

to determine the regression coefficient of the logistic model by extracting the learning sample's characteristics, then applying to the multi-parameter fusion identification of the samples to be recognized. For each learning sample, depending on the statistic analysis, characteristic values of the identification indexes in 5-second time window may easily be determined. According to the attributes of the learning samples (200 lane keeping samples, 200 lane changing intent samples), binary logistic regression model is built by the SPSS statistic analysis software, the expression is as follows:

$$p = \frac{1}{1 + e^{-(-17.88 + 1.547x_1 - 0.223x_2 + 0.239x_3 + 2.228x_4)}} \quad (4)$$

where, p is the probability of drivers having lane changing intent, x_1 is the number of saccades, x_2 is visual search extent, x_3 is saccade amplitude, x_4 is std of head rotation degree in the horizontal direction.

After establishing the intent identification model, we should verify the efficiency of the model before its real application. Goodness-of-fit test results based on Cox & Snell RCS² and Nagelkerke RN² are shown in Table 1. The

nearer RCS² and RN² approximates to 1, the better model's fitting effect to be. Results show that both RCS² and RN² are greater than 0.75, which verifies the excellent classification efficiency of the logistic model.

TABLE 1 Validation of the model

Step	Cox & Snell RCS ²	Nagelkerke RN ²
1	0.783	0.923

After the efficiency verification of the logistic model, then we can carry out the lane changing intent identification process depending on the built model. For the lane keeping and lane changing intent samples to be identified, supposing that the properties of the samples are unknown, we scramble them to the "gray box". For any one of those samples in the "gray box", extract the characteristic values of the indicators in the identification time window, then putting into the logistic model, finally the identification properties of the samples could be determined. By comparing samples' identification results and their real properties, we can verify the identification performance of the logistic model, which is shown in Table 2.

TABLE 2 Classification efficiency of the logistic model

Samples to be identified	Identification results		Success rate
	Lane changing intent	Lane keeping	
Lane changing intent	True Positive (TP)	False Negative (FN)	91.75%
206	189	17	
Lane keeping	False Positive (FP)	True Negative (TN)	89.05%
201	22	179	
Sum	Positive samples	Negative samples	90.42%
407	211	196	

When lane changing intent samples are precisely identified to be "lane changing intent", we define it as "true positive". If lane keeping samples are mistakenly identified to be "lane changing intent", the identification results may be defined as "false positive". Similarly, "true negative" and "false negative" could be defined in turn. Generally, we use "sensitivity", "specificity" and "accuracy" (overall identification success rate) to evaluate identification performance of the model, they are calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

$$Specificity = \frac{TN}{FN + TN} \times 100\% \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (7)$$

According to the identification results in Table 2, sensitivity of the logistic model is about 91.75%, specificity is 89.05%, and overall identification success rate is about 90.42%, which indicates that the logistic model could effectively identify drivers' lane changing intent.

The existing lane changing assistance system regards turn signal as the main basis to identify drivers' lane-changing intention. The research group obtains drivers' use of turn signals based on the experiment under real road environment, which is shown in Figure 9. Origin of the horizontal axis represents the initial time of the lane changing behavior, the negative values corresponds to the time of lane changing intent stages, and the positive values corresponds to the time after lane changing operation. Figure 9 shows that by the initial time of lane changes, turn signals usage is about 48.6%, subsequently increase to 76% by 5-second after the lane changing operation. If we regard turn signals as the main basis to identify drivers' lane changing intent, the identification success rate even below 50%, which is far lower than its counterpart of the logistic model built in this paper.

Comparing to other lane changing intent identification methods, the logistic model built in this paper could effectively avoid the uncertainty caused by overly dependent on drivers' maneuvering characteristics [15-16], and the identification success rate surpasses 90%. Once the research results are applied in the development process of lane change auxiliary system, both working performance and reliability of the system could be remarkably improved, thereby guaranteeing the safety of the lane changing process.

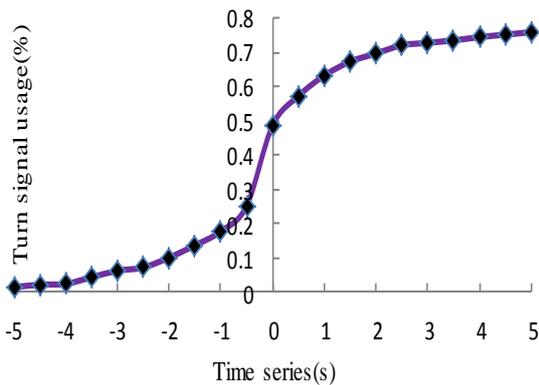


FIGURE 9 Turn signal usage rates in time axis

6 Conclusions

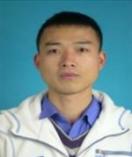
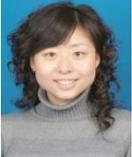
In this research, we propose a logistic model to identify lane changing intent by monitoring drivers' eye and head movements. Lane changing time window is determined by extracting drivers' fixation characteristics of the rearview mirrors before lane changing operation. Based on the differential analysis of the visual characteristics and head movements between lane keeping and lane changing intent stages, characteristic index system is further constructed. By building the logistic model, drivers' lane changing intent is precisely identified, and the identification success rate may reach 90.42%. The research results may provide important theoretical foundation for the improvement of intelligent vehicles' active safety technology, as well as the optimization of lane changing assistance system.

Acknowledgments

This work is supported by National Natural Science Foundation of China (51178053) and Natural Science Foundation of Chongqing (cstc2013jcyjA30015).

References

[1] Tijerina L 1999 *Transportation Human Factors* 1(2) 159-76
 [2] Peng J S, Guo Y S, Shao Y M 2013 *Applied Mechanics and Materials* 361-363 1875-79
 [3] Salvucci D D, Liu A 2002 *Traffic Psychology and Behavior* 5(8) 123-32
 [4] Salvucci D D, Mandalia H M, Kuge N, Yamamura 2007 *Journal of the Human Factors and Ergonomics Society* 49(3) 532-42
 [5] Tijerina L, Garrott W R 2005 *Journal of the Transportation Research Board* 1937 37-43
 [6] Liu A 1998 *Eye Guidance in Reading and Scene Perception* Press: Elsevier chapter 20
 [7] Doshi A, Trivedi M M 2009 *Intelligent Transportation Systems* 10(3) 453-62
 [8] Lethaus F, Rataj J 2007 *Intelligent Transport System* 1(3) 199-204
 [9] Henning M J, Georgeon O, Krems J F 2007 The quality of behavioral and environmental indicators used to infer the intention to change lanes *Proceeding of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* July 9-12 2007 1-7
 [10] Olsen E C B, Lee S E, Wierwille W W 2005 *Journal of the Transportation Research Board* 1937 44-50
 [11] Lee S E, Olsen E C B, Wierwille W W 2004 *A Comprehensive Examination of Naturalistic Lane-Changes* chapter 3
 [12] Guo Y S 2009 *The Effect of Traffic Environment and Driving Experience on Drivers' Eye Movement and Workload* University Press: Chang'an chapter 2 (in Chinese)
 [13] Underwood G, Chapman P, Bowden K, Crundall D 2002 *Transportation Research Part F: Traffic Psychology and Behaviour* 5(2) 87-97
 [14] Pearce J, Ferrier S 2000 *Ecological Modelling* 133(3) 225-45
 [15] Zhang L L 2011 *Research on Motorists' Intention Recognition for Traffic Safety Precaution* University Press: Wuhan University of technology chapter 5 (in Chinese)
 [16] Hou H J 2013 *Research on Lane Changing Intention Recognition Method for Freeway Drive* University Press: Jilin chapter 4 (in Chinese)

Authors	
	<p>Jinshuan Peng, born on June 27, 1982, Anhui, China</p> <p>Current position, grades: assistant professor in transportation at Chongqing Jiaotong University. University studies: Ph.D. (2012) in vehicle operation engineering at Chang'an University. Scientific interest: driving behavior characteristics, traffic safety engineering. Publications: 27.</p>
	<p>Lei Xu, born on April 7, 1982, Shanxi, China</p> <p>Current position, grades: assistant professor in transportation at Chongqing Jiaotong University. University studies: Ph.D. (2011) in Logistics Engineering at Chang'an University. Scientific interest: supply chain emergency management, dangerous goods transport. Publications: 18.</p>

The application of time domain and frequency domain statistical factors on rolling bearing performance degradation assessment

Yujing Wang^{1, 2}, Yicheng Jiang^{1*}, Shouqiang Kang²

¹*School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001*

²*School of Electrical and Electronic Engineering, Harbin University of Science and Technology, Harbin 150080*

Received 2 May 2014, www.tsi.lv

Abstract

Rolling bearing performance degradation assessment is a predict and prevent technology. In order to assess the performance degradation degree of the rolling bearing, and make the time domain and frequency domain statistical factors be applied more effectively in rolling bearing performance degradation assessment, a comprehensive analysis method is proposed based on time domain and frequency domain statistical factors. Time domain and frequency domain statistical factors are calculated and analysed for the life cycle data of the rolling bearing. Outer raceway moderate fault and severe fault of the rolling bearing can be distinguished well by peak-to-peak level, the root-mean-square (RMS) value, and kurtosis value of time domain factors; normal state and mild fault can be distinguished better by frequency centroid, F_3 , F_4 and F_5 of frequency domain factors than each time domain factor. The outer raceway performance degradation condition of the rolling bearing can be monitored well by using the proposed comprehensive analysis method, which uses partly frequency domain factors to analyse mild fault and partly time domain factors to analyse moderate fault and severe fault.

Keywords: rolling bearing, life cycle, statistical factor, performance degradation assessment

1 Introduction

Rolling bearing is the important rotating base element of machinery device and applied in many fields of national production, but while it is main fault source of machinery device [1-3]. Once the rolling bearing occurs faults, the machinery device may be damaged, the stability and safety of the integral production system will be influenced and enormous economic loss will be caused, even if casualty [4, 5]. So, it has become a research focus of fault diagnosis and condition assessment field to monitor and maintain the working condition of rolling bearing [6].

At present, proposed time domain and frequency domain factors usually are used for diagnosing whether the rolling bearing causes faults. For example reference [7] adopted peak-to-peak level, the root-mean-square (RMS) value, crest factor and kurtosis value to detect the fault information of the rolling bearing. The location of the fault can be detected by using spectrum analysis method and a well effect can be obtained; aiming to the rolling element damage of rolling bearing, a damage severity assessment method was proposed based on RMS by reference [8]. And compared with kurtosis value, crest factor and frequency domain amplitude of FFT transform, the validity and accuracy of proposed method is proved. In addition, based on time domain, frequency domain factors and time-frequency domain characteristic factors, some intelligent diagnosis methods for rolling bearing fault were proposed by some references [9-12].

However, there are relatively few cases to use time domain and frequency domain factors into rolling bearing performance degradation assessment. Generally speaking, when a rolling bearing is running, it passes through different stages of degradation until it is no longer functional. Because of different degradation degrees, the rolling bearing usually passes a series of different performance degradation conditions [13]. According to the integral change trend of bearing degradation degrees, the bearing degradation can be roughly divided into four conditions, that is normal condition, minor fault, moderate fault and severe fault [14]. The performance degradation of rolling bearing is a predict and prevent technology, that supported by device life cycle data and identifying the performance degradation degrees during the device performance degradation processing. By through detecting the running conditions of the device, the maintenance planning of the production device can be carried out targeted, and achieve the predict and prevent maintenance, avoid sudden fault and reach high efficiency and safety production.

The time domain and frequency domain factors are used for assessing rolling bearing performance degradation degrees in this paper. By analysing the life cycle data, the advantages and disadvantages of each statistical factor can be obtained. A comprehensive analysis method is proposed based on partly time domain and frequency domain statistical factors. Comparing the assessment curves, the validity of the proposed method is

* *Corresponding author* e-mail: jiangyc@hit.edu.cn

proved for performance degradation assessment of rolling bearing outer raceway.

2 Time domain and frequency domain statistical factors of the rolling bearing

The vibration signal of rolling bearing contains its running condition information which reflected by the time domain and frequency domain statistical factors. The calculation method of each factor is shown as follows:

1) time domain statistical factors [9]

a) peak-to-peak Level, which reflects the impact strength produced by bearing local fault point, the calculation formula is:

$$X_{p-p} = X_{\max} - X_{\min} \tag{1}$$

b) RMS, which reflects the total energy of the signal, the calculation formula is:

$$X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \tag{2}$$

c) Shape Factor, the calculation formula is:

$$S_f = \frac{X_{rms}}{|X|} \tag{3}$$

d) Impulse Factor, the calculation formula is:

$$I_f = \frac{X_{\max}}{|X|} \tag{4}$$

e) Kurtosis Value, which reflects the statistic of vibration signal distribution characteristic, the calculation formula is:

$$K_v = \frac{\beta}{X_{rms}^4} \tag{5}$$

f) Crest Factor, which describes the sharp peak degree, the calculation formula is:

$$C_f = \frac{X_{\max}}{X_{rms}} \tag{6}$$

g) Clearance Factor, the calculation formula is:

$$CL_f = \frac{X_{rms}}{X_r} \tag{7}$$

For Equations (1)-(7), $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$, $X_{\max} = \max\{|x_i|\}$,

$$X_{\min} = \min\{|x_i|\} \quad , \quad i=1,2,\dots,N, \quad X_r = \left[\frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right] \quad ,$$

$$\beta = \frac{1}{N} \sum_{i=1}^N x_i^4 \tag{7}$$

Above factors, peak-to-peak level and RMS are dimension factors and other five factors are dimensionless factors.

2) Frequency domain statistical factors

The frequency domain statistical factors of reference [9] are selected, that is F_1 - F_5 (No. 1-5 in Table 1). And the frequency domain statistical factors of reference [15] are selected, that is the frequency centroid, the mean square frequency, RMS frequency and the frequency variance (No. 6-9 in Table 1).

TABLE 1 Frequency domain statistical factors

No.	calculation formulas	No.	calculation formulas	No.	calculation formulas
1	$F_1 = \frac{\sum_{k=1}^K s(k)}{K}$	4	$F_4 = \frac{\sum_{k=1}^K (s(k) - F_1)^4}{(K(F_2)^2)}$	7	$MSF = \frac{\sum_{i=1}^N \bar{x}(i)^2}{4\pi^2 \sum_{i=1}^N x(i)^2}$
2	$F_2 = \frac{\sum_{k=1}^K (s(k) - F_1)^2}{K - 1}$	5	$F_5 = \frac{\sum_{k=1}^K f_k^2 s(k)}{\sqrt{\sum_{k=1}^K s(k) \sum_{k=1}^K f_k^4 s(k)}}$	8	$RMSF = \sqrt{MSF}$
3	$F_3 = \frac{\sum_{k=1}^K (s(k) - F_1)^3}{(K(\sqrt{F_2})^3)}$	6	$FC = \frac{\sum_{i=1}^N \bar{x}(i)x(i)}{2\pi \sum_{i=1}^N x(i)^2}$	9	$VF = MSF - (FC)^2$

where, in the calculation formulas of No. 1-5, $s(k)$ is the spectrum value when $k=1,2,\dots,K$, K is the total number of spectrum lines; in the calculation formulas of No. 6-9, $\bar{x}(i) = [x(i) - x(i-1)] \times f_s$, f_s is the sampling frequency.

3 The experiment and the analysis

3.1 THE LIFE CYCLE DATA OF ROLLING BEARING

The life cycle vibration signals related to the rolling bearing and the paper investigation were provided by university of Cincinnati IMS laboratory, the experimental device is shown in Figure 1.

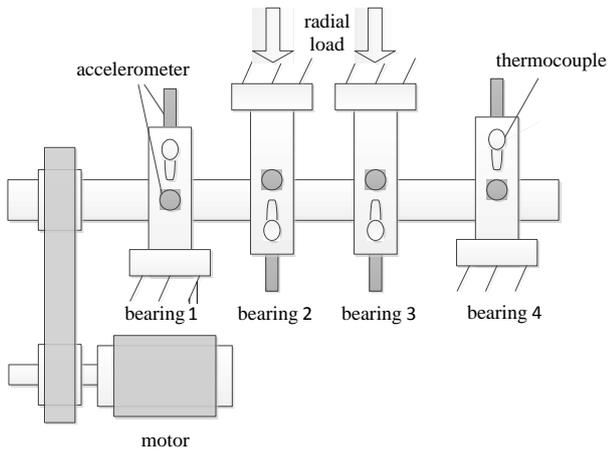


FIGURE 1 Schematic diagram of the experimental device.

Four bearings were mounted on the same shaft, bearing rotating speed remained 2000 rpm, PCB 353B33 high sensitivity accelerometer sensor is mounted on each bearing. The vibration data of the experiment are collected once every ten minutes using NI 6062E data acquisition card, the sampling frequency is 20 kHz, the data collecting time is about 164 hours. There are 984 files and each file consists of four rows and 20480 columns data, that is four passages data and each passage data is 20480 points.

3.2 THE ASSESSMENT CURVES ANALYSIS OF TIME DOMAIN AND FREQUENCY DOMAIN STATISTICAL FACTORS FOR THE EXPERIMENTAL DATA

After the experiment, according to each passage data respectively, the time domain and frequency domain statistical factors of every data segment can be calculated. Then the statistical factors of all data segments, that is the life cycle data, are drawn and the life cycle assessment curve of rolling bearing could be obtained.

1) The assessment curves and analysis of the domain factors.

The peak-to-peak level assessment curves of rolling bearing 2, 3 and 4 are shown in Figure 2, the assessment curves of each time domain statistical factor of rolling bearing 1 are shown in Figure 3.

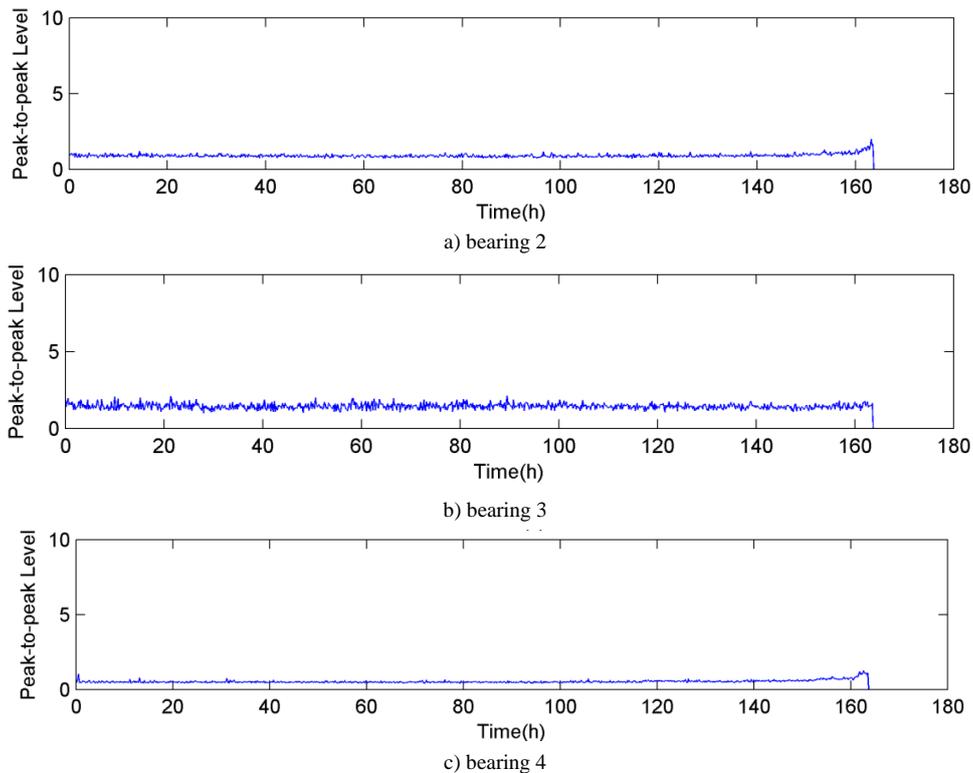


FIGURE 2 Peak-to-peak Level assessment curves of bearing 2, 3 and 4

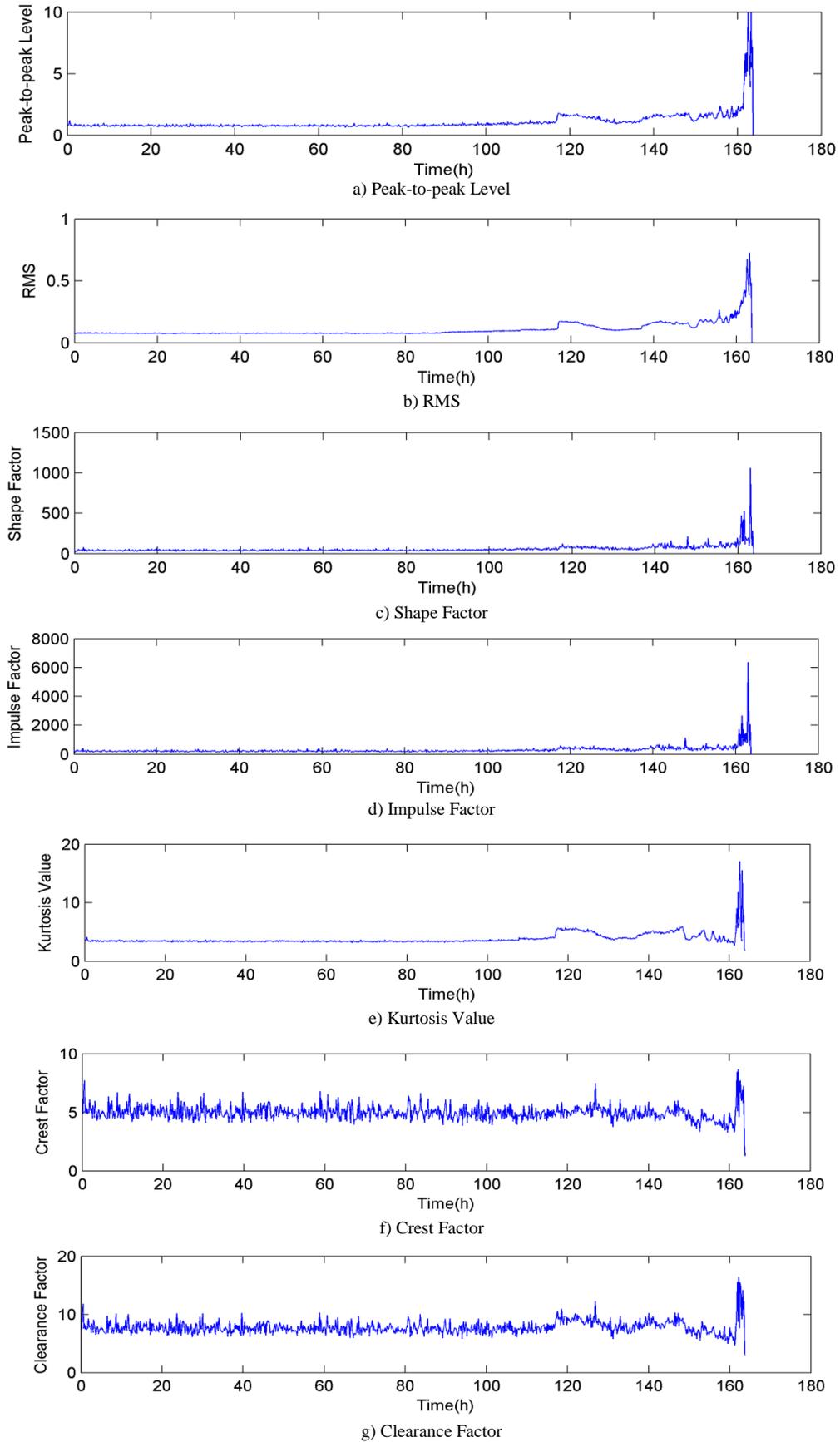


FIGURE 3 The assessment curves of time domain statistical factors of bearing 1

From the assessment curves of time domain statistical factors in Figure 2 and Figure 3, the change trends of peak-to-peak level factor of bearing 2, 3 and 4 are relatively consistent and keep stably. The similar curves can be obtained for other time domain statistical factors. So, it can be preliminarily judged that bearing 2, 3 and 4 are normal condition and no severe fault; but comparing with other bearing curves, the latter stage of each factor assessment curve of bearing 1 fluctuates violently, so the bearing 1 occurs fault. And with the increase of fault degree, the performance of the bearing degrades faster and faster that is consistent with the actual condition. The assessment curve amplitudes of bearing 2, 3 and 4 increase slightly about at 160h around, in fact, it is influenced by the fault of the bearing 1. When the end of the experiment, the outer of the bearing 1 occurs fault and finally the bearing becomes failure because of serious fault.

Next, the bearing 1 with the fault is analysed detailed. From Figure 3 (a) and Figure 3 (b), it can be seen that peak-to-peak level and RMS of the bearing 1 all remain consistent and stable condition at the beginning of a long period of time, which shows that the rolling bearing runs well; at 88.5h around, the amplitudes of peak-to-peak level and RMS increase slightly, the rolling bearing occurs minor fault and can work stably at this stage; until 116.7h around, the amplitudes of peak-to-peak level and RMS increase obviously, which represents the bearing degradation degree becomes serious and comes into the moderate fault stage. After the fault becoming serious, the

degradation speed of the bearing accelerates and the amplitudes fluctuate stronger; at 158.3h around, it can be seen that the fluctuation becomes further stronger from the two factors' change, and the amplitudes rise obviously, the bearing comes into the severe fault stage; until 163.5h around, the factors rise rapidly, the bearing could not run and finally it is no longer functional, the experiment finishes. From Figure 3 (c), (d) and (e), it can be seen that the shape factor, impulse factor and kurtosis value all show sensitively reflection to the moderate fault and the severe fault. Among these three factors, the amplitudes of kurtosis value change more seriously and have better validity for identifying the moderate fault. From Figure 3 (f), the entire change trend of crest factor assessment curve fluctuates bigger and only reflects to the severe fault. From Figure 3 (g), clearance factor reflects to the moderate fault and severe fault, but cannot distinguish them easily.

Summary, peak-to-peak level and RMS factors reflect relatively weakly for outer raceway minor fault of rolling bearing and the sensitivity is low; other time domain statistical factors do not reflect to outer raceway minor fault of rolling bearing. Except for crest factor and clearance factor, other time domain statistical factors are sensitive to the moderate fault, the severe fault and the failure condition.

(2) The assessment curves and analysis of the frequency factors

The assessment curves of each frequency statistical factor of the bearing 1 are shown in Figure 4.

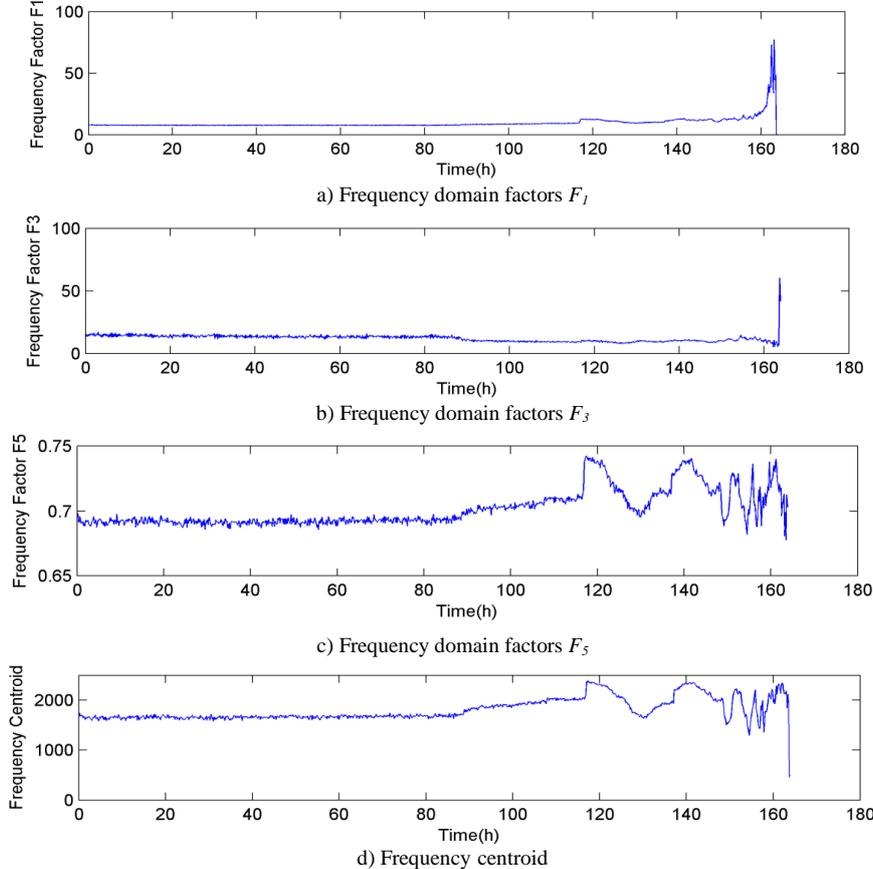


FIGURE 4 The assessment curves of frequency domain statistical factors of bearing 1

From Figure 4 (a), it can be seen that the frequency factor F_1 fluctuates slightly at 88.5h around and increases obviously at 116.7h around, the growth rate increases at 158.3h around and continues to rise until the bearing failure at 163.5h around. So, the frequency factor F_1 is very sensitive to rolling bearing outer raceway moderate fault, severe fault and the failure condition, and slightly sensitive to the minor fault. The similar result can be obtained for the frequency factor F_2 and the Figure is omitted. From Figure 4 (b), the frequency factor F_3 obviously decreases at 88.5h around which represents the rolling bearing comes into the minor fault stage, then the amplitude fluctuates slightly until 158.3h around becomes stronger and the bearing final failure at 163.5h around. So, the frequency factor F_3 is very sensitive to rolling bearing outer raceway minor fault, generally sensitive to the severe fault and relatively lack to distinguish moderate fault. The similar result can be obtained for the frequency factor F_4 and the Figure is omitted. From Figure 4 (c) and (d), the frequency factor F_5 and frequency centroid are relatively stable before 88.5h and increase obviously at 88.5h which represents the bearing comes into minor fault stage, then the amplitude increases slowly until 116.7h around further increases obviously which represents the bearing comes into moderate fault stage, next the curve amplitude strongly waves. So the frequency factor F_5 and frequency centroid are very sensitive to the minor fault and moderate fault, and nearly can't distinguish severe fault. The curve trends of the mean square frequency, RMS frequency and the frequency variance factors are similar to factor F_5 besides the amplitude, the curves are omitted.

It is visible that frequency centroid, F_3 , F_4 and F_5 factors are more sensitive to outer raceway minor fault of the bearing than time domain statistical factors, but obviously lack to severe fault and the failure condition.

Based on above time domain and frequency domain factors analysis for rolling bearing life cycle data, time domain factors and frequency domain factors respectively have different assessment ability aiming to different degree faults. So, in practical application, the advantages of each factor can be used for comprehensive assessment.

References

- [1] Wang G, He Z, Chen X 2013 Journal of Mechanical Engineering 49(1) 63–72 (in Chinese)
- [2] Wong M L D, Jack L B, Nandi A K 2006 Modified self-organising map for automated novelty detection applied to vibration signal monitoring *Mechanical Systems and Signal Processing* 20(3) 593–610
- [3] Kankar P K, Sharma S C, Harsha S P 2010 Expert Systems with Applications (2010) 1–11 (in Chinese)
- [4] Cocconcelli M, Bassi L, Secchi C, Fantuzzi C, Rubini R 2012 An algorithm to diagnose ball bearing faults in servomotors running arbitrary motion profiles *Mechanical Systems and Signal Processing* 27 667–82
- [5] Boumahdi M, Dron J P, Rechak S, Cousinard O 2010 On the extraction of rules in the identification of bearing defects in rotating machinery using decision tree Expert Systems with Applications 37(7) 5887–94
- [6] Carnero M C, Pedregal D J 2011 Forecasting turbine problems by means of the state space framework *Journal of Loss Prevention in the Process Industries* 24(4) 432–9
- [7] Karacay T, Akturk N 2009 Experimental diagnostics of ball bearings using statistical and spectral methods *Tribology International* 42(6) 836–43
- [8] Li Yuqing, Wang Rixin, Xu Minqiang 2013 Journal of vibration and shock 32(18) 169–73 (in Chinese)
- [9] Lei Y, He Z, Zi Y 2007 Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas *Mechanical Systems and Signal Processing* 21(5) 2280–94
- [10] Samanta B, Al-Balushi K R 2003 Artificial neural network based fault diagnostics of rolling element bearings using time-domain features *Mechanical Systems and Signal Processing* 17(2) 317–28
- [11] Nikolaou N G, Antoniadis I A 2002 Rolling element bearing fault diagnosis using wavelet packets *NDT & E International* 35(3) 197–205

4 Conclusions

By analysing time domain and frequency domain factors curves of experimental data, the change trend of the curve is basically consistent with the practical outer raceway fault condition of rolling bearing. But the assessment ability of each factor is different for outer raceway different fault degrees.

Comparing with other time domain factors, peak-to-peak level, RMS and kurtosis value can well distinguish outer raceway moderate fault and severe fault of rolling bearing. Comparing with each time domain factor, F_3 , F_4 , F_5 and frequency centroid of the frequency domain factors are more sensitive to outer raceway initial fault of rolling bearing and the stability is well.

By using the proposed comprehensive analysis method, partly frequency domain statistical factors (F_3 , F_4 , F_5 and frequency centroid) are used for analysing outer raceway minor fault of the rolling bearing, and partly time domain statistical factors (peak-to-peak level, RMS and kurtosis value) are used for analysing moderate fault and severe fault, the outer raceway performance degradation condition of the rolling bearing can be monitored well.

Although the experiment shows that the proposed comprehensive analysis method can monitor the outer raceway performance degradation condition of the rolling bearing, it is inconvenient for user to use multiple statistical factors. So, the further research focus is to study a unified and effective assessment factor for rolling bearing performance degradation.

Acknowledgments

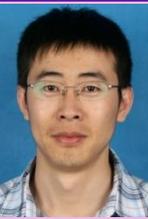
Project supported by National Natural Science Foundation of China (Grant No. 51305109), the Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20122303120010), and the Specialized Foundation for Technological Innovation Talents of Harbin City (Grant No. 2013RFLXJ019).

[12]Sejdic E, Djurovi I, Jiang J 2009 ime-frequency feature representation using energy concentration: An overview of recent advances *Digital Signal Processing* 19(2009) 153–83

[13]Pan Y, Chen J, Li X 2010 Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy c-means *Mechanical Systems and Signal Processing* 24(2) 559–66

[14]Guo Lei, Li Xinglin, Wu Shen 2012 *Bearing* 8(2012) 46–50 (in Chinese)

[15]Yang Huibin 2011 Hubei: Hubei Industry University 24–29 (in Chinese)

Authors	
	<p>Yujing Wang, born in 1983, Henan, China</p> <p>Current position, grades: lecturer of College of electrical and electronic engineering, Harbin University of Science and Technology. Ph.D. student of Harbin Institute of Technology.</p> <p>University studies: B.Sc. degree in electronic information engineering. M.Sc. degree in signal and information processing from Harbin University of Science and Technology, Heilongjiang, China, in 2004 and 2007 respectively.</p> <p>Scientific interest: non-stationary signal processing, fault diagnosis technology.</p> <p>Experience: Since July 2007 electrical and electronic engineering, Harbin University of Science and Technology, Heilongjiang, China.</p>
	<p>Yicheng Jiang, born in 1964, Harbin, China</p> <p>Current position, grades: professor and supervisor for Ph. D. students in Harbin Institute of Technology.</p> <p>University studies: Ph.D. degree in communication and information system from Harbin Institute of Technology, Heilongjiang, China, in 1997.</p> <p>Scientific interest: non-stationary signal processing, target recognition technology.</p> <p>Experience: Since 2001 he has been a professor, and since 2003 he has been a supervisor for Ph. D. students.</p>
	<p>Shouqiang Kang, born in 1980, Heilongjiang, China</p> <p>Current position, grades: associate professor. College of electrical and electronic engineering, Harbin University of Science and Technology.</p> <p>University studies: M.Sc. degree in signal and information processing both from Harbin University of Science and Technology, Heilongjiang, China in 2007. Ph.D. degree in system analysis, information control and processing from Belarusian State University, Minsk, Belarus, in June 2011.</p> <p>Scientific interest: non-stationary signal processing technology.</p> <p>Experience: since 2011 he has been an associate professor and supervisor for M. Sc. students.</p>

Evolution of the Nanjing urban green land based on GIS analysis

Hao Xu*

College of Landscape Architecture, Nanjing Forestry University, Nanjing, China

Received 01 May 2014, www.tsi.lv

Abstract

Nanjing is one of the important central cities in the Yangtze River Delta, which goes through dramatic urbanization development in recent 40 years. In this research, green land distribution data in 1966, 1981 and 2004 were extracted from topographic maps, aerial imagery, and so on. Time series analysis has been conducted on the GIS platform; characteristics of the evolution of urban green land from 1966 to 2004 were summarized from the perspective of the scale, function, pattern analysis. The results reveal that an environmental function plays an important role, status of recreational function began to rise, the spatial pattern of green land has a tendency of specialization and complication, the downward of its total size accelerates. Woodland which maintain the basic pattern of the recent 40 years green land system in Nanjing remains stable structure.

Keywords: green land, evolution, GIS, analysis, Nanjing

1 Introduction

The research of urban green land not only includes the qualitative and quantitative study of its distribution and functional characters, but also contains a deep understanding of the historical evolution characteristics of the green, only in this way, the crisis and mechanisms could be fully diagnosed, planning and construction methods and goals will also be further cleared. Analysis of the evolution of the characteristics of urban green land should be based on three aspects, namely scale, function and pattern, The scale reflects the total size of all kinds of green lands, the function reflects whether the role of the relationship between urban structure of green land and human dwelling is reasonable, the pattern implies the structural characteristics represented by different functions and different scales of green space. The evolution features of green land can be comprehended essentially from size, function and pattern. The purpose of this research aims at grasping the changing nature of the characteristics of various types of urban green lands development through the analysis of the evolution, and accumulating research data for scientific planning.

Former research related to the evolution of green was basically qualitative, because of limited data and methods. In recent years, Landscape ecology methods and information technology had been taken to analyse quantitatively the green pattern, but the evolution analysis of urban green land lacked. In the aspect of Quantitative research, Suzuki M, etc. use different years ancient map data to conduct a time-sequence to analyse the centre green area in Tokyo on a GIS platform, and to explore the 100 years of evolution characteristics in the Tokyo Green Area [1-9]. In this research, we focus on the

green lands in Nanjing, analyse its evolution quantitatively from scale, function and pattern, sum up the characteristic variation, and grasp the mechanism of green land variation.

2 Materials and methods

Nanjing is located in the lower reaches of the Yangtze River Ningzhen hills and mountains area, north latitude 31°14' to 32°37', longitude 118°22'~119°14'. Urban population reached 5.48 million in 2011, which is one of the important central cities in the Yangtze River Delta region.

Nanjing has a distribution of low hills around the city with the Yangtze River cutting through and has environmental characteristics of dense rivers and lakes. In the past half century, the velocity of urbanization in Nanjing was rapid, which exert a serious impact on the ecological environment. Therefore, it is necessary to study the changes in the urban green land system.

The scope of this research include the south of the Yangtze River, west of Rao city highway, south to first-line Andemen, North qixia avenue - East shogunate first-line of the city, involving the main city of Nanjing Gulou district, all parts of Baixia, Qinhuai district, the most parts of Xuanwu, jianye, and Xiaguan district, part of Qixia, Yuhuatai district in Nanjing, all of which are dense areas for peoples' living and dwelling. These areas are the main city now, which are the typical regions that witness the changes of urban green land for nearly 4 decades.

Analysis of the historical development of the green land must begin with the time-section that is needed for concrete explanation. Three time cross-sections were selected and analysed in 1966, 1981, 2004, due to the lack

* *Corresponding author* e-mail: xuhao73@vip.163.com

of the Nanjing map data before the founding. Aviation image and topographic maps of 1966, 1981 and 2004 were performed by vector processing in MapInfo software, green land distributions for these three time sections were extracted and analysed. Characters of green land evolvement of the 1966-1981 and 1981-2004 period were further compared from the difference analysis of size,

function and pattern.

Figure 1 shows the extracted green lands, which included parks, neighbourhood green spaces, squares, woodland, affiliated green land, waterfront green land, agricultural land and other patches of green land. Table 1 shows types of green lands selected in this study.

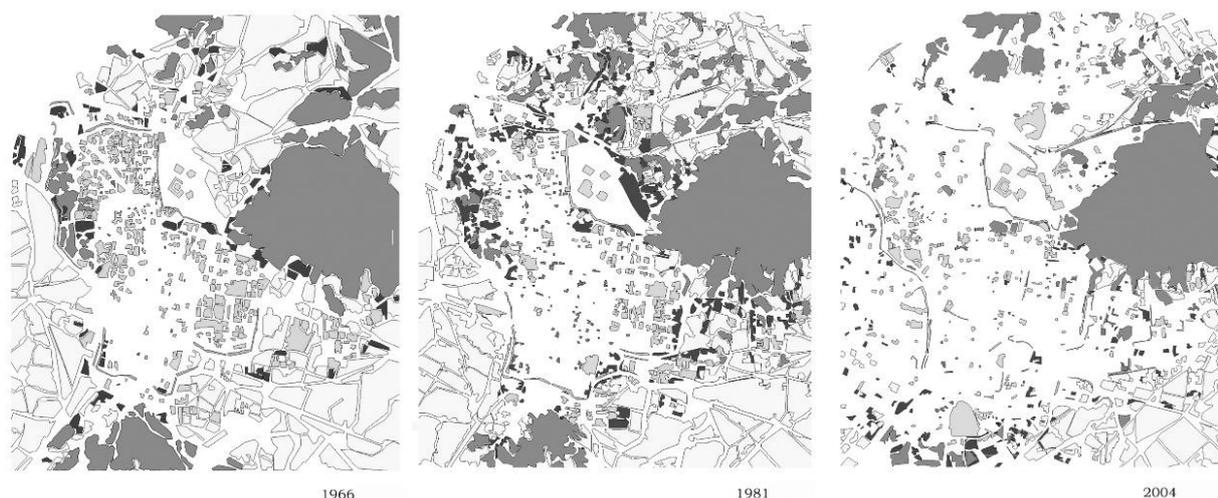


FIGURE 1 Green lands extracted using GIS

TABLE 1 Types of green lands selected in this study

green land types	scope
park	various types of parks specified by city government, exclusive of lakes in the parks
road green belt	green belt in the middle and side of the road
block green belt	concentrated green land in the residential area
square	city square, street square
waterfront green land	belt-like waterfront green land
woodland	acres of forest land containing facilities inside
affiliated green land	affiliated green land of constitution, schools, factories, and research institutes
agricultural land	Farms, farmland
other green land	other patches of greened open land

3 Analysis and results

Green space in the 1966 can be divided into parks, woodland, agricultural land, affiliated green land, block green belt, waterfront green land and others. Patches of complete agricultural land were distributed in north, west, south of Qinhuai River and northeast, north of Xuanwu Lake, The woodland was mainly distributed in the hilly land of the peripheral urban area. The city is littered with affiliated green land patches. Hunan Road, Xinjiekou area was a traditional residential centre with high dense living units and relatively sparse affiliated green land. There were chunks of affiliated green land in the military sites located in the south of Purple Mountain. Even in the suburbs areas, such as South of Qinhuai River, north of Xuanwu Lake, some factories and some affiliated green lands scattered there. Block green belt were very few, dotting around the city. There were intermittent waterfront green lands in the west of Xuanwu Lake and both sides of the Qinhuai River.

In terms of size, agricultural land was the largest part

of Nanjing green lands in 1966, accounting for 50% of the weak. Followed by woodland, it accounted for 34% of the total. The affiliated green land accounted for 10.4% of the total. Other types of green lands accounted for less than 6% of the total size. The affiliated green land constituted the main green land of the city.

Green spaces in 1981 included parks, woodland, agricultural land, affiliated green land, block green belt, waterfront green land and others. The more fragmented setup of green land was in consistent with that in 1966, surrounded by large tracts of agricultural land and woodland. The affiliated green land was still the subject of urban green space, accounted for 7% of the total size. There was an expanding tendency for affiliated green land compared with that in 1966. Factory in the suburbs showed an increasing tendency, while factory in the residential land decreased, so did the affiliated green land in urban areas. But peripheral affiliated green land increased and revealed decentralized and miniaturize. There was also an increase for block green belt within the city. As major

outdoor recreation places for the locals, park green land and block green belt accounted merely for relatively low ratio of 2%. Waterfront Green land was located on both sides of Xuanwu Lake and Qinhuai River. Open green land in urban fringe increased significantly, indicating that the city-building activities in 1981 became common.

Green land in 2004 included parks, road green belt, block green belt, square, waterfront green land, woodland, agricultural land, affiliated green land and unknown others. Woodland was a major component of the Nanjing urban green space, and its distribution was mainly concentrated in the hilly land of the eastern outskirts, woodland in the main urban area was basically depleted. The outskirts distributed large numbers of agricultural land, large proportions of it lied in southeast, less and fragmented in southwest and northeast, its distribution ratio gradually reduced from the countryside to urban built-up area. The affiliated green land was the most widely distributed type within the city, but with generally small patches and a high degree of fragmentation. Park, block green belt and squares played the role of urban leisure. Green parks were mainly concentrated in the areas that were famous for its natural history and culture; however, serious shortage existed within the residential areas. block green belt was generally small, sparse and distributed unevenly. Numbers of square and waterfront green lands were few.

From 1966 to 1981, the scale of agricultural land, and waterfront green land was essential flat. Woodland and affiliated green land had decreased slightly; there were also small increases in the size of the green parks and block green belt. The green open land for unknown use (other green lands) increase larger. Although the total amount of green lands did not change significantly, the ebb and flow of the various types of green scale resulted from joint effect of various urbanization factors, the result indicated that Nanjing city was on the eve of new expansion cycle.

From 1981 to 2004, due to the effect of urbanization, the green land reduced sharply in total. Recreational functions-based and government-invested green land was significantly increased in various types of green land. Increment of block green belt and waterfront green was weak; Decrement of woodland was relatively small. The expansion of the city's eroded agricultural land around the original town, due to the increment in building density, the primary affiliated green land was occupied by the large numbers of buildings; the original green open lands were substituted by others, resulting in the half reduction of affiliated green land, agricultural land and others.

Figure 2 shows scales of various green lands of Nanjing city in 1966, 1981, and 2004. From 1966 to 2004, the total green lands reduced by 21%. Affiliated green land and agricultural land reduced most in various types of green lands, decreased by 70% and 61% respectively. Size of woodland decreased by 12%. Greater growth existed in

parks and block green belt with an increment of 5.8 times and 8.5 times separately. But the waterfront green land increased by 13%. Overall, the effect of urbanization green deepened since 1981, resulting in larger variation of the green land than that in 1981.

Functional evolvement of green land could be discovered from changes in the size of the various types of green lands. From the year 1966 to 1981, agricultural land and woodland based on production function and environmental protection were the main component of green land. Recreational function was still taken on by parks and block green belt despite the increment of its size, the fact that its overall proportion was less indicated that the recreation was not the main function, but the dominant role of urban green land remained. The affiliated green land can only provide limited recreational and ecological role due to the absence of public properties.

Recreational function was valued and enhanced with the proportional increment of parks and block green belt since 1981. Affiliated green land declined accordingly. Under the effect of expanding urban, the proportion of agricultural land decreased from 50% to 32%. The productive function of green land gradually weakened. The proportion of woodland scale rose from 32% to 48.8% in spite of its slightly decrement, the environmental protection in green land strengthened accordingly.

In general, environmental protection remained the major function of Nanjing green land from 1966 to 2004, with the expansion of the city and the necessities of habitat living productive function gradually weakened and recreational functions gradually increased.

Figure 3 shows patch numbers of various green lands of Nanjing in 1966, 1981 and 2004. In this research, patch density index (PD) and landscape shape index (LSI) were adopted, the interferential and fragmented extent of each time cross-sectional area was compared, and the evolve mental characteristics of green land patterns from 1966 to 2004 were analysed [10]

Figure 4 shows PD value changes of various green lands of main urban area of Nanjing from 1966 to 2004. Figure 5 shows LSI value changes of various green lands. From 1966 to 1981, the PD value of agricultural land plummeted, LSI value rose sharply before 1981, and then both showed a downward trend since 1981, the fact indicated that the urbanization process continuously exert a effect on the size agricultural land, which had been seriously interfered, crushed and then disappearing. The spatial structure of block green belt fragmented increasingly with its growing scale, PD values and LSI values continued rising. Firstly, the affiliated green land had been disturbed, and then showed a fragmentation trend, after that, this suffix block disappear faster, manifesting as PD values firstly increased and then decreased and LSI values continued to decline. PD and LSI value of Waterfront Green land and Park was essentially

flat before 1981, and then improved. The fact indicated that there is no change in these two types of green space before 1981; structure complexity was enhanced with its increment in size and patch since 1981. The PD value of the woodland remained stable, LSI values firstly increased and then declined, this indicated that the body of spatial structure showed an ordered trend and a lowering fragmentation despite the certain interference.

Some green lands showed larger changes such as

agricultural land, affiliated green land, block green belt and others before 1981, changes of LSI value did exist in agricultural land and affiliated green land. Greater variation happened in PD and LSI value of parks and waterfront green land, which indicated that urbanization process had greater influence on agricultural land, affiliated green land, block green belt and others, however park and waterfront green land were promoted by human conscious construction activities since 1981.

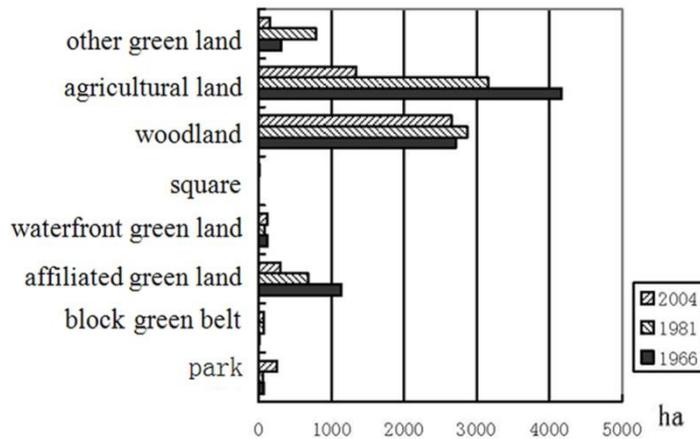


FIGURE 2 Scales of various green lands of Nanjing city in 1966, 1981, and 2004

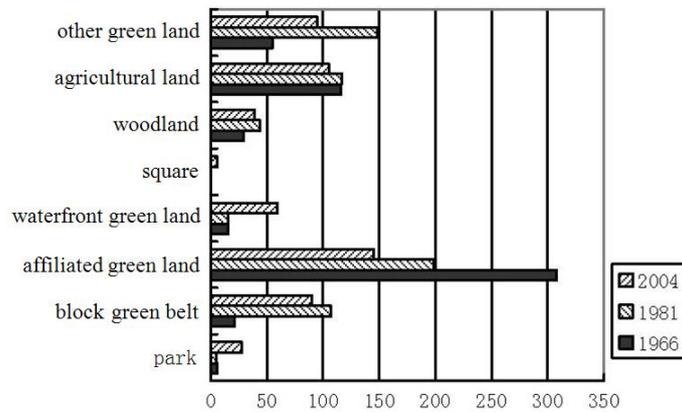


FIGURE 3 Patch numbers of various green lands of Nanjing in 1966, 1981, 2004

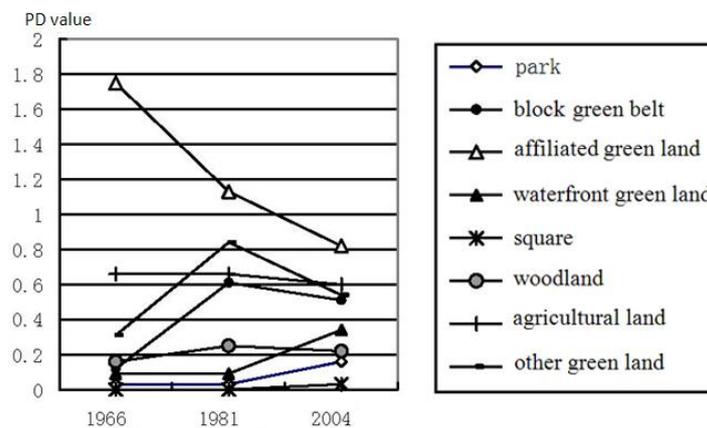


FIGURE 4 PD value changes of various green lands of main urban area of Nanjing from 1966 to 2004

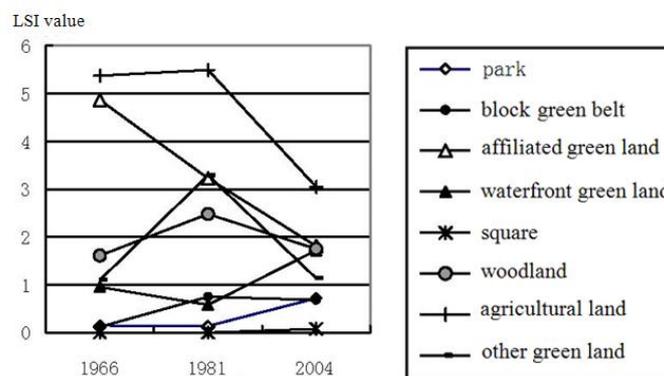


FIGURE 5 LSI value changes of various green lands of main urban area of Nanjing from 1966 to 2004

It was concluded from the variation curves of overall PD and LSI that there were greater variations in the agricultural land, affiliated green land, block green belt and others, and milder variations in waterfront green land and park. Variations in the woodland showed the smoothest character. The fact indicated that spatial structure of woodland is the most stable one, better structural stability were showed in the waterfront green land and parks. Agricultural land, affiliated green land, block green belt and others possessed the least structural stability.

4 Conclusions

Prior to the 1980s, the scale changes of the Nanjing urban green land are small. Proportions of all kinds of green land were stable, but the spatial structure was susceptible to be interfered, and showed a general tendency of fragmentation. In this stage, the main functions of the green land were production and environmental protection.

After the 1980s, the total size of the Nanjing urban green lands reduced sharply, influenced by the rapid urbanization process. Suburban agricultural land had been eroded greatly and then disappeared, because of urban sprawl and the impact of human activities. The size of

affiliated green land suffers from such a significant reduction that isolated island emerged. A mild size increment of park, waterfront green land and block green belt suggested that recreational function was valued gradually.

During the evolvement of urban green land in Nanjing, the most stable structure was the woodland, which was the main factor to keep the green pattern, environmental and recreational function that woodland embodied had important implications for city of Nanjing.

Overall, during the evolvement of urban green land in Nanjing, environmental function had been a major role; role of recreational function began to rise, spatial pattern of green land showed ever-growing fragmentation and complexity trend, the size of the total green lands declined accelerated.

The green land distributing information was extracted from the Nanjing Aeronautics photographs and topographic maps on the GIS platform, green land information in 1966, 1981 and 2004 was gathered and quantitatively compared for the first time, the result indicated that woodland which showed an stable structure maintained the basic 40-year-green land pattern of Nanjing, green land space pattern fragmented continuously, environmental protection was the main

function of the Nanjing Greenland, GIS data processing and analysis capabilities contributed to the accurate grasp of green land evolution in qualitative and quantitative analysis method, and also provided serialized reference for the analysis of urban green land and the construction.

Acknowledgments

This paper is supported by the fund of Nanjing Forestry University (G2014019) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

- [1] Steiner F 2000 *The Living Landscape: An Ecological Approach to Landscape Planning McGraw-Hill Professional* 200–9
- [2] Shi X, Li M, Zhang H 2001 Application of Remote Sensing Technology in the Overall Planning of the Urban Green space System in Guangzhou *Science of Surveying and Mapping* 42-4 (in Chinese)
- [3] Han H, Gao J, Liu G 2003 Assessment of the Ecological Benefits of Urban Vegetation under the Support of Remote Sensing and GIS *Chinese Journal of Applied Ecology* 2301-4 (in Chinese)
- [4] Bai L, Wu W, Wu Z 2001 Application of RS and GIS in the Surveying of the Green space System in Hefei *Journal of Northwest Forestry University* 59-63 (in Chinese)
- [5] Hasebe G, Suzuki M 1996 A study on mapping land use transition in the process of urbanization-case study of open space of Edo-Tokyo *Papers and Proceedings of the Geographic Information Systems Association* 73-8
- [6] Segl K, Kaufmann H 2001 *IEEE Transactions on Geoscience and Remote Sensing* 39(9) 2080-3
- [7] Shailesh K, Joydeep G, Melba M C 2001 *IEEE Transaction on geoscience and Remote sensing* 39(7) 1368-79
- [8] Suzuki M 2003 *Geographical Information Systems for landscape Architecture Tokyo Soft Science* 27-45
- [9] Xu H, Suzuki M 2005 Application of RS, GIS and GPS technology on green land analysis and planning *Journal of Nanjing Forestry University* (5) 115–8 (in Chinese)
- [10] Wu J. 2000 *Landscape Ecology Beijing: Higher Education Press* 99-119 (in Chinese)

Author



Hao Xu

Current position, grades: Associate professor at Nanjing Forestry University.

Scientific interest: GIS, RS and landscape analysis

Publications: 30

A comparative analysis on circulation efficiency of different banana circulation modes in Zhangzhou city based on DEA model

Yaoting Chen¹, Junyu Dai^{2*}

¹*School of history and social development of Minnan Normal University, No.36, Xian Qian Road Zhangzhou in fujian province, China*

²*School of Minnan Normal University No.36, Xian Qian Road Zhangzhou in fujian province, China*

Received 6 July 2014, www.tsi.lv

Abstract

In the paper, the software DEAP was used to evaluate relative efficiency of 8 kinds of banana circulation modes in Zhangzhou City. CCR model and BCC model were employed. The related index included input (unit circulation cost, circulation time, and circulation loss rate) and output (net profit). The result showed that, the comprehensive efficiency, pure technology efficiency, scale efficiency in Mode 1 (farmers-third party logistics–supermarket–consumers) and Mode 8 (farmers-banana sales stalls–consumers) were relatively efficient. Therefore, it was judged that the main factors affecting the circulation efficiency were the compression of circulation level and the ascension of scale and professional level.

Keywords: DEA, circulation mode, circulation efficiency

1 Introduction

Technology efficiency was firstly proposed by the British economist M. J. Farrell of Cambridge University in 1957 based on Debreu's research work. Farrell proposed that technology efficiency refers to the ratio of the actual output and the production boundary of an enterprise under the given inputs; while the ratio of the gap between the production boundary and the actual output and the production boundary is the technology inefficiency. In 1972, S. N. Afriat firstly employed the maximum likelihood approach to establish the frontier production function model with statistical properties, which opened a new phase of the study using econometric model to analyse technology efficiency; however, the result of this approach was greatly affected by the residual distribution form. Different assumptions often lead to different estimation results. In 1974, J. Richmond firstly proposed the re-modified ordinary least squares approach to study the frontier production function.

However, there is a common drawback in the above models: they assumed that frontier production function of each production unit was the same, and the technology efficiency lead to the difference between actual output and boundary output. Actually, that is not the case. In view of this problem, in 1977 D. J. Aigner C. A. Knox Lovell in America and W. Meeusen in Belgian proposed stochastic frontier production function, which made the measuring of technology efficiency possible. At the same time, the famous strategist W. A. Charnes and W. Cooper proposed Data Envelopment Analysis in 1978 (abbreviated as

DEA), a new approach for evaluating the efficiency developed based on the concept of relative efficiency.

Based on variable returns to scale (VRS), the technology efficiency was further decomposed into pure technology efficiency and scale efficiency by Banker (1984), so that the application of this technology became more widely. Since then, based on the theory and practice in the use and development of logistics, and it gradually formed a non parametric approach used for economic quantitative analysis, mainly dependent on the linear programming technique. However, the endeavour of famous American strategists A. Charnes and W. W. Cooper (1985) made the non parametric approach popular in the form of data envelopment analysis (DEA) in the early 1980s; therefore, sometimes DEA was called non parameter approach or Farrel efficiency analysis.

On the basis, DEA began to be widely employed by research staff in various fields, for example: analysis on economic situation in cities, efficiency analysis of financial institutions, and management assessment of the public utilities. Besides efficiency evaluation, DEA is also employed for forecasting and early warning in economic system (Sheng Zhaohan et al., 1996; Klimberg et al, 2009). It reveals the unknown information in the system, to provide decision support for managers.

The economic meaning of DEA efficiency is: the output level in the existing system, structure and technology level; if the optimal value of the model is less than 1, it is considered that the economic activity is inefficient, thus the input can be reduced to achieve the current output; conversely, if the optimal value of the

* *Corresponding author* e-mail: 170000369@qq.com

model is equal to 1, the economic activity is efficient, and the output of the current input is the maximum output.

2 A brief introduction of banana circulation modes in Zhangzhou city

In the paper, price data in each link in each circulation mode in Zhangzhou market from January 1, 2008 to December 30, 2012 were collected, and the average unit circulation cost, average net profit, circulation loss rate, circulation time of each circulation mode in each year were calculated. The banana produced from Tianbao Town, Xiangcheng District, Zhangzhou City in different sales modes were chosen as the object of the study. The reasons were there exist obvious differences in time to market of the same banana from different production places at the same final consumption market, and the actual situation in the research was that the local banana played a great part in Zhangzhou Market. According to the survey, the result showed that there were 8 kinds of banana sales modes in Zhangzhou. They are:

- Mode 1: Farmers – third party logistics – supermarket – consumers.
- Mode 2: Farmers – assemblers – wholesalers – supermarket – consumers.
- Mode 3: Farmers – assemblers – wholesalers – fruit shops – consumers.

Mode 4: Farmers – assemblers – wholesalers – farmers market – consumers.

Mode 5: Farmers – assemblers – wholesalers – fruit shops – mobile stalls – consumers.

Mode 6: Farmers – fruit shop – consumers.

Mode 7: Farmers –farmers market – consumers.

Mode 8: Farmers – banana sales stalls – consumers

3 DEA model

Data envelopment analysis (DEA) is a linear programming model, mainly used for the efficiency evaluation of complex system with multiple inputs and multiple outputs. It does not need dimensionless treatment to the data, so this approach is more reliable. At present, in the academic circle, a relatively complete data envelopment analysis system has been formed and widely used in many fields. DEA has many different models, which represent different economic significance. The CCR model is the most basic model, which contains non-Archimedean infinitesimal, and the size of decision-making unit does not affect the relative efficiency. When the evaluation results show that the decision-making unit is DEA efficient, and the decision-making unit is technically efficient and scale efficient.

CCR model and input and output of its basic variables are shown in Table 1:

TABLE 1 CCR model and input and output of its basic variables

		DMU_1	DMU_2	...	DMU_n		
v_1	1	x_{11}	x_{12}	...	x_{1n}		
v_2		x_{21}	x_{22}	...	x_{2n}		
...		
v_m	m	x_{m1}	x_{m2}	...	x_{mn}		
		y_{11}	y_{12}	...	y_{1n}	1	u_1
		y_{21}	y_{22}	...	y_{2n}	2	u_2
...		
		y_{s1}	y_{s2}	...	y_{sn}	3	u_n

In the Table 1, x_{ij} indicates the amount of No. i input for No. j DMU. When $x_{ij} > 0$, y_{rj} indicates the amount of No. r input for No. j DMU. When $y_{ij} > 0$, v_i represents a measure of No. i input; while v_r represents a measure of No. r input.

Among them, the input amount of DMU_1 is $X_j = (x_{1j}, x_{2j}, \dots, x_{mj})^T$, while the output amount is $Y_j = (y_{1j}, y_{2j}, \dots, y_{sj})^T$, $j = 1, 2, \dots, n$. The corresponding weight coefficient is $v_j = (v_1, v_2, \dots, v_m)^T$, $u_j = (u_1, u_2, \dots, u_s)^T$, h_j is called the efficiency evaluation index of DMU_j .

$$h_j = \frac{u^T Y_j}{v^T X_j} = \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}}, j = 1, 2, \dots, n. \tag{1}$$

Based on the principle that efficiency evaluation index of each decision-making unit should not be more than 1, the efficiency of No. j_0 decision making unit is evaluated. Select the right coefficient u and v to make h_0 highest, which is shown as following:

$$\left\{ \begin{array}{l} \max \frac{\sum_{r=1}^s u_r y_{rj_0}}{\sum_{i=1}^m v_i x_{ij_0}} \\ s.t. \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1 \\ u_r \geq 0 \\ v_i \leq 0 \end{array} \right. , \tag{2}$$

where $j = 1, 2, \dots, n$, $r = 1, 2, \dots, s$, $i = 1, 2, \dots, m$. When Charnes-Cooper is employed, ($t = \frac{1}{v^T x_0}$, $w = tw$, $u = tu$) is

changed, slack variables s^- , s^+ are added, and the duality programming and non-Archimedean infinitesimal ε are introduced, the Equation (1) can be converted to:

$$\begin{cases} \min [\theta - \varepsilon (e_1^T s^- + e_2^T s^+)] \\ s.t. \sum_{j=1}^n \lambda_j x_j + s^- = \theta x_0 \\ \sum_{j=1}^n \lambda_j y_j - s^+ = y_0 \\ \lambda_j \geq 0, j = 1, 2, \dots, n \\ s^+ \geq 0, s^- \geq 0 \\ e_1 = (1, 1, \dots, 1)^T \in R^m \\ e_2 = (1, 1, \dots, 1)^T \in R^s \end{cases} \quad (3)$$

Among them, according to the requirements on the accuracy of calculation model, non-Archimedean infinitesimal ε can select different range (10^{-5} to 10^{-10}). Compared to DMU_0 , λ_j is combined ratio of DMU_j in reconstructed effective DMUs. Based on non-Archimedean infinitesimal, the optimal solutions of CCR model λ_0 , s_0^+ , s_0^- , θ_0 are calculated, through which it can be determined whether the research object is of technology efficiency and scale efficiency. The judgment basis is as following:

a) When $\theta_0 = 1$, $s_0^+ = 0$, $s_0^- = 0$, the decision making unit j_0 is DEA efficient, and economic activity of decision-making unit is of technology efficiency and scale efficiency.

b) When $\theta_0 = 1$, $s_0^+ \neq 0$, $s_0^- \neq 0$, the decision making unit j_0 is weak DEA efficient, and economic activity of decision-making unit is not both at the best of technology efficiency and scale efficiency.

c) When $\theta_0 < 1$, the decision making unit j_0 is not DEA efficient, and economic activity of decision-making unit is neither at the best of technology efficiency nor scale efficiency.

In the CCR model, the research object is the same type of decision making units in a system; it analyses the relatively effective production frontier in a certain input or output level, and compared it with each decision making unit, so as to evaluate their effectiveness.

The BCC model is based on the CCR model. In BBC model, it divided the technology efficiency into pure technology efficiency (PTE) and scale efficiency (SE). After adding constraints on weight λ , it evaluates the relative efficiency of decision making units in the case of variable returns to scale. Among them, the pure technology efficiency (PTE) refers to the ratio of the actual input and

the optimal input. If PTE = 1, it indicates that the decision-making unit is of pure technical efficiency; if PTE < 1, the decision-making unit is of pure technical inefficiency. Scale efficiency (SE) refers to the scale efficiency level of each individual decision making unit. If SE = 1, it indicates that the decision making unit is of scale efficiency; if SE < 1, it shows that the decision making unit is scale inefficiency.

4 Indicators construction

According to the requirements of data envelopment model, in the paper, unit net profit is selected as output indicators, and unit circulation cost, circulation time and circulation loss rate are selected as input indicators.

a) Unit net profit. It is the output index to measure the banana circulation efficiency, including sales and profits. In the paper, unit net profit is used to measure the banana circulation efficiency.

b) Unit circulation cost. It refers to the total cost in the banana circulation process, including picking, packing and transportation cost, market cost and management cost.

c) Circulation time. It refers to the time that it takes for the goods to transfer from production to consumption. Because after picking, the banana has to be ripened in the warehouse for about 3-5 days, and the ripening time in each banana circulation mode is the same. To be more accurate calculation, the circulation time deducts its ripening time in the course. That is the time when banana in the course of circulation (not including the ripening time).

d) Circulation loss rate. It refers to the amount of loss in the circulation process or the proportion of loss accounted for the total number.

TABLE 2 Input and output index of each banana circulation mode in Zhangzhou City in 2008 (based on the survey)

Circulation Mode	output		input	
	Unit net profits (Yuan/kg)	Unit circulation cost	circulation time (hour)	Circulation loss rate (%)
Mode 1	0.8	0.82	16	13%
Mode 2	0.72	0.9	28	16%
Mode 3	0.58	0.75	24	16%
Mode 4	0.58	0.69	24	17%
Mode 5	0.48	0.65	24	17%
Mode 6	0.63	0.70	12	35%
Mode 7	0.62	0.65	12	38%
Mode 8	0.6	0.53	12	38%

Note: In order to make the data comparable, unit net profit, unit circulation cost in each link are assessed in banana prices. Loss rate is calculated in percentage, and circulation time is calculated in hour. Data sources: according to the survey data.

TABLE 3 Input and output index of each banana circulation mode in Zhangzhou city in 2009 (based on the survey)

Circulation Mode	output		input	
	Unit net profits (Yuan/kg)	Unit circulation cost	circulation time (hour)	Circulation loss rate (%)
Mode 1	0.83	0.78	16	12%
Mode 2	0.79	0.82	28	16%
Mode 3	0.62	0.76	24	14%
Mode 4	0.51	0.63	24	16%
Mode 5	0.42	0.66	24	16%
Mode 6	0.67	0.71	12	30%
Mode 7	0.52	0.62	12	36%
Mode 8	0.59	0.49	12	36%

Note: In order to make the data comparable, unit net profit, unit circulation cost in each link are assessed in banana prices. Loss rate is calculated in percentage, and circulation time is calculated in hour. Data sources: according to the survey data.

TABLE 4 Input and output index of each banana circulation mode in Zhangzhou City in 2010 (based on the survey)

Circulation Mode	output		input	
	Unit net profits (Yuan/kg)	Unit circulation cost	circulation time (hour)	Circulation loss rate (%)
Mode 1	0.88	0.75	16	11%
Mode 2	0.79	0.84	28	17%
Mode 3	0.69	0.79	24	14%
Mode 4	0.58	0.65	24	16%
Mode 5	0.54	0.69	24	16%
Mode 6	0.75	0.73	12	30%
Mode 7	0.57	0.66	12	36%
Mode 8	0.73	0.50	12	36%

Note: In order to make the data comparable, unit net profit, unit circulation cost in each link are assessed in banana prices. Loss rate is calculated in percentage, and circulation time is calculated in hour. Data sources: according to the survey data.

TABLE 5 Input and output index of each banana circulation mode in Zhangzhou City in 2011 (based on the survey)

Circulation Mode	output		input	
	Unit net profits (Yuan/kg)	Unit circulation cost	circulation time (hour)	Circulation loss rate (%)
Mode 1	0.99	0.73	16	10%
Mode 2	0.86	0.86	28	18%
Mode 3	0.67	0.86	24	14%
Mode 4	0.55	0.67	24	16%
Mode 5	0.54	0.70	24	16%
Mode 6	0.79	0.74	12	30%
Mode 7	0.57	0.65	12	36%
Mode 8	0.73	0.51	12	36%

TABLE 7 Efficiency evaluation results of each banana circulation mode in Zhangzhou city

Circulation Mode	Decision-making unit	Comprehensive efficiency	Pure technology efficiency	Scale efficiency	Returns to scale
Mode 1	1	1	1	1	constant returns to scale
Mode 2	2	0.774	0.839	0.923	increasing returns to scale
Mode 3	3	0.682	0.912	0.748	increasing returns to scale
Mode 4	4	0.682	1	0.682	increasing returns to scale
Mode 5	5	0.651	0.948	0.687	increasing returns to scale
Mode 6	6	0.994	1	0.994	increasing returns to scale
Mode 7	7	0.838	1	0.838	increasing returns to scale
Mode 8	8	1	1	1	constant returns to scale
mean		0.828	0.962	0.859	

Note: the comprehensive efficiency is technology efficiency calculated based on the CCR model; pure technology efficiency is technology efficiency calculated based on the BCC model; scale efficiency = comprehensive efficiency / pure technology efficiency.

Note: In order to make the data comparable, unit net profit, unit circulation cost in each link are assessed in banana prices. Loss rate is calculated in percentage, and circulation time is calculated in hour. Data sources: according to the survey data.

TABLE 6 Input and output index of each banana circulation mode in Zhangzhou City in 2012 (based on the survey)

Circulation Mode	output		input	
	Unit net profits (Yuan/kg)	Unit circulation cost	circulation time (hour)	Circulation loss rate (%)
Mode 1	1.01	0.71	16	10%
Mode 2	0.84	0.88	28	18%
Mode 3	0.69	0.81	24	17%
Mode 4	0.57	0.68	24	23%
Mode 5	0.54	0.73	24	25%
Mode 6	0.75	0.75	12	31%
Mode 7	0.56	0.69	12	35%
Mode 8	0.74	0.53	12	35%

Note: In order to make the data comparable, unit net profit, unit circulation cost in each link are assessed in banana prices. Loss rate is calculated in percentage, and circulation time is calculated in hour. Data sources: according to the survey data.

5 Model evaluation and calculation results

In the paper, the reverse tracking approach was employed, combined with depth interviews and questionnaire survey. A large number of first-hand information was obtained, which provides necessary data support for the empirical analysis. Due to the multiple circulation links in the circulation process, the corresponding index data of each circulation link is added, in order to get the overall index data. Specifically, first of all, the interview records and questionnaires are classified according to the different circulation modes, and then subdivided into different circulation links, and input and output data of each link in each circulation mode are added, to get the input-output data of the whole circulation mode, finally the average of each input and output index in each circulation mode is calculated for the empirical analysis. As Table 1 and Table 5 showed, 8 kinds of circulation modes listed in the Table 8 were considered as 8 decision making units. Combined with the data in the Table 6 and the software DEAP, CCR model and BCC model based on the input-oriented were employed to evaluate efficiency of the circulation mode.

6 Conclusions

Results were shown in Table 6. According to the evaluation results, it came out the following conclusions:

a) Among the 8 banana circulation modes, 2 modes were efficient, while 6 were inefficient. Namely, most banana circulation modes with retail terminal were in the inefficient state. After the comprehensive efficiency was divided into pure technology efficiency and scale efficiency, it could be seen that pure technology inefficiency and the scale inefficiency resulted in the inefficient state of most banana circulation modes, and scale inefficiency was the main reason.

b) Table 8 shows that the relative efficiency of circulation mode with minimum circulation level were higher than that with more circulation level, that is, the

comprehensive efficiency of Mode 1 and Mode 8 were higher than that of Mode 3, Mode 4, Mode 6 and Mode 7. Thus, it could be seen, in the case that the supermarket organized banana supply from the third party logistics, the supplier had played a positive role, or the behaviour of banana farmers who skipped the process of purchase, wholesale promoted the circulation efficiency.

c) The relative efficiency of Mode 5: Farmers – assemblers – wholesalers – fruit shops – mobile stalls – consumers was the lowest, only 0.651.

Acknowledgement

This work was supported by Fujian Science & Technology Project (No: 2014R0080)

References

[1] Afriat S N 1972 Efficiency estimation of production functions *International Economic Journal* 13(3) 568-98

[2] Aigner D J, Lovell C A, Schmidt P 1977 Formulation and estimation of stochastic frontier production function *Journal of Econometrics* 6(1) 21-37

[3] Banker R D, Charnes A, Cooper W W, Swarts J, Thomas D 1989 An introduction to DEA with some of its models and their uses *Research Governmental and Nonprofit Accounting* 5 125-63

[4] Charnes A, Huang Z M, Rousseau J J, Wei Q L 1990 Cone external solutions of multi-payoff games with cross-constrained strategy *Sets Optimization* 21(1) 51-69

[5] Charnes A, Cooper W W, Rodes E 1978 Measuring the efficiency of decision making units *European Journal of Operational Research* 2(6) 429-44

[6] Charnes A, Cooper W W, Golany B 1985 Foundations of data envelopment analysis for pareto - koopmans efficient empirical production functions *Journal of Econometrics* 30(1) 91-107

[7] Charnes A, Cooper W W, Li S L 1989 Using DEA to evaluate relative efficiency in the economic performance of Chinese cities *Socio-economic planning sciences* 11(23) 325-44

[8] Debreu G 1951 The coefficient of resource utilization *Journal of Econometric Society* 19(3) 273-92

[9] Farrell M J 1957 The measurement of productive efficiency *Journal of the Royal Statistic Society* 120(3) 253-90

[10] Klimberg R K, Lawrence K D, Lai T 2009 Using data envelopment analysis (DEA) to forecast bank performance, In Lawrence K D, Klimberg R K (ed.) *Advances in business and management forecasting* 6 Emerald Group Publishing Limited 53-61

[11] Sheng Z, Zhu Q, Wu G 1996 DEA theory, approach and application *Beijing Science Press* 54-60 (In Chinese)

Authors	
	<p>Yaoting Chen, born in September, 1979, Xiangcheng District, Fujian Province, China</p> <p>Current position, grades: Associate Professor of School of history and social development Minnan Normal University, China. University studies: PH.D in management Fujian Agricultural and forestry University in China. M.Sc. in management Tianjin University in China. Scientific interest: circulation of agricultural products. Publications: 13 papers. Experience: teaching experience of 9 years, 10 scientific research projects.</p>
	<p>Junyu Dai, born in April, 1984, Xiangcheng District, Zhangzhou City, Fujian Province, China</p> <p>Current position, grades: Associate Professor of Minnan Normal University, China. University studies: PH.D in management from Fujian Agricultural and forestry University in China. Scientific interest: circulation of agricultural products Publications: 4 papers. Experience: teaching experience of 6years, 2 scientific research projects.</p>

A Comprehensive FAHP evaluation model on domestic sports economy development mode

Tao Shi¹, Zhijie Xing^{2*}

¹Handan College, Handan city, Hebei province, China

²Hanshan District Experimental Primary School, Handan city, Hebei province, China

Received 19 July 2014, www.tsi.lv

Abstract

Sports economy refers to rational allocation of various resources to develop sports related functions and economy. This study identified the bottleneck factors constraining the development of sports economy. A fuzzy hierarchical evaluation model was developed using FAHP. The results indicated internal competitiveness, development of market-oriented economy and social development played important role to sports economy. Thus, policy recommendations on sustainable and sound sports economic development were proposed. The future prospect and development mode were analysed.

Keywords: sports economy, fuzzy analytic hierarchy process, evaluation model

1 Introduction

Sports economy has made great contributions to the development of domestic economy. With the development of domestic economy, its contributions have become increasingly prominent. Sports economy evaluation often requires integrated qualitative and quantitative analysis on a variety factors. Currently, the development of domestic sports economy has encountered certain bottleneck. Detailed study on various factors restricting the development of domestic sports economy will be helpful to give a clear and objective understanding on development of domestic sports economy. This study attempted to develop a hierarchy fuzzy evaluation model to analyse various constraints. Based on the results, recommendations on sustainable, rapid and robust development of domestic sports economy and its prospects in the future were proposed.

2 The meaning and positioning of domestic sports economy

Sports economy refers to rational allocation of various resources to develop sports related functions and economy. More simply, sports economy is a special industry integrated public sports and relevant economic behaviour [1]. The sports related industries are collectively known as sports service industry. The sports industry is mainly targeted to three pillar industries referring to development process in developed countries and actual conditions in China: competition performance, fitness and entertainment, and intangible assets. However, the sizes and scale of these three industries are limited, especially the intangible assets. Although China is a strong country

regarding sports, a few kinds of mass sports are popular, such as table tennis and badminton. In addition, the operation of such main industry is not standardized. Thus, the following goals are proposed based on the needs of sports economy development and actual situation:

2.1 TO IMPROVE ECONOMIC BENEFITS OF SPORTS INDUSTRY

Currently, the output of sports economy is not high and also the profits are low. Therefore, improvement of economic benefits is identified as primary goal. Sports economy can be promoted from three ways:

- 1) direct economic benefits, i.e. the sports venues, sports facilities and broadcast/TV right.
- 2) indirect economic benefits, i.e. the benefits from organization of large events and games, etc., as well as investments in associated infrastructures. Such as the Beijing Olympic Games hold in 2008, the investments in venues and related infrastructure were as high as 134 billion CNY. The total investment reached 280 billion CNY. Such investments have brought more than 600 billion CNY direct benefits.
- 3) Derivative economic benefits, such as advertising, lottery and football lottery. The benefits will be doubled during big events. For example, football lottery sales are about eight times than usual during 2014 World Cup. Sports itself is an industry integrated watching, information and economy. The growth of industry economy will be enhanced by promoting its social and economic benefits.

* *Corresponding author* e-mail: 1051153583@qq.com

2.2 TO ACCELERATE URBAN CONSTRUCTION AND CREATE JOB OPPORTUNITIES

According to the statistics, large events such as Olympic Games and World Cup would allow the host city to improve the infrastructures to 20-50 years in advance. In addition, the construction of venues and related facilities will create employment opportunities for thousands labor force, which will bring great significance for social benefits. According to statistics, Shanghai has more than 6000 fitness centres that providing a large number of jobs. There are nearly 200 sports industry relevant enterprises in Shenzhen. These enterprises not only create economic output but also provide a large number of jobs and contribute to social stability. Some world champions founded companies in his/her own name, such as Li Ning and Deng Yaping. These companies have promoted the brand name and also created a lot of jobs. At presents, China is in a critical period of developing a moderately prosperous society. It will have great significance to promote social harmony and stability if vigorously develop sports economy, accelerate urbanization and create a large number of job opportunities.

2.3 TO OPTIMIZE INDUSTRIAL STRUCTURE AND PROMOTE RAPID DEVELOPMENT OF TERTIARY INDUSTRY

The sports industry has experienced rapid growth with the development of domestic economy, especially the promotion of Beijing 2008 Olympic Games. The major part of sports industrial chain concentrates in the tertiary industry. Therefore, its robust and orderly development is important to optimize industrial structure and promote the tertiary industry. In addition, with the expansion of the tertiary industry, the related industries in the secondary and tertiary industry will also be promoted. The increasing awareness of sports and pursuit of health have greatly promoted the development of sports apparel and sports equipment producers. In addition to contributing to the domestic economic output, the development of sports economy would promote industrial structure and provide foundation for sustained economic growth.

3 Constraints to sports economy

In recent years, the domestic economy has been growing rapidly. The reasons can be summed up to two aspects. On the one hand, the sports consumption affordability improved constantly. With the sustained economic development and construction of the well-off society, residents' living has been improved continuously. Once physiological, security and other basic demands are satisfied, people start to pursue social interaction, self-realization, respects and other high-level demands. Meanwhile, with rising incomes, residents have more spare money that can be used for sports. According to relevant statistics, money spent on sports and sports

services have been increasing in recent years. On the other hand, the sports market in China consisting of competition performance, sports tourism, fitness and entertainment has gradually expanded to certain scale. However, there are various inevitable constraints and bottleneck during the development of sports economy. It is necessary to identify the factors restricting development of sports economy to find an appropriate model for future development.

3.1 INSTITUTIONAL CONSTRAINTS

In the past, sports events were entirely arranged by the National Sports Commission without any market-oriented mechanism. With the process of reform and opening – up, the constraints from conventional institution have been recognized [2]. The government attempted to transform from government behaviour to corporate behaviour gradually. However, it still not completely got rid of the long-term impact of planned economy. In addition, with the promotion of market-oriented economy, China has opened up a way to organize sports event jointly by corporation and sports institutions. However, the sports institutions were often in dominant position. Enterprises were hard to get adequate autonomy so that they were not very willing to be involved. Generally, enterprises only obtained few profit after the events. Usually enterprises participated in sports events for two purposes:

- 1) to enhance visibility and social benefits;
- 2) economic benefits. However, due to long term institutional constraints, enterprises were not very positive to invest in sports.

3.2 LEGISLATION CONSTRAINTS AND LACK OF POLICY SUPPORT

Currently, there is few legislation regarding sports market and lack of authorization legislation. The management authority, role and responsibility are not clearly defined. It is found that in Britain, the United States and other developed countries the governments often give proper policy supports to develop sports economy, such as certain tax relief for the revenue from sports event and operation of stadium, preferential policies for land acquisition and low-interest loans for stadiums etc. Such supports in China are far from enough. The sports industry has suffered “cold reception” in the preferential policies, which also caused enterprises reluctant to enter into sports industry.

3.3 SPORTS MARKETS STILL IN THE INITIAL STAGE

The response speed of sports market and ability to create demand are highly related with the needs of the market. Currently, people's consumption concept on sports has not been completely converted, which mainly presented in limited consumption capacity and irrational consumption structure. In addition, there is lack of high-quality sports management talent. The irrational knowledge structure

and incomprehensive capability of the sports management staff have severely restricted the development sports market so that the sports market in China is still at initial stage.

4 FAHP model for sports economy development mode

The FAHP model has been introduced to sports economics due to its strong applicability. For the evaluation of socio-economic system, integrated qualitative and quantitative FAHP is required in the case of many subjective qualitative indicators. For sports economics, single indicator evaluation can be used to identify strengths and weaknesses and support decision-making. AHP and FAHP can be used for overall comprehensive evaluation. The qualitative factors are described by quantitative means and then transformed back to qualitative factors through fuzzy comprehensive evaluation so that qualitative factors and quantitative factors are fully integrated and the subjective factors are excluded during this process.

4.1 AHP MODEL

4.1.1 Features of analytic hierarchy process

The Analytic Hierarchy Process (AHP) is a structured technique developed in 1970s by T.L. Saaty, an American operational researcher. This technique has the capability to integrate qualitative analysis and quantitative analysis effectively. The procedure for using the AHP can be summarized as five steps. Step 1: propose problem; step 2: establish hierarchy analysis model; step 3: develop judgment matrix; step 4: establish priorities among the elements of the hierarchy; step 5: yield a set of overall priorities for the hierarchy. The comprehensive evaluation value can be generated through calculating the weight of each element of each hierarchy to the overall goal thereby providing basis for alternative selection.

AHP is useful to make effective, reliable and rational decision and it is applicable to wide field, such as resources allocation, alternative selection and evaluation.

4.1.2 Establish hierocracy model and judgment matrix

A multi-stage hybrid hierarchy structure is usually adopted for sports economy evaluation. It is necessary to note that the elements selected should be independent. The

complex problem is decomposed into several sub-problems comprised of different elements which can be further broken down so that the hierarchy is built.

The judgment matrix is the basic information of AHP, which is also an important basis for calculation of weight. In this step, the experts and researchers invited need to repeatedly answer the question: which one is more important to achieve the goal b_i or b_j ? How important is it (using 1-9 to measure the importance)? Then, a pairwise comparisons matrix containing n elements for hierarchy B is established.

4.1.3 Determine the weight of each element and the maximum eigenvector of judgment matrix

1) Determine relative weight using uniform criteria:

$$W_i = \frac{\left(\prod_{j=1}^n b_{ij}\right)^{\frac{1}{n}}}{\sum_{k=1}^n \left(\prod_{j=1}^n b_{kj}\right)^{\frac{1}{n}}}, (i = 1,2,3,\dots,n);$$

eigenvector $W = [W_1, W_2, W_3, \dots, W_n]^T$;

2) Calculate the maximum eigenvector of the judgment matrix:

$$\lambda_{\max} = \frac{1}{n} \sum_{j=1}^n \frac{\sum_{i=1}^n b_{ij} W_j}{W_i}$$

4.1.4 Test consistency of judgment matrix

The consistency index CI is used to test the judgment matrix:

$$CI = \frac{\lambda_{\max} - n}{n - 1}, CR = \frac{CI}{RI}$$

When $CI = 0$, the judgment matrix is fully consistent; when $CI < 0.1$, the consistence is at acceptable level; when $CI \geq 0.1$, (RI is the average random consistency index of which the values are given in Table 1), it needs to adjust the judgment matrix until satisfactory.

TABLE 1 RI matrix from 1to 9 order

order	1	2	3	4	5	6	7	8	9
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45

4.2 FAHP MODEL

The AHP is straightforward that the judgment matrix does not consider fuzziness. However, fuzzy judgment is always used unconsciously when solve complex decision problem. Therefore, Fuzzy Analytic Hierarchy Process

(FAHP) has been developed to apply AHP in fuzzy conditions.

When comparing two elements, quantitative importance is give to both elements, i.e. represented by membership function A . The fuzzy judgment matrix $A = (a_{ij})_{n \times n}$ has the following two properties:

$$a_{ii} = 0.5, \forall i \in N,$$

$$a_{ij} + a_{ji} = 1, a_{ij} \geq 0, \forall i, j \in N (i \neq j).$$

Such judgment matrix is called fuzzy complementary judgment matrix. For the fuzzy complementary judgment matrix, the weight is determined by:

$$W_i = \frac{\sum_{j=1}^n a_{ij} + \frac{n}{2} - 1}{n(n-1)}, i \in N.$$

Similarly, it needs to test the consistency of the weight values determined. The compatibility of the fuzzy judgment matrix is used to test the consistency. The procedures are as follows:

4.2.1 Determine characteristic matrix for fuzzy judgment matrix A

The weight vector of fuzzy judgment matrix A is

$$W = [W_1, W_2, W_3, \dots, W_n]^T \text{ and } \sum_{i=1}^n W_i = 1, W \geq 0 (i \in N).$$

The element of characteristic matrix is:

$$W_{ij} = \frac{W_i}{W_i + W_j}, \forall i, j \in n.$$

Then the characteristic matrix of judgment matrix A can be presented as $W = (W_{ij})_{n \times n}$.

4.2.2 Test consistency of fuzzy complementary judgment matrix and characteristic matrix

The attitude of decision-maker is denoted by α . When

$$I(A, W) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |a_{ij} + W_{ji} - 1| \leq \alpha.$$

The consistency of judgment matrix is regarded as satisfactory. The smaller α indicates the decision maker has high requirement on consistency of fuzzy judgment matrix. Usually α is taken as 0.1.

Establish a Multi-dimensional Sports Economy Development Model.

Based on the identification of bottleneck restricting development of domestic sports economy and results of FAHP evaluation, a multi-dimensional model is established (Figure 1).

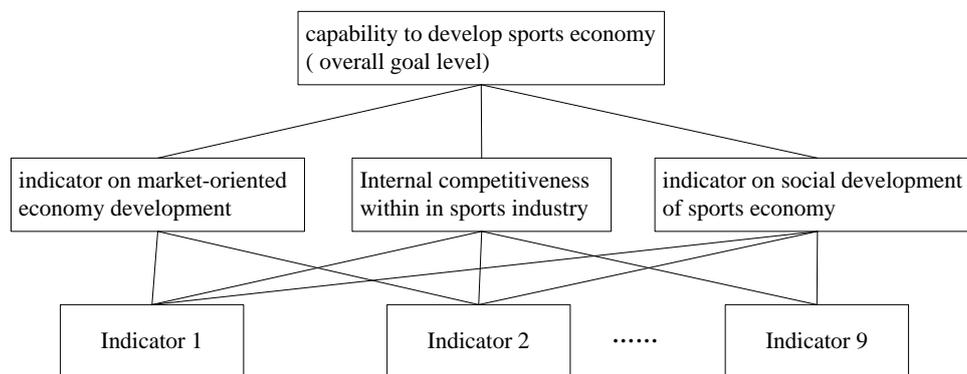


FIGURE 1 Multi-dimensional model on development of sports economy

For the middle hierarchy (criteria level), these three indicators are pairwise compared for preference by experts to construct judgment matrix for the evaluation system of sports economy development:

$$U = \begin{bmatrix} 1 & \frac{1}{5} & 3 \\ 5 & 1 & 7 \\ \frac{1}{3} & \frac{1}{7} & 1 \end{bmatrix}.$$

The corresponding weights of each element are: [0.1884, 0.7307, 0.0808]. The consistency of weights is tested: $\lambda_{max} = 3.0649$; $CI = 0.03245$; $CR = 0.056 < 0.1$, indicating the consistency is satisfactory.

Using FAHP the fuzzy judgment matrix is determined as:

$$U = \begin{bmatrix} 0.5 & 0.3 & 0.6 \\ 0.7 & 0.5 & 0.8 \\ 0.4 & 0.2 & 0.5 \end{bmatrix}.$$

The corresponding weights of each element are: [0.3167, 0.4167, 0.2666]. The consistency index $I = 0.0842 < 0.1$, concluding the consistency is satisfactory.

It is shown that internal competitiveness is the most critical factor to sports economy development, followed by market-oriented economy and social development. The results from FAHP are satisfactory and comply with actual situation.

5 Recommendations on policy and development model for domestic sports economy

It is concluded from the FAHP evaluation that internal competitiveness is the most important factor to development of sports economy. Therefore, it is critical to establish regulated market. In view of this, recommendations on policy and model selections are proposed.

5.1 RECOMMENDATIONS ON POLICY FOR DEVELOPMENT OF DOMESTIC SPORTS ECONOMY

5.1.1 Establish and improve socialistic market system

China, a big country with 1.3 billion populations, has huge demands on sports and also is a huge market. With the development of reform and opening-up, living conditions and incomes have been greatly improved. The consumption habits were also changed. There is increasing demands for fitness, leisure, sports and entertainment [4]. People are willing to spend more time and money in physical exercise to pursuit for health. National wide fitness programs have reached a climax and brought more business opportunities. Together with more commercialized competition sports, these will promote the prosperity of sports goods market [3].

The sports market in China like a huge “cake” that many other counties covet. Since joining WTO, some international brands have entered into Chinese market. To establish and improve socialistic market system, as well as strengthen domestic brands will not only further motivate sports market but also promote the prosperity of domestic sports goods market.

5.1.2 Strengthen marketization of sports industry

At present, China’s sports economy has broken the limits of original planned economy. However, the marketization is not completed. Thus, it is essential to accelerate the pace of sports innovation and strengthen market-oriented sports industry. The sports institutions at each level should strengthen cooperation with enterprises to maximize the effectiveness of diverse sports. Most sports institutions are public institutions. With the promotion of public institution reform program in 2014, the sports institutions are encouraged to transfer to enterprise if possible and accelerate the pace of industrialization. However, this is a long term and systematic project that has to be phased. First stage: from full funding to balance allocation. Second stage: from balance allocation to self-supported [5]. Third stage: from self-supported to enterprise. Through the staged reform the sports institutions could find multi funding source and don’t need to rely on government allocations. The sports could cooperate with strong enterprises to jointly promote sports industry and market-oriented process so that the sports institutions and enterprises are complementary with each other.

5.1.3 Enhance policy support

Sports industry is a kind of quasi-public goods, i.e. the public and private properties coexist in sports industry. The government cannot ignore the public welfare when vigorously promoting development of sports economy, while private enterprises often only pursue for profits. Sports industry usually benefits “externalities” so that private enterprises are not interested in. Therefore, the government is required to afford the input in sports infrastructures and perform management responsibilities. At present, the sports market system in China is mainly for the purpose of cultivating various sports markets. On the one hand, sports goods, talent and capital are three basic elements for sports economy. These three elements interact with each other and have become three pillars. On the other hand, sports technology, property and equipment ensure material and services supplies for these three pillar industries. Therefore, it is critical to provide appropriate basic supporting industries timely. In addition, reasonable and orderly development of sports market can promote consumption and also create a large number of job opportunities thereby making great contributions to social benefits. Therefore, the government needs to actively provide guidance and supports to the development of sports industry and establish sound regulation to promote diverse development especially. In particular, the industry reform is still at initial stage. Many enterprises are still waiting to see the further prospect. The relevant governments should give certain supports, such as tax incentives, to attract more substantial enterprises and promote sustainable development jointly.

5.2 FUTURE DEVELOPMENT MODEL FOR DOMESTIC SPORTS ECONOMY

Phelan Gary, the general secretary of World Tourism Organization predicted that China would become the largest tourist destination in 2015. The authority considered tourism would become a pillar industry and new growth point. Currently, sports tourism has become a hot industry. Sports tourism integrates tourism and sports industry that sports resources and tourism resources complementary and mutually beneficial with each other [6]. Tourism and sports industry enjoy similar social and cultural background and have similar product features. Sports tourism contains various functions, including leisure, fitness, entertainment, sightseeing and participation.

Sports economy, making use of both sports resources and tourism resources, will become new economic growth point. It can improve living, promote inheritance and propagation of modern civilization and also drive other related industries and create more employment opportunities, thereby improve regional economy.

Reference

- [1] He G 2012 Research on the evaluation of the coordinated development between china regional mass sports and economy *Journal of Beijing Sport University* **35**(6) 23-7 (in Chinese)
- [2] Lei X, Li S, Luan F, et al 2012 Application of improved AHP in evaluation on sustainable sports industry *Journal of Shandong Institute of Physical Education and Sports* **4** 12-3 (in Chinese)
- [3] Yu C, Liu Z, Ding H, et al 2001 Research on evaluation indexes of sustainable development and human resources *Journal of Xi'an Institute of Physical Education* **18** 24-5 (in Chinese)
- [4] Li J, Guo L 2012 Research on sports industry model China *Sport Science and Technology* **3** 47-8 (in Chinese)
- [5] Hu X 2013 The breakthrough of Chinese sport reform: changing the institutional bottleneck of the transformation of the mode of sports development *Journal of Sports and Science* **01** 26-7 (in Chinese)
- [6] Yu L 2013 Public Service supply mode transformation and realistic choice since 30 years of reform and opening-up in China *China Sports Science* **33**(2)11-21 (in Chinese)

Authors	
	<p>Tao Shi, born on December 25, 1978, Handan city, Hebei province, China</p> <p>Current position, grades: lecturer of Institute of Physical Education, Handan College, Handan, China. Scientific interest: physical education and sports coaching. Publications: more than 3 papers. Experience: teaching experience of 8 years, 3 scientific research projects.</p>
	<p>Zhijie Xing, born on April 20, 1980, Handan city, Hebei province, China</p> <p>Current position, grades: senior teacher of Hanshan District Experimental Primary School, Handan, China. Scientific interest: physical education and sports coaching. Publications: more than 1 paper. Experience: teaching experience of 7 years, 2 scientific research projects.</p>

A study on the determinants of e-commerce customer satisfaction

Xinling Du^{1, 2}, Lingyi Meng^{1*}, Yujian Wu¹

¹*School of Management, Beijing Normal University Zhuhai Campus, Zhuhai 519087, China*

²*School of Humanities and Economic Management, China University of Geosciences (Beijing), Beijing 100083, China*

Received 6 June 2014, www.tsi.lv

Abstract

Based on existing research, a model and evaluation index systems are developed for assessing e-commerce satisfaction of customers. We centralizes associated questionnaires for six topics, namely, convenient operation for online shopping, product information, distribution service, safety and reliability of the system, handling of customer complaints, and staff services, that are administered among college students in Guangdong Province of China who have experienced purchasing items online. The proposed model explores the factors that affect customer satisfaction. The effects of these factors are then analysed using descriptive statistics and factor analysis method to verify the accuracy of the model and to draw Safety and reliability of the system has the highest correlation coefficient with customer satisfaction, different gender and grade factors significantly affect customer satisfaction.

Keywords: e-commerce, customer satisfaction, factor analysis model, evaluation index system

1 Introduction

With its new web-based economic role, e-commerce has become a new part of the lives of the people. The improvement of customer satisfaction to promote the development of e-commerce poses a challenge and has a higher requirement for corporate marketing. The number of Internet users and online shoppers increased to 564 and 242 million, respectively, by the end of 2012 [1]. Online transactions in China have rapidly amounted to 1.304 trillion Yuan [2]. Online shoppers are primarily composed of young people, particularly college students and white collar workers. Students have become the main force and the potential consumers of online stores. Therefore, the online shopping behaviours of college students must be investigated, and the factors that influence their satisfaction must be determined. This paper selects six dimensions with 30 factors that may affect the E-Commerce Customer Satisfaction.

2 Summary of customer satisfaction

Cardozo (1965) introduced the concept of customer satisfaction in the marketing field and argued that the willingness of customers to purchase more products in the future could be easily predicted if their satisfaction is positive [3]. Westbrook (1981) described satisfaction as an emotional state that could evaluate the interaction between customers and staff members [4]. Oliver (1981) defined customer satisfaction as the psychological reaction that customers feel whenever they evaluate the realization of a service against a certain standard [5]. Tse and Wilton

(1988) argued that customer satisfaction arises from the gap between the expectations and the realizations of customers [6]. Fornell (1992) defined customer satisfaction as the general evaluation of a purchased product or service [7]. Athanassopoulos (2000) defined customer satisfaction as satisfying the expectations of customers [8]. Woodruff (1997) defined customer satisfaction as the result of comparing product attributes with product performance and the expectations of the customer [9]. This paper discusses the factors that affect college students' online shopping satisfaction.

3 E-commerce customer satisfaction evaluation index system and measurement model

3.1 E-COMMERCE CUSTOMER SATISFACTION MEASUREMENT MODEL

This paper generates an e-commerce customer satisfaction model (Figure 1) based on the customer satisfaction theory, the Swedish Customer Satisfaction Index Model (SCSB) [10], the American Customer Satisfaction Index Model (ACSI) [11], the European Customer Satisfaction Index Model (ECSI) [12], and other scientific customer satisfaction index models that can be integrated with the actual situation in China. The following hypotheses are proposed based on the model illustrated in Figure 1:

* *Corresponding author* e-mail: 15919151886@163.com

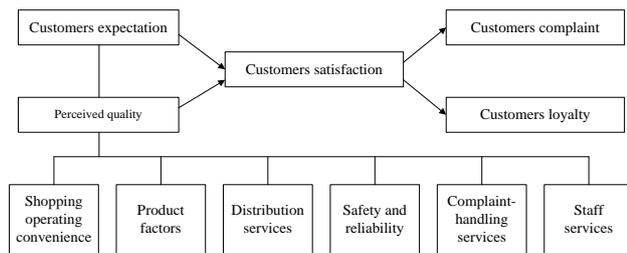


FIGURE 1 Online shopping satisfaction measurement model

- H1: Shopping operating convenience has a significant effect on online shopping satisfaction;
- H2: Product factors have a significant effect on online shopping satisfaction;
- H3: Distribution services have a significant effect on online shopping satisfaction;
- H4: Safety and reliability of the system have a significant effect on online shopping satisfaction;
- H5: Complaint-handling services have a significant effect on online shopping satisfaction;
- H6: Staff Services have a significant impact for online shopping satisfaction;
- H7: Gender has no significant effect on online shopping satisfaction; and
- H8: Grades have no significant effect on online shopping satisfaction.

3.2 BUILD E-COMMERCE CUSTOMER SATISFACTION EVALUATION INDEX SYSTEM

Based on the e-commerce customer satisfaction model, this paper selects six dimensions with 30 factors that may affect the service quality of the e-commerce trading platform. These factors include online shopping environment, friendly interface and definition, reaction rate of interface, shopping convenience, comprehensive of the search function, product categories, product cost and performance, product quality, detailed instructions of products, product price, updating of product information, detailed description of products, consistent level of products, promotion measures and efforts, logistics, timeliness of product distribution, accuracy of delivery location, stability of website, safety of the website, safety of transaction process, convenience of online payments, security on the privacy of customers, fulfilment of commitments, returns rate of goods and refunds, valuation of customer perception, timeliness of handling complaints, after-sales services, patency of customer platform, communication skills of staff members, and attitude of staff members. These factors are denoted with the variables to and are arranged into six dimensions based on the following principles:

1) Operating convenience, which denotes how customers can easily shop on the e-commerce trading platform.

2) Product factors from which the customers can determine the price, classification, description, and other basic information of a specific product.

3) Distribution services, such as logistics, timeliness of distribution, and other services.

4) Safety and reliability of the system, which include on the privacy of the customer, safety and convenience of payment, and other issues that are related to how customers use the e-commerce trading platform to purchase products online.

5) Complaint-handling services, which focus on the timely response to customer reviews and complaints.

6) Staff services, which refers to the attitude of staff members when serving their customers and their ability to communicate with their customers.

4 Questionnaire design and data collection

4.1 QUESTIONNAIRE DESIGN

Following the evaluation index system, the questionnaire is divided into two parts. The first part collects the basic information of customers (i.e., college students), such as their gender, grade, and major, to determine their background and for the conduct of univariate analysis. The second part analyses the comments on and the satisfaction of students with online shopping. Using a five-point Likert (1932) scale, [13] the online shopping satisfaction of customers is rated on a scale ranging from 5 to 1, which respectively denotes completely satisfactory, satisfactory, average, dissatisfactory, and completely dissatisfactory.

4.2 DATA COLLECTION

College students in Guangdong Province of China with online shopping experience are selected for the survey. A total of 400 questionnaires are distributed, of which 373 are considered valid questionnaires. A total of 100 e-questionnaires are sent via e-mail to students from nine colleges, and 94 valid e-questionnaires are returned to the researchers. A total of 467 valid questionnaires are collected and a 93.4% response rate is achieved.

5 Empirical research and statistical analysis of the e-commerce customer satisfaction evaluation index system

5.1 DESCRIPTIVE STATISTICAL ANALYSIS

Descriptive statistical analysis is performed to analyse the gender, grade, and other variables of each respondent as well as those of the entire sample. A total of 467 students have participated in the survey, and the ratio of males to females is approximately 4:6. Given the online shopping experience of the participants who range from freshman to senior students, they can clearly evaluate their online shopping satisfaction. The sample is primarily composed of junior students, who account for 37.9% of the entire sample. Seniors account for 26.8% of the sample, while the freshmen and sophomores have a relatively small ratio in the entire sample.

The minimum and maximum values of the index are set to 1 and 5, respectively, which demonstrate the wide distribution of the data. The mean value of the variables is between 3.0 and 4.0, with only three variables having values that are less than 3.0. This observation demonstrates that the respondents are generally satisfied with the factors in the scale and that the sample structure is ideal. The standard deviation is one of the most common quantized forms that can reflect the discreteness level and accuracy of the data. The standard deviation of the survey data ranges from 0.6 to 1.1, which indicates that the survey data have a certain discrepancy and a relative discreteness. In accordance with our expectations, these results also show that the satisfaction of the respondents depends on various factors. Corresponding to the measure metrics, the mean value of overall customer satisfaction is 3.46, which indicates the satisfaction of the college students with their previous online shopping experience. However, this result does not indicate that the students are very satisfied with their online shopping experience. This limited degree of satisfaction may be attributed to several factors present during online shopping.

5.2 RELIABILITY ANALYSIS AND VALIDITY ANALYSIS

Reliability represents the consistency or stability of the scale as well as serves as an index for homogeneity testing. Cronbach’s alpha coefficient is used in this paper to test scale reliability. The Cronbach’s alpha coefficient is computed as follows:

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K \sigma_{Y_i}^2}{\sigma_x^2} \right), \tag{1}$$

where σ_x^2 is the variance for the total sample and $\sigma_{Y_i}^2$ is the variance for the collected sample.

The statistical analysis software SPSS19.0 is used to analyse the reliability of the data. Table 1 shows the reliability of the entire scale and of each variable. α of the entire scale is 0.898, which indicates a favourable reliability by exceeding the 0.8 reliability threshold. α of each variable in the scale is greater than 0.7, which denotes the high reliability of the survey.

TABLE 1 Questionnaire reliability test

	N of Items	Cronbach's Alpha
Shopping operating convenience	5	0.743
Product factors	9	0.785
Distribution services	3	0.715
Safety and reliability of the system	6	0.782
Complaints services	4	0.738
Staff services	3	0.728
Total	30	0.898

Reliability and validity are two aspects of the questionnaire evaluation. Validity refers to the extent to which the tools can measure the psychological or behavioural traits of the measured object. A high validity

suggests that the results can reflect the characteristics of the measured object more accurately. Validity is divided into construct validity, criterion-related validity, and content validity. Given that the criterion-related and content validities require expert qualitative research validity or an accepted criterion measure, they are difficult to be applied in practice. Therefore, the validity of the questionnaire is measured in this study by construct validity and factor analysis. The Kaiser–Meyer–Olkin (KMO) test and the Bartlett test of sphericity are performed to examine the suitability of the factor analysis for the sample data. The KMO value ranges from 0 to 1. A KMO value that is close to 1 indicates a large amount of common factors among variables and confirms the suitability of factor analysis for the data. According to Kaiser (1974), the KMO value should be larger than 0.6 before performing a factor analysis [14]. As shown in Table 2, the KMO value for this study is 0.844, which is larger than 0.6. Moreover, the approximate chi-square distribution of the Bartlett's test of sphericity is 5590.500 (df is 435), which is significant at a level of 0.000. Therefore, factor analysis is considered suitable for the sample data.

TABLE 2 Results of the KMO test and the Bartlett’s test of sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.844
Bartlett's Test of Sphericity	Approx. Chi-Square	5590.5
	df.	435
	Sig.	0.000

5.3 FACTOR ANALYSIS MODEL AND APPLICATION

Factor analysis is a statistical method that divides variables into several groups based on the strength of their correlation. Variables within the same group have a high level of correlation, while the correlation of variables between groups is lower. Each group of variables represents a basic structure that is called a common factor.

The e-commerce satisfaction of customers is discussed in this paper. Each observed component is examined using the sum of a specific factor and the linear function with the unobservable factor of the least number.

Given n samples that observe p variables, the observed data are standardized by using the equation $zx_i = (x_i - \bar{x}_i) / s_i$, where zx_i is a standard index value, x_i is the initial index value, \bar{x}_i is a mean value, s_i is a standard deviation, and zx_i is the new standard variable.

$zx = (zx_1, zx_2, zx_3, \dots, zx_p)'$ is an observable random variable, whose covariance matrix is the same as its correlation matrix. $F = (F_1, F_2, F_3, \dots, F_m)'$ is an unobservable vector where $m < p$, F_i and F_j are independent of $i \neq j$. $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_p)'$ is also independent of F , ε_i and ε_j and of $i \neq j$. Therefore, the factor analysis model is constructed as follow:

$$zx_i = \alpha_{ij}F_j + \varepsilon_i, (i = 1, 2, 3, \dots, p; j = 1, 2, 3, \dots, m). \quad (2)$$

This equation is expanded further as follow:

$$\begin{cases} zx_1 = \alpha_{11}F_1 + \alpha_{12}F_2 + \dots + \alpha_{1m}F_m + \varepsilon_1 \\ zx_2 = \alpha_{21}F_1 + \alpha_{22}F_2 + \dots + \alpha_{2m}F_m + \varepsilon_2 \\ \dots \\ zx_p = \alpha_{p1}F_1 + \alpha_{p2}F_2 + \dots + \alpha_{pm}F_m + \varepsilon_p \end{cases} \quad (3)$$

The matrix of this factor analysis model can be described as follow:

$$zx = AF + \varepsilon, \quad (4)$$

where $zx = (zx_1, zx_2, zx_3, \dots, zx_p)'$, $F = (F_1, F_2, F_3, \dots, F_m)'$,

$$A = \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{p1} & \dots & \alpha_{pm} \end{pmatrix}.$$

Here $F_1, F_2, F_3, \dots, F_m$ are called latent factors, and $m < p$, $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_p$ are called the specific factors or the special factors of zx_i as the component of zx . Matrix A is the factor loading matrix. Each α_{ij} in matrix $A = (\alpha_{ij})$ is called a factor loading, which is the correlation coefficient of zx_i and F_j .

Table 3 shows the cumulative variance devoting rates and the rotated loading matrix that is obtained by varimax rotation on all variables. Following the principle that the eigenvalues must be greater than 1, we select six common factors, and the sum of their cumulative variance devoting rates amounts to 55.67%. Factor analysis is then conducted based on the factor loading and the correlation coefficient of zx_i and F_j . A larger correlation coefficient indicates a larger amount of observable common traits between two variables, with zx_i as the most powerful coefficient. Table 3 shows all the correlation coefficients between the common factors that are drawn from the customer satisfaction evaluation indicators. These indicators are all higher than 0.5, which indicates that the six factors have a favourable explanatory index. That is, the six common factors, namely, shopping convenience, product factors, distribution services, safety and reliability of the system, complaint-handling services, and staff services, which are represented as $F_1, F_2, F_3, F_4, F_5, F_6$, can sufficiently explain the evaluation indicators, which is verified by the factor analysis.

TABLE 3 Factor loadings of the perception of customer satisfaction evaluation

	Component
	1
F1 Shopping operating convenience	0.711
F2 Product factors	0.574
F3 Distribution services	0.638
F4 Safety and reliability of the system	0.824
F5 Complaint-handling services	0.760
F6 Staff services	0.791

Extraction method: Principal component analysis. A) One extracted component.

Table 4 shows that the factor loadings of F_1 on zx_1 , zx_2 , zx_3 , zx_4 and zx_5 are 0.501, 0.592, 0.742, 0.788, and 0.409, respectively, which are also much greater than those of other indicators. F_1 reflects the meaning of these five indicators synthetically and is called "shopping operating convenience". The factor loadings of F_2 on zx_6 , zx_7 , zx_8 , zx_9 , zx_{10} , zx_{11} , zx_{12} , zx_{13} and zx_{14} are 0.428, 0.603, 0.579, 0.570, 0.485, 0.598, 0.609, 0.704, and 0.451, respectively, which are also much greater than those of other indicators. F_2 reflects the meaning of these nine indicators synthetically and is called "product factors". The factor loadings of F_3 on zx_{15} , zx_{16} and zx_{17} are 0.834, 0.760, and 0.557, respectively, which are also much greater than those of other indicators. F_3 reflects the meaning of these three indicators synthetically and is named as "distribution services". The factor loadings of F_4 on zx_{18} , zx_{19} , zx_{20} , zx_{21} and zx_{22} are 0.584, 0.760, 0.731, 0.421, and 0.512, respectively, which are much greater than those of other indicators. F_4 reflects the meaning of these five indicators synthetically and is called "safety and reliability of the system". The factor loadings of F_5 on zx_{23} , zx_{24} , zx_{25} , zx_{26} , zx_{27} and zx_{28} are 0.616, 0.673, 0.745, 0.684, 0.692, and 0.585, respectively, which are much higher than those of other indicators. F_5 reflects the meaning of these six indicators synthetically and is called "complaint-handling services". The factor loadings of F_6 on zx_{29} and zx_{30} are 0.534 and 0.537, respectively, which are also much greater than those of other indicators. F_6 reflects the meaning of these two indicators synthetically and is named as "staff services".

With the effects of specific factors disregarded, the factor analysis model can be expressed as follows:

$$\begin{cases} zx_1 = 0.501F_1 + 0.089F_2 - 0.115F_3 + 0.201F_4 + 0.200F_5 + 0.384F_6 \\ zx_2 = 0.592F_1 + 0.069F_2 + 0.070F_3 + 0.097F_4 + 0.308F_5 + 0.201F_6 \\ \dots \\ zx_{30} = -0.097F_1 + 0.005F_2 + 0.295F_3 + 0.242F_4 + 0.490F_5 + 0.537F_6 \end{cases} \quad (5)$$

TABLE 4 Factor loading matrix and the cumulative variance contribution rate

	Component					
	F ₁	F ₂	F ₃	F ₄	F ₅	F ₆
z _{x1} Online shopping environment	0.501	0.089	-0.115	0.201	0.200	0.384
z _{x2} Friendly interface and definition	0.592	0.069	0.070	0.097	0.308	0.201
z _{x3} Reaction rate of interface	0.742	0.124	0.093	0.184	0.150	-0.070
z _{x4} shopping Convenience	0.788	0.054	0.126	0.149	0.047	-0.052
z _{x5} Comprehensiveness of the search function	0.409	0.008	0.189	0.326	0.104	0.349
z _{x6} Product categories	0.352	0.428	0.099	0.191	-0.202	0.512
z _{x7} Product cost-performance	0.249	0.603	-0.021	0.234	0.012	0.418
z _{x8} Product quality	0.143	0.579	-0.060	0.118	0.335	0.209
z _{x9} Detailed instructions of products	0.250	0.570	0.085	0.216	0.204	0.178
z _{x10} Product price	0.181	0.485	-0.050	0.086	-0.170	-0.157
z _{x11} Updating rate of product information	-0.122	0.598	0.168	0.053	0.035	-0.059
z _{x12} Detailed description of products	-0.023	0.609	-0.017	0.041	-0.004	-0.120
z _{x13} Consistency of products	0.107	0.704	0.116	-0.081	0.035	0.112
z _{x14} Promotion measures and efforts	-0.061	0.451	0.172	0.118	0.315	0.181
z _{x15} Logistics	0.164	0.154	0.834	0.031	0.182	0.047
z _{x16} Timeliness of product distribution	0.158	0.146	0.760	0.107	0.247	-0.037
z _{x17} Accuracy of delivery location	-0.100	-0.044	0.557	0.440	0.013	0.315
z _{x18} Stability of the website	0.256	0.104	0.219	0.584	0.157	0.165
z _{x19} Safety of the website	0.263	0.117	0.008	0.760	0.291	0.075
z _{x20} Safety of transaction process	0.218	0.117	0.032	0.731	0.250	0.120
z _{x21} Convenience of the online payment	0.365	0.114	0.043	0.421	0.018	0.165
z _{x22} Security on the privacy of customers	-0.055	0.254	0.177	0.512	0.411	-0.256
z _{x23} Fulfillment of commitments	0.164	0.160	-0.021	0.328	0.616	-0.065
z _{x24} Speed of the return of goods and refund	0.149	0.035	0.103	0.142	0.673	0.008
z _{x25} Valuation of customer perception	0.087	0.024	-0.012	0.082	0.745	-0.139
z _{x26} Timeliness of handling complaints	-0.018	0.014	0.104	0.106	0.684	0.172
z _{x27} After-sales services	0.172	0.043	0.136	0.120	0.692	0.094
z _{x28} Potency of customer platform	0.223	0.059	0.362	-0.001	0.585	0.177
z _{x29} Communication skills of staff members	0.095	-0.014	0.124	0.132	0.529	0.534
z _{x30} Attitude of staff members	-0.097	0.005	0.295	0.242	0.490	0.537
Eigenvalues	7.994	2.760	1.972	1.569	1.235	1.171
% of Variance	26.647	9.200	6.575	5.229	4.116	3.903
Cumulative %	26.647	35.847	42.422	47.651	51.767	55.670

5.4 CORRELATION ANALYSIS

The Pearson correlation coefficient and the two-tailed test analysis are conducted in this study to examine the correlation of each factor with customer satisfaction. The results are shown in Table 5.

TABLE 5 Correlations between influence factors and customer satisfaction

		Customer satisfaction
F1 Shopping operating convenience	Pearson Correlation	0.714**
	Sig. (2-tailed)	0.000
	N	467
F2 Product factors	Pearson Correlation	0.701**
	Sig. (2-tailed)	0.000
	N	467
F3 Distribution services	Pearson Correlation	0.608**
	Sig. (2-tailed)	0.000
	N	467
F4 Safety and reliability of the system	Pearson Correlation	0.797**
	Sig. (2-tailed)	0.000
	N	467
F5 Complaints service	Pearson Correlation	0.753**
	Sig. (2-tailed)	0.000
	N	467
F6 Staff services	Pearson Correlation	0.670**
	Sig. (2-tailed)	0.000
	N	467

** . Correlation is significant at the 0.01 level (two-tailed).

As can be seen in Table 5, shopping operating convenience, product factors, distribution services, safety and reliability of the system, complaint -handling services, and staff services are positively correlated with customer satisfaction at significant level of 0.01.

The factor on safety and reliability of the system has the largest correlation coefficient with customer satisfaction of 0.830. This value indicates the presence of a strong linear relationship between these two factors as well as demonstrates that safety and reliability of the system has the most powerful influence on customer satisfaction. Therefore, H4 is verified.

The correlation coefficients of shopping convenience, product factors, complaint-handling services, and staff services with customer satisfaction are approximately 0.7, which exceeds the significance level. Therefore, H1, H2, H5, and H6 are verified.

Lastly, the correlation coefficient between distribution services and customer satisfaction is 0.608, which indicates a positive correlation at a relatively significant level. Therefore, H3 is verified.

5.5 ANALYSIS OF VARIANCE (ANOVA)

ANOVA is performed to examine whether a control variable has a significant effect on the observed variables. Gender and grade are used as control variables in this paper.

TABLE 6 ANOVA results of different genders satisfaction

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.356	1	2.356	13.398	0.000
Within Groups	81.783	465	0.176		
Total	84.139	466			

Table 6 shows the ANOVA result of gender on customer satisfaction. The sum of square deviation of gender from customer satisfaction is 84.139. When only the effect of gender is considered, gender can explain 2.356 of the variance within the sum of square deviation, and the 81.783 variance is caused by the sampling. The mean square errors of these variances are 2.356 and 0.176, respectively. Correspondingly, the F statistic is 13.398 and the P value is 0. The null hypothesis H7 is rejected when the significance level α is 0.05 and when the P value is less than the significance level α . Therefore, gender has a significant effect on customer satisfaction. As shown in Table 7, the ANOVA result of grade on customer satisfaction is similar to that of gender on customer satisfaction. The sum of square deviation of the grade is 84.139. When only the effect of the grade is considered, 2.453 of the variance can be explained within the sum of square deviation, while the variance of 81.687 is caused by the sampling. The mean square errors of these variances are 0.818 and 0.176, respectively. Accordingly, the F statistic is 4.634 and the corresponding P value is 0.003. Given that the significance level α is 0.05, the null hypothesis H8 is rejected. Therefore, customer satisfaction is highly affected by grade.

TABLE 7 ANOVA results of different grades satisfaction

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.453	3	0.818	4.634	0.003
Within Groups	81.687	463	0.176		
Total	84.139	466			

References

- [1] China Internet Network Information Center (CNNIC) 2013 *33th Statistical Report on Internet Development in China*. http://www.cnnic.net.cn/hlwfzjy/hlwxzbg/hlwtjbg/201301/t20130115_38508.htm / 29 May 2014
- [2] iResearch 2013 *The Scale of market transactions is more than 130 million* <http://ec.iresearch.cn/shopping/20130125/192011.shtml> / 1 June 2014
- [3] Cardozo R N 1965 An Experimental Study of Consumer Effort, Expectation and Satisfaction *Journal of Marketing Research* **2**(3) 244-249
- [4] Westbrook R A 1981 Sources of Consumer Satisfaction with Retail Outlets *Journal of Retailing* **57**(3) 68-85
- [5] Oliver R L, Linda G 1981 Effect of Satisfaction and its Antecedents on Consumer Preference and Intention *Advances in consumer research* **8**(1) 88-93
- [6] Tse DK, Wilton P C 1988 Models of Consumer Satisfaction Formation: An Extension *Journal of Marketing Research* **25**(2) 204-12
- [7] Fornell C 1992 A National Customer Satisfaction Barometer: The Swedish Experience *Journal of Marketing* **56**(1) 6-21
- [8] Athanassopoulos A D 2000 Customer satisfaction cues to support market segmentation and explain switching behaviour *Journal of Business Research* **47**(3) 191-207

6 Conclusion

A survey is administered among college students to examine their online shopping satisfaction using the e-commerce customer satisfaction evaluation model. The conclusions are listed as follows:

First, online shopping satisfaction can be affected by six factors, namely, shopping convenience, product factors, distribution services, safety and reliability of the system, complaint-handling services, and staff services.

Second, the hypotheses that are formulated based on the customer satisfaction evaluation model are all verified. Shopping operating convenience, product factors, distribution services, safety and reliability of the system, complaint-handling services, and staff services are all positively and significantly correlated with online shopping satisfaction. Safety and reliability of the system has the highest correlation coefficient with customer satisfaction, indicating that a high level of safety and reliability of the system leads to a high level of customer satisfaction.

Third, the descriptive statistical analysis shows that the overall online shopping satisfaction of college students is still in the middle level and is far from satisfactory. Given the rapid development of the cyber economy, all e-commerce trading platforms must strive to improve the quality of their services to attract more consumers and to maintain their status in the market.

Finally, the ANOVA results reflect that different gender and grade factors significantly affect customer satisfaction.

This empirical research involved only college students as the sample. This limited scope may affect the generalization of the research results. Several other issues also need further studies and discussions.

Acknowledgments

The authors would like to acknowledge the financial support provided by the National Statistics Research Projects of China under Grant 2012LY086 for this research. We also express our gratitude to the Soft Science Projects in Guangdong Province, China for their valuable research findings and our professor, Gao Yongduan, for her insightful comments.

- [9] Woodruff R B 1997 Customer Value: The Next Source for Competitive Advantage *Journal of the Academy of Marketing Science* 25(2) 139-153
- [10] Johnson M D, Fornell C 1991 A framework for comparing customer satisfaction across individuals and product categories *Journal of Economic Psychology* 12(2) 267-86
- [11] Fornell C, Johnson M.D, Anderson E W, Cha J, Bryant B E 1996 The American customer satisfaction index: nature, purpose, and findings *Journal of Marketing* 60(4) 7-18
- [12] Kristensen K, Martensen A, Gronholdt L 2000 Customer satisfaction measurement at Post Denmark: results of application of the European Customer Satisfaction Index methodology *Total Quality Management* 11(7) 1007-15
- [13] Likert R 1932 A Technique for the Measurement of Attitudes. *Archives of Psychology* 22(140) 5-55
- [14] Kaiser H F 1974 An index of factorial simplicity *Psychometrika* 39(1) 31-6.
- [15] Wu M 2010 *SPSS Operation and Application the Practice of Quantitative Analysis of Questionnaire Data* Chongqing: Chongqing University Press 216-257
- [16] He X 2011 *Multivariate Statistical Analysis (3rd edition)* Beijing: China Renmin University Press 110-45

Authors	
	<p>Xinling Du, born in October, 1971, Zhuhai City, Guangdong Province, P.R. China</p> <p>Current position, grades: vice associate professor of School of Management, Beijing Normal University Zhuhai Campus, China. University studies: B.Sc. in Statistics at Lanzhou University of Finance and Economics in China. M.Sc. in Statistics at Renmin University of China. Scientific interest: survey and statistical analysis, statistical assessment. Publications: more than 30 papers. Experience: Teaching experiences of 20 years, 20 scientific research projects.</p>
	<p>Lingyi Meng, born in March, 1993, Zhuhai City, Guangdong Province, P.R. China</p> <p>Current position, grades: student at the School of Management, Beijing Normal University Zhuhai Campus, China. University studies: Beijing Normal University Zhuhai Campus. Scientific interest: survey and statistical analysis, statistical assessment.</p>
	<p>Yujing Wu, born in June, 1992, Zhuhai City, Guangdong Province, P.R. China</p> <p>Current position, grades: student at the School of Management, Beijing Normal University Zhuhai Campus, China. University studies: Beijing Normal University Zhuhai Campus. Scientific interest: survey and statistical analysis, statistical assessment.</p>

A study on the efficiency of public culture service based on DEA cross evaluation

Liping Fu¹, Juan Li^{1, 2*}, Zuting Zheng²

¹Public Resource Management Research Center, College of Management and Economics, Tianjin University, Tianjin, China

²Hebei United University, Tangshan, Hebei, China

Received 14 June 2014, www.tsi.lv

Abstract

This paper analysed the inputs and outputs of public cultural services in 31 provinces of China in 2012, including municipalities and autonomous regions, based on the CCR model of DEA and the DEA cross-evaluation model. The DEAP2.1 software was also used for this empirical analysis to probe the performance and problems of the inputs and outputs of public cultural services. From analysis, this paper reached some conclusions, including the development of cultural services should be based on the increase of industrial inputs, as well as the optimization of resource allocation, so as to achieve the optimum state of inputs and outputs. The research results can provide a reference for the further improvement of the quality and inputs and outputs efficiency of public cultural services in all regions.

Keywords: Public Culture Services, DEA Model, Inputs and Outputs, Performance Evaluation

1 Introduction

DEA is a new field of cross study of operations research, management science, and mathematical economics, and it is used to study the relative effectiveness of decision-making units through mathematical programming calculation [1-5]. Compared with traditional methods, it has many advantages. By standardizing different variables of different units, DEA is very objective, and it can avoid many problems caused by various indicators and dimensions [6]. Initially, DEA is used by non-profit organizations to evaluate the efficiency of inputs and outputs, and it permits placing emphasis on choosing inputs and outputs [7]. Public cultural services cover a wide range, and involve more indicators' inputs and outputs, so it is hard to conduct effective evaluations of all the inputs and outputs [8]. Therefore, this paper adopted DEA cross-evaluation model and selected indicators to analyse the efficiency of inputs and outputs of public cultural services in 31 provinces of China in 2012.

2 The establishment of DEA model

2.1 THE EFFECTIVENESS JUDGMENT AND ECONOMIC IMPLICATION OF DEA MODEL

The relative efficiency of many input and output sectors and units, is evaluated by DEA through the mathematical programming model [9]. The effectiveness of a decision-making unit (DMU), is defined by the ratio between the weighed multi-index output and multi-index input of the unit [10].

$$DMU_L = \begin{bmatrix} x_i \\ y_i \end{bmatrix} (1 \leq i \leq n),$$

where $x_i = [x_{1i}, x_{2i}, \dots, x_{mi}]^T$, $y_i = [y_{1i}, y_{2i}, \dots, y_{si}]^T$ are the vectors of input index m of DMU_i $x_{1i}, x_{2i}, \dots, x_{mi}$ and the vectors of output index s of $y_{1i}, y_{2i}, \dots, y_{si}$ ($x_{ji}, y_{ji} > 0$) respectively, denoted: $X = [x_1, x_2, \dots, x_n]$, $Y = [y_1, y_2, \dots, y_n]$, X is called the multi-index input matrix and Y is the multi-index output matrix.

Definition 1: $T = \{(x, y) \text{ output } y \text{ can be produced by } x\}$ is called the input and output combination of DMU_i and set v_i as a weighting measurement of category i input, u_k is the measurement of category k output, then the weighting coefficient vector is:
 $v = \{v_i, i = 1, 2, \dots, m\}$; $u = \{u_k, k = 1, 2, \dots, s\}$;

Definition 2:
$$h_j = \frac{u^T y_{kj}}{v^T x_j} = \frac{\sum_{k=1}^s u_k y_{kj}}{\sum_{i=1}^m v_i x_{ij}}, \quad j = 1, 2, \dots, n$$
 is the

effectiveness index of evaluation unit j DMU_j . Select appropriate u and v , make $h_j \leq 1$; the increase of h_{j_0} indicates that the evaluation unit DMU_{j_0} can obtain more output with less input. Therefore, investigating the evaluation unit DMU_{j_0} in multi-evaluation units is not the

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

best way, change u and v as far as possible to obtain the maximum h_{j_0} . Construct the following CCR model:

$$(\bar{p}) \begin{cases} \max \frac{\sum_{k=1}^s u_k y_{kj_0}}{\sum_{i=1}^m v_i x_{ij_0}} = V_{\bar{p}} \\ s.t. \frac{\sum_{k=1}^s u_k y_{kj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1, j = 1, 2, \dots, n \\ u_k \geq 0, k = 1, 2, \dots, s \\ v_i \geq 0, i = 1, 2, \dots, m \end{cases} \quad (1)$$

Set $t = \frac{1}{v^T x_0}$, $\varpi = tv$ and $\mu = tu$, and then Equation (1) can be transformed to the linear programming model:

$$(p) \begin{cases} \max \mu^T y_0 = V_p \\ s.t. \varpi^T x_0 - \mu^T y_j \geq 0, j = 1, 2, \dots, n \\ \varpi^T x_0 = 1 \\ \varpi \geq 0, \mu \geq 0 \end{cases} \quad (2)$$

According to the duality theory of linear programming, the dual programming model of (P) is:

$$\begin{cases} \min \theta \\ s.t. \sum \lambda_j x_j + s^- = \theta x \\ \sum \lambda_j y_j - s^+ = y_0 \\ \lambda_j \geq 0, s^- \geq 0, s^+ \geq 0, j = 1, 2, \dots, n \end{cases} \quad (3)$$

The concept of Archimedes infinitesimal is introduced for getting solution conveniently in simplex method of linear programming and simplifying the inspection work:

$$(D_e) = \begin{cases} \min [\theta - \varepsilon (\hat{e}^T s^- + \hat{e}^T s^+)] \\ s.t. \sum \lambda_j x_j + s^- = \theta x_0 \\ \sum \lambda_j y_j - s^+ = y_0 \\ \lambda_j \geq 0, s^- \geq 0, s^+ \geq 0, j = 1, 2, \dots, n \end{cases} \quad (4)$$

where: $\hat{e} = (1, \dots, 1)^T \in R^m$, $e = (1, \dots, 1) \in R^s$. The optimal solution of programming problem (D_e) is λ^* , s^{*-} , s^{*+} , θ^* ; if $\theta^* = 1$, then DMU_{j_0} is weak DEA is efficient; if $\theta^* = 1$, $s^{*-} = 0$ and $s^{*+} = 0$, DMU_{j_0} is DEA-efficient, and the models above are the CCR models of evaluation unit.

In CCR model θ is the effective value of the decision-making unit DMU_0 (refers to the effective application

degree of input relative to output); X_i is the input factor combined with DMU_i and can be expressed by $(X_{i1}, X_{i2}, \dots, X_{im})$; Y_i is the output factor combined with DMU_i and can be expressed by $(y_{i1}, y_{i2}, \dots, y_{ip})$; λ is the combination ratio that is the decision-making unit i of a reconstructed effective DMU combination relative to DMU_0 and s^- and s^+ are slack variables.

Its economic implication is:

a. If $\theta = 1$ and $s^- = s^+ = 0$, DMU_0 is called DEA-efficient, in the economic system which is composed by n decision-making units, on the basis of original input X_0 , the output Y_0 has reached to the optimum condition;

b. If $\theta = 1$ and $s^- \neq 0$ or $s^+ \neq 0$, DMU_0 is called the weak DEA efficiency and in the economic system which is composed by n decision-making units, reduce the s^- of the input X_0 and do not change the output Y_0 or don't change input X_0 and increase the s^+ of the output;

c. If $\theta < 1$, DMU_0 is called DEA-inefficient, in the economic system which is composed by n decision-making units, through combination reduce the proportion of original input X_0 and don't change the original output Y_0 .

2.2 THE VALUE OF RETURNS TO SCALE IN DEA METHOD

Set $k = \sum \lambda_j$ and then k is called the value of returns to scale of DMU_0 , where:

a. If $k = 1$, the returns to scale of DMU_0 do not change and now DMU_0 reaches the maximum output scale point;

b. If $k < 1$, the returns to scale are increasing and if the k decreases faster, the scale will increase faster; it indicates that the proper increase of input can bring the higher output on the basis of DMU_0 input X_0 ;

c. If $k > 1$, the returns to scale are decreasing and if the k increases faster, the scale will decrease faster; it indicates that the increase of input cannot bring the higher output on the basis of DMU_0 input X_0 and now it is unnecessary to increase the decision-making unit.

2.3 THE DEFINITION OF INPUT REDUNDANCY RATIO AND OUTPUT INEFFICIENCY RATIO

Set the input redundancy ratio as α_{ij} and $\alpha_{ij} = S_{ij}^- / X_{ij}$ indicates the proportion of the weighted indicators, which can be saved. In the same way, set $\beta_{ij} = S_{ij}^+ / y_{ij}$ and β_{ij} is called as output inefficiency ratio indicating the proportion of weighted indicators, which shall be added.

The comparison of the input redundancy ratio or the output inefficiency ratio in different years in the economic system can show which aspects have been improved and should be strengthened in management. Moreover, the horizontal comparison can also be conducted to analyse the input redundancy ratio and output inefficiency ratio among different related economic systems in the same period of time.

3 The empirical research on the performance evaluation of public culture service based on DEA

3.1 THE SELECTION OF DECISION-MAKING UNITS

The DEA method is principally applied to the evaluation of Relative Merits among same kinds of samples, so the selected samples for evaluation must be of the same kind as DMU. In recent years, the Party Central Committee and the State Council have paid high attention to the construction of public culture service. The outline of the 12th Five-Year Plan has pointed out “to enhance public cultural products and services and establish and improve the public cultural services system” [11]. In recent years, the investment of public finance in the cultural construction has increased continuously, and the national cultural undertakings expense has increased in the latest five years with 20% annual growth rate and reached RMB 29.14 per capita in 2011 [12]. In this paper, the inputs and outputs data of 31 provinces, including municipalities and autonomous regions, in china, were selected as the decision-making units for the DEA evaluation, and the horizontal comparison on the public culture service performances were conducted to evaluate their advantages and disadvantages of various provinces, municipalities and autonomous regions in China. The inputs and outputs data in this study were taken from China Cultural Relics Statistics Yearbook 2013.

3.2 THE SELECTION OF EVALUATION INDICATORS

Different evaluation indicators of DEA system will result in different effectiveness evaluation results [13]. Therefore, the following aspects, including the realization of evaluation purpose, the complete reflection the evaluation purpose and the input vectors are connected

with the output vectors, should be principally considered for the selection of evaluation indicators. The evaluation of the efficiency of public culture service is a complex system, involving multiple inputs and outputs. The inputs contain physical factors, such as people, money, and commodity, can be applied to quantitative calculation, and intangible factors, like national policy support and technical support and so on. The multiple outputs contain both various cultural activities aiming at the improvement of people's livelihood and the contributions to social and economic development.

According to the data size of unity, the principle of comparability, at the same time, the data availability, the representative of indicators and the principle of mutual independence of indicators, this paper chose three input indicators, and they were per capita cultural operating expense (RMB), cultural workers number, and public cultural institutions.

The number of the public cultural institutions is the sum of the number of art performance venues, libraries, museums and mass cultural centres; the number of cultural-practitioners is the total number of staff in the art performance venues, libraries, museums and mass cultural centres; the cultural undertakings expense per capita embodies the capital investment in the cultural undertakings by the government, which is the core indicator reflecting the cultural undertakings development. The output indicators selected in this paper were based on the forms of cultural activities and the output quantities, including the number of cultural activities, and the number of participating people, in which the number of cultural activities was the sum of the number of art groups' performances held in art stadium, and art exhibitions, cultural activities and training classes held by cultural centres; and the number of participating people was the sum of the number of audience of the art group performances, audience going to the art stadium or museums and attending art exhibitions, people participating in art exhibitions, cultural activities and training classes held by cultural centres.

3.3 DATA REDUCTION

The data reduction and analysis were made according to China Cultural Relics Statistics Yearbook 2013. The data of various indicators is unified and comparable. Table 1 presents the details of specific data:

TABLE 1 DMU input and output data of public culture service investment performance

Region	Cultural undertakings expense per capita (Yuan)	Numbers of cultural workers	Number of public cultural institutions	Number of cultural activity participators (10,000)	Number of cultural Activities (10,000)
Beijing	110.55	13361	486	3010.8656	10.5727
Tianjin	56.11	4893	135	1393.2055	3.2731
Hebei	15.74	21113	846	5356.566	10.2954
Shanxi	36.34	20545	660	5069.8109	13.5354
Inner Mongolia	65.12	10463	349	2681.9817	3.547
Liaoning	33.53	12878	481	3578.638	6.7711
Jilin	34.53	6615	226	1715.9153	3.5798
Heilongjiang	24.4	9342	355	3001.0366	3.6978

Shanghai	120.65	17981	380	4205.669	15.2522
Jiangsu	37.59	26692	1043	12275.8601	45.9412
Zhejiang	65.2	30551	1155	10006.7174	15.0313
Anhui	15.08	22564	1344	3802.5818	13.4117
Fujian	33.9	16491	585	4182.5967	7.148
Jiangxi	17.71	11267	496	4118.7341	5.4668
Shandong	21.36	18405	752	9402.1821	10.2993
Hainan	15.99	27708	864	11977.6866	12.0508
Henan	24	16711	577	5653.0646	8.0448
Hubei	19.12	12445	486	6107.1678	8.4259
Hunan	36.3	26251	757	11118.4997	11.5049
Guangdong	25	6439	294	2905.308	6.5208
Guangxi	64.5	3810	116	893.49	0.9768
Chongqing	41	7899	358	3051.2564	3.156
Sichuan	34.04	21870	961	6978.9988	11.6886
Guizhou	27.81	5058	258	1761.1226	2.2668
Yunnan	28.07	11841	534	3170.7065	6.0521
Tibet	88.09	3175	201	299.0573	0.8063
Shanxi	40.87	16647	531	5501.2832	4.8185
Gansu	35.38	9641	397	3508.3203	3.5298
Qinghai	89.8	2316	143	304.3684	1.259
Ningxia	68.84	1833	65	669.3121	1.0033
Sinkiang	56.01	8108	346	2399.3285	9.5304

4 Model calculation and analysis

4.1 SOLUTION TO CCR MODEL

The software DEAP 2.1 was applied in this paper to solve CCR model and sort the results of 31 provinces (including municipalities and autonomous regions) according to scale efficiency. See Table 2 for the results.

TABLE 2 Results of solution to CCR model

Region	Overall efficiency	Pure technical efficiency	Scale efficiency	Evaluation conclusion
Jiangsu	1	1	1.000 -	DEA-efficient
Shandong	1	1	1.000 -	DEA-efficient
Henan	1	1	1.000 -	DEA-efficient
Guangdong	1	1	1.000 -	DEA-efficient
Zhejiang	0.679	0.679	1.000 -	Weak DEA efficiency
Hunan	0.995	1	0.995 irs	DEA-inefficient, progressive increase of returns to scale
Sichuan	0.64	0.651	0.983 irs	DEA-inefficient, progressive increase of returns to scale
Shanghai	0.932	0.987	0.945 irs	DEA-inefficient, progressive increase of returns to scale
Shanxi	0.741	0.799	0.928 irs	DEA-inefficient, progressive increase of returns to scale
Guangxi	0.926	1	0.926 irs	DEA-inefficient, progressive increase of returns to scale
Beijing	0.517	0.562	0.92 irs	DEA-inefficient, progressive increase of returns to scale
Chongqing	0.756	0.836	0.905 irs	DEA-inefficient, progressive increase of returns to scale
Hubei	0.736	0.814	0.904 irs	DEA-inefficient, progressive increase of returns to scale
Gansu	0.712	0.804	0.886 irs	DEA-inefficient, progressive increase of returns to scale
Shanxi	0.601	0.706	0.852 irs	DEA-inefficient, progressive increase of returns to scale
Inner Mongolia	0.566	0.676	0.838 irs	DEA-inefficient, progressive increase of returns to scale
Fujian	0.548	0.654	0.838 irs	DEA-inefficient, progressive increase of returns to scale
Liaoning	0.587	0.704	0.834 irs	DEA-inefficient, progressive increase of returns to scale
Tianjin	0.787	1	0.787 irs	DEA-inefficient, progressive increase of returns to scale
Ningxia	0.785	1	0.785 irs	DEA-inefficient, progressive increase of returns to scale
Sinkiang	0.683	0.875	0.781 irs	DEA-inefficient, progressive increase of returns to scale
Jiangxi	0.738	1	0.738 irs	DEA-inefficient, progressive increase of returns to scale
Yunnan	0.723	1	0.723 irs	DEA-inefficient, progressive increase of returns to scale
Heilongjiang	0.542	0.757	0.716 irs	DEA-inefficient, progressive increase of returns to scale
Guizhou	0.659	0.957	0.688 irs	DEA-inefficient, progressive increase of returns to scale
Hebei	0.687	1	0.687 irs	DEA-inefficient, progressive increase of returns to scale
Hainan	0.651	1	0.651 irs	DEA-inefficient, progressive increase of returns to scale
Hainan	0.54	0.927	0.582 irs	DEA-inefficient, progressive increase of returns to scale
Jilin	0.579	1	0.579 irs	DEA-inefficient, progressive increase of returns to scale
Qinghai	0.316	0.853	0.370 irs	DEA-inefficient, progressive increase of returns to scale
Tibet	0.197	0.717	0.274 irs	DEA-inefficient, progressive increase of returns to scale

Note: the overall efficiency in the table refers to the technical efficiency with no consideration of the returns to scale; the pure technical efficiency refers to technical efficiency with the consideration of the returns to scale; and the scale efficiency refers to the scale efficiency with consideration of returns to scale.

According to the result of solution to CCR model in Table 2, we can see that the scale efficiency in Jiangsu, Shandong, Henan and Guangdong is 1, reflecting the relative optimality of input and output of public culture service. These four regions put more in the public culture service area, which is also started relatively earlier and has already stepped into a mature phase. Although the scale

efficiency of Zhejiang is also 1, the overall efficiency and pure technical efficiency are not 1, indicating that a higher investment in public culture service is made after awareness of its importance, which has not achieved a good effect and needs further adjustment.

There are 26 DEA-inefficient regions, with efficient pure technical efficiency (the pure technical efficiency is 1) and inefficient scale efficiency (the scale efficiency is

less than 1) in 9 regions of Hunan, Guangxi, Tianjin, Ningxia, Anhui, Jiangxi, Guizhou, Hebei and Jilin, indicating the weak mean technical level in these regions and demand for further projection analysis of its reasons.

In order to know the reasons for weak scale efficiency of public culture service in the inefficient DEA regions, the projection analysis was carried out and the results were listed in Table 3.

TABLE 3 Rankings of various provinces and projection analysis of decision-making unit of inefficient DEA

Region	Rank	DEA efficiency value (θ)	Input redundancy			Insufficient output	
			Cultural undertakings expense per capita (Yuan)	The number of cultural workers	The number of public cultural institutions	The number of the cultural activity participators (10,000)	The number of cultural activities (10,000)
Jiangsu	1	1.000 -	0	0	0	0	0
Shandong	1	1.000 -	0	0	0	0	0
Henan	1	1.000 -	0	0	0	0	0
Guangdong	1	1.000 -	0	0	0	0	0
Zhejiang	1	1.000 -	39.127	9804.798	370.677	0	0
Hunan	6	0.995 irs	0	0	0	0	0
Sichuan	7	0.983 irs	11.875	7629.674	389.193	0	0
Shanghai	8	0.945 irs	61.718	8265.713	4.896	143.845	0
Shanxi	9	0.928 irs	8.216	4873.498	106.751	0	2.303
Guangxi	10	0.926 irs	0	0	0	0	0
Beijing	11	0.92 irs	48.383	6231.31	212.703	130.26	0
Chongqing	12	0.905 irs	6.738	1298.195	69.37	0	2.524
Hubei	13	0.904 irs	4.461	4881.733	107.259	28.089	0
Gansu	14	0.886 irs	6.934	1889.581	77.81	0	2.962
Shanxi	15	0.852 irs	10.699	9262.899	194.315	180.556	0
Inner Mongolia	16	0.838 irs	21.124	3596.463	113.212	0	1.301
Fujian	17	0.838 irs	11.713	7396.247	202.134	25.283	0
Liaoning	18	0.834 irs	9.93	5093.557	142.45	0	0.093
Tianjin	19	0.787 irs	0	0	0	0	0
Ningxia	20	0.785 irs	0	0	0	0	0
Sinkiang	21	0.781 irs	7.026	1017.025	57.504	754.663	0
Anhui	22	0.738 irs	0	0	0	0	0
Jiangxi	23	0.723 irs	0	0	0	0	0
Yunnan	24	0.716 irs	6.808	2871.867	134.329	378.67	0
Heilongjiang	25	0.688 irs	1.05	1810.192	15.278	178.928	2.584
Guizhou	26	0.687 irs	0	0	0	0	0
Hebei	27	0.651 irs	0	0	0	0	0
Hainan	28	0.582 irs	4.72	714.312	8.488	52.178	0.707
Jilin	29	0.579 irs	0	0	0	0	0
Qinghai	30	0.370 irs	21.138	341.55	72.435	430.986	0
Tibet	31	0.274 irs	24.904	897.602	109.405	520.704	0.371

The projection analysis in Table 3 shows that the actual values of various input and output indexes are the same as the project values in the four regions of Jiangsu, Shandong, Henan and Guangdong with the relatively optimum input and output of public culture service, so the corresponding adjustment amount is 0, indicating that the input and output level in these regions has reached the optimum utilization state. As a weak DEA efficient region, Zhejiang Province is faced with serious input redundancy, so the investment in the public culture service should be reduced appropriately, which will not result in the decrease of output.

According to Table 3, it is not hard to discover that with the pure technical efficiency 1, the input redundancy and insufficient output in 9 regions is 0, indicating that the input and output of the public culture service in these regions are relatively balanced. However, the input can be increased appropriately due to the smaller scale of the

public culture service, which can result in the incensement of the output and the scale efficiency and gradually achieve a mature public culture service. Other regions like Sichuan and Shanghai belong to the inefficient regions in pure technical efficiency and scale efficiency. Analysed from an input point, these regions are all faced with the redundant input in cultural undertakings expense per capita, cultural-practitioner quantity and public cultural institution quantity. Under a low technical level, with the other factors unchanged, increasing the input of cultural undertakings expense per capita, cultural-practitioner quantity and public cultural institution quantity excessively will result in inefficient production. Therefore, analysed from a public culture service performance point, the optimization of resource allocation and incensement of industrial input should be carried out together, with the expectation of achieving the optimum state of input and output. Analysed from an output point, there exists the

insufficient person-time and quantity of the cultural activities in most regions, which needs to be improved in the future.

4.2 DEA CROSS EVALUATION MODEL

The DEA cross evaluation model is proposed to solve the problem of failure to distinguish the advantages and disadvantages of decision-making units by the traditional DEA evaluation method [14]. In the new DEA cross evaluation, the DEA is used as a kind of ranking tool for multi-criteria decision making, with the main function of distinguishing the efficiency of efficient decision-making units so that these units can be ranked.

Basic idea of cross evaluation: calculate the efficiency value of other DMU_k by the optimum weight $w_i^* = \begin{bmatrix} v_i^* \\ u_i^* \end{bmatrix}$ of each DMU_i to obtain the cross evaluation value:

$$E_{ik} = \frac{y_k^T u_i^*}{x_k^T v_i^*} \tag{5}$$

The higher value of E_{ik} is, more beneficial for DMU_k but the more adverse for DMU_i .

Since the optimal solutions u_i^* and v_i^* are not unique, the cross evaluation value E_{ik} is uncertain. According to E_{ik} and $i \in \{1, 2, \dots, n\}$, $k \in \{1, 2, \dots, n\}$, the linear programming below may be solved by the aggressive cross evaluation:

$$\begin{cases} \min y_k^T u \\ s.t. y_j^T u \leq x_j^T v \quad (1 \leq j \leq n), \\ y_i^T u = E_{ik} x_i^T v, \\ x_k^T v = 1, \\ u \geq 0, v \geq 0 \end{cases} \tag{6}$$

The cross evaluation value can be worked out by the optimal solution u_{ik}^* and v_{ik}^* in Equation (6):

$$E_{ik} = \frac{y_k^T u_{ik}^*}{x_k^T v_{ik}^*} = y_k^T u_{ik}^* \tag{7}$$

The cross evaluation matrix is constituted by the cross evaluation value:

$$E = \begin{bmatrix} E_{11} & E_{12} & \dots & E_{1n} \\ E_{21} & E_{22} & \dots & E_{2n} \\ \dots & \dots & \dots & \dots \\ E_{n1} & E_{n2} & \dots & E_{nn} \end{bmatrix}$$

In the above matrix, the elements E_{ii} on the principal diagonal are the self-evaluation value and the non-principal diagonal $E_{ik} (k \neq i)$ is the cross evaluation value. E is the evaluation value of DMU_i by decision-making units i , indicating that the higher values account for the superior of DMU_i ; row i of E (except the elements on the principal diagonal) is the evaluation value of DMU_i for other decision-making units, indicating that the lower the values are, the more beneficial for DMU_i :

$$e = \frac{1}{n} \sum_{k=1}^n E_{ki} \tag{8}$$

Take the average value of column i of E as an index for measurement of advantage or disadvantage of DMU_i and regard e_i as the overall evaluation of decision-making units for DMU_i , and the higher e_i accounts for the superior of DMU_i . The calculation results by DMU_i MATLAB software are shown in Table 4.

TABLE 4 Results of e value in cross evaluation model

e_i	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
Value	0.422	0.553	0.488	0.505	0.441	0.493	0.467	0.541
e_i	e_9	e_{10}	e_{11}	e_{12}	e_{13}	e_{14}	e_{15}	e_{16}
Value	0.556	0.962	0.570	0.319	0.461	0.604	0.873	0.866
e_i	e_{17}	e_{18}	e_{19}	e_{20}	e_{21}	e_{22}	e_{23}	e_{24}
Value	0.625	0.856	0.819	0.738	0.401	0.576	0.543	0.502
e_i	e_{25}	e_{26}	e_{27}	e_{28}	e_{29}	e_{30}	e_{31}	
Value	0.444	0.130	0.584	0.572	0.198	0.591	0.536	

According to the table above: $e_{10} > e_{15} > e_{16} > e_{18} > e_{19} > e_{20} > e_{17} > e_{14} > e_{30} > e_{27} > e_{22} > e_{28} > e_{11} > e_9 > e_2 > e_{23} > e_8 > e_{31} > e_4 > e_{24} > e_6 > e_3 > e_7 > e_{13} > e_{25} > e_5 > e_1 > e_{21} > e_{12} > e_{29} > e_{26}$.

The new ranking for 31 regions obtained according to the above ranking of values of e is shown in Table 5.

According to Table 5, various input and output performances of public culture service in Jiangsu rank the first place but the e value has not reached 1 yet, indicating that insufficient input possibly exists at some aspects except cultural undertakings expense per capita. Therefore, cultural-practitioner quantity and public cultural institution quantity etc. and the further improvement shall be carried out.

TABLE 5 Analysis of the results of cross evaluation

Region	Jiansu	Shandong	Henan	Hunan	Guangdong	Guangxi	Hubei	Jiangxi
Ranking	1	2	3	4	5	6	7	8
Region	Ningxia	Shaanxi	Chongqing	Gansu	Zhejiang	Shanghai	Tianjin	Sichuan
Ranking	9	10	11	12	13	14	15	16
Region	Heilongjiang	Sinkiang	Shanxi	Guizhou	Liaoning	Hebei	Jilin	Fujian
Ranking	17	18	19	20	21	22	23	24
Region	Yunnan	Inner Mongolia	Beijing	Hainan	Anhui	Qinghai	Tibet	
Ranking	25	26	27	28	29	30	31	

5 Research conclusion

This paper made a comprehensive assessment and analysis on the performance and efficiency of public cultural service of 31 provinces (including municipalities and autonomous regions) in China. In order to avoid subjective judgment, the evaluation of the inputs and outputs of different regions' public cultural service was made in a quantitative way, which made the results more objective. On the basis of DEA analysis results, the deep analysis was made for the 31 regions with adequate application of projecting analysis and the overall efficiency ranking was carried out in the 31 decision-making units through DEA cross evaluation models. The conclusions were drawn as follows:

Conclusion 1: all of the 31 provinces (including municipalities and autonomous regions) have increased the investment in public cultural service to some degree, in accordance with the requirement of the outline of the 12th Five-Year Plan, which advocates enhancing the supply of public cultural product and service, and establishing and perfecting the system of public cultural services. The industrial scale in Jiangsu, Shandong, Henan and Guangdong province has been greatly improved, occupying the top list of 31 provinces (including municipalities and autonomous regions) in China. Their system of public cultural service becomes more and more mature, while the other regions still need to make more advancement in increasing the input and output efficiency of public cultural service, perfecting the allocation of resources and using efficiency.

Conclusion 2: statistics show that the scale efficiency in Zhejiang province is 1, but both its overall efficiency

and pure technical efficiency are not 1, which indicate that although a higher investment in public culture service has been made after realizing its importance, the higher investment has not achieved better results and Zhejiang needs further adjustment to the quantity of output value.

Conclusion 3: despite the inputs and outputs of public cultural service in 9 regions, including Hunan and Guangxi, are relatively balanced, their scale efficiency is not high. Therefore, more investment should be properly made to increase the scale efficiency, gradually making the public cultural service industry become mature.

Conclusion 4: the pure technical efficiency and scale efficiency in Shanghai and Sichuan and other regions do not make any sense, which means restricted by the technical factors, if they do not change other conditions, excessive input of people, money and materials in public cultural service will lead to inefficient results. So considering the efficiency and performance of public cultural service, these regions should increase the input of public cultural service as well as allocating resources optimally, to achieve a best state of input and output. Analysed from the perspective of output, most regions perform insufficiently in output, so the improvement and promotion should be carried out in the future.

Conclusion 5: although Jiangsu province tops the comprehensive list of every input and output performance index of public culture service, its e value has not reached 1 through DEA cross efficiency evaluation. It indicates that insufficient input exists in some other aspects, besides cultural undertakings expense per capita, cultural-practitioner quantity and public cultural institution quantity and so on, and needs to be improved in the future.

References

- [1] Deraman N A 1982 Data envelopment analysis (DEA) evaluation framework of hotel services *Recent Researches in Economics* 56(1) 154-9
- [2] Jun T 2000 Research of multiple criteria random DEA model and application on evaluation investment Chinese Quarterly *Journal of Mathematics* 15(2) 70-5 (in Chinese)
- [3] Caves W D 1982 The economic theory of index numbers and the measurement of input output and productivity *Econometrics* 50(6) 1393-1414
- [4] Charnes A, Cooper W 1982 Measuring the efficiency of decision making units *European Journal of Operational Research* 73(2) 429-44
- [5] Mohammadi A, Ranaei H 2011 The Application of DEA based malmquist productivity index in organizational performance analysis *International Research Journal of Finance and Economics* 62(6) 68-76
- [6] Renshou Z et al. 2011 Research on performance of cultural industry in 13 provinces of China based on DEA model-based on input-output data of 2007 from 13 cities of Guangdong *China Soft Science* 26(2), 183-92 (in Chinese)
- [7] Xiuge T, Fenghu W 2011 Research on the efficiency of public investment in culture based on DEA *Development Studies* 28(2) 90-3 (in Chinese)
- [8] Yanxin Z 2013 Performance evaluation of public culture service: empirical research based on DEA *Journal of Shandong Administration Institute* 25(2) 33-8 (in Chinese)
- [9] Wenyan W, Yulan Z 2013 Empirical research on the tendency of culture industry input and output efficiency based on DEA-Based on the culture industry input and output data from 2004-2010 of Jiangsu *Nanjing University of Finance and Economics* 30(9) 51-55 (in Chinese)

[10] Jiating W, Rong Z 2009 Research on efficiency of cultural industry in 31 provinces of China based on three-stage DEA model *China Soft Science* 24(9) 75-82 (in Chinese)

[11] Jianmei J 2008 A Study on the performance evaluation of government common culture service system *The Journal of Shanghai Administration Institute* 9(4) 60-5 (in Chinese)

[12] Han X, Ma X 2012 A Study on the cultural industry development efficiency of China based on DEA model *Journal of Yunnan University of Finance and Economics* 28(3) 146-153 (in Chinese)

[13] Xu J, Xu X, Chen X, Cui L, Lu C, Xue B 2013 Evaluation on urban efficiencies of Gansu province based on DEA-cross model *Journal of Natural Resources* 28(4) 618-624 (in Chinese)

[14] L. Biao 2011 Research on the scientific and technological resources allocation efficiency based on DEA cross model *Modern Management Science* 28(12) 35-8 (in Chinese)

Authors	
	<p>Liping Fu, born in February, 1963, Tianjin City, P.R. China</p> <p>Current position, grades: Professor at the College of Management and Economics, Tianjin University, China. University studies: B.Sc. in Political Economics at Beijing Normal University in China. M.Sc. and Ph.D. in World Economy at Nankai University in China. Scientific interest: public management, technological innovation. Publications: more than 50 papers. Experience: teaching experience of 29 years, more than 10 scientific research projects.</p>
	<p>Juan Li, born in March, 1979, Tangshan City, Hebei Province, P.R. China</p> <p>Current position, grades: Associate professor at the College of Yisheng, Hebei United University, China. University studies: B.Sc. in Industrial Analysis at Jilin Institute of Chemical Technology in China. M.Sc. at Yanshan University in China. Scientific interest: public management, technological innovation. Publications: more than 20 papers. Experience: teaching experience of 8 years, 3 scientific research projects.</p>
	<p>Zuting Zheng, born in May, 1978, Tangshan City, Hebei Province, P.R. China</p> <p>Current position, grades: Associate professor at the College of Economics, Hebei United University, China. University studies: B.Sc. in Industrial Foreign Trade at Harbin Engineering University, China. M.Sc. in industrial economics at Hebei Polytechnic University in China and PhD in Technological Economics and Management at Tianjin University in China. Scientific interest: technological innovation and industrial economics. Publications: more than 20 papers. Experience: teaching experience of more than 10 years, more than 20 scientific research projects.</p>

Application of CCD-model-based DEA analysis method in research on agricultural economic growth

Lixian Jing*, Juan Li, Yi An

Hebei United University, 46 Xinhua Road, Tangshan, Hebei, China

Received 1 August 2014, www.tsi.lv

Abstract

This paper presents research into the statistical data related to China's economic growth from 1990 to 2011, obtained by application of the DEA analysis method to the data in the China Statistical Yearbook 2012. First, a brief introduction is given to the principle structure of the Data Envelopment Analysis mathematical model. Second, the CCD model is established, based on modelling of the decision-making unit DMU_j in the Data Envelopment Analysis model; and a detailed discussion is made of the corresponding linear program.

Finally, the optimum value of this program is obtained and the results evaluated through study of the data collected by the mathematical model established in the paper. The main conclusions are as follows: the per capita income of China's farmers is closely related to the national total agricultural economic output and the total power of agricultural machinery, but is not that closely related to the crop planting area, area of affected crops or crops disaster area. Further, China's agricultural development shall lead to an advanced level of huge mechanization.

Keywords: CCD model, DEA analysis method, programming model, non-Archimedean infinitesimal

1 Introduction

At present in China, to construct a new socialist countryside and to realize the ambitious goal of a well-off society, the Three Rural Issues must be solved. The agricultural economy in particular has a key role to play in national development and maintenance of the national life. However, there is a difficulty in adjusting the agricultural economy as only the farmers who have adapted to free economy can provide the essential driving force needed for its development [1]. Nonetheless, the agricultural economy must be developed, and thus great importance must be attached to the associated technology. In this paper, research is conducted into the data on development of the agricultural economy from 1990 to 2011 by the Data Envelopment Analysis (DEA) analysis method, with the expectation of providing feasible technology and advice for the future development of the agricultural economy.

Many people have greatly contributed to research on the DEA analysis method and agricultural economic development, and consequently significant progress has been made in both the analysis method and the broader research field. We highlight the following three as amongst the most outstanding contributions. In Analysis on the Rate of Contribution Made by Rural Education to Zhejiang Agricultural Economic Growth, by Jianfeng Hu and other teachers of Zhejiang Institute of Science and Technology, the rate of contribution made by rural education to Zhejiang's agricultural economic growth is calculated using the C^2R model on statistical data concerning agricultural input and output of Zhejiang Province from

1996 to 2005. Their analysis indicates that the total power of agricultural machinery and fertilizer use have little effect on the total agricultural output value, and the main driving forces of agricultural economic growth are the agricultural acreage and the degree of labor education [1]. In Mathematical Model of Density and Fertilizer in Agricultural Production Function and Its Optimization, CHEN Rongying and other teachers of Henan Vocation-Technical Teachers College establish, by the principle of calculus, a mathematical model of density and fertilizer in agricultural production and its optimization. They provide a scientific analysis of a method for determining the high-yield optimum density and the optimal amount of fertilizer use, which demonstrates the important role of advanced mathematics in modern agricultural scientific research [2]. Huang Liping of the College of Economics and Management of Southwest University has made a detailed analysis of the current income status of Chinese farmers in his Masters thesis, Analysis in the Components and Increasing Factors of the Farmers Income, by means of times-series analysis. He proposed that China's national income growth is fundamentally the result of adjustment of the redistribution policy of social wealth, with farmers' income increasing through promotion of social innovation, development and total wealth accumulation whilst maintaining macroeconomical sustainable growth.

By applying the DEA analysis method and building on previous research, this paper gives an analysis of the statistical data on China's agricultural economy and the changing relationships through time between the average net income of farmers, national gross agricultural

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

production, acreage under cultivation, crops disaster area and total power of agricultural machinery, with a view to obtaining the solution to development of the agricultural economy.

2 Theory of the DEA mathematical model

DEA is a new kind of analysis that integrates management science, mathematical economics, operational research and mathematics. Its underlying mathematical model is linear programming, and it aims to evaluate the relative efficiency of comparable decision-making units of the same kind. Thus, it provides a method for the quantitative analysis of many input indicators and output indicators. The DEA method includes the *CCR* model, *BCC* model, *C²GS²* model, *C²W* model and *C²WH* model. The *CCR* model is mainly applied for contribution rate calculation and efficiency evaluation; the *BCC* and

C²GS² models are applied for the evaluation of the technical efficiency between production departments; the *C²W* model is applied for the evaluation of an infinite number of decision-making units; and the *C²WH* model is applied in the case of many inputs and outputs. The following is a brief description of the above five models and their application to obtain the optimal analysis model of agricultural economic growth.

2.1 MATHEMATICAL EXPRESSION OF THE DEA MODEL

The data expressions of Equations (1) and (2) can be obtained according to the corresponding definitions of the quantities of decision-making units, input types, output types, input weight coefficients and output weight coefficients, as shown in Table 1.

TABLE 1 Definition of parametric variables of model decision-making units

Quantity of decision-making unit	Quantity of input types		Quantity of output types		Quantity of input weight coefficient	Quantity of output weight coefficient		
<i>n</i>	<i>m</i>		<i>s</i>		<i>m</i>	<i>s</i>		
	1	2	3	...	<i>j</i>	...	<i>n</i>	
<i>v</i> ₁	1	<i>x</i> ₁₁	<i>x</i> ₁₂	<i>x</i> ₁₃	...	<i>x</i> _{1<i>j</i>}	...	<i>x</i> _{1<i>n</i>}
<i>v</i> ₂	2	<i>x</i> ₂₁	<i>x</i> ₂₂	<i>x</i> ₂₃	...	<i>x</i> _{2<i>j</i>}	...	<i>x</i> _{2<i>n</i>}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>v</i> _{<i>i</i>}	<i>i</i>	<i>x</i> _{<i>i</i>1}	<i>x</i> _{<i>i</i>2}	<i>x</i> _{<i>i</i>3}	...	<i>x</i> _{<i>i</i><i>j</i>}	...	<i>x</i> _{<i>i</i><i>n</i>}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>v</i> _{<i>m</i>}	<i>m</i>	<i>x</i> _{<i>m</i>1}	<i>x</i> _{<i>m</i>2}	<i>x</i> _{<i>m</i>3}	...	<i>x</i> _{<i>m</i><i>j</i>}	...	<i>x</i> _{<i>m</i><i>n</i>}
	1	2	3	...	<i>j</i>	...	<i>n</i>	
<i>y</i> ₁₁	<i>y</i> ₁₂	<i>y</i> ₁₃	...	<i>y</i> _{1<i>j</i>}	...	<i>y</i> _{1<i>n</i>}	1	<i>u</i> ₁
<i>y</i> ₂₁	<i>y</i> ₂₂	<i>y</i> ₂₃	...	<i>y</i> _{2<i>j</i>}	...	<i>y</i> _{2<i>n</i>}	2	<i>u</i> ₂
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>y</i> _{<i>r</i>1}	<i>y</i> _{<i>r</i>2}	<i>y</i> _{<i>r</i>3}	...	<i>y</i> _{<i>r</i><i>j</i>}	...	<i>y</i> _{<i>r</i><i>n</i>}	<i>r</i>	<i>u</i> _{<i>r</i>}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>y</i> _{<i>s</i>1}	<i>y</i> _{<i>s</i>2}	<i>y</i> _{<i>s</i>3}	...	<i>y</i> _{<i>s</i><i>j</i>}	...	<i>y</i> _{<i>s</i><i>n</i>}	<i>s</i>	<i>u</i> _{<i>s</i>}

$$h_j = \frac{u^T y_j}{v^T x_j} = \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}}, j = 1, 2, \dots, m. \tag{1}$$

For appropriate selection of the input and output weight coefficients, Equation (3) takes values in the range (0,1).

2.2 THE CCR MODEL

If the efficiency index of decision-making unit *j*₀ is defined as the target, and the efficiency indices of all decision-making units are used as restriction, then the *CCR* model can be constructed as shown in Equation (4).

$$\left\{ \begin{array}{l} \max \quad h_{j_0} = \frac{\sum_{r=1}^s u_r y_{rj_0}}{\sum_{i=1}^m v_i x_{ij_0}} \\ \text{s.t.} \quad \frac{\sum_{r=1}^s u_r y_{rj_0}}{\sum_{i=1}^m v_i x_{ij_0}} \leq 1 \quad j = 1, 2, \dots, s \\ v \geq 0 \quad u \geq 0 \end{array} \right. \tag{4}$$

In Equations (1) and (2): *x*_{*ij*} is the total input of decision-making unit *j* to category *i* input; *y*_{*rj*} is the total output of decision-making unit *j* to category *r* output; *v*_{*i*} is a measurement of category *i* input, the input weight coefficient; *u*_{*r*} is a measurement of category *r* output, the output weight coefficient; *i* = 1, 2, ..., *m* ; *r* = 1, 2, ..., *s* ; *j* = 1, 2, ..., *n* ; and *x*_{*ij*} and *y*_{*rj*} are greater than zero.

*DMU*_{*j*} then has the corresponding efficiency evaluation index for each decision-making unit as shown in Equation (3).

where $v = (v_1, v_2, \dots, v_m)^T, u = (u_1, u_2, \dots, u_s)^T$ and the model in Equation (4) can be seen to be a time-sharing program by the use of the Charnes–Cooper conversion, as shown in Equation (5).

$$\left. \begin{aligned} t &= \frac{1}{v^T x_0} \\ w &= tv \\ \mu &= tu \end{aligned} \right\} \Rightarrow w^T x_0 = 1. \tag{5}$$

Therefore, the fractional model shown in Equation (4) can be transformed into the linear program shown in Equation (6).

$$\left\{ \begin{aligned} \max \quad & h_{j_0} = \mu^T y_0 \\ \text{s.t.} \quad & w^T x_j - \mu^T y_j \geq 0 \quad j = 1, 2, \dots, n \\ & w \geq 0 \quad \mu \geq 0 \\ & w^T x_0 = 1 \end{aligned} \right. \tag{6}$$

The efficiency evaluation relative target of decision-making unit j_0 is defined by the optimal solution to the linear program for all other decision-making units, which is the reason for applying the CCR model. By the principle of duality, the microanalysis of agricultural economy can be conveniently made from a theoretical and economic perspective, provided that the dual model in Equation (6) is established.

The dual program corresponding to Equation (6) is shown in Equation (7).

$$\left\{ \begin{aligned} \max \quad & \theta \\ \text{s.t.} \quad & \sum_{j=1}^n \lambda_j x_j \leq \theta x_0 \\ & \sum_{j=1}^n \lambda_j y_j \leq \theta y_0 \\ & \lambda_j \geq 0 \quad , j = 1, 2, \dots, n \\ & \theta \in (-\infty, +\infty) \end{aligned} \right. \tag{7}$$

TABLE 2 List of methods for DEA efficiency judgment

Parameter characteristic	DEA efficiency judgment
optimum value $h_{j_0} = 1$	Weak DEA efficiency
$w^* > 0, u^* > 0^*$ and the optimum value $h_{j_0}^* = 1$	DEA efficiency

Theorem 2. The equivalent condition of DMU_{j_0} for weak DEA efficiency is that the optimum value of the linear program given by Equation (6) is $\theta^* = 1$; and the equivalent condition of DMU_{j_0} for DEA efficiency is that additionally each optimal solution λ^* should correspond to $s^{*+} = 0, s^{*-} = 0$.

The judgment of technical efficiency and scale efficiency can be carried out at the same time by the application of the CCR model, as follows:

Criterion 1. If $\theta^* = 1$ and $s^{*+} = 0, s^{*-} = 0$, then the decision-making unit j_0 can be judged as DEA efficient and the economic activity of the decision-making unit is

For convenience in calculation and discussion of Equation (7), it is necessary to introduce the slack variable s^+ and residual variable s^- . This transforms the inequality in Equation (7) into the Equation (8).

$$\left\{ \begin{aligned} \max \quad & \theta \\ \text{s.t.} \quad & \sum_{j=1}^n \lambda_j x_j + s^+ = \theta x_0 \\ & \sum_{j=1}^n \lambda_j y_j - s^- = \theta y_0 \\ & \lambda_j \geq 0 \quad , j = 1, 2, \dots, n \\ & \theta \in (-\infty, +\infty) \quad s^+ \geq 0, s^- \leq 0 \end{aligned} \right. \tag{8}$$

2.3 PROPERTIES OF THE CCR MODEL

Theorem 1. A feasible solution exists for both the linear program given in Equation (6), and its dual program given in Equation (7). Further, the optimal solution also exists; if the optimal solution to Equation (6) is $h_{j_0}^*$ and the optimal solution to Equation (7) is θ^* , then $h_{j_0}^* = \theta^*$.

For the purpose of evaluating the efficiency of decision-making units, the concepts of weak DEA efficiency and DEA efficiency are defined as shown in Table 2, which also gives the parameter settings for Equation (6).

the result of both technical efficiency and scale efficiency at the same time;

Criterion 2. If $\theta^* = 1$ and at least one input or output is greater than 0, then the decision-making unit j_0 can be judged as weakly DEA efficient, but the economic activity of the decision-making unit is not the result of both technical efficiency and scale efficiency;

Criterion 3. If the optimum value has $\theta^* < 1$, then the decision-making unit j_0 can be judged as not being DEA efficient, and the economic activity is neither the result of optimal technical efficiency nor of optimal scale efficiency.

3 Research on empirical analysis and result

3.1 INDEX SYSTEM OF AGRICULTURAL ECONOMICAL GROWTH EVALUATION MODEL

The research in this paper is based on China’s national agricultural economy from 1990 to 2011, with all input and output index systems of the evaluation model as shown below.

Input: the total crops planting area (m^2) is expressed by

X_1 ; the area of affected crops (m^2) is expressed by X_2 ; the crops disaster area (m^2) is expressed by X_3 ; and the total power of agricultural machinery (10,000kw) is expressed by X_4 .

Output: the per capita net income of rural households (Yuan) is expressed by Y_1 and the total agricultural output value (0.1 billion Yuan) is expressed by Y_2 .

The variation of the index-related data each year is shown in Table 3.

TABLE 3 List of data related to agricultural economy from 1990 to 2011

Year	Per capita net income of rural household	Total agricultural output value	Total crops planting area	Area of affected crops	Crops disaster area	Total power of agricultural machinery
1990	686.30	4954.30	148362.2667	38474.0000	17819.3333	28707.7000
1991	708.60	5146.40	149585.8000	55472.0000	27814.0000	29388.6000
1992	784.00	5588.00	149007.1000	51332.0000	25893.0000	30308.4000
1993	921.60	6605.10	147740.7000	48827.0000	23134.0000	31816.6000
1994	1221.00	9169.20	148240.6000	55046.0000	31382.0000	33802.5000
1995	1577.70	11884.60	149879.3000	45824.0000	22268.0000	36118.1000
1996	1926.10	13539.75	152380.6000	46991.0000	21234.0000	38546.9000
1997	2090.10	13852.50	153969.2000	53427.0000	30307.0000	42015.6000
1998	2162.00	14241.88	155705.7000	50145.0000	25181.0000	45207.7000
1999	2210.30	14106.22	156372.8100	49979.5000	26733.7000	48996.1200
2000	2253.40	13873.60	156299.8460	54688.0000	34374.0000	52573.6063
2001	2366.40	14462.80	155707.8615	52214.6000	31793.1000	55172.1000
2002	2475.60	14931.54	154635.5134	46946.1000	27159.9000	57929.8500
2003	2622.20	14870.10	152414.9623	54505.8000	32516.3000	60386.5410
2004	2936.40	18138.36	153552.5454	37106.2562	16297.3165	64027.9100
2005	3254.90	19613.37	155487.7289	38818.2251	19966.0600	68397.8486
2006	3587.00	21522.28	152149.0000	41091.4100	24631.9400	72522.1234
2007	4140.40	24658.10	153463.9305	48992.3533	25063.8200	76589.5635
2008	4760.62	28044.15	156265.6989	39990.0340	22283.4700	82190.4132
2009	5153.17	30777.50	158613.5498	47213.6900	21234.2633	87496.1000
2010	5919.01	36941.11	160674.8135	37425.9000	18538.1000	92780.4757
2011	6977.29	41988.64	162283.2204	32470.5000	12441.3000	97734.6585

3.2 RESEARCH RESULTS AND ANALYSIS

If the CCR model with non-Archimedean infinitesimal ϵ is adopted, then the model shown in Equation (9) can be established for evaluating China’s national agricultural economy in 2011.

$$\begin{cases}
 \min [\theta - \epsilon (s_1^- + s_2^- + s_3^- + s_4^- + s_1^+ + s_2^+)] \\
 s.t. \quad X_1 \lambda^T + s_1^- = 162283.2204\theta \\
 \quad \quad X_2 \lambda^T + s_2^- = 32470.5000\theta \\
 \quad \quad X_3 \lambda^T + s_3^- = 12441.3000\theta \\
 \quad \quad X_4 \lambda^T + s_4^- = 97734.6585\theta \\
 \quad \quad Y_1 \lambda^T - s_1^+ = 6977.29\theta \\
 \quad \quad Y_2 \lambda^T - s_2^+ = 41988.64\theta \\
 \lambda \geq 0 \\
 s_i^- \geq 0 \quad i = 1, 2, 3, 4 \\
 s_r^+ \geq 0 \quad r = 1, 2
 \end{cases} \quad (9)$$

X_i, Y_i in Equation (9) indicate the 22- dimension indexed row vector, where the definition of each index is shown in Table 1 and λ is a 22- dimension row vector.

In the same way, the CCR model with non-Archimedean infinitesimal ϵ in 2010 and other years can be obtained, with the solutions shown in Table 4.

The evaluation result shows that the total power of agricultural machinery has greater influence on the rural household per capita net income and the total agricultural output value, while the crops planting area, area of affected crops and crops disaster area have less influence.

TABLE4 List of variation of the optimum value and evaluation result of CCR model by year

Year	Optimum Value	Evaluation Result	Year	Optimum Value	Evaluation Result
1990	M=0.8432, L1=0.14	B	2001	M=0.8894, L1=0.09	B
1991	M=1, L1=1	A	2002	M=0.9946, L4=0.05	B

1992	M=1, L1=1	A	2003	M=0.7854, L1=0.23	B
1993	M=1, L1=1	A	2004	M=1, L1=1	A
1994	M=1, L1=1	A	2005	M=1, L1=1	A
1995	M=0.7433, L4=0.19	B	2006	M=1, L4=1	A
1996	M=1, L4=1	A	2007	M=0.9742, L1=0.08	B
1997	M=1, L1=1	A	2008	M=0.9143, L1=0.22	B
1998	M=1, L1=1	A	2009	M=1, L1=1	A
1999	M=0.8744, L4=0.16	B	2010	M=1, L4=1	A
2000	M=1, L1=1	A	2011	M=1, L1=1	A

Note: Evaluation result A indicates DEA efficient and scale benefit unchanged; and evaluation result B indicates not DEA efficient and scale benefit increasing. In the optimum value column, M indicates θ and L_i indicates λ_i .

4 Conclusion

In this paper, research was conducted into the statistical data related to China's economic growth from 1990 to 2011 by means of the CCR-based DEA analysis method, using the data in the China Statistical Yearbook 2012. From the calculations, the conclusion is drawn that the rural household per capita net income is closely related to

the growth of national gross agricultural production and the total power of agricultural machinery, and receives less influence from the crops planting area, crops disaster area and area of affected crops. Moreover, the model is established for various years through use of the CCR model, and the optimum value is obtained for the linear program, which leads to a rational and reliable result.

References

- [1] Jianfeng H et al. 2007 Analysis on the Rate of Contribution Made by Rural Education to Zhejiang Agricultural Economic Growth *Journal of Zhejiang Institute of Science and Technology* 24(6) 670-7 (in Chinese)
- [2] Martić M, Savić G 2001 An application of DEA for comparative analysis and ranking of regions in Serbia with regards to social-economic development *European Journal of Operational Research* 132(2) 343-56
- [3] Rongjiang C 1997 Mathematical Model of Density and Fertilizer in Agricultural Production Function and Its Optimization *Journal of Henan Vocation-Technical Teachers College* 25(3) 32-5 (in Chinese)
- [4] Banker R D, Chames A, Cooper W W 1984 Some models for estimating technical and scale inefficiencies in data envelopment analysis *Management Science* 30(9) 1078-92
- [5] Quanling W, Ming Y 1989 An introduction to DEA CCR model *DEA (I) Systems Engineering: Theory&Practice* 9(1) 58-69
- [6] Quanling W, Yugang C 1989 Some important DEA models for evaluating relative efficiency: DEA (II) *Systems Engineering: Theory&Practice* 9(2) 55-68
- [7] Xiaoya L, Jinchuan C 2007 Arithmetic of extra resource allocation based on DEA method *Journal of Systems Engineering* 22(1) 57-61
- [8] Cook W D, Seiford L M 2009 Data envelopment analysis (DEA): Thirty years on *European Journal of Operational Research* 192(1) 1-17
- [9] Abeles T 1996 Sustainable Agriculture in the US *Journal of Sustainable Agriculture* 17(4) 3-8
- [10] Lijun D, Pingyu Z, Ping L 2010 Equilibrium of population and economic development in the top ten urban agglomerations in China *Journal of the Graduate School of the Chinese Academy of Sciences* 27(2) 154-62 (in Chinese)
- [11] Jianwen X 2003 On CPC.s Outlook of Agricultural Economic Structure since the Founding of PRC *Journal of Shan Xi Finance and Economics University* 25(6) 41-6 (in Chinese)
- [12] Ting L 2009 The Low-level Saturation Phenomenon in Chinese Peasant Houeholds Agricultural Investment and Its Causes *Journal of South China Agricultural University (social science edition)* 8(1) 19-23 (in Chinese)
- [13] Guohua T 2009 Empirical Analysis on the Relationship between Rural Public Goods Supply and Farmers' Income Growth *Journal of Hunan Agricultural University (Social Sciences)* 10(2) 19-24 (in Chinese)
- [14] Zhihua Y, Lijun J 2009 Relationship between bending property and density of wheat stem *Agricultural Science&Technology* 10(5) 100-1 (in Chinese)
- [15] Fuqing A et al 2008 Effects of nitrogen application rate, density and seedling age on dry matter accumulation of not tillage rape in seedling stage *Agricultural Science& Technology* 9(6) 93-6 107 (in Chinese)
- [16] Zhisun Z et al. 2009 Effects of planting density, duration of disclosing plastic film and nitrogen fertilization on the growth dynamics of rapeseed under no-tillage cultivation *Agricultural Science& Technology* 10 (1) 130-4 139 (in Chinese)

Authors



Jing Lixian, born in September 1985, Tangshan City, Hebei Province, P.R. China

Current position, grades: lecturer of Yisheng College, Hebei United University, China.
University studies: B.Sc. in Social Sports at Hebei Normal University in China, M.Sc. at Tianjin Institute of Physical Education in China.
Scientific interest: social management, the humanities and social sciences.
Publications: more than 10 papers.
Experience: 3 years of teaching experience, more than 3 scientific research projects.



Juan Li, born in March 1979, Tangshan City, Hebei Province, P.R. China

Current position, grades: Associate professor at the College of Yisheng, Hebei United University, China.
University studies: B.Sc. in Industrial Analysis at Jilin Institute of Chemical Technology in China. M.Sc. at Yanshan University in China.
Scientific interest: public management, technological innovation.
Publications: more than 20 papers.
Experience: teaching experience of 8 years, 3 scientific research projects.



Yi An, born in July 1982, Tangshan City, Hebei Province, P.R. China

Current position, grades: Engineer at Yisheng College, Hebei United University, China.
University studies: B.Sc. in Electronic and Information Engineering at Hebei University of Technology in China. M.Sc. at Hebei University of Technology in China, Control Theory and Control Engineering.
Scientific interest: control theory and control engineering, and teaching management.
Publications: more than 10 papers.
Experience: 5 years of teaching experience, more than 7 scientific research projects.

Impact analysis of trade openness on China's economic growth

Yan-Liang Liu*

Bohai University, BoHai City, LiaoNing Province, China, 121013

Received 21 July 2014, www.tsi.lv

Abstract

Trade openness plays can affect China's economy significantly. In this paper, effect of trade openness on China's economic growth was discussed by using the error correction model, Granger causality test and impulse-response function. Moreover, an in-depth analysis on strategies that can facilitate China's economic development was carried out. The empirical research results demonstrated that: 1) there's a short-run equilibrium relationship rather than a long-run one between trade openness and economic growth; 2) the import trade is related with the export trade. Due to the hedging, their collaborative effect is smaller than their independent effects; 3) trade openness can facilitate economic growth within a certain time period, which presents a "U-shaped" influence mode.

Keywords: trade openness, economic growth, effect model

1 Introduction

Macroscopically, trade openness refers to proportion of the total volume of imports and exports of a region in its GDP. It reflects the contributions of imports and exports to regional GDP. China has achieved an economic boom since the reform and opening up. China's economic development accelerates with the increasing openness to the world. Currently, China's trade openness has already exceeded 60%. In the past, China's trade focused on labour-intensive products, such as textile, clothes, etc. On one hand, such trade structure is difficult to win international competitive edges. On the other hand, it is easy to lose abundant China's resources and causes various anti-dumping policies. With the economic development, China's trade pattern changes accordingly. The current China's trade pattern is focused on technology-intensive and processing industries, benefited from China's continuously increasing openness to the world. Considering its close relationship with economic growth, trade openness can affect economic development significantly.

2 Relative theories and researches

The neoclassical growth theory declared that the higher the trade openness is, the higher the regional trade volume will be and the better effect of the regional economies of scale will be. Romer (1986) believed that high trade openness is beneficial for regional technical progress and thereby promotes economic growth [1]. Helpman (1991) analysed from the perspective of political economy and concluded that trade openness increases transparency of economic activities, reduces rent-seeking activities and makes resources used for production activity to promote economic growth [2]. Rogoff (1996) believed that

countries with higher trade openness have higher capacity to absorb and digest new ideas and technical progress of advanced countries [3]. Based on the general equilibrium theory, Upadhyay (2002) discovered that trade openness is an important factor that determines industrial specialization level [4]. These foreign researches demonstrate the important positive effect of trade openness on regional economic development.

Chinese researchers have been arguing about the effect of trade openness on economic growth. Some scholars argue that trade openness is a kind of dependence: higher trade openness indicates higher dependence of regional economic growth on foreign trade, while smaller trade openness indicates lower dependence. Some believes that trade openness reflects regional participation in foreign trade: higher trade openness indicates the active participation of regional economy in international trade, while smaller trade openness indicates the inactive participation. On this basis, many Chinese scholars have explored effect of China's trade openness on its economy. Some scholars reported that trade openness is insignificant to economy. For example, Zhang Liguang et al. (2004) explored effect of China's trade openness on the long-run economic equilibrium, finding that the trade openness is insignificant in term of direct GDP promotion [5]. Bao Qun (2003) also believed that China's trade openness is insignificant to economic growth [6]. Some scholars reported that trade openness can affect economy to a certain extent, but such effect is indirect instead of direct. For example, Huang Xinfei (2007) deemed that increasing trade openness will accelerate technical progress of competitive industries, thus improving specialized production technology and promoting the long-term economic growth [7]. Some other scholars thought that trade openness can affect economy to a certain extent, but such effect is very complicated. For example, Zhang

* *Corresponding author* e-mail: liuyanliang-333@126.com

Qingjun (2008) proposed the U-shaped growth model [8]. He believes that effect of trade openness on economy is not a simple straight line, but presents a U-shaped variation law. In other words, effect of trade openness on economy often increases firstly and then weakens gradually. This paper attempts to give explore effect of China's trade openness on economic growth thoroughly from various perspectives by using the error correction model, Granger causality test and impulse-response function.

3 Correlation analysis between trade openness and economic growth

3.1 VARIABLES AND DATA

For China, the trade openness of one year (TO) refers to the proportion of China's total volume of imports and exports in the GDP of this year. To analyse more specifically, the author divided TO into import trade openness (ITO) and export trade openness (ETO). ITO refers to the proportion of China's total imports of one year in GDP of this year, while ETO refers to the proportion of China's total exports of one year in GDP of this year. To sum up, indexes involved in this quantitative analysis includes TO, ITO, ETO and GDP.

Data used in the empirical analysis were annual data on www.tjcn.org, including total imports, total exports, total volume of imports and exports, and GDP from 1978 ~2012. To eliminate heteroscedasticity while maintaining same quantitative analysis relationship, GDP data with high values were converted into log data (LGDP). Through a simple calculation of TO, necessary indexes for quantitative analysis can be gained: $ITO_i = \frac{\text{Total-imports}_i}{GDP_i}$,

$$ETO_i = \frac{\text{Total-exports}_i}{GDP_i}$$

3.2 ERROR CORRECTION MODEL

Unit root test was conducted to TO, ITO, ETO and LGDP. The unit root test in this paper used ADF test. The test results demonstrated that under no differential, all four indexes were unstable and had roots of unity. However, they became stable and their roots of unity were eliminated after the first-order difference. ADT test results are listed in Table 1. All indexes have similar variation trend, thus making co-integration analysis feasible.

TABLE 1 ADF test results of TO, ITO, ETO and LGDP

Index	t-Statistic	Prob.*	Test critical values:	
LGDP	-3.8064	0.0072	1% level	-3.6463
TO	-4.7303	0.0006	5% level	-2.9540
ETO	-5.0538	0.0002	10% level	-2.6158
ITO	-4.3953	0.0014		

Error correction term has to be determined before the establishment of error correction model. Therefore, the least square regression was implemented by taking GDP as the dependent variable while TO, ETO and ITO as independent variables. The results of three independent variables are 8.9388, 15.8890 and 19.7386, respectively. Then, the error correction term can be defined by using GENR:

$$ECM_1 = LGDP(-1) - 8.9388 \cdot TO(-1),$$

$$ECM_2 = LGDP(-1) - 15.8890 \cdot ETO(-1),$$

$$ECM_3 = LGDP(-1) - 19.7386 \cdot ITO(-1).$$

The ECM estimation results of TO, ETO, ITO and LGDP are listed in Table 2. For the error correction model established with TO and GDP, one of its tail probability valued 0.1571 (>0.1), indicating their insignificant correlation. Therefore, TO and GDP could not be used to establish the error correction model. Similarly, ETO and GDP also have insignificant correlation and could not be used to establish the error correction model. Since the error correction model reflects a long-run equilibrium relationship, the insignificant correlation means there is no long-run equilibrium relationship between two variables. However, they still may have a short-run equilibrium relationship. As a result, Granger causality test and impulse-response function are needed to confirm such short-run equilibrium relationship. With respect to the ITO-GDP relationship, both tail probabilities were smaller than 0.1, but the determination coefficient was 0.2613, over smaller under two variables. This implies that there is a weak correlation between ITO and GDP.

TO includes ITO and ETO. Although there is a weak correlation between ITO and GDP, the relationship between TO and GDP in China is even weaker. This is because there is no long-run equilibrium relationship between ETO and GDP, and exports take the dominant role in the economic growth of China, a trade surplus country. The weak correlation between ITO and GDP may be caused by the design and technological progresses through imports. This consolidates economic base for the economic boom.

The error correction model verifies the long-run equilibrium relationship between variables. The short-run equilibrium relationship can be further analysed through Granger causality test and impulse-response function.

TABLE 2 ECM estimation results of TO, ETO, ITO and LGDP as well as relative test results

No	Variable	Coefficient	Std. Error	t-Statistic	Prob.	R-squared	Adjusted R-squared
1	D(TO)	0.3092	2.13E-01	1.45E+00	0.1571	0.2060	0.1547
	ECM	-0.0418	1.60E-02	-2.62E+00	0.0135		
	C	0.4564	1.20E-01	3.808707	0.0006		
2	D(ETO)	0.4226	3.90E-01	1.08E+00	0.2864	0.1337	0.0078
	ECM	-0.0359	1.74E-02	-2.06E+00	0.0474		
	C	0.4214	1.35E-01	3.129868	0.0038		
3	D(ITO)	0.74903	4.18E-01	1.79E+00	0.0828	0.2613	0.2137
	ECM	-0.041276	1.37E-02	-3.02E+00	0.0050		
	C	4.44E-01	1.00E-01	4.433938	0.0001		

4 Granger causality test and impulse-response analysis

4.1 GRANGER CAUSALITY TEST

Granger causality test was implemented to TO, ETO, ITO and LGDP (Table 3). Under one year lagged economic growth, one tail probability of every index is smaller than 0.1. Therefore, under 5% significance level, “TO is not the cause of LGDP change”, “ITO is not the cause of LGDP change” and “ETO is not the cause of LGDP change” are denied. However, it also could not to verify that “LGDP is the cause of TO, ITO and ETO changes”.

The Granger causality test reveals a short-run equilibrium relationship of ITO, ETO and TO with GDP. Such relationship is unidirectional. In other words, TO can cause GDP changes, but GDP could not influence TO. This is mainly because TO can affect import-export volume which is an important component of GDP, whole TO is controlled by the government. Government policies and decisions (e.g. tariff and export subsidy) can influence TO directly, but GDP could not affect TO significantly.

TABLE 3 Granger causality test results of TO, ETO, ITO and LGDP

Null Hypothesis	F-Statistic	Prob.
LGDP does not Granger Cause TO	0.58229	0.4512
TO does not Granger Cause LGDP	5.98753	0.0203
LGDP does not Granger Cause ITO	0.62023	0.4369
ITO does not Granger Cause LGDP	8.04044	0.0080
ETO does not Granger Cause LGDP	3.62451	0.0663
LGDP does not Granger Cause ETO	0.96198	0.3343

4.2 IMPULSE-RESPONSE ANALYSIS

The impulse-response function analyses the dynamic characteristics between dependent variable and independent variables based on the VAR model. Such dynamic characteristics refer the effect mechanism of every independent variable change or impact on the dependent variable. Since the VAR model is based on the Granger causality test, only unidirectional impulse-response analysis can be implemented to TO, ETO, ITO and LGDP according the Granger causality test results. The analysis results are shown in Figure 1-3.

Figure 1 shows response of LGDP to TO innovation. It remains basically same at the very beginning. However, such effect increases continuously as time goes on, reaching the peak at the 8th year. Subsequently, it becomes equilibrium and begins to weaken. Figure 2 shows

response of LGDP to ITO innovation. Similar variation law of LGDP with that in Figure 1 is observed. Figure 3 shows response of LGDP to ETO innovation, which still presents similar variation law. Viewed from the Y-axis of Figure 1, Figure 2 and Figure 3, TO is significantly smaller than ETO and ETO is significantly smaller than ITO. Therefore, effect of TO on GDP increases gradually, reaching the peak at about the 6th year. The maximum positive impact of TO is about 1, while the maximum positive impacts of both ITO and ETO are about 2. TO is smaller than ITO and ETO as well as their sum. Therefore, they can form a certain angle (>120°). This explains the smaller collaborative effect of import and export than their independent effects.

Response of LGDP to Nonfactorized One Unit TO Innovation

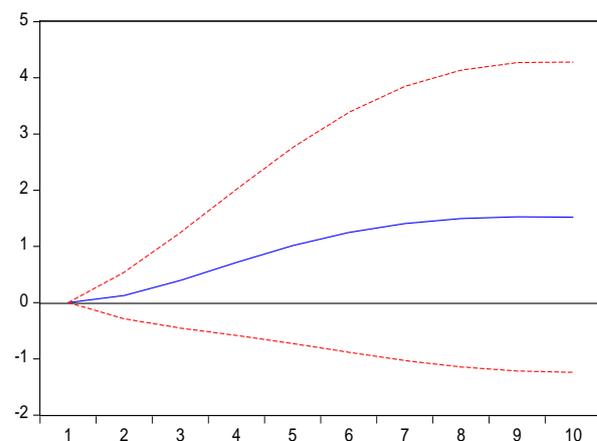


FIGURE 1 Impulse response of LGDP to TO

Response of LGDP to Nonfactorized One Unit ITO Innovation

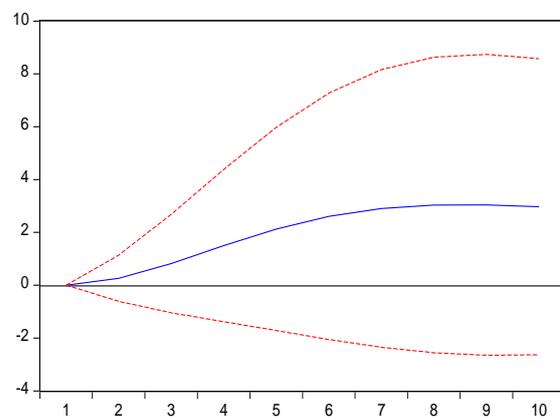


FIGURE 2 Impulse response of LGDP to ITO

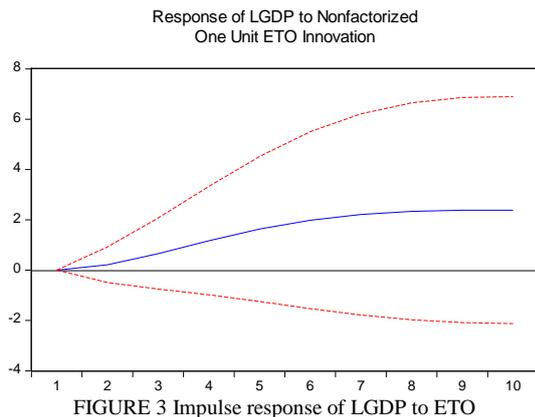


FIGURE 3 Impulse response of LGDP to ETO

5 Result analysis

Based on the error correction model, Granger causality test and impulse-response maps, we can conclude that:

1) Instead of the long-run equilibrium relationship, there is a short-run equilibrium relationship between TO and GDP. The error correction model confirms the weak correlation between ITO and GDP. However, as a trade surplus country, China’s export volume is higher than import volume and is significantly non-correlated with GDP. This causes the insignificant long-run equilibrium relationship between TO and GDP. According to the Granger causality test results, there is a certain causality between TO and GDP. TO can increase GDP viewed from the impulse-response maps. Such positive effect can be manifested by various aspects:

Firstly, imports can introduce in advanced technologies and equipment, stimulate improvements of local technologies and equipment, and intensify regional competitive edges. At the beginning of reform and opening up, China had backward technologies and equipment. Most technologies and equipment had to be introduced or learned from foreign countries. Technology import saves time for technical innovation and brings China advanced production level, thus facilitating the rapid economic development. On the other hand, China invents more advanced technologies based on these imported technologies, which further facilitates China’s economic development to a certain level. Nowadays, China has become an important technology exporter. For instance, China is planning to export the high-speed rail technology to other countries in the world, which can save a lot of time for foreign countries to achieve economic breakthroughs quickly.

Secondly, exports can increase domestic production. Domestic technologies and equipment also can be improved by participating in international competition. Trade openness includes import and export. Import is important to regional economic level and export also plays an important role in regional economy. China has been a trade surplus country since 1992. During this period, China’s trade structure changed. Exports shifted from labour-intensive products with low value and production level to technology-intensive products with increasing

added value. China is not a backup producer any more, but participates in international competition truly. For example, China’s Huawei technology has competed with IT technologies of various developed countries like the United States.

2) There is a hedging relationship between import and export, which causes the weaker collaborative effect of import and export than their independent effects. It can be seen from the impulse-response maps that independent ITO and ETO influence GDP more significantly than TO. This reflects that effect of import and export on economy could not be added simply. A collaborative relationship larger than 90° may exist, which weakens the collaborative effect (Figure 4). Viewed from the perspective of China, TO was less than 10% in 1978, but increased to the peak (65.17%) in 2006 and decreased to 47.00% in 2012. This may be caused by following reasons:

Firstly, the import and export structural change will adjust impact angle. As China has entered into the WTO, many imports could not pass policy constraints any more, thus diversifying import products. These imported products not only offset domestic shortage, but also compete with domestic products directly. Similarly, exports shifted from labour-intensive products to technology-intensive products. Such import and export structural change influences their direction, thus affecting the overall TO of China.

Secondly, the abovementioned TO variation is caused by China’s sound economic strategy. China adjusted its economic strategy after the financial crisis in 2008. Key attentions were shifted from exports to domestic demands. It facilitates domestic demand development vigorously, but slows down the foreign trade development. This is because internal consumption of domestic products increases with the development of domestic demand, thus slowing down export growth. Generally, TO decreases when the import and export growth is slower than China’s economic development.

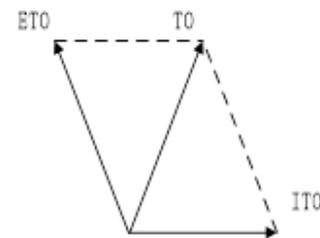


FIGURE 4 Collaborative effect of import and export

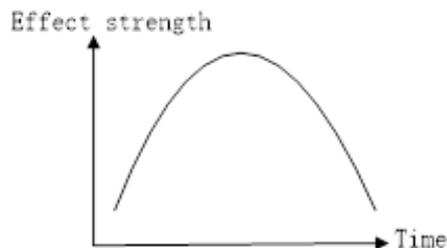


FIGURE 5 Effect strength of TO on economy

Thirdly, TO has a positive effect on GDP within a certain time period. According to the impulse-response analysis, effect of TO on GDP intensifies gradually, reaching the peak at the 8th year and then decreased. It presents an inverse “U-shaped” effect mode (Figure 5). This confirms the correctness of China’s domestic demand-oriented economic strategy:

Firstly, China’s TO has reached 60%, which is very high although it is smaller than some wide-open countries. China is a big country and could not open fully like common small countries. This is caused by the more complicated economy of big countries, which is difficult to be controlled and solved upon financial crisis. This is why wide-economic-open countries or regions in the world are small countries and regions. Economy can be influenced by both political competition and economic competition. Moreover, considering the frequent financial crises of developed countries, higher TO will bring more serious negative impacts. As a result, it is correct for China to adopt domestic demand-oriented economic strategy for the sake of its sound economic development and national political security.

Secondly, the domestic demand-oriented economic strategy is driven by the unbalanced regional economic development in China. An evident urban-rural economic development disharmony exists in China. The over lower rural consumption not only influences the living standard of rural residents, but also affects the rural economic or even the whole economic development. As a result, the domestic demand-oriented economic development strategy is necessary considering of the economic security and harmonious economic development in China.

6 Conclusions and suggestions

Based on abovementioned analysis, trade openness plays an important positive role in China’s economic growth. It improves China’s production level significantly and accelerates China’s economic development. Effect of ITO and ETO on economy differs from that of TO. ITO and ETO form a certain angle, making their collaborative effect smaller than the sum of their independent effects. Furthermore, trade openness affects economic growth with a certain time period. Meanwhile, effect of world economy on China’s economy increases as trade openness expands continuously. Good world economy can promote economic growth of China, while world economic crisis will bring China’s economy negative effects. Such negative effect is proportional to the trade openness. Considering the sound development and security of China’s economy, the author gives some suggestions:

Firstly, trade openness shall be enhanced in the period when it can promote economic growth. The empirical research demonstrates an inverse “U-shaped” effect mode of trade openness on economic growth. This may be because trade openness will facilitate domestic technology improvement and production, but after a certain period,

such facilitation effect will be weakened as the technical gap between regions narrows. At the beginning of reform and opening up, China mainly imported technologies and equipment. With the technical progress, China not only imports but also exports technologies and equipment. It imports less and less technologies and equipment, but exports more and more.

Secondly, import and export may form a certain collaborative effect. The angle of such collaborative effect may be determined by import-export content and level. To enhance positive effect of import and export trades on China’s economic growth, the import-export trade structure and content shall be perfected and trade quality shall be improved. Although China’s trade structure has been improved to a certain extent in recent years, export is still dominated by products with lower added value, which is against resource protection, collaborative effect of import and export, and the achievement of China’s foreign trade strategic goal – increase productivity and facilitate sound economic development of China.

Thirdly, China’s trade openness increases steadily and gradually. Although such steady development strategy contributes to economic growth, the maximum trade openness shall be limited. Higher trade openness enables world economy to occupy higher proportion in regional economy and exert larger effect on China’s economy, especially during the world economic crisis. Therefore, both advantages and disadvantages of trade openness shall be considered during enhancing foreign trade. To keep economic stability, China has been using conservative economic policies, which are beneficial for the sound economic development. Since the reform and opening up, China has achieved rapid economic development with backward systems, thus causing many economic and social problems [8]. On this basis, China shall adopt sound economic development strategy to prevent uncontrollable economic instability caused by over high trade openness.

Fourthly, equal attentions shall be paid to enhancing foreign trade and promoting domestic demand. As China has entered into the WTO, its economy will surely be influenced by the world economy. Hence, China shall take measures to avoid adverse effect but enhance positive effect. Although China is expanding the domestic demand continuously and slowing down the foreign trade development in recent years, this will not affect the importance of trade. Therefore, China shall view domestic demand promotion and foreign trade development equally [9]. Currently, China is strongly recommended to adjusting its import and export structures simultaneously with economic and industrial restructuring.

Acknowledgement

Social science fund project of Liaoning Province (The coordinated development of Liaoning coastal economic belt port logistics research L12BJL013).

References

- [1] Romer P M 1986 Increasing returns and long-run growth *Journal of Political Economy* **94** 1002-37
- [2] Grossman, Helpman 1991 Innovation and Growth in the Global Economy *Cambridge MIT press* 39-45
- [3] Obstfeld M, Rogoff K 1996 Foundations of International Macroeconomics *Cambridge MIT Press* 143-75
- [4] Gopinath M, Upadhyay M 2002 Human capital, technology, and specialization: a comparison of developed and developing countries *Journal of Economics* **75**(2) 161-79
- [5] Bao Q, Xu H, Lai M 2003 Trade openness and economic growth: theory and empirical research in China *The Journal of World Economy* **02** 10-8
- [6] Zhang L, Shi Y, Lihua 2004 Long-run equilibrium relationship between trade openness and economic growth *Finance Economics* **01** 78-82
- [7] Huang X, Shu Y 2007 Research on trade openness, industrial specialization and China's economic growth *Journal of International Trade* **12** 11-7
- [8] Zhang Q 2008 Empirical analysis of trade openness and economic growth *Journal of International Trade* **08** 23-7
- [9] Xue J 2013 Effect of trade openness on China's economic growth *Master Dissertation of Nanjing University Nanjing China* 21-8

Author



Yan-liang Liu, born in August, 1963, JinZhou City, LiaoNing Province, China

Current position, grades: the professor of Bohai university, Bohai city, China.

Scientific interest: business administration.

Publications: 11 papers.

Experience: teaching experience of 28 years, 9 scientific research projects.

Influence of China's trade imbalance on economy in the background of great nation

Chong Qian¹, Zhan-ao Wang^{2*}

¹First Author, City College of Wenzhou University, Wenzhou city, Zhejiang province, China, 325035

²Corresponding Author, City College of Wenzhou University, Wenzhou city, Zhejiang province, China, 325035

Received 7 July 2014, www.tsi.lv

Abstract

The developed countries have been accusing China for its trade surplus for a long time. In order to verify the influence of trade imbalance on the economy, our research applied techniques like error correction function, Granger test and impulse response function. The research outcomes manifest that: the influence of trade imbalance on the economy is not significant. On the contrary, the economy imposes certain impact on trade imbalance. Both imports and exports significantly promote the development of economy. Furthermore, imports remarkably facilitate exports. This can also be attributed as the reason of China's trade deficit before 1992.

Keywords: great nation, trade imbalance, error correction model

1 Introduction

Since the implementation of the reform and opening up policy, China has witnessed rapid economic growth based on the constant introduction of international technology, talent and knowledge as well as the continuous output of Chinese talents and products. Therefore, trades have imposed monumentally positive impact on the economy of China. However, since 1994, China has maintained the trade surplus, which is constantly enhanced. To 2012, the trade surplus of China reached 1455.829 billion Yuan, accounting for 5.96% of total volume of imports and exports, and 12.68% of the gross exports. Many developed countries consider that the long-term trade surplus is the root reason of the rapid growth of China's economy, and it also generates adverse effects on the economy of other countries, such as the investment and employment of some regions and the development rate of regional economy. As a result, years of trade surplus in China have triggered an increasing amount of trade disputes. Domestically, many point out that the long-term trade surplus is highly harmful for China, a country with extremely limited per capita hold of resources. Besides, trade surplus may lead to the massive brain drain and resource loss of China, thus harming the future development of China. Our research, however, attributes the root cause of trade surplus and deficit to the development pattern of the world economy, and both surplus and deficit are to serve the regional economic development, thus showing more benefits than harms for regional economic development. From 1978 when China adopted the reform and opening-up policy to 1992, China maintained a long period of trade deficit. During this period, China primarily introduced technology, equipment and talent in the country.

Meanwhile, the output was very limited. This situation was determined by the position of China in the then international configuration.

2 Relevant theories and research

Many scholars believe that the serious trade imbalance seriously influences the economy of China. However, most studies in China focus on the influencing factors and improvement measures of trade imbalance. Research on the specific influence of trade imbalance on the economy was highly limited. Lihua Lang (2006) explored into the influence of trade imbalance between China and U.S.A, indicating that trade imbalance would lead to trade conflicts and trade conflicts would inevitably influence the economy [1]. Jianqiang Wu (2009) pointed out that the continuing expansion of trade surplus in China would eventually bring about the prominence of negative influences [2]. For example, the increased degree of dependence upon foreign trade weakens the independence of China's economy; Appreciation of the Yuan will enhance the deterioration of the independence of China's monetary policy; the trade conflicts between China and the rest of the world are enhanced. Yanling Wu (2008) thought that the long-term trade imbalance of China not only intensifies trade conflicts between China and other countries but also enhances China's shielding ability against international financial risks [3]. Specifically, it ensures the stability of China's economy. However, with the development of trade surplus, the economy of China grows to be increasingly dependent on foreign trade, thus influencing the macroscopic readjustment and control of the economy by the Chinese government. Linjuan Jia (2013) researched the influence of global trade imbalance

* Corresponding authors e-mail: wzawhu@163.com

on the international economic order, and considered that the trade imbalance is mainly reflected in the long coexistence of trade surplus and deficit between China and U.S.A [4]. This manifests the economic status of developed country and developing country, and the trade imbalance is actually intensified by economic globalization and financial hegemony of U.S.A. A range of studies also analysed the causes of the long-term trade surplus in China. Jianying Chu (2006) analysed reasons of trade surplus between China and U.S.A, considering that trade surplus is not necessarily beneficial and trade deficit is not necessarily harmful [5]. The trade surplus in China is subject to the industrial transfer in East Asia and the advantage of labour cost due to the underdeveloped economic status of China. The trade surplus between China and U.S.A has produced positive influence on American consumers and the national industrial restructuring of China. However, China loses the opportunity of investment due to the substantive counterpart of foreign exchange reserves and foreign exchange reserve, and the government's subsidies to export enterprises are actually indirect subsidies to U.S.A. Wanqing Lu (2009) believed that the international division of labour determines the international trade, and China's status in the international division of labour of products in East Asia generates voluminous trade surplus of the processing industry in China [6]. At present, although the growth rate of imports and exports on year-on-year basis in China has declined drastically due to the global economic crisis, the remarkable foreign trade surplus of China would not change as long as the labour division pattern in East Asia remains unchanged.

Our research indicates that in terms of trade imbalance, trade surplus and deficit would generate negative influence

on China's economic environment but also positively boost the development of the entire economy. Hence, based on data and data analysis, error correction model and impulse response function was used to argue and analyse the influence of trade imbalance on the economy of China. Besides, the essence behind the influence was probed.

3 Empirical research

3.1 VARIABLES AND DATA

In order to identify the relationship between trade imbalance and the economy and highlight characteristics of trade imbalance, our research adopted the trade imbalance rate (MCB). Trade imbalance rate refers to the proportion of trade margin in the total volume of imports and exports. In order to reflect the directionality of trade imbalance, trade margin is obtained by using the gross exports of a region to deduct the gross imports. If the gross exports are larger than gross imports, MCB is a positive value, and vice versa. Regional GDP was used as the regional economy data. Since import proportion (IP), export proportion (EP) and GDP are large time series data, logarithmics were carried out to eliminate heteroscedasticity without changing its linear relationship. Therefore, our research contains four variables: LGDP, import proportion (IP), export proportion (EP) and trade imbalance rate (MCB).

Data in the empirical analysis were obtained from the annual data released by the official website of Statistical Bureau of China. Data from 1978 to 2012 were included. Index data was calculated and shown in Table 1.

TABLE 1 GDP and MCB Data

Year	GDP (a hundred million yuan)	Total Export-Import Volume (a hundred million yuan)	Gross Export (a hundred million yuan)	ETO
1978	3645	355	167.6	0.0460
1979	4063	454.6	211.7	0.0521
1980	4546	570	271.2	0.0597
1981	4892	735.3	367.6	0.0751
1982	5323	771.3	413.8	0.0777
1983	5963	860.1	438.3	0.0735
1984	7208	1,201.00	580.5	0.0805
1985	9016	2,066.70	808.9	0.0897
1986	10275	2,580.40	1,082.10	0.1053
1987	12059	3,084.20	1,470.00	0.1219
1988	15043	3,821.80	1,766.70	0.1174
1989	16992	4,155.90	1,956.10	0.1151
1990	18668	5,560.10	2,985.80	0.1599
1991	21782	7,225.80	3,827.10	0.1757
1992	26923	9,119.60	4,676.30	0.1737
1993	35334	11,271.00	5,284.80	0.1496
1994	48198	20,381.90	10,421.80	0.2162
1995	60794	23,499.90	12,451.80	0.2048
1996	71177	24,133.80	12,576.40	0.1767
1997	78973	26,967.20	15,160.70	0.1920
1998	84402	26,849.70	15,223.60	0.1804
1999	89677	29,896.20	16,159.80	0.1802
2000	99215	39,273.20	20,634.40	0.2080
2001	109655	42,183.60	22,024.40	0.2009
2002	120333	51,378.20	26,947.90	0.2239

2003	135823	70,483.50	36,287.90	0.2672
2004	159878	95,539.10	49,103.30	0.3071
2005	184937	116,921.80	62,648.10	0.3388
2006	216314	140,974.00	77,597.20	0.3587
2007	265810	166,863.70	93,563.60	0.3520
2008	314045	179,921.47	100,394.94	0.3197
2009	340903	150,648.06	82,029.69	0.2406
2010	401513	201,722.15	107,022.84	0.2665
2011	473104	236,401.99	123,240.60	0.2605
2012	519470	244,160.21	129,359.25	0.2490

3.2 ERROR CORRECTION MODEL

Firstly, we carried out the unit root test. Test results presented that LGDP and MCB sequences were not stabilized without difference. However, after first order difference, unit roots of sequences were eliminated and sequences were stabilized. Therefore, a co-integration relationship might exist between LGDP and MCB, and the error correction model could be established. Items of error correction were defined first, with LGDP as dependent variable and MCB as independent variables. On this basis, the least squares method was adopted to generate the regression coefficient of 12.5672. Therefore, GENR was used for definition:

$$ECM1 = LGDP(-1) - 12.5672 \cdot MCB(-1),$$

$$ECM2 = LGDP(-1) - 0.7384 \cdot LEP(-1),$$

$$ECM3 = LGDP(-1) - 0.7563 \cdot LIP(-1),$$

$$ECM4 = LEP(-1) - 1.0243 \cdot LIP(-1).$$

Next, with $d(LGDP)$ as dependent variable, ECM and $d(MCB)$, $d(LIP)$ and $d(LEP)$ as independent variables, we

carried out the least square regression. Results are shown in Table 2. According to the results, it can be found that the tailed probability of MCB and GDP exceeded 0.1 significantly. Therefore, MCB and GDP were not correlated. However, the tailed probability of LIP, LEP and GDP was much smaller than 0.01, showing strong correlation. The insignificant correlation between MCB and GDP indicated that the relationship between MCB and economy was not a long-term effect. It might also impose an effect so that could not be reflected in the model under the strong effects of other factors. It also manifested that the promoting or hindering influence of MCB on the economy was limited. LEP was significantly and positively correlated with LGDP, demonstrating that exports promoted the economic growth or economic growth facilitated exports. The significantly positive correlation between LIP and LGDP presented that imports greatly promoted the economic growth or economic growth promoted imports. The significant correlation between imports and exports verified the existence of a positive relationship between imports and exports. Or else, imports were conducive to exports or exports were conducive to imports. The relationship should be further verified through Granger test.

TABLE 2 ECM estimation and relevant test results

No.	Variable	Coefficient	Std. Error	t-Statistic	Prob.	R-squared	Adjusted R-squared
1	D(MCB)	-0.2596	0.1850	-1.4031	0.1705	0.0652	0.0048
	ECM1	0.0080	0.0094	0.8499	0.4019		
	C1	0.0640	0.0977	0.6545	0.5176		
2	D(LEP)	0.1835	0.0547	3.3535	0.0021	0.4360	0.3996
	ECM2	-0.2104	0.0577	-3.6487	0.0010		
	C2	0.9820	0.2391	4.1071	0.0003		
3	D(LIP)	0.2208	0.0502	4.3956	0.0001	0.5510	0.5221
	ECM3	-0.2060	0.0495	-4.1637	0.0002		
	C3	0.8195	0.1728	4.7439	0.0000		
4	D(LIP)	0.8826	0.0660	13.3783	0.0000	0.8784	0.8705
	ECM4	-0.5331	0.1456	-3.6614	0.0009		
	C4	-0.4569	0.1300	-3.5131	0.0014		

3.3 GRANGER TEST

Granger test is a way to measure the causal relationship between variables, and an essential requirement to establish the VAR model and impulse response function. We thus conducted Granger test of LGDP, LIP, LEP and MCB sequences. Through several times of tests, it was found that time-delayed first-order showed the optimal performance. Results are shown in Table 3. According to the table, GDP is the reason of changes in MCB. Besides,

LIP and LEP generate changes in LGDP and LIP is the cause leading to changes in LEP.

According to the error correction model and Granger test, economic growth can give rise to changes in MCB to a certain degree. However, their correlation is not significant so that the specific direction was unable to be determined. In addition, imports and exports positively promoted the economic growth, indicating that any forms of trades impose certain impact on the economy. It is unscientific to simply define trade surplus or trade deficit as a harmful thing. Imports are beneficial to the

domestically technological development, management improvement and to the compensation of domestic disadvantages. For example, China lacks energy resources so that imports of energy are of great importance for the economic development in China compared with other developed countries. Besides, a country can participate in the international competition through imports and exports, which could improve the social division of labour and reflect the comparative advantage of a region, such as the advantageous labour power in China and the advantageous biotechnology of U.S.A. Meanwhile, imports greatly promote exports, which can be attributed to the highly limited productivity of each country or region. When a region does not produce one kind of products, the production of another kind of products will be enhanced, thus leading to the maintenance or increase of the productivity in the entire region. This reflects the theory of comparative advantage and the theory of social division of labour in a region. Furthermore, imports of excellent technologies and equipment remarkably promote the development of regional productivity, as well as the growth of imports and exports to a certain extent.

effects of imports on exports also commence at a high level, and remarkably and positively promote exports. However, the effects would gradually decrease, showing a trend of returning to zero.

TABLE 3 Granger test results of LGDP, LIP, LEP and MCB

Null Hypothesis:	F-Statistic	Prob.
LGDP does not Granger Cause MCB	3.6383	0.0658
MCB does not Granger Cause LGDP	0.8026	0.3772
LGDP does not Granger Cause LIP	0.1381	0.7127
LIP does not Granger Cause LGDP	11.5369	0.0019
LEP does not Granger Cause LIP	1.5909	0.2166
LIP does not Granger Cause LEP	6.8796	0.0134
LEP does not Granger Cause LGDP	9.3653	0.0045
LGDP does not Granger Cause LEP	0.0011	0.9736

3.4 IMPULSE RESPONSE FUNCTION

Based on results of Granger test, the VAR model and a diagram of impulse response function were established, as shown in Figure 1-4.

According to Figure 1, the impact of economic growth on MCB is very complex, and the complexity negatively influences the model accuracy. In the figure, when an impact is imposed on GDP, its effects on MCB are not immediate but decreasing first, rising and then reducing to zero. We assumed that MCB is a positive value at first. In other words, when the volume of exports is larger than that of imports, the economic growth would reduce the gap between exports and imports, and then expand the gap. Eventually, effects of the economic growth would gradually disappear. The time limit of the effects is approximately 7 years. It can be seen from Figure 2 that exports could dramatically drive the economic growth. The effects are not formed suddenly but increased gradually. In Figure 3, firstly, effects of imports on economic growth commence at a high level, and keep enhancing until reaching the maximum value, and then gradually decrease, showing a tendency of returning to zero. This pattern indicates that effects of imports cannot last long as effects of exports. According to Figure 4,

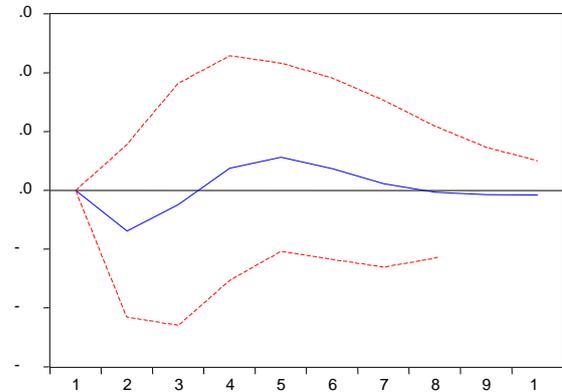


FIGURE 1 Pulse response of MCB to GDP

Response of LGDP to Cholesky
One S.D. LEP Innovation

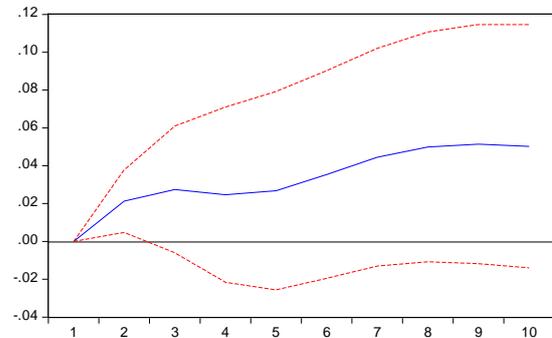


FIGURE 2 Pulse response of GDP to EP

Response of LGDP to Cholesky
One S.D. LIP Innovation

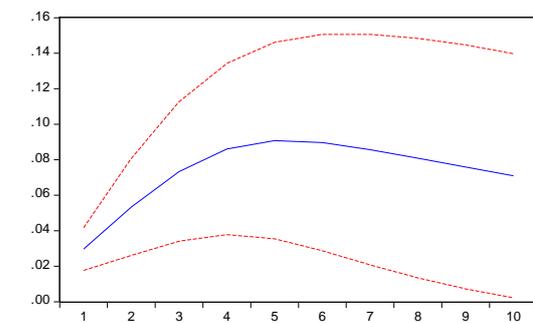


FIGURE 3 Pulse response of GDP to IP

Response of LEP to Cholesky
One S.D. LIP Innovation

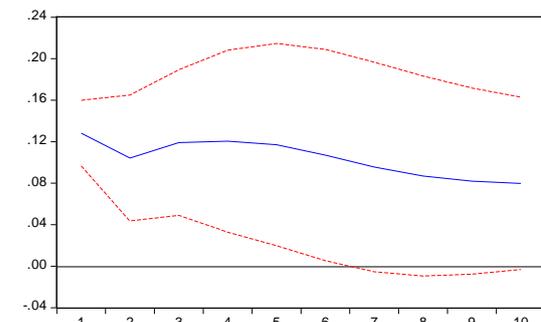


FIGURE 4 Pulse response of EP to IP

4 Analysis of empirical findings

4.1 RELATIONSHIP BETWEEN GROWTH AND MCB

Empirical findings verified that the impact of trade imbalance on economic growth is highly limited, which cannot be proved based on analytic demonstration through error correction model, Granger test and impulse response functions. This indicated that it is useless for any country to rely on trade surplus to promote the economic growth. Besides, the accusation of China's adoption of trade surplus for high-speed economic development made by developed countries is not scientifically justified. Contrarily, economic growth influences MCB, showing a pattern of decreasing first, followed by rising. This demonstrates that MCB is correlated with the status of economic development. With the rapid development of China's economy and the increase of productivity, trade surplus has become a trend. At the initial stage of the economic development in developed countries, the trade surplus was remarkable. However, due to the currently constrained rate of economic development, the duration of effects of economic growth on MCB is also limited. Meanwhile, developing countries develop rapidly, leading to a change in the direction of trade imbalance. To this end, trade surplus is a trend of economic development, rather than a means of achieving economic growth by developing countries. With the comprehensive analysis, the status of trade imbalance is determined by the following aspects.

First of all, it is affected by the international economic division of labour. For the trade condition of a country, factors that dominated the trade condition, such as outputs or inputs of technology, equipment, capitals, resources and talents are subject to the international economic division of labour. If a region has abundant resources like rare earth, the region will inevitably output rare earth. Similarly, if a region has rich capitals, in addition to domestic investment, the capital would be substantially exported to other countries. Contrarily, if a region lacks technology, imports of technology would dominate the region. If a region lacks capitals, foreign capital will be imported to the region. Under such circumstances, the international economic division of labour is formed, which determines the trade pattern. For example, when capital is scarce and resources are abundant in a country, international capitals will be imported and resources are applied to make products. In this way, the product exports exceed imports, forming trade surplus.

Secondly, it is affected by the status of region in the arena of international economy. According to the trade imbalance process of China, at the initial stage of implementing the reform and opening-up policy, this economically underdeveloped country primarily imported international technology, equipment, talent and capital, with less exports. On this basis, the situation of trade deficit was formed. With the economic development, a wide range of international technology, equipment, talent and capital have been introduced into China and become

relatively saturated. Exports of products manufactured by these techniques and equipment changed the status of trade and formed trade surplus.

Thirdly, advantages and disadvantages of a region are relative to the control country. For two countries, the status of trade imbalance should be determined by advantages and disadvantages of two countries. Therefore, for China and U.S.A, the advantage of labour force in China is very obvious, so is the disadvantage of technology, equipment and other factors in China. To this end, China primarily exported talents and products, while U.S.A focusing on exports of technology and equipment. Moreover, the labor force in China is highly abundant compared with the limited exports of technology and equipment in U.S.A, thus forming a long period of trade surplus.

4.2 RELATIONSHIP BETWEEN IMPORTS AND ECONOMIC GROWTH

Trade imbalance has little impact on economic growth, but imports and exports positively promote the economic development, indicating that the influence of imports or exports on the economy is monumental. This also points out that both imports and exports remarkably promote the economic development. It is unscientific to adopt strategies and views about the promoting role of larger volume of imports or exports in promoting the economic development. According to the respective impact of imports and exports on the economic development, imports often impose a greater impact on the economy at the initial stage but exports' influence on the economy last longer. In general, the influence should be balanced. For China, the country remained a state of trade deficit at the beginning of implementing the reform and opening policy. This trend is conducive to improving the technological level of China, level of productivity and promoting the economic development. With the technological advancement and increased productivity, effects of exports would continue, thus gradually enhancing the exports. On this basis, the influencing trend of imports and exports on the economy can be verified. With the rapid economic development, the dependency of a region on imports of advanced technology and designs decreases, gradually reaching the advanced level. At this moment, imports cannot further improve the level of productivity. The constrained rate of productivity increases also leads to the limitation of exports, thus altering the trade tendency. To this end, with the slowdown of the economic growth rate in China, exports may be constrained and the development tendency of trades may be altered, making China a country with trade deficit. Hence, China should gradually expand the domestic demand to relieve the trade crisis. Therefore, both imports and exports positively promote the economy. This is attributed to the following aspects:

First of all, imports can supplement the inadequate aspects of a region. At the beginning of implementing China's reform and opening up policy, the county was faced with insufficient productivity, technology,

equipment and capital. Therefore, at that moment, China primarily imported the above mentioned elements. Even now, China still faces the problem of insufficient capital and technology. Therefore, China needs to purchase or import the advanced production technologies from other countries at a frequent basis.

Secondly, imports are beneficial to promoting the renewal of regional technology. For one thing, imports of advanced technology and equipment could improve the level of productivity in China. For another, imports of foreign products could promote both domestic products and foreign products, thus facilitating the technological upgrades of Chinese enterprises.

Thirdly, exports are the redistribution of regional resources. If a region has abundant capitals, exports of capital could redistribute the regional capitals. If a region has rich labour resources, talents can be exported to realize the redistribution of labour power. If a region has sufficient material resources, the region can adopt exports of material exports, thus realizing the redistribution of material resources. The redistribution of material resources can improve the utilization efficiency of resources and the rapid development of regional economy.

4.3 RELATIONSHIP BETWEEN EXPORTS AND EXPORTS AND IMPORTS

According to the empirical analysis, imports positively promote exports, yet showing a gradual decreasing influence. For China, the initial imports were conducive to the technological advancement. Besides, the introduction of advanced equipment improved the then level of productivity, thus promoting the economic growth of China. For example, the manufacturing industry in China was a primary industry for exports, and its initial equipment and production technologies were imported. However, exports could not influence imports. Thus, the unidirectional relationship between imports and exports demonstrates the importance of imports. Thus, the initial trade deficit in China was a key process for the later economic growth at a fast pace. When a region maintains trade deficit for a long period, the economic development is hindered to a certain extent because of an insufficient imports of technology and a focus on exports of products. At this moment, if the technological innovation of the region is not effective, the economic development will be confined. This can be attributed to one of the reasons for the gradual slowdown of the current economic growth rate in China. Hence, the view held by developed countries on China's benefiting from trade surplus is not supported by scientific proofs. The trade surplus of China primarily originates from the insufficiency in technological innovation, imports of technology, and exports of cheap products from the manufacturing industry. As a result, the long-term trade surplus is harmful for the economic development of China. Specifically, the relationship between imports and exports is mainly caused by the following aspects:

Firstly, imports for insufficiencies. From a microscopic perspective, the purchasing choice between domestic technology and foreign technology made by an enterprise is subject to price as well as the comparison between domestic and foreign technologies. In other words, if the domestic technologies are equivalent or slightly inferior to foreign technologies, an enterprise would give the priority of purchases to domestic technologies. From the macroscopic aspect, insufficiencies are the principal drive of imports. If a region has insufficiencies in product imports, imports of technologies and talents could compensate for the insufficiencies, thus promoting the rapid economic development.

Secondly, exports for surplus. According to the exporting status of China, most exported products are abundant commodities in China. In other words, under general circumstances, products of a region should satisfy the needs of the region. Under the premise of satisfying the domestic needs, products would be exported to foreign markets.

Thirdly, the regional macro-control cannot ensure the completely clear trade passage. With the traffic development, gaps of prices between domestic products would be narrowed and the market is greatly intervened by the macro-control of other countries. For one thing, a country needs to protect the domestic products to a certain extent. For another, the exports of domestically insufficient resources need to be prevented so as to avoid the influence on civil life and economic development of the region.

5 Policy suggestions

According to the above analysis, it can be found that the influence of trade imbalance on the economy is limited, and the long-term trade surplus reflects the insufficiency of technological innovation in China. This insufficiency partly originates from the insufficient imports of advanced technologies from other countries and partly stems from the limited technological reform of China. Moreover, with the economic development, the tendency of foreign trade in China would gradually change from trade surplus to trade deficit. On this basis, in order to promote the economic development, the government may proceed from two aspects:

Firstly, vigorously promoting domestic demand and easing trade conflicts. This measure is proposed based on three factors. Firstly, in a recessing concept of great nation, the long-term trade surplus will cause remarkable trade conflicts and hinder the economic development; Secondly, with the decreasing rate of China's economic development, the altering of trade tendency becomes an inevitable outcome. Since the reliance on foreign trade would lead to altered process of the economic development. However, the promotion of domestic demands can reduce the reliance on other countries, thus eliminating the influence of the alteration on the economy.

Secondly, highlighting the increase of productivity and changing the currently embarrassing trade situation. In recent years, developed countries have been accusing the commodity dump of China. This is mainly attributed the concentration of China’s exports on labour-intensive manufacturing industries. At present, exports in China also depend on the advantage of labour force, which is gradually diminishing. To change the situation, the productivity of China should be improved. To realize the goal, China needs to enhance the domestic innovation and imports of advanced technology and equipment. Nowadays, China primarily exports products at a low level of productive forces. Under this context, profit margin is low and the resource protection in China may be undermined. Furthermore, China cannot obtain the competitive edge in trades, and more significant trade conflicts may be generated.

Thirdly, enhance communications with other countries. All countries should pursue the true nature of trade imbalance, and China needs to clarify the view on trade

imbalance. International laws and regulations should be applied to protect the legal rights and interests in trades between China and other countries. In this way, China can avoid influence and harms on legal trade freedom and economic development of China from sophism of other countries. Some developed countries attribute the long-term trade surplus as an intentional measure of China, and jeopardize the smooth trades of China by anti-dump cases. To this end, China should adopt reasonable and legal measures to protect the normalization of trades with other countries.

Acknowledgement

This material is based upon work funded by Zhejiang Provincial Natural Science Foundation of China under Grant No.LQ14G030015, and MOE (Ministry of Education in China) Project of Humanities and Social Sciences (No.13YJC790137 and 14YJC790163).

References

[1] Lang L 2006 Reasons, impacts and policy of trade unbalance between China and U.S.A *Research on Economics and Management* 07 31-49

[2] Wu J 2009 Research on China’s Trade Surplus in the Context of Global Economy Imbalance *Master Dissertation of Southwestern University of Finance and Economics Nanjing China* 21-2 (in Chinese)

[3] Wu Y 2008 Research on China’s Trade Imbalance in the Context of Global Economy Imbalance *Master Dissertation of Xiamen University Xiamen China* 13-8 (in Chinese)

[4] Jia L, Yang H 2013 The effects of global trade imbalances to the international economic order and its countermeasure *Journal of Heman Institute of Science and Technology (Social Sciences Edition)* 03 6-9 (in Chinese)

[5] Chu J 2006 Reasons and Countermeasures of Trade Imbalance between China and U.S.A *Master Dissertation of Capital University of Economics and Business Beijing China* 27-78 (in Chinese)

[6] Lu W 2009 Reasons and trends of trade surplus in China – a new perspective based on international division of labour for products in East Asia *International Trade Issue* 07 61-70 (in Chinese)

Authors	
	<p>Chong Qian, born in October, 1983, Wenzhou city, Zhejiang province, China</p> <p>Current position, grades: the lecturer of City College of Wenzhou University, Wenzhou city, Zhenjiang province, China. Scientific interest: international finance. Publications: 2 papers. Experience: teaching experience of 4 years, 2 scientific research projects.</p>
	<p>Zhan-ao Wang, born in October, 1982, Wenzhou city, Zhejiang province, China</p> <p>Current position, grades: the lecturer of City College of Wenzhou University, Wenzhou city, Zhenjiang province, China. Scientific interest: world economy and international electronic commerce. Publications: 1 paper. Experience: teaching experience of 3 years, 2 scientific research projects.</p>

Nonlinear time series of deformation forecasting using improved BP neural networks

Cai-yun Gao^{1, 2*}, Xi-min Cui¹

¹College Of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing 100083, China

²College of Surveying and Mapping Engineering, Henan University of Urban Construction, Pingdingshan 467044, China

Received 6 July 2014, www.tsi.lv

Abstract

Although the back propagation neural network has been successfully employed in various fields and demonstrated promising results, literatures show its performance still could be improved. Therefore, we present a comprehensive comparison study on the application of different BP algorithm in time series of deformation forecasting. Four types of typical improved BP algorithm, namely, momentum, conjugate gradient, Quasi-Newton and Levenberg-marquardt algorithms, are investigated. An illustrative example of high-rise building settlement deformation is adopted for demonstration. Results show that the improved BP algorithms can increase the prediction accuracy and have faster convergence speed.

Keywords: artificial neural networks, back propagation, deformation forecasting, learning rate, convergence, improved

1 Introduction

Artificial neural networks are often used to model and forecast complicated nonlinear time series, especially deformation systems (such as dam deformation, land subsidence in coal mine, settlement of subway tunnel etc.). Theoretically, multilayer feed-forward neural networks can accurately approximate almost any nonlinear function and thus error back propagation algorithm (BP) have been widely researched and applied in time series of deformation forecasting [1-5].

Although there are many successful applications of standard BP algorithm, it has many drawbacks, such as:

- 1) require a long time to train the networks,
 - 2) depending on the choice of the initial weight and number of hidden neurons,
 - 3) being sensitive to the learning rate,
 - 4) having poor generalization for complicated nonlinear functions [6-14].
- To overcome these drawbacks, a large number of researchers concentrate upon improvements of BP in two aspects. On the one hand, a number of researchers focus mainly on improvements of based on standard gradient descent, including automatically adjusting the learning rate algorithm as training, additional momentum factor, and resilient of BP algorithm etc. On the other hand, a number of researchers focus mainly on based on the standard numerical optimization, including Quasi-Newton algorithm, LM (Levenberg-marquardt) algorithm. For example, the Quasi-Newton training algorithm improved the convergence rate of the standard BP algorithm but requires computing the Hessian matrix, so this leads to a large computational burden and storage expense [1-14]. In this paper we

present a comprehensive comparison study on the application of different BP algorithm in time series of deformation forecasting, four types of typical improved BP algorithm, namely, momentum, conjugate gradient, Quasi-Newton and Levenberg-Marquardt algorithms, are investigated.

The remaining paper is organized as follows. In Section 2, the fundamentals of standard back propagation are introduced and discussed. Based on the concepts in Section 2, the improved BP algorithms are presented in Section 3. Then, an illustrative example of high-rise building settlement deformation is adopted to demonstrate the adaptability and effectiveness of the improved BP algorithms in Section 4. Finally, Section 5 concludes the paper.

2 Standard back propagation algorithm (SDBP)

Multilayer feed forward BP network is one of the most popular techniques in the field of ANN. Standard back propagation is the generalization of the Widrow-Hoff learning rule to multilayer networks and nonlinear differentiable transfer functions. The common topology of a BP neural network model is illustrated in Figure 1. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined by you [1-5].

In general, the BP algorithm includes the forward and the backward process. In the forward process, a vector is added to the input layer, which is then spread along the network, finally, an output vector is

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

obtained as a response of the input vector, in which the synaptic weights cannot be changed.

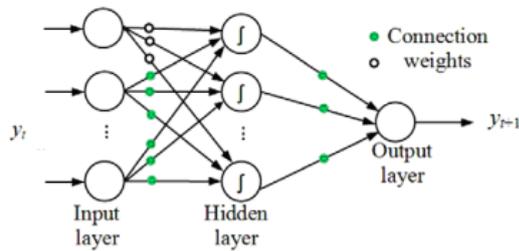


FIGURE 1 The structure of BP neural networks

Then the backward process, an error signal will be obtained by comparing the output signal with the defined output; the error signal is then forward-spread to modify the weight from one output layer to another. The forward and the backward process alternate and constantly circulate, then will be convergent with the defined output in some states. The BP algorithm is carried out as follows:

$$\hat{y}_k(t) = \sum_{j=1}^n v_{jt} \cdot f\left[\sum_{i=1}^m w_{ij} \cdot y_i(t) + \theta_j\right] + r_t, \quad (1)$$

where, f is activation function, i and j are the number of neurons of hidden layer and output layer, respectively, y_i is input vector, w_{ij} and v_{jt} are weights between the input/hidden layers and hidden/output layers, respectively, θ_j and r_t are the bias of neurons. Let $y_k(t)$ be desired output of neural network. There is an error between actual output and desired output, this error, named mean square error, can be expressed as follows:

$$E = \frac{1}{2} \sum_{k=1}^{N_1} \sum_{t=1}^n [y_k(t) - \hat{y}_k(t)]^2. \quad (2)$$

Standard back propagation is a gradient descent algorithm, as is the Widrow-Hoff learning rule, in which the network weights moved along the negative of the gradient of the performance function. The updating rules are as follows:

$$x(k+1) = x(k) - ag(k), \quad (3)$$

$$g(k) = \frac{\partial E(k)}{\partial x(k)}. \quad (4)$$

where, $\Delta x(k)$ is weight and bias matrix from the input layer to the hidden or the hidden layer to the output layer at the k -th learning step, a is the learning rate, $g(k)$ is the grads of E at the k -th learning step, $E(k)$ is the error function of the network at the k -th learning step, k is the step of training iteration.

3 Improved back propagation algorithm

There are a number of variations on the standard back propagation algorithm that are based on other optimization

techniques, such as momentum back propagation, conjugate gradient and Newton methods etc. In this section, describes the learning procedures of several improved back propagation algorithm [7-12].

3.1 MOMENTUM BACK ALGORITHM (MOBP)

Gradient descent with momentum, allows a network to respond not only to the local gradient, but also to recent trends in the error surface. MOBP depends on two training parameters: the learning rate and the amount of momentum, momentum is set between 0 (no momentum) and values close to 1 (lots of momentum). In SDBP, the learning rate α has a small value because it can decrease with a change of weighted value at the learning step. Consequently, the learning becomes very slow. The MOBP can accelerate the learning step of BP. The learning method of the MOBP is the same to BP-ANN, but it introduces additional momentum factor η , $0 < \eta < 1$.

The change of weighted value at the $k+1$ -th learning step of MOBP can be expressed as follows:

$$\Delta x(k+1) = \eta \Delta x(k) + \alpha(1-\eta) \frac{\partial E(k)}{\partial x(k)}, \quad (5)$$

where, $\Delta x(k)$ is the change of weighted values at the k -th learning step. Therefore, the relevant weighted value $x(k+1)$ is given as:

$$x(k+1) = x(k) + \Delta x(k+1). \quad (6)$$

3.2 CONJUGATE GRADIENT BACKALGORITHM (CGBP)

The SDBP algorithm adjusts the weights in the steepest descent direction. This is the direction in which the performance function is decreasing most rapidly, however, this does not necessarily produce the fastest convergence. To produces generally faster convergence than steepest descent directions, the search of MOBP algorithm is performed along conjugate directions. All of the conjugate gradient algorithms start out by searching in the steepest descent direction on the first iteration

$$p(0) = -g(0). \quad (7)$$

A line search is then performed to determine the optimal distance to move along the current search direction:

$$x(k+1) = x(k) + ap(k), \quad (8)$$

$$p(k) = -g(k) + \beta(k)p(k-1) \quad (9)$$

where, $p(k)$ is the search direction at $k+1$ -th iteration, the constant $\beta(k)$ is computed by various versions of conjugate gradient such as Fletcher-Reeves update, Polak-Ribière update, Powell-Beale Restarts, Scaled Conjugate

Gradient. Take Fletcher-Reeves of conjugate gradient for example, the formula is follows:

$$\beta(k) = \frac{g^T(k)g(k)}{g^T(k-1)g(k-1)}. \tag{10}$$

3.3 QUASI-NEWTON ALGORITHM (QNP)

Newton's method is an alternative to the conjugate gradient methods for fast optimization. The basic step of Newton's method can be expressed as follows:

$$x(k+1) = x(k) - A^{-1}(k)g(k), \tag{11}$$

where, $A(k)$ is the Hessian matrix (second derivatives) of the performance index at the current values of the weights and biases but is often complex to compute. There is a class of algorithms are called quasi-Newton (or secant) methods which is based on Newton's method, but doesn't require calculation of second derivatives. The most successful in studies is the Broyden, Fletcher, Goldfarb, and Shanno update.

3.4 LEVENBERG-MARQUARDT ALGORITHM (LMBP)

Levenberg-marquardt is the fastest method for training moderate-sized feed forward neural networks. It was designed to approach second-order training speed without having to compute the Hessian matrix. The LM algorithm's update formula is given as:

$$x(k+1) = x(k) - (J^T J + \mu J)^{-1} J^T e, \tag{12}$$

where, J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors.

4 Experimental results

To demonstrate the effectiveness of the SDBP and improved BP models, we use deformation time series of a high-rise building as an illustrating example, the data in Figure 2 quoted from reference literature [15]. In this study, we use the historical deformation from 1 to 36 as our research data. There are 36 observations, where 1-30 are used for model fitting and 31-36 are reserved for testing.

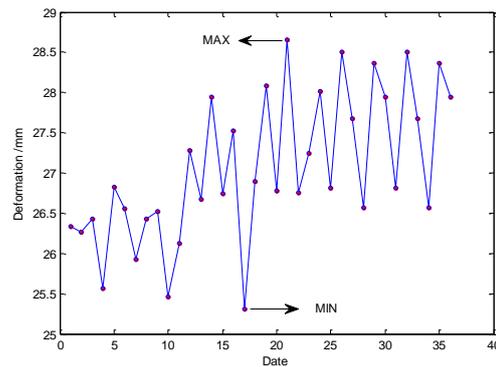


FIGURE 2 The deformation data set

4.1 COMPARISON BETWEEN CONVERGENCE OF FIVE BP MODELS

It has showed in Section 2 and 3 that the BP network has great differences in learning rate, convergence speed and iteration times under the different learning algorithm. For prediction of deformation time series, we can't make sure that in which learning algorithm, the forecast effect of BP network could reach the optimum, and therefore, it is necessary to discuss how the BP networks influence prediction accuracy in different learning algorithm. In calculation, set the preconditions as follows:

The neural network consists of three layers, i.e., input layer, hidden layer, and output layer. The neuron number of the input layer is selected to be 5, and the input to the input layer is a vector of the known historical deformation values, which are upgraded constantly by a fixed-size sliding data window according to the delay deformation time obtained newly. Through the optimization of one-dimensional region search algorithm of BP network [5], the neuron number of the hidden layer is selected to be 13, the output layer has only one neuron, and its output is just the predicted deformation time delay. The activation function of a node is a sigmoid function. Figure 3-7 shows the convergence curve simulated by matlab.

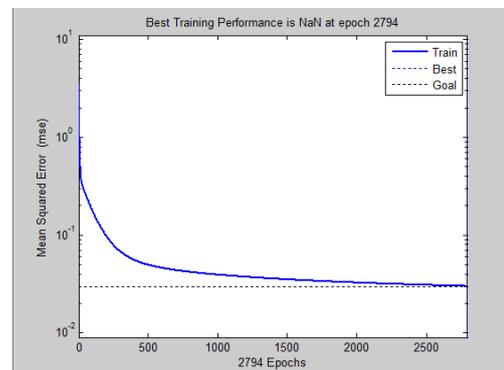


FIGURE 3 Effect of training speed according to SDBP

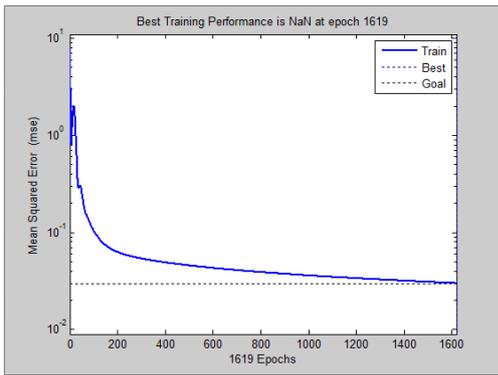


FIGURE 4 Effect of training speed according to MOBP

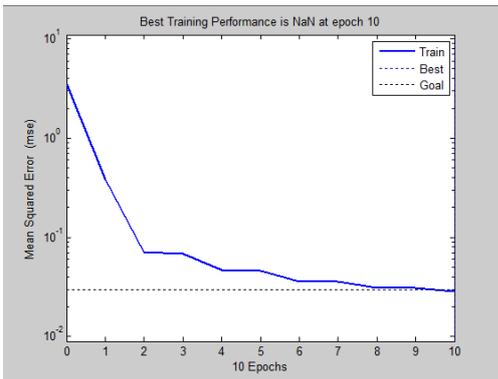


FIGURE 5 Effect of training speed according to CGBP

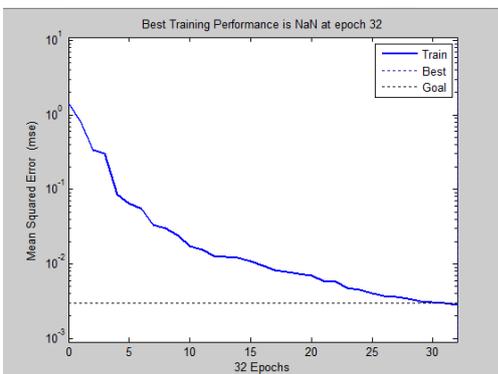


FIGURE 6 Effect of training speed according to QNBP

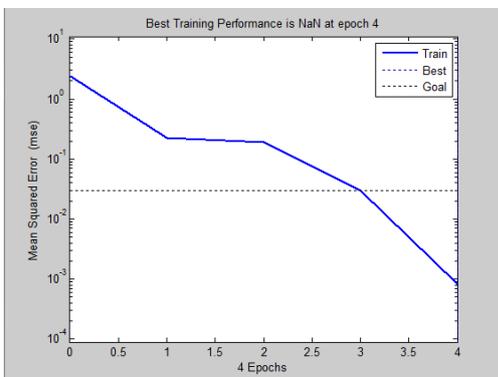


FIGURE 7 Effect of training speed according to LMBP

From the results, it is seen that the training convergence speed of the MOBP, CGBP, QNBP and LMBP are obviously faster than that of the SDBP. It is obvious that the convergence speed is greatly improved.

4.2 COMPARISON BETWEEN THE PREDICTION OF FIVE BP MODELS

Data from 1 to 30 is simulated and data of 31 and 36 is the prediction. Then, error of simulation and prediction can be calculated, and the results are in Tables 1 and 2. The simulation and prediction results show that improvement BP models are higher than SDBP.

TABLE 1 Simulation error of five BP models (units: mm)

	SDBP	MOBP	CGBP	QNBP	LMBP
AAE	0.6009	0.5177	0.5273	0.5384	0.3049
ARE	0.0224	0.0193	0.0195	0.0199	0.0114

(AAE: Average absolute error; ARE: Average relative error)

TABLE 2 Prediction error of five BP models (units: mm)

Date	SDBP	MOBP	CGBP	QNBP	LMBP
31	0.0691	0.4867	0.4220	0.4324	0.1359
32	0.1605	0.0299	0.3491	0.4808	0.3398
33	-0.2013	0.1906	0.1445	0.1548	-0.008
34	-0.5593	0.2535	0.0449	0.1042	0.5342
35	-0.2878	0.2827	0.1984	0.5736	0.2093
36	0.7427	0.39240	0.4916	0.2803	0.1029
AAE	0.3368	0.2727	0.2751	0.3377	0.2216
ARE	0.0122	0.0099	0.0099	0.0121	0.0081

(Note: AAE, Average Absolute Error; ARE: Average Relative Error)

5 Conclusions

From the results in this study, the following three conclusions can be drawn:

1) ANN is a jumped-up interceptive subject, using it to predict deformation time series, it is feasible and practically.

2) The improvement BP learning algorithm is much better than standard BP algorithm for deformation time series prediction. The LM algorithm's convergence rate is the quickest one and has higher prediction precision.

3) The BP neural network under the different learning algorithm has different characteristics in network training, when we use BP neural network to do the modelling prediction, we should select the best learning algorithm to set up the network according to the actual situation of prediction problem.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (41071328) and Key Laboratory of Mine Spatial Information Technologies of SBSM (State Bureau of Surveying and Mapping) (KLM201306).

References

- [1] Xu H, Li G 2005 Application of BP neural network to observed data processing for dam based on matlab *Engineering Journal of Wuhan University* **38**(3) 50-4
- [2] Wu Y P, Teng W F, Li Y W 2007 Application of grey neural network model to landslide deformation prediction *Journal of Rock Mechanics and Engineering* **26**(3) 632-6
- [3] Peng X G, Wang J, Sun L H 2009 Application of optimal dimension modeling for BP network in prediction of times of deformation *Journal of Geodesy and Geodynamics* **29**(6) 113-6
- [4] Cao Y B, Yang E C, Xie L F 2012 Study of landslide deformation prediction based on gray model-evolutionary neural network model considering function of environmental variables *Rock and Soil Mechanics* **33**(3) 847-52
- [5] Gao C Y, Gao N 2012 Comparison with RBF and BP neural network used in deformation prediction *Yellow River* **34**(7)131-4
- [6] Gao N, Cui X M, Gao C Y 2012 An effective hybrid approach for processing deformation monitoring data *Advanced Materials Research* 446-9, 3247-51
- [7] Zhao X Y, Lai K S, Dai D M 2007 An improved BP algorithm and its application in classification of surface defects of steel plate *Journal of iron and steel research* **14**(2) 52-5
- [8] Hu Q C, Hu B, Jiang H F 2013 Application of BP artificial neural network to the displacement prediction of deep foundation pile *Safety and Environment Engineering* **20**(3) 154-8
- [9] Yi Q L, Zeng H E, Huang H F 2013 Reservoir landslide deformation forecast using BP neural network *Hydrogeology and Engineering Geology* **40**(1) 124-7
- [10] Miao X Y, Chu J K, Du X W 2011 Application of LM-BP neural network in predicting dam deformation *Computer Engineering and Applications* **47**(1) 220-2
- [11] Lera G, Pinzolas M *IEEE Transactions on Neural Networks* **13**(5) 1200-3
- [12] He Z Y, Zheng W 2008 Deformation prediction of deep foundation pit based on bp neural network *Journal of South China University of Technology* **36**(10) 92-6
- [13] Gao N, Wu L C, Gao C Y 2006 The problems of GPS height conversion by artificial neural network method *Engineering of Surveying and Mapping* **15**(3) 67-9
- [14] Hu W S, Sha Y J 2003 Converting GPS height by a new method based on neural network *Journal of Southeast University(English Edition)* **19**(1) 53-7
- [15] Gao C Y, Pan C J, Gao N 2012 An effective combined approach based on GM and AR for deformation analysis of high-rise buildings *Advanced Materials Research* 368-73 2123-7

Authors



Cai-Yun Gao, born in October, 1980, Jinzhou County, Hebei Province, China

Current position, grades: Ph. D. student of college of geoscience and surveying engineering, China University of Mining and Technology, Beijing, China.
University studies: Master of Engineering in Geodesy and Surveying Engineering from East China Institute Of Technology in China. Bachelor of Engineering in Surveying and mapping engineering from Institutes Of Technology Of Hebei in China.
Scientific interest: intelligence algorithm in the application of geodesy, surveying engineering, disaster prediction.
Publications: 15 papers.
Experience: teaching experience of 8 years, 6 scientific research projects.



Xi-Min Cui, born in June, 1967, KuanDian County, LiaoNing Province, China

Current position, grades: the professor of College Of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing.
University studies: Doctor of Engineering from China University of Mining and Technology (Beijing) in China.
Scientific interest: deformation monitoring and surveying and mapping data processing.
Publications: 50 papers.
Experience: teaching experience of 30 years, 15 scientific research projects.

Relative humidity prediction of northern greenhouse environmental factors on the basis of a radial basis function neural network

Chunling Chen^{1, 2*}, Long Wang², Tongyu Xu², Jiawei Qi²

¹Key Laboratory of Protector Horticulture, Ministry of Education, Shenyang Agriculture University, 120 Dong Ling Road, Shenyang, China

²School of Information and Electrical Engineering, Shenyang Agriculture University, 120 Dong Ling Road, Shenyang, China

Received 10 July 2014, www.tsi.lv

Abstract

With its advantages of abundant resource, popularity, and efficiency, solar greenhouse is the only type of greenhouse that is widely used in Northern China. This study proposes a simulation prediction model that is based on a radial basis function artificial neural network. This model is suitable for dealing with humidity in northern solar greenhouses. We select 600 groups of training data to establish the network model and to verify its accuracy. We then randomly select 80 groups for validation. With a 7.35% average error rate, the prediction model shows satisfactory performance. Thus, the results can be used to predict the relative humidity curve in a greenhouse, as well as provide a scientific basis for reasonable regulation and control of a greenhouse environment.

Keywords: solar greenhouse, relative humidity, predict model, radial basis function neural network

1 Introduction

Daily food is important for Chinese people. In Liaoning Province, people traditionally preserve a sufficient amount of Chinese cabbage and radish before winter. Solar greenhouses enable people in Liaoning to buy various vegetables during winter. A solar greenhouse is the only type of greenhouse that is widely used in Northern China. Given their large area, low cost, simple structure, high light transmittance, thermal insulation, and heat storage capacity, solar greenhouses have become suitable for the cold weather in Northern China. Almost 30% of solar greenhouses in China are built in Liaoning Province, and these greenhouses have already improved the production capability of main fruits under the northern weather conditions. Furthermore, solar greenhouses guarantee the supply of winter vegetables in China. The production efficiency of solar greenhouses is also the largest agricultural planting benefit industry in more than 20 years [1].

Air humidity is an important environmental factor affecting greenhouses. High and low-humidity environments are unsuitable for crop growth [16]. To improve crop growth and prevent diseases, people must understand indoor humidity changing rules, as well as appropriate forecasting and reasonable measurements for adjustment. Considerable research has focused on greenhouse humidity forecasting models. Guo Zhenghao [2] established the solar greenhouse humidity of the air dynamic prediction model for Northern China on the basis of the heat balance equation and the water quality

dynamic balance relationship. According to the moisture balance inside the greenhouse, He Fen et al.[3] established the greenhouse dynamic prediction model under indoor and outdoor meteorological conditions and greenhouse structure. Guo Qingchun et al. [4] introduced a relative humidity prediction model on the basis of a back propagation (BP) artificial neural network. Xue Xiaoping established the soil moisture forecast model on the basis of the support vector machine method.

This study proposes a prediction model of greenhouse relative humidity on the basis of a radial basis function (RBF) artificial neural network and previous research in Northern China. The remainder of this paper is organized as follows: Section 1 introduces the prediction model. Section 2 presents the test data analysis of prediction results. Section 3 discusses the model testing and validation. Finally, the discussion and conclusion are presented.

2 Selection of neural network

An artificial neural network [5] imitates the structure and function of a biological neural network on the basis of nonlinear mathematical models. This network contains a set of input and output units that are connected with weight. A neural network can be divided into a forward neural network, a feedback neural network, a self-organizing neural network, and so on. A neural network can be used for prediction, classification, and pattern recognition. BP and RBF neural networks are the most popular network models.

* Corresponding author e-mail: sncc1@163.com

2.1 BP NEURAL NETWORK

A BP neural network is a network model of the error BP algorithm. Such network is a one-way transmission of a multi-layer forward neural network, which has high non-linear mapping capability, self-learning and adaptive capability, generalization capability, and fault tolerance. The structure of the network is shown in Figure 1.

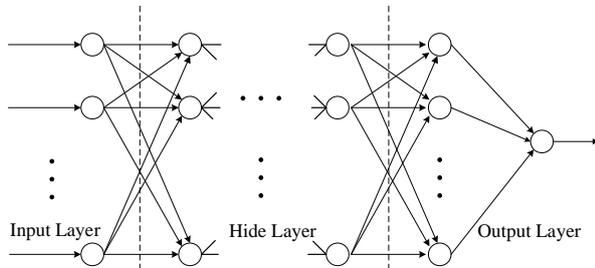


FIGURE 1 Structure of BP neural network

The BP algorithm, which is based on the gradient descent algorithm, is an instructor learning algorithm. The main idea is that the learning process is divided into two stages. The first stage (mode propagation) is input information flow through the input, hidden, and output layers through a layer-upon-layer transfer process. The second stage (error BP) obtains the desired output value and the error signal along the original path layer BP, as well as adjusts the weights and threshold. The BP algorithm has a slow convergence speed, local minimum, predictive capability, and training capability of contradiction.

2.2 RBF NEURAL NETWORK

An RBF neural network is a feed-forward network with satisfactory performance. With the use of traditional techniques of interpolation in multidimensional space, this network can perform identification and modelling on almost all systems. An RBF neural network is based on the RBF with a hidden layer unit base, hidden layer, hidden layer to the input vector transform, and pattern of low-dimensional transform input data into a high-dimensional space. Hence, this network enables the linear low-dimensional space, non-separable problem to be separated in a linear high-dimensional space.

An RBF network is a three-layer feed forward network. The mapping from input to output is nonlinear, whereas the mapping from the hidden layer space to the output space is linear. The structure of the RBF network is shown in Figure 2.

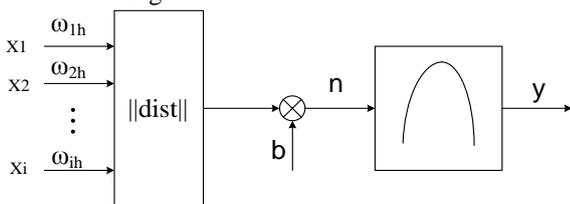


FIGURE 2 Structure of RBF neural network

2.3 COMPARISON AND SELECTION

Poggio and Girosi proved that the RBF network is the best approximation of continuous functions. With local activation function, RBF networks have more advantages than BP networks in terms of convergence speed. The learning method of RBF avoids the local optimal solution. Thus, an RBF network can approximate any nonlinear function with arbitrary precision after a full study of a sufficient number of hidden layer nodes. In addition, this network has the approximation capability for fast convergence rate and strong capability to resist noise, aside from its repair capacity.

After comparing the advantages and limitations of the two networks, we set the RBF neural network as the predictive network and the network activation function as the Gauss RBF. This relationship can be expressed as:

$$h(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right). \tag{1}$$

3 Test data selection of prediction

3.1 OUTLINE OF EXPERIMENTAL GREENHOUSE

The typical northern greenhouse (123.57°E, 41.83°N) has the following dimensions: shoulder height, 1.2 m; ridge height, 2.8 m; span, 8 m; and length, 50 m. The greenhouse has no indoor heating or ventilation equipment.

Data acquisition was performed on 22 to 24 May 2014. The sampling frequency was 1 time/min. The acquisition parameters were as follows: indoor temperature, indoor relative humidity, greenhouse light intensity, greenhouse soil temperature, dew point temperature, CO₂ concentration, outdoor temperature, and humidity. Acquisition equipment was as follows: WEMS-RHT-3/780 wireless temperature and humidity sensor, WEMS-CO₂/780 wireless CO₂ concentration sensor, WEMS-L2/780 wireless outdoor light sensor, and WEMS-ST/780 wireless soil temperature sensor.

3.2 DATA SELECTION

A greenhouse environment has multi-factors that can interact with one another. Indoor relative humidity is related to crop photosynthesis, crop transpiration, irrigation condition, indoor temperature, ventilation, and other processes.

Principal component analyses focus on condensing large original variables into a few factors with minimal information loss and on enabling the factor to obtain a certain explanatory method of multivariate statistical analysis. Principal component analysis is a mathematical method of data dimensionality reduction. The basic idea is to recombine the numerous relevant indicators X_1, X_2, \dots, X_p (such as p indicators) to a group with a lesser number of unrelated composite indicators F_m than the

original target. He Fen [6] measured greenhouse environmental factors that affect the air humidity of data samples. Principal component analysis was conducted on the sample data. Results showed that the important factors include indoor temperature, outdoor humidity, outdoor temperature insulation, outdoor sun shade on degree, outdoor wind speed, skylight window angle, and side window opening angle.

External weather conditions influence greenhouse temperature and humidity. Xin Zhihong et al. [8] analyzed the influence of different sky conditions and different external weather conditions on the temperature and humidity inside the greenhouse. Results showed that the greenhouse thermal insulation performance is satisfactory and that the outdoor greenhouse temperature and humidity are related to the sky conditions and the variations in meteorological elements. They concluded that internal and external greenhouse humidity values possess a good positive correlation. During the closed period, relative humidity is less affected by the external factors and remains stable. The indoor and outdoor humidity difference tends to decrease during the ventilation period, indicating a positive correlation.

In sum, in cold weather, no heating and humidifying equipment are required, and greenhouses are mainly characterized by a sealed insulating state and relatively less ventilation time, while indoor water is mainly sourced for crop physiological function and irrigation. Thus, we select the indoor temperature, indoor light intensity, and dew point temperature prediction as the input factors. We select 600 groups of training data to establish the network model and to verify model accuracy. We then randomly select 80 groups for validation.

3.3 PRELIMINARY ANALYSIS OF DATA

An analysis of 22 to 24 monitoring records within 4 d revealed that indoor temperature reached the maximum at 4 p.m. to 5 p.m. The average maximum temperature was 38.41°C. The indoor temperature reached the lowest value of 10°C to 12°C. The average minimum temperature was 15.77°C. The greenhouse temperature curve is shown in Figure 3.

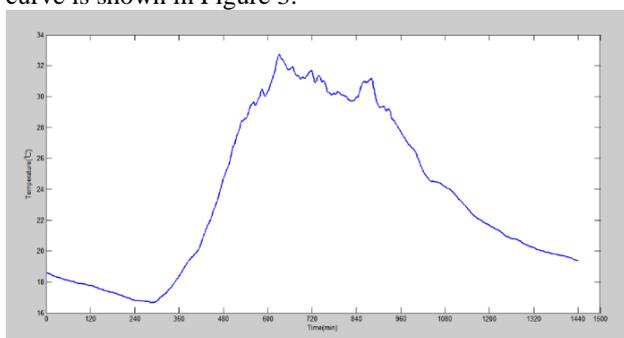


FIGURE 3 Greenhouse temperature curve

The greenhouse indoor relative humidity reached the highest value at 4 p.m. to 6 p.m. The average maximum humidity was 88.22%. The indoor relative humidity

reached the lowest value of 10°C to 15°C. The minimum average humidity was 36.44%. The greenhouse indoor humidity curve is shown in Figure 4.

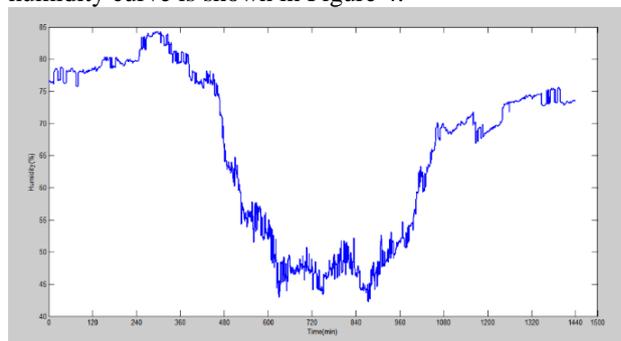


FIGURE4 Curve of humidity in greenhouse

The greenhouse light intensity reached the maximum at 9 to 11. The average maximum light intensity was 45559.375 lux. The greenhouse light curve is shown in Figure 5.

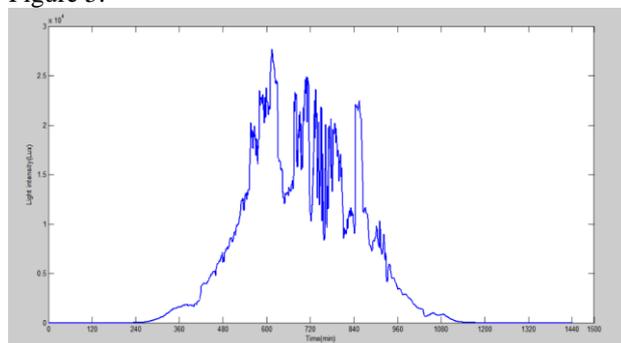


FIGURE 5 Curve of light intensity in greenhouse

By comparing the greenhouse temperature, relative humidity, and light intensity, the indoor temperature was found to correlate negatively with humidity. The indoor relative humidity and light intensity were also negatively correlated with each other. A comparison of the greenhouse indoor temperature, relative humidity, and light intensity curves is shown in Figure 6.

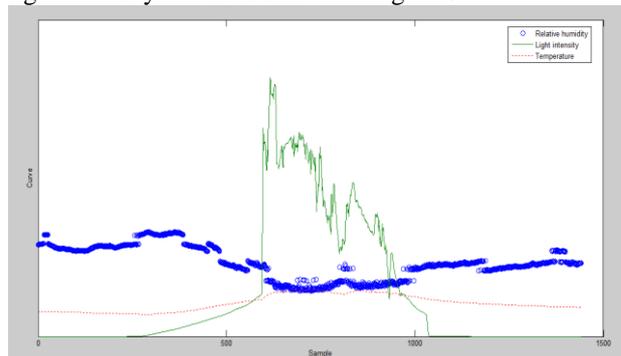


FIGURE 6 Comparison of humidity and temperature or light intensity

4 Model testing and validation

By using MATLAB program with 0.01 mean square error and 0.8 propagation velocity of RBF, we input the sample and set up the network. We then simulate and test the model. The simulation and actual values of the average

relative error were both 0.19%, and the predictive and actual values of the average relative error were both 7.35%. The fitting contrast diagram of the network simulation value and actual value is shown in Figure 7. The fitting contrast diagram of the network forecasting value and actual value is shown in Figure 8. These results show that the network output can simulate the indoor relative humidity and predict its trend.

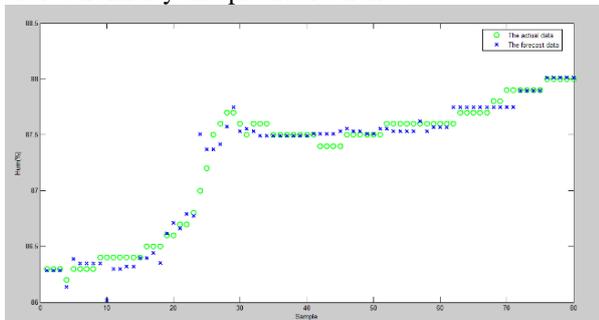


FIGURE 7 Simulation curve

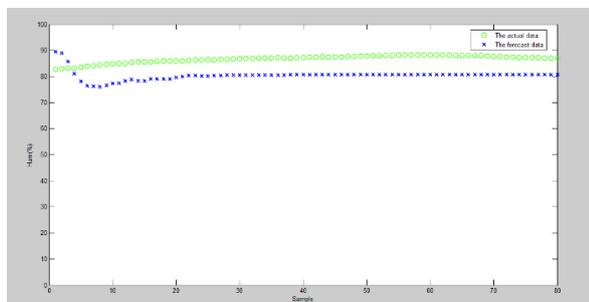


FIGURE 8 Prediction curve

5 Conclusions

Neural networks have been used to establish the prediction model of greenhouse environmental factors. In study, we used an RBF artificial neural network to establish the forecast simulation model for predicting and analysing the humidity of environmental factors in a greenhouse in Northern China. After testing, verification, and offline prediction research by the RBF neural network, results showed that the network convergence speed is significantly faster, the network set-up time is relatively shorter, and the function approximation capability is significantly higher than previous values. Therefore, the simulation model can simulate the basic trend with satisfactory prediction effect. This prediction can be used to control feed-forward greenhouse environments and to provide a scientific basis for the reasonable regulation and control of greenhouse environments.

Acknowledgements

This work was supported by the financial grant from China postdoctoral science foundation (2012M510838) and the science research project of Liaoning Province Education Department (L2014249) and the open issues project of Beijing Information Centre.

References

- [1] Li T 2005 Current situation and development prospect of solar greenhouse industry in China *Journal of Shenyang Agricultural University* **36**(2) 131-8 (in Chinese)
- [2] Guo Z 2012 Study on the ecological dynamics of Northern Greenhouse *Jilin University* (in Chinese)
- [3] He F, Ma C, Zhang J 2009 Dynamic prediction model of humidity in greenhouse *Transactions of the Chinese Society for Agricultural Machinery* **40**(10) 173-7 (in Chinese)
- [4] Guo Q 2013 Research on the application of artificial neural network in the prediction of relative humidity *Modern food science and technology* **12**(9) 1297-301 (in Chinese)
- [5] Xue X 2007 Research on prediction model of soil moisture based on the method of SVM *Chinese Journal of Soil Science* **38**(3) 427-33 (in Chinese)
- [6] Jiang Z 2001 Introduction to artificial neural network *Beijing Higher Education Press* (in Chinese)
- [7] He F, Ma C 2008 Study on the principal component analysis of greenhouse air humidity based on artificial neural network modeling *Journal of Shanghai Jiaotong University (Agricultural Science)* **24**(1) 492-5 (in Chinese)
- [8] Xin Z 2009 Effects of external weather conditions on the temperature and humidity inside greenhouse *Meteorological and Environmental Sciences* **32**(2) 30-2 (in Chinese)
- [9] Ferreira P M, Faria E A, Ruano A E 2002 Neural networks models in greenhouse air temperature prediction *Neurocomputing* **43** 51-75
- [10] Wang X, Wei R, Kang X 2010 Application of seasonal time series model in forecasting greenhouse daily humidity *Chinese Agricultural Science Bulletin* **26**(22) 407-12 (in Chinese)
- [11] Wang X, Wang J, Zhao B 2012 Study on changes of phases and the temperature and humidity inside and outside environment of pear cultivation shed greenhouse *Chinese Agricultural Science Bulletin* **28**(1) 201-6 (in Chinese)
- [12] Ferreira P M 2002 Neural networks models in greenhouse air temperature prediction *Neurocomputing* **43** 51-75
- [13] Liu Y, Zhang L 2007 Realization and performance of BP network and RFB neural network *Research on electronic measurement technology* **4** 77-80
- [14] Zhu D 2004 Artificial neural network research present situation and Prospect *Journal of Jiangnan University* **3** 103-8 (in Chinese)
- [15] Hui Qiang Zhi 2005 BP network and RBF network in function approximation of the comparative study of *Bulletin of science and technology* **21**(2) 193-6
- [16] Chen Z, Yuan X, Yu G 2014 Experimental research on moisture transmission rate between humidity sensitive material and the external environment *INMATEH Agricultural Engineering* **43**(2) 97-106

Authors	
	<p>Chen Chunling, born in January, 1971, Long Yan County, Fu Jian Province, China</p> <p>Current position, grades: the Associate Professor of school of information and electrical engineering, Shenyang Agriculture University, China. University studies: Doctor' degree in Agricultural Electrification and Automation from Shenyang Agriculture University of Liaoning in China. Scientific interest: agricultural information technology, electrical engineering. Publications: 20 papers. Experience: teaching experience of 20 years, 2 scientific research projects.</p>
	<p>Wang Long, born in April, 1989, Zhang Ye County, Gan Su Province, China</p> <p>Current position, grades: Graduate student of information and electrical engineering, Shenyang Agriculture University, China. University studies: Bachelor 's in Computer Science and Technology from Shenyang Agriculture University of Liaoning in China. Scientific interest: computer science and agricultural information technology. Experience: several researches and development works about Agricultural Information Technology.</p>
	<p>Xu Tongyu, born in November, 1967, Shenyang County, Liaoning, China</p> <p>Current position, grades: the Professor of School of information and electrical engineering, Shenyang Agriculture University, China. University studies: Doctor' degree in Agricultural Electrification and Automation from Shenyang Agriculture University of Liaoning in China. Scientific interest: agricultural information technology, electrical engineering. Publications: 30 papers. Experience: teaching experience of 24 years, 4 scientific research projects.</p>
	<p>Qi Jiawei, born in July, 1991, Zhengzhou City, Henan Province, China</p> <p>Current position, grades: postgraduate course (Power System Engineering, MSC) in the university of Manchester in Britain. University studies: Bachelor' degree in Electrical Engineering and Automation from Shenyang Agriculture University of Liaoning in China. Scientific interest: electrical engineering, power system protection. Experience: postgraduate course (Power System Engineering, MSC) in the university of Manchester in Britain.</p>

Research on stock analysis methods based on fluid mechanics

Zhang Jilin^{1, 2*}

¹Research Center for Data Analysis, Fujian University of Technology, Fuzhou, Fujian Province, China, 350118

²Fujian Collaborative Innovation center for Beidou Navigation and Intelligent Traffic, Fuzhou, Fujian Province, China, 350118

Received 6 July 2014, www.tsi.lv

Abstract

In this study, the leading indicator is considered as a new technical analysis method for evaluating stocks. This paper is written using the special characteristics and features of the principles of fluid motion, which can be applied to microscopic and macroscopic aspects of stock prices. Microscopic aspect: the stock that changes fastest is considered the leading indicator, which can be easily found by auto-filtering in Excel after working out the speed rate via program Matlab. The trend of other stocks will be predicted by analysing the leading indicator. Macroscopic aspect: determine the leading indicator, which are the top 10 stocks listed in the speed analysis results in the microscopic aspect. The trend of A-shares in a certain time period will be predicted according to the trend of the leading indicator. In order to verify the viability of the leading indicator analysis method in microscopic and macroscopic aspects, we do the following research. We collect 4 days of stock price data to find out the leading indicator via the microscopic or macroscopic leading indicator method. When the corresponding leading indicator is found, then we predict the trend of other stocks and A-shares in a certain time period, so that the feasibility of the leading indicator methods can be proved.

Keywords: financial physics, leading indicator, stock analysis, Matlab

1 Introduction

Securities investment analysis is a comprehensive study of the information that affects the value or the price of securities. It is the professional way to estimate the value of stocks as well as any changes in stock value, and as an essential part of securities investment it cannot be overlooked. Through proper securities investment analysis, investors can make a scientific assessment and value stocks correctly thereby reducing their risks and maximizing the net profit of an investment [1, 2].

Three methods of conducting securities analysis are as follows:

1) The basis analysis method. This method derives results mainly according to the theories of economics, finance, and investment;

2) The technical analysis method. This method derives results based on the changing laws of the securities market;

3) The portfolio analysis of securities. The point of this analysis method is to reduce non-systematic risks by creating diversity investments, whose outstanding feature is to be quantified [3-5].

The securities investment believes that the change of stock prices depends on the supply-demand relationship in the market. All the factors that affect stock prices can be found by noting price changes and transactions instead of analysing the basic data. Technical analysis is intuitive, but time-limited. This method is mostly used for analysing short-term fluctuations in the stock market to find the right time for investing [6]. The theory of

technical analysis is actually the theory of market behaviour, which can be generally classified into 6 parts: K-line theory, Tangent theory, Morphology theory, Technical index theory, Wave theory, and Circulation period theory.

Using the features of fluid motion, the author of the presents study has created a new stock analysis method - the leading indicator analysis method derived from microscopic and macroscopic aspects. This paper is divided into four parts:

1) the introduction;

2) the micro leading indicator and macro leading indicator are briefly introduced with empirical analysis;

3) analysis of the advantages of the leading indicator analysis;

4) the conclusion.

2 A newly discovered indicator for technical analysis - leading indicator

2.1 MICROSCOPIC ASPECT

The main concern of leading indicator analysis method is as follows: The leading indicator stock will be found after analysing the data of stock prices using the leading indicator analysis method. The trend of other stocks during a period will be predicted by analysing the trend of the leading indicator stock.

* Corresponding author e-mail: jilinzhang1976@126.com

2.1.1 Background and definition

The main concepts of this method focus on the mechanics of the fluid particle as described by Lagrange in his study of fluid motion. The motion of each fluid particle is read from the very beginning to the end; that is, the position of each fluid particle is changing over time. From the principles of fluid mechanics, we learn that the fluid particle is normally treated as the smallest research object, and the fluid is the medium made up of numerous fluid particles that occupy the whole fluid space consecutively without gaps [7].

Firstly, consider the fluid as a composition made up of numerous fluid particles that occupy the whole fluid space consecutively without gaps. Secondly, allow that at the very same moment, the fluid particles become streamlines, and that the numerous streamlines turn into clusters. Since the streamlines never join or go out, the cluster is composed of the streamlines of the same shape in sequential order. Thirdly, allow that the velocity vector of the fluid particles on the curve line is tangent to it. Each fluid particle is at a speed, and given the three conditions mentioned above, we will see that the velocity of high-speed value fluid at time point t is the same as that of low-speed value fluid at time point $t+k$; that is to say, the fluid of high-speed value will present its velocity direction one step earlier than the low speed value fluid (shown as Figure 1). Hence, we only need to figure out the velocity direction of high-speed value fluid at any time point, and the velocity direction of low speed value fluid will be predicted. The amount of time will be decided later by the speed value.

The fluid of highest speed value is considered the leading indicator, and its velocity direction will appear earlier than any other fluids. We will figure out the velocity direction of other fluids after a period of time as long as we analyse the velocity direction of the leading indicator. Meanwhile, the speed of the other fluids determines how much time it takes to form their velocity direction—the higher speed value it has, the less time it spends, and vice versa.

In order to have a direct impact of the quantity of fluids on the speed value, we build a two-dimensional rectangular coordinate, from which the greatest speed value fluid will be easily seen. The result is roughly shown in the Figure 2 as follows:

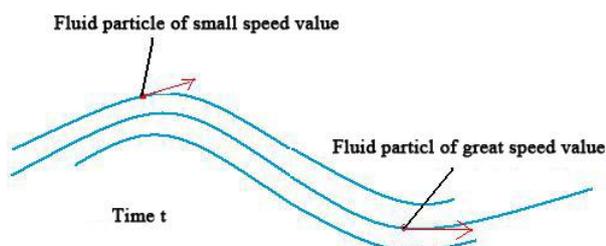


FIGURE 1 Fluid particle movement

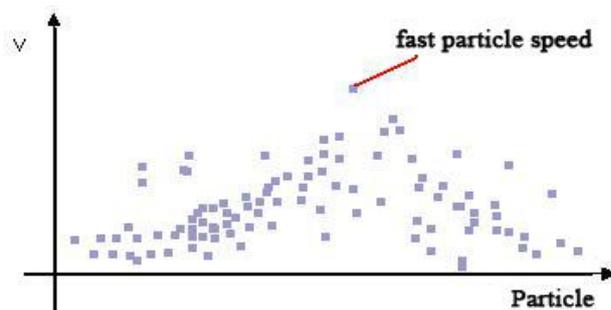


FIGURE 2 Fluid particle speed

A change in the supply-demand relationship is the direct reason that changes the stock price occur. Any change or policy of politics, economy, finance, banking, business transactions, diplomacy, military, or the social situation will influence the investors' prediction during an afternoon session. Whether the result is positive or not directly influences the investors' behaviour, which will influence the supply-demand relationship in the stock market and change the stock price. Since we can see the changes in stock price, we will know the market better only by analysing the changes of stock price [8-10].

To project the leading indicator of the fluid motion into the stock price, the stock price will be taken as the fluid, the change of stock price as the speed, and the trend of stock price as the velocity direction. We can then see that the stock whose price is fluctuating more quickly than normal will show its direction earlier than the others. And it will therefore be regarded as the leading indicator, the one which presents its direction earlier than other stocks. From its trend, we will predict the others after a period time. The time when other stocks show their trends depends on their speed; the higher speed they have, the earlier they appear, and vice versa.

Hence, Microscopic Leading Indicator is the method to figure out the stock price that fluctuates quickest by analysing the historic stock price data, the so-called the stock leading indicator. The other's price trend will be predicted by analysing the price trend of the leading indicator. Since the speed of the other stocks is different from one another, the time needed for prediction will also be different.

2.1.2 Analysis on speed and time

The leading indicator method not only can be used in the study of daily stock prices, but also other stock price studies of different time periods, such as a week, a month, or even a year. The study results below are listed based on daily stock price; the others can also be figured out in the same way.

Every day, the stock market is open from 9:25 am to 3:00 pm. During this period, the opening price and the closing price are formed by Call Auction. Since the price forming has an effect on the result, we will remove the stock price formed by Call Auction when the MATLAB project performs the data analysis. That is to say, only the stock price formed by Continuous Auction will be kept

for speed analysis.

1) Speed analysis

From the Figure 3, we can see that the daily stock price changes in a vibrating way. To take the stock price results picked every L seconds as the time sequence of daily stock price, $P_1, P_2, P_3, \dots, P_m$ stand for the stock price picked every L second, we will see 4 speeds. They are:

1) Average increasing speed v_1 : when $P_i \leq P_j, (i < j \text{ and } i, j = 1, \dots, m), v_1 = \sum (P_j - P_i) / Ln_1, n_1$ refers to P_j is equal or greater than the number of P_i ;

2) Average decreasing speed v_2 : when $P_i > P_j, (i < j \text{ and } i, j = 1, \dots, m), v_2 = \sum (P_j - P_i) / Ln_2, n_2$ refers to P_j is less than or equal to the number of P_i .

3) Non-absolute value integrated average speed v_3 : when $P_i \leq P_j, (i < j \text{ and } i, j = 1, \dots, m), \text{ and } P_k > P_h, (k < h \text{ \& } k, h = 1, \dots, m), v_3 = [\sum (P_j - P_i) + \sum (P_h - P_k)] / L(n_1 + n_2), n_1$ refers to P_j is equal or greater to the number of P_i , and n_2 refers to P_h is less than or equal to the number of P_k .

Three explanations for speed v_3 as follows: when $v_3 > 0$, it means that the increase speed of the stock price is greater than decrease speed, and it will move to the high price in a short period; $v_3 < 0$ means that the decrease speed is greater than the increase speed, and it will move to the low price in a short period; and $v_3 = 0$ means that the increase speed is almost equal to the decrease speed, and the stock price will be almost on a horizontal level in a short period. Hence, the speed v_3 reflects the trend of the stock price.

4) Absolute value integrated average speed v_4 : when $P_i \leq P_j, (i < j \text{ \& } i, j = 1, \dots, m), \text{ and } P_k > P_h, (k < h \text{ \& } k, h = 1, \dots, m), v_4 = [|\sum (P_j - P_i)| + |\sum (P_h - P_k)|] / L(n_1 + n_2), n_1$ refers to P_j is equal or greater to the number of P_i , and n_2 refers to P_h is less than or equal to the number of P_k .

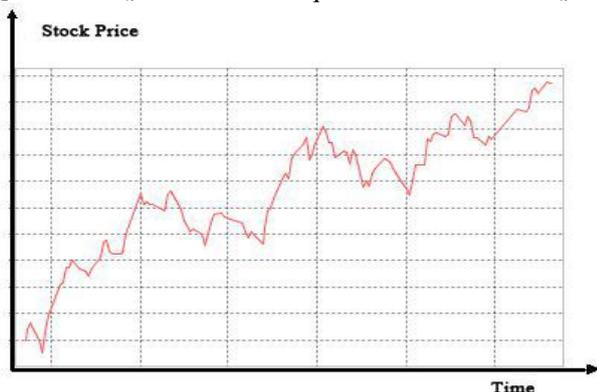


FIGURE 3 Daily stock price

Speed v_4 is the absolute value between the increased speed and decreased speed. The stock of highest speed v_4 is considered as the leading indicator. To the non-leading indicator stocks, speed v_4 is the exact factor that decides when the trend of the leading indicator will appear. The greater the speed is, the less time it takes. Easily can we read the stock quantity of each speed rate from the Figure

4 shown below by putting the stock codes into the two-dimensional rectangular coordinate along with the speed v_4 .

From the four speeds, we can learn the logic of the microscopic leading indicator method; that is, speed v_4 is the key to figuring out the leading indicator of stocks, while speed v_3 is the point to predict the trend of leading indicator.

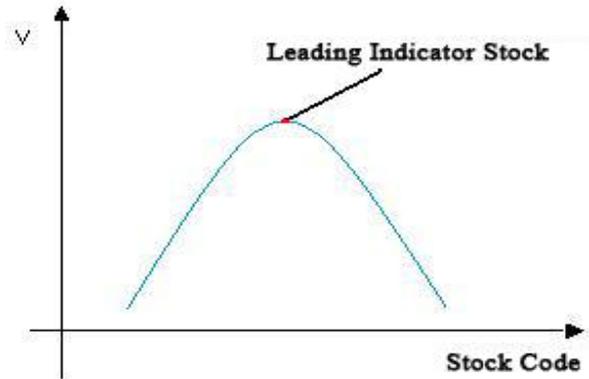


FIGURE 4 Leading indicator stock

2) Time analysis

When we figure the leading indicator and its trend via the above speed analysis, the trend of other stocks after a period time will also be known. The length of the “period time” will depend on speed v_4 of the other stocks. To assume that the other stocks are moving at the speed of v_4 within the “period time”, under this perfect and ideal condition, we will work out the ratio between the “period time” and the time needed to form the trend.

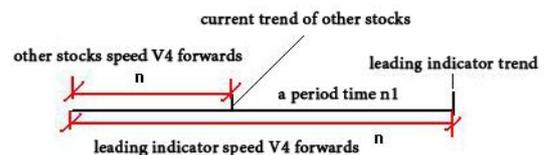


FIGURE 5 Time of leading indicator

Let “ n ” be the time the other stocks spend to form the current trend; that is, the time that the leading indicator needs to form the predicted trend. “ n_1 ” stands for the “period time” that the other stocks need to form the trend that the leading indicator predicts shown as the Figure 5. Then:

$$\begin{aligned} \text{other stocks speed } v_4 \times (n + n_1) &= \\ \text{leading indicator speed } v_4 \times n & \end{aligned} \tag{1}$$

With a little calculation, we will get the time ratio:

$$\begin{aligned} \frac{n_1}{n} &= \\ \frac{\text{Leading Indicator Stock Speed } v_4 - \text{Other Stock Speed } v_4}{\text{Other Stocks Speed } v_4} & \end{aligned} \tag{2}$$

Following the above time-ratio equation, under an ideal condition, we will without doubt work out the “period time” that other stocks need to form the trend

predicted by the leading indicator.

However, in fact, the speed v_4 of other stocks in the “period time” is not always the same, and it will change as the time moves on. Even though there are differences between theory and a real situation, the law figured out in theory is also workable. Therefore, I would like to make it clear here that the leading indicator in this paper is analysed under an ideal condition.

2.1.3 Leading indicator method based on program Matlab

We learn from the above study that the leading indicator and time are decided by speed v_4 , while the trend of leading indicator is decided by speed v_3 . Hence, we can predict the trend of other stocks after a period of time by working out the speed v_3 and speed v_4 of the leading indicator and the speed v_4 of the other stocks. The leading indicator can be found by the auto-filtering function available in Excel after working out the speed v_3 and speed v_4 using the program Matlab.

The working theory of the program Matlab is to read repeatedly from Excel forms the stock price data in the current folder and try to determine the speed of the stocks in the folder, from which the data formed by Call Auction will be deducted. When the results are completed, the file will be saved to the Excel form named “conclusion” in the folder, using the auto-filtering function in Excel to do the final study [11].

```
locate=2; //see result at the initial position of Excel
//y1 as the initial file name & y2 as the final file name
for Name=y1:1:y2
    Namestr=[num2str(Name),'.xls'];
    //to check the filename available or not
    p=exist(Namestr,'file');
    //Speed calculation is processing
    if p==2
        //put the data into program Matlab rectangularly
        Original=xlsread(Namestr);
        //read stock prices in the rectangular
        Cut=Original(:,2);
        //select the number of stock price per second
        [count,n]=size(Cut);
        j=1; k=1;
        //deduct the stock price formed by Call Auction
        for i=3:(count-3)
            if Cut(i)>=Cut(i+1)
                //increase speed matrix
                increase(j)=(Cut(i)-Cut(i+1))/L;
                j=j+1;
            else
                //decrease speed matrix
                decrease(k)=(Cut(i)-Cut(i+1))/L;
                k=k+1;
            end
        end
        //elements number of increase speed matrix
        [count1,n1]=size(increase);
        //elements number of decrease speed matrix
        [count2,n2]=size(decrease);
        //average increase speed
```

```
avgincrv=(sum(increase))/n1;
//average decrease speed
avgdecrv=(sum(decrease))/n2;
//non-absolute value integrated average speed
avgv1=(sum(increase)+sum(decrease))/(n1+n2);
//absolute value integrated average speed
avgv2=(sum(increase)+abs(sum(decrease)))/(n1+n2);
V=[Name,avgincrv,avgdecrv,avgv1,avgv2];
//to confirm the place where the result is saved
locatestr=['A',num2str(locate)];
locate=locate+1;
//save the speed result in the Excel for further study
xlswrite('conclusion', V, 'conclusion', locatestr);
else
    continue;
end
end
```

The program indicates that the speed result begins saving from the very second line; that is to say, the content in first line should be described clearly enough to indicate the meaning of each line to make sure that the form is expressing perfectly and precisely.

In order to filter the stock of greatest speed v_4 , the leading indicator stock, we set the filter function for the four columns. And we can then predict the price trend of the other stocks after a period time according to the speed v_3 of the leading indicator. The “period time” can be figured out by the time ratio equation from the above time analysis via the speed v_4 of the other stocks.

2.1.4 Application and verification on microscopic leading indicator method

1) Data collection and process:

The data collection includes the opening price and closing price formed by Call Auction, which will be removed when the program Matlab calculates the change speed of stock prices, and the stock price formed by Continuous Auction will remain. Therefore, we only need to collect the stock price data between 27/06/2011 and 30/06/2011, codes from 000001 to 601999, and save them respectively to Excel in four files (Figure 6). We set the Excel name as “20110627 + stock code” to make it more convenient when moving the data to the program Matlab. While collecting the data, we select the time sequence of stock price for every 3 seconds.



FIGURE 6 Stock price data file

2) Results analysis and verification

We will integrate the speed calculation results in four label-Excel, and analyse it in order. As we know, the stock price whose speed v_4 is greatest is the leading indicator, and stock price of greatest speed will be found by setting the auto-filtering function, descending order, for the list of speed v_4 .

1) Result analysis and verification for the data on

2011-6-27:

using Excel, we built a two-dimensional rectangular coordinate consisting of speed v_4 and the corresponding stock codes. The stock quantity and the speed rate of the leading indicator belonging to each speed value can be viewed directly from the Figure 7.

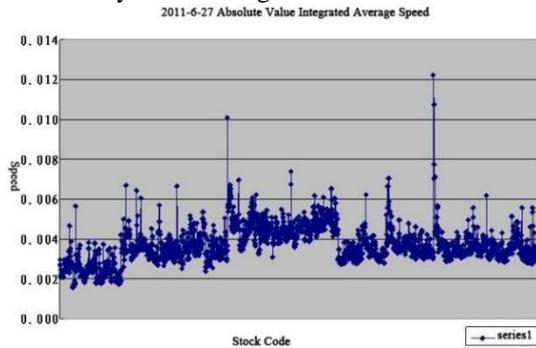


FIGURE 7 Absolute value integrated average speed

To pick out the Top 10 results of Speed v_4 in descending order, shown as the Table 1:

TABLE 1 v_3 and v_4 of stock on June 27

stock code	avg increase speed	avg decrease speed	speed v_3	speed v_4
20110627600519	0.00810	-0.02361	-0.00028	-0.01220
20110627600520	0.00688	-0.02146	-0.00061	0.01073
20110627002304	0.00676	-0.01938	-0.00015	0.01009
20110627600521	0.00441	-0.01706	-0.00126	0.00776
20110627300006	0.00495	-0.01422	-0.00012	0.00740
20110627600523	0.00375	-0.01661	-0.00163	0.00715
20110627600522	0.00346	-0.01702	-0.00196	0.00704
20110627600261	0.00482	-0.01314	-0.00008	0.00702
20110627002353	0.00444	-0.01396	-0.00043	0.00695
20110627300007	0.00451	-0.01291	-0.00009	0.00673

Table 1 show that the stock code No 600519 is the leading indicator, whose speed v_3 is -0.00028. That is to say, the prices of other stocks are tending to decline after a period of time. We can select two stocks from the data on 28th to verify the leading indicator method:

- verification of the stock code No 600520: As per the time ratio equation determined by the time analysis, we will calculate the time ratio $(0.01220-0.01073)/0.01073=0.13662 \approx 7/50$, which indicates that it takes 7/50 of the initial time for the stock to show the decline trend. Shown in the Figure 8, we can see the trend of stock code No 600520 on 28th. The part in the red circle is near 7/50, the point at which the stock is moving to the low price, and it proves that the leading indicator is feasible.

- verification of the stock code No 002304: Following the time ratio equation, we will figure out that the time ratio is $(0.01220-0.01009)/0.01009=0.20854 \approx 5/24$, which indicates that it takes 5/24 of the initial time for the stock to show the decline trend. From the Figure 9 showing the trend of stock code No 002304 on 28th, we will see that the part in the red circle is near 5/24, where the stock is moving to the low price. Even though it is a small movement, it still proves that the leading indicator is feasible.

In conclusion, the stock code No 600519 is the leading indicator, and the microscopic leading indicator method is proved feasible.



FIGURE 8 Absolute value integrated average speed

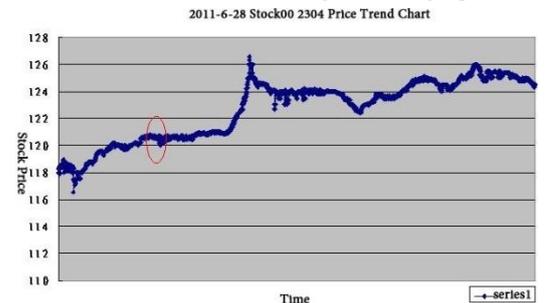


FIGURE 9 Price trend of stock 002304 on June 28

2) Result analysis and verification for the data on 2011-6-28: using Excel, we built a two-dimensional rectangular coordinate consisting of speed v_4 and the correspondent stock codes. The stock quantity and the speed rate of the leading indicator belonging to each speed value can be viewed directly from the Figure 10.

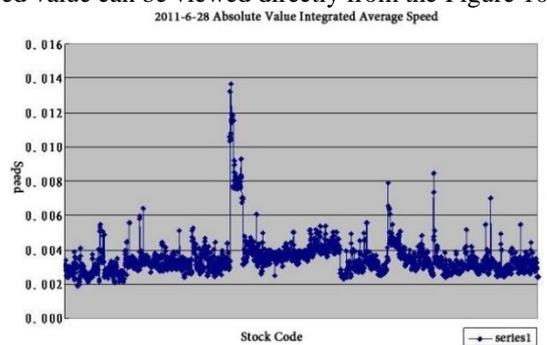


FIGURE 10 Absolute value integrated average speed

To pick out the Top 10 results of Speed v_4 in descending order, shown as the Table 2:

TABLE 2 v_3 and v_4 of stock on June 28

stock code	avg increase speed	avg decrease speed	speed v_3	speed v_4
20110628002309	0.00915	-0.02631	-0.00022	-0.01368
20110628002304	0.00908	-0.02472	-0.00015	0.01321
20110628002313	0.00775	-0.02330	-0.00046	0.01186
20110628602317	0.00780	-0.02314	-0.00037	0.01186
20110628302315	0.00774	-0.02306	-0.00039	0.01179
20110628602311	0.00764	-0.02272	-0.00038	0.01163
20110628602316	0.00757	-0.02269	-0.00043	0.01157
20110628602319	0.00752	-0.02265	-0.00045	0.01152
20110628002312	0.00753	-0.02264	-0.00044	0.01152
20110628302318	0.00747	-0.02263	-0.00048	0.01148

From the Table 2, the stock No 002309 is the leading indicator, whose speed v_3 is - 0.00022. That is to say, the prices of other stocks are tending to decline after a period of time. We can select two stocks from the data on 29th to verify the leading indicator method:

- verification of the stock code No 002304: As per the time ratio equation determined by the time analysis, we will calculate the time ratio $(0.01368-0.01321)/0.01321=0.03545\approx 3/85$, which indicates that it takes 3/85 of the initial time for the stock to show the decline trend. Shown in the Figure 11, we can see the trend of stock code No 002304 on 29th. The part in the red circle is near 3/85, at which point the stock is moving to the low price, and it proves that the leading indicator is feasible.

- verification of the stock code No 002313: Following the time ratio equation, we will figure out that the time ratio is $(0.01368-0.01186)/0.01186=0.15370\approx 6/39$, which indicates that it takes 6/39 of the initial time for the stock to show the decline trend. From the Figure 12, showing the trend of stock code No 002313 on 29th, we will see that the part in the red circle is near 6/39, where the stock is moving to the low price, and it proves the leading indicator is feasible.

In conclusion, the stock code No 002309 is the leading indicator, and the microscopic leading indicator method is proved feasible.

2011-6-29 Stcok 002304 Price Trend Chart

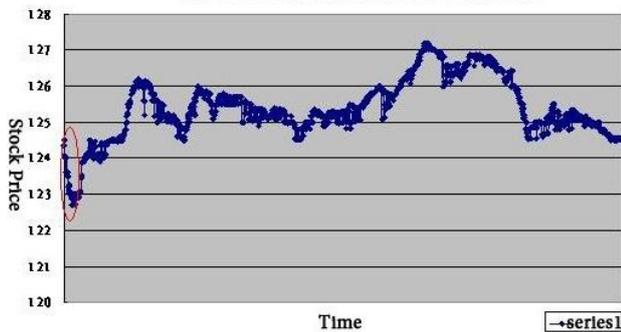


FIGURE 11 Price trend of stock 002304 on June 29

2011-6-29 Stock 002313 Price Trend Chart

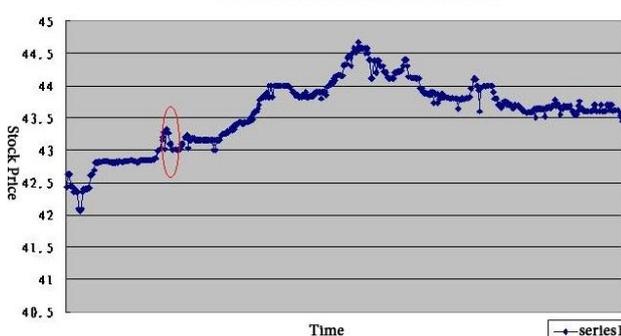


FIGURE 12 Price trend of stock 002313 on June 29

3) Result analysis and verification for the data on 2011-6-29:

using Excel, we built a two-dimensional rectangular coordinate consisted of speed v_4 and the correspondent stock codes. The stock quantity and the speed rate of the leading indicator belonging to each speed value can be

viewed directly from the Figure 13.

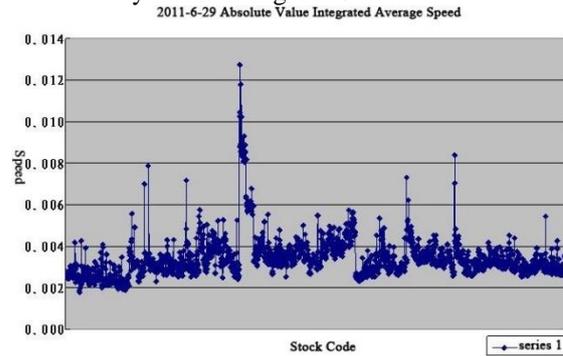


FIGURE 13 Absolute value integrated average speed

To pick out the Top 10 results of Speed v_4 in descending order, shown as the Table 3:

TABLE3 v3 and v4 of stock on June 29

stock code	avg increase speed	avg decrease speed	speed v3	speed v4
20110629002304	0.00863	-0.02416	-0.00004	-0.01273
20110629002310	0.00790	-0.02261	-0.00016	0.01179
20110629002306	0.00677	-0.02071	-0.00049	0.01045
20110629602305	0.00664	-0.02040	-0.00051	0.01027
20110629302311	0.00685	-0.01963	-0.00014	0.01023
20110629602322	0.00590	-0.01875	-0.00061	0.00930
20110629602313	0.00574	-0.01863	-0.00070	0.00914
20110629602315	0.00559	-0.01831	-0.00072	0.00895
20110629002309	0.00557	-0.01823	-0.00072	0.00892
20110629302327	0.00557	-0.01798	-0.00065	0.00885

From the Table 3, the stock, code No 002304, is shown as the leading indicator, whose speed v_3 is - 0.00004. That is to say, the prices of other stocks are tending to decline after a period time. We can select two stocks from the data on 30th to verify the leading indicator method:

- verification of the stock code No 002310: As per the time ratio equation determined by the time analysis, we will calculate the time ratio $(0.01273-0.011-9)/0.01179=0.08031\approx 20/249$, which indicates that it takes 20/249 of the initial time for the stock to show the decline trend. Shown in the Figure 14, we can see the trend of stock code No 002310 on 30th. The part in the red circle is near 20/249, at which the stock is moving to the low price, and it proves that the leading indicator is feasible.

- verification of the stock code No 002306: Following the time ratio equation, we will figure out that the time ratio is $(0.01273-0.01045)/0.01045=0.21810\approx 100/458$, which indicates that it takes 100/458 of the initial time for the stock to show the decline trend. From the Figure 15 showing the trend of stock code No 002306 on 30th, we will see that the part in the red circle is near 100/458, where the stock is moving to the low price, and it proves the leading indicator is feasible.

In conclusion, the stock code No 002304 is the leading indicator, and the microscopic leading indicator method is proved feasible.



FIGURE 14 Price trend of stock 002310 on June 30



FIGURE 15 Price trend of stock 002306 on June 30

4) Result analysis and verification for the data on 2011-6-30: using Excel, we built a two-dimensional rectangular coordinate consisted of speed v_4 and the correspondent stock codes through Excel. The stock quantity and the speed rate of the leading indicator belonging to each speed value can be viewed directly from the Figure 16.

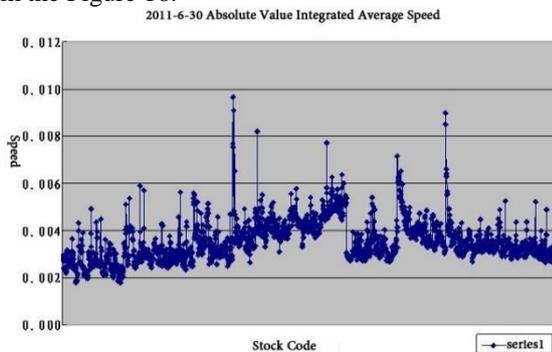


FIGURE 16 Absolute value integrated average speed

To pick out the Top 10 results of Speed v_4 in descending order, shown as the Table 4:

TABLE 4 v_3 and v_4 of stock on June 30

stock code	avg increase speed	avg decrease speed	speed v_3	speed v_4
20110630002304	0.00647	-0.01848	-0.00012	-0.00965
20110630002310	0.00615	-0.01726	-0.00004	0.00909
20110630600519	0.00577	-0.01796	-0.00050	0.00900
20110630600520	0.00533	-0.01728	-0.00064	0.00849
20110630002407	0.00587	-0.01467	-0.00044	0.00819
20110630300124	0.00502	-0.01520	-0.00032	0.00771
20110630002306	0.00490	-0.01533	-0.00044	0.00766
20110630002305	0.00462	-0.01564	-0.00073	0.00754
20110630600259	0.00480	-0.01372	-0.00098	0.00716
20110630600523	0.00372	-0.01456	-0.00111	0.00659

From the Table 4, the stock, code No 002304, is the leading indicator, whose speed v_3 is -0.00012. That is to say, the prices of other stocks are tending to decline after a period time. But the speed is declining more slowly since the speed v_4 is 0.00965.

Though few stocks are taken from the above 4-day data to verify the leading indicator method, it turns out that the leading indicator is surely workable.

2.2 MACROSCOPIC LEADING INDICATOR METHOD

Differing from microscopic aspect research on each stock price's trend, macroscopic aspect research is used to analyse the whole stock market using the method of the leading indicator. Actually, the macroscopic leading indicator is a method based on the microscopic leading indicator since the stock market is made up of many single stocks, and the change of the whole market is the sum of each stock's change.

2.2.1 Definition

As we know, the sample stocks of A-shares are all listing stocks that reflect the price change status of A-shares. Hence, when the trend of A-shares is figured out, the trend of the whole stock market will be predictable. Shanghai composited securities index series uses the Paasche Composite Index equation for calculating, and the equity quantity of sample stock's for weighting it [12]. The equation is:

$$\text{Report Period Index} = \frac{\text{Total Report Period Value}}{\text{Total Based Period Value}} = \text{Based Period Index} \quad (3)$$

In the above equation, the amount = \sum (stock price * authorized shares). The equation makes it clear that the stock price change effect the report period index, and finally, that A-shares follows the trend of stock price change.

There is a description about the microscopic leading indicator: the share of greater speed v_4 will show the trend earlier than the smaller one. Therefore, we can pick out the shares whose speed v_4 is listed among the top 10 shares every day according to the microscopic leading indicator analysis. The share showing its trend earlier every day than the others represents the trend of whole stock market.

We can conclude that the macroscopic leading indicator aims at finding the shares whose speed v_4 are listed among top 10 based on the analysis of microscopic leading indicator, and the share always listing in the top 10 is the leading indicator. What's more, the whole market trend in the future will be predictable when we work out the trend of stock index A after a period time. The trend of stock index A can be known by figuring out the whole market trend after a period time based on the

speed v_3 trend of the macroscopic leading indicator. However, I would like to draw to your attention that the leading indicator figured out by macroscopic leading indicator method is different from that of microscopic leading indicator method.

Two situations can happen when using the macroscopic leading indicator method. In one situation, some shares of smaller speed v_4 show the trend earlier than some of leading indicator of higher speed v_4 , in the other situation, the leading indicator of smaller speed v_4 shows earlier than other shares of smaller speed v_4 . In this case, we need to select an intermediate time value. It is the time point when stock index A mostly presents the trend of leading indicator. It can be determined by selecting the average value of the time ratio between the shortest and the greatest time that the speed v_4 smaller than macroscopic leading indicator takes to show the trend of leading indicator. It will neither take too much time for the shares of great speed to leave, nor too much time for the shares to show the trend.

2.2.2 Application and verification on microscopic leading indicator method

It is easy to see from the above four-day speed analysis results based on microscopic leading indicator method that the stock code No 002304 always ranks in the top 10, even the top 3. According to the theory of the macroscopic leading indicator, the stock code No 002304 is the leading indicator, whose current speed v_3 trend is exactly the trend of A-shares after a period time. We will work out the medium time ratio according to the time calculated by the progressive analysis:

1) 27th: the speed v_3 of stock code No 002304 is -0.00015, which shows that the stock price is going to decline. The time ratio for other stocks to show the same trend earliest is:

$$(0.01009-0.00776)/0.00776=0.30148.$$

The time ratio for other stocks to show the same trend latest is:

$$(0.01009-0.00156)/0.00156=5.49026,$$

$$(0.30148+5.49026)/2\approx 3.$$

It means that the A-shares is supposed to show the same decline trend as stock code No 002304 after 3 days (with 27th included), that is on 29th.

2) 28th: the speed v_3 of stock code No 002304 is 0.00015, which shows that the stock price is going to increase. The time ratio for other stocks to show the same trend earliest is:

$$(0.01321-0.01186)/0.01186=0.11419.$$

The time ratio for other stocks to show the same trend latest is:

$$(0.01321-0.00190)/0.00190=5.94365.$$

The medium time ratio is:

$$(0.11419+5.94365)/2\approx 3.$$

It means that the A-shares is supposed to show the same increase trend as stock code No 002304 after 3 days (with 28th included), that is on 30th.

3) 29th: the speed v_3 of stock code No 002304 is -0.00004, which shows that the stock price is going to decline. The time ratio for other stocks to show the same trend earliest is:

$$(0.01273-0.01179)/0.01179=0.08031.$$

The time ratio for other stocks to show the same trend latest is:

$$(0.01273-0.00180)/0.00180=6.08081.$$

The medium time ratio is:

$$(0.08031+6.08081)/2\approx 3.$$

It means that the A-shares is supposed to show the same decline trend as stock code No 002304 after 3 days (with 29th included), that is on 1st.

4) 30th: the speed v_3 of stock code No 002304 is -0.00012, which shows that the stock price is going to decline. The time ratio for other stocks to show the same trend earliest is:

$$(0.00965-0.00909)/0.00909=0.06153.$$

The time ratio for other stocks to show the same trend latest is:

$$(0.00965-0.00178)/0.00178=4.41990.$$

The medium time ratio is:

$$(0.06153+4.41990)/2\approx 2.$$

It means that the A-shares is supposed to show the same decline trend as stock code No 002304 after 2 days (with 30th included), that is on 1st.

A-shares status collected between 27/06/2011 and 01/07/2011 is as follows:

TABLE 5 Opening price of A-shares of 5 days

date	opening price	closing price
01/07/2011	2899.58	2890.61
30/06/2011	2859.94	2893.53
29/06/2011	2888.61	2858.35
28/06/2011	2891.14	2890.51
27/06/2011	2877.62	2889.59

The table verifies what we predict above, on 29th, the stock code No 002304 declines as what predicts on 27th; on the 30th, it goes up as was predicted on 28th; on 1st, it declines again as was predicted on 29th and 30th. In a word, the stock code No 002304 is the leading indicator, which proves that the macroscopic leading indicator method is feasible.

2.3 THE CONCLUSION

The first two parts clearly describe the definition of the microscopic or macroscopic leading indicator method, which was also proved workable by analysing the 4-day stock price data based on the leading indicator method.

The microscopic leading indicator method is used to analyse a single stock by figuring out the trend of leading indicator to predict the trend of other stocks after a period time; while the macroscopic leading indicator is used to analyse the whole stock market status by figuring out the trend of the leading indicator to predict that of the A-shares after a period time. The stock investors will get more reference information by combining the two methods to avoid the risky investment and so maximize profit.

Therefore, it is better for investors to have a general idea of what the market will be after a period time through the macroscopic leading indicator. The trend of decline or increase is predictable by the speed value and is unique and fixed. Investors can analyse the trend of the stock they count on based on the microscopic leading indicator method. Since the speed is unique and fixed, it is easy to choose a right time to buy or sell in order to reduce the risks on investment.

However, the analysis result concluded by leading indicator method is better used only for a reference since it is processed under a perfect condition, which is far away different from a real situation. Hence, investors are better off analysing the market situation not only with reference to the leading indicator analysis results, but also by a comprehensive study in other ways in order to make a better decision [13, 14].

3. Advantages of leading indicator method

3.1 ADVANTAGE OF MICROSCOPIC LEADING INDICATOR METHOD

The way that initial analysis methods select a right time for buying or selling stocks or predicting the future of stock prices is according to ready figure rules from the records, line charts, or curve charts. Even though we use the figure rules that are already known to analyse the trend charts, we will soon realize that the same trend can be explained in different ways when analysed from different aspects.

For example, according to the morphology theory, the head-and-shoulder pattern is actually the reversal pattern. However, when observed from a wide range, it will probably turn out to be a mid-way continuous pattern (shown in Figure 17).

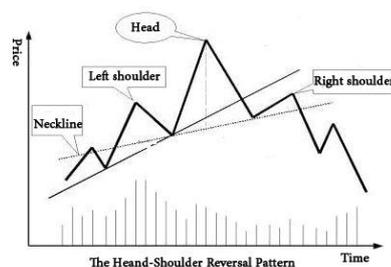


FIGURE 17 Head and shoulders shape

When we use the wave theory to analyse the trend charts, people holding different opinions have different descriptions of the same pattern, and they all make sense. A falling wave can be considered as the second wave or “a” wave. If it is the second wave, the third one will be attractive. If it turns out to be an “a” wave, it will be falling sharply (shown as Figure 18).

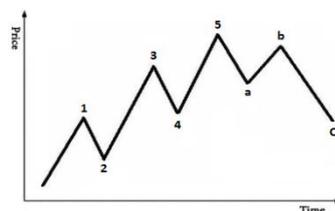


FIGURE 18 Stock price changing trend

Ambiguous opinions from different viewpoints stop people from selecting a right investment time. What makes it worse is that people cannot read the stock price speed simply from the figure laws, and people are too depressed to make a right decision.

As we know what the speed of the physical quantity to the moving level of an object and is same in stock speed to the changing speed of stock price in securities market. A big speed value always means a quick change on stock price, while a small speed value means a slow change. Microscopic leading indicator method focuses on analysing the change speed of stock price. That the speed value is unique and fixed keeps the ambiguities away to help the investors choose a right time for investment. What’s more, investors can figure out the precise trend and time to make a better decision by working out the leading indicator through Matlab program and the filtering function of Excel.

For instance, from the results of speed analysis and time analysis of data from the 29th, we know that stock code No 002304 is the leading indicator whose trend indicates that the price of other stocks will be declining after a period time. The time ratio of stock code No 002310: $(0.01273-0.01179)/0.01179 \approx 20/249$ expresses that it takes around 20/249 of the initial time to start the trend of decline, which is proved to be true from the trend Figure 19 of the 30th. Therefore, investors can make a better decision with the analysis results shown by the leading indicator method instead of original analysis methods, for it is more clearly with no ambiguities.



FIGURE 19 Price trend of stock 002310 on June 30

3.2 ADVANTAGES MACROSCOPIC LEADING INDICATOR

The summary of leading indicator methods indicates that the original analysis methods only focus on a certain stock instead of the whole market. The one-side analysis is good for those investors who are interested in one certain stock, while not suitable for the investors who prioritize the whole market. Macroscopic leading indicator perfectly corrects the shortcomings of the original analysis methods and microscopic leading indicator by presenting the whole market situation. It helps investors to get a better and clearer understanding of the stock market situation through the method of

References

- [1] Epps T W, Epps M L 1976 The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis *Econometrica* **44**(2) 305-21
- [2] Ni S X Y, Pearson. N D, Poteshman. A M 2005 Stock price clustering on option expiration dates *Journal of Financial Economic* **78**(1) 49-87
- [3] Kumar S, Managi S, Matsuda A 2012 Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis *Energy Economics* **34**(1) 215-26
- [4] Zhai A M, Zhou T 2011 The stock market price volume relation analysis based on behavioral assumption for market actor chinese *Journal of Management Science* **19** 31-7
- [5] Zhang J L 2013 Dynamic analysis on trend of stock price based on elasticity coefficient model *Computer Modelling and New Technologies* **17**(4) 260-9
- [6] Di A M, Wang X F 2010 Model of relations between stock price and trading volume based on theoretical plasticity model *Statistics and Decision* **3** 144-6
- [7] Chauvet M, Potter S 2000 Coincident and leading indicators of the stock market *Journal of Empirical Finance* **7**(1) 87-111
- [8] Celik S, Ergin H 2014 Volatility forecasting using high frequency data: Evidence from stock markets *Economic Modelling* **36** 176-90
- [9] Joe H 2005 Asymptotic efficiency of the two-stage estimation method of copula-based models *Journal of Multivariate Analysis* **94** 401-19
- [10] Xiao W, Fei Q, Chen W 2008 Forecasting Chinese stock markets volatility based on neural network combining *Natural Computation* **7** 23-7
- [11] Pan W 2010 Empirical analysis of stock returns volatility in China market *Financial Theory and Engineering (ICFTE)* 17-21
- [12] Wei W X, Gao F M 2001 An empirical analysis on the interaction of price and trading volume of Shanghai's stock market *Prediction* **20**(6) 63-8
- [13] Horng W-J, Lee J-Y 2008 An impact of U.S. and U.K. Stock return rates' volatility on the stock market returns: An evidence study of Germany's stock market returns *Convergence and Hybrid Information Technology* **2** 1159-63
- [14] Huang Y Y 2010 Application of non-linear regressive model in stock price forecast *Journal of Xiaogan University* **26**(3) 62-4
- [15] Zhang J L 2013 Dynamic analysis on trend of stock price based on elasticity coefficient model *Computer Modelling and New Technologies* **17**(4) 260-9

combining the macroscopic leading indicator with the original analysis method or the microscopic leading indicator.

Similar to the advantages of microscopic leading indicator, the leading indicator can be figured out by studying the speed. The trend and time are predictable based on the leading indicator and can help investors make a specific judgment on the trend of whole stock market without any ambiguities [15].

4 Conclusion

The paper briefly introduces the leading indicator as well as its feasibility and advantages. Investors can figure out the leading indicator by analysing the historic data information via Matlab program. Studying the leading indicator, investors will predict the trend of other stocks or the trend of whole stock market. In this way, they will choose a perfect time for investment to reduce the risks to a minimum.

Acknowledgement

This paper is financially supported by Fujian Social Fund 2013B206.

Author



Jilin Zhang, born in 1976, Zhouning County, Fujian Province, China

Current position, grades: Associate professor of Research Center for Data Analysis, Fujian University of Technology, China.

University studies: M.Sc. in Mathematics in Fujian Normal university, Ph.D. degree in management science and engineering from University of Science and Technology Beijing, China, in 2008.

Scientific interest: data analysis, financial mathematics and statistics.

Publications: 23 papers.

Experience: teaching experience of 11 years, 4 scientific research projects.

Rural logistics service providers pricing and competition-cooperation research considering the 3PL accessibility

Jing Xu^{1, 2*}, Guanxin Yao^{1, 2}

¹School of Management, Jiangsu University, Zhenjiang, Jiangsu, China, 212013

²Yancheng Institute of Technology, Yancheng, Jiangsu, China, 224051

Received 1 July 2014, www.tsi.lv

Abstract

Rural logistics service providers pricing and competition-cooperation problems are researched in this article considering the 3PL accessibility in which the post logistics dominants. In the competition stage a model of master-slave Stackelberg game is to be established, and in the stage of cooperation, a cooperation model under different circumstances and the optimal pricing scheme without government intervention are discussed. This study could provide some economic phenomena existing explanations, as well as a reference and theoretical support for the decision-making act of Postal Logistics and other rural logistics-related subject.

Keywords: agricultural products logistics, post logistics, pricing, game theory, rural logistics

1 Introduction

With the rapid development of industrialization of agriculture and rural economic and social integration, it puts forward higher requirements of modernization of production, supply and consumption patterns on functions, methods and efficiency of agro-food logistics. The rapid rise of Third Party Logistics (3PL) provides more convenient for the exchange of goods between urban and rural areas in developed regions and greatly reduces the cost of circulation of agricultural products and agricultural production. At the same time, it also brought greater competition and challenges to traditional postal logistics. However, in the less developed areas of the rural economy, coverage of 3PL is still very low in addition to the postal logistics, and then the relationship between Postal Logistics and third party logistics companies faces not only competition but also a lot of possible cooperation. Therefore, under the rapid development of third-party logistics background, the study of competing and pricing problem between agricultural logistics and multi-service providers has a theoretical reference value, which can ensure effective express of urban agricultural logistics needs and reasonable conduct of rural logistics supply.

2 Theoretical review

The third-party logistics refers to a business model that the logistics service is provided by companies beside the demand side and the supply side. In other countries, it is also known as contract logistics, logistics or logistics alliance externalizing [1]. With the development of the social division of labour, more and more companies

focused on their core business, logistics outsourcing to more specialized companies. Currently, 3PL is widely used in various industries, including pricing method such as the transaction affecting pricing, operating cost pricing, cost-plus pricing method, pricing and revenue sharing pricing as well as benchmark price method [2]. These methods have advantages and disadvantages. For example, the transaction affecting pricing is simple but not flexible, and operating cost pricing calculates accurate but more complex, while cost-plus pricing method is likely to cause controversy. In practical applications, one or several methods (or deformation) are often used, and some also include other incentives or punitive measures, which is not conducive to effective communication, management and coordination between the main supply chain members [3]. Therefore, some scholars start from different systems state, and consider balancing strategy of 3PL pricing decisions in different ways. In recent years, pricing issues of logistics provider's service studied by scholars have increased gradually. Lei L (2006) [4] studied the optimal pricing of third-party logistics service providers and their downstream when the needs of price-sensitive. Bernstein [5] discussed the issue of price equilibrium under price and service competition. Liu and Wu (2011) [6] analysed the coordination of the two-stage supply chain system including the logistics service providers to participate and discussed the wholesale contract price in the two-step charges. Most of these studies are based on the expanded supply chain. In addition, there have been many scholars who began to focus on the interaction problem between the main of supply chain logistics. Xiao [7] and other agricultural products from the behaviour of the main logistics and cooperation tend to prefer starting the main

* *Corresponding author* e-mail: xujing1990mail@126.com

agricultural products between the logistics demand, revenue between the main supply, cooperation and non-cooperation between the supply and demand sides in case of a comparative analysis. Although their study covering both supply and demand of agricultural products, but not yet involved in agricultural logistics service providers. Based on evolutionary game theory, Li Lihue [8] studied the problem of cooperation and coordination of logistics of the same business logistics alliance in the state below the asymmetric nature. Guo [9] analysed the distribution of benefits with the idea of using non-cooperative game method when both supply and demand sides in logistics alliance faced different costs of information and consumption coefficients. The research mainly considered the optimal allocation of interests in cooperation and non-cooperative cases, but the reality is often more complicated. The situation between cooperation and non-cooperation is not identical same in different scenarios, even the same subject reacts differently in different geographical conditions.

This article conducts in-depth and refinement research based on this, specifically discusses the pricing and competing issues of supply chain logistics service providers. Service pricing and competition and cooperation issues of rural logistics service providers (in agricultural logistics, for example) are researched under the circumstances that the third-party logistics(excluding postal logistics) has or not owned outlets of logistics in remote rural areas, respectively. Assuming logistics providers' demand for agricultural products has certain needs rigidity. The total amount of logistics is fixed. Products are sent to buyers through third-party logistics, producing a certain loss in transit. Agricultural products demanders determine their optimal pricing strategies and decisions selected through the game theory, in order to provide some theoretic support of decision making for Postal Logistics and other agricultural products logistics related subjects.

3 Research model

The flow of agricultural products is from farmers to urban demanders. For agricultural logistics service provider, its cost function is $C(x) = F + v(x, l)$, and x – transport tonnage, l – routes, $v(x, l)$ – changes in freight costs. In the formula, $v(x, l) = \eta_1 x + \eta_2 l$ and η_1 – the affected factor of shipping volume on changes in freight cost, η_2 – the affected factor of transport distance on changes in freight cost. The revenue function is $\pi(x) = px$, and p – tonne-kilometres freight.

For agricultural products logistics demanders, the cost to choose a logistics service provider is $Cx = \pi x + t\delta x + Tt$ and t – transport time, δ – loss of agricultural units per unit of time, $Tt = \lambda t$ is psychological and other costs for waiting t to get the products. Its size may reflect the level of agricultural

logistics service, and λ – the cost of per unit of time.

There are two logistics service providers on the market, Postal Logistics and third party logistics. To show the difference, note Postal Logistics subscript 1, and the third-party logistics 2.

3.1 PRICING AND CASE SOLVING IN COMPETITION

Between the agricultural products demander and supplier, there is third-party logistics besides Postal Logistics network. That is to say, other third party logistics server can cover both, and the road is good. Assuming the amount of agricultural products logistics needs is fixed, i.e. $x_0 = x_1 + x_2$. So for two logistics companies operating independently and competing with each other, want to attract more suppliers of agricultural products (or the needs of providers) to choose their own service by adjusting price under certain conditions. If it comes to the equilibrium state, it should be met:

$$C_1 = C_2; x_1 > 0, x_2 > 0, \tag{1}$$

$$C_1 > C_2; x_1 = 0, x_2 > 0, \tag{2}$$

$$C_1 < C_2; x_1 > 0, x_2 = 0. \tag{3}$$

Equation (1) represents that the two logistics companies make the same price, and there are people who commission them to transport agricultural products. Equation (2) means that all agricultural products are issued by the third party Logistics Company as a result of a lower price compared with the Post Logistics. Equation (3) is just the opposite to Equation (2). Obviously, as a rational man, any logistics company will not let the latter two cases be true. It can be calculated from Equation (1):

$$x_1 = \frac{p_2 x_0 l + t_2 \delta x_0 + \lambda(t_1 - t_2)}{(p_1 + p_2)l + (t_1 + t_2)\delta}, \tag{4}$$

$$x_2 = \frac{p_1 x_0 l + t_1 \delta x_0 + \lambda(t_1 - t_2)}{(p_1 + p_2)l + (t_1 + t_2)\delta}. \tag{5}$$

Conclusion 1. In circumstances of certain total logistics demand, to attract more customers, an alternative strategy is to cut price or increase the reaction rate (the waiting time is shorter), and lower customer waiting costs (also reduces losses during circulation of agricultural products).

Proof: From the partial derivative of Equation (4) and (5), we get

$$\frac{\partial x_1}{\partial p_1} = \frac{-x_1 l}{(p_1 + p_2)l + (t_1 + t_2)\delta} \leq 0,$$

$$\frac{\partial x_1}{\partial t_1} = \frac{-(\lambda + \delta x_1)}{(p_1 + p_2)l + (t_1 + t_2)\delta} \leq 0.$$

Similarly, $\frac{\partial x_2}{\partial p_2} \leq 0, \frac{\partial x_2}{\partial t_2} \leq 0$.

The transportation volume of each logistics company is the decreasing function of its shipping rates and transit time, so the above conclusion holds.

Conclusion 2. When the two logistics companies change their strategy, one's increase speed of logistics amount is the reduce speed of the other.

Proof: Through derivation, we know

$$\frac{\partial x_2}{\partial p_2} = -\frac{\partial x_1}{\partial p_1}, \frac{\partial x_1}{\partial p_1} = -\frac{\partial x_2}{\partial p_1}, \frac{\partial x_1}{\partial t_1} = -\frac{\partial x_2}{\partial t_1}, \frac{\partial x_1}{\partial t_2} = -\frac{\partial x_2}{\partial p_2}$$

The problem is proved.

Due to the postal logistics' larger size, multi-network, wide area coverage, so it dominates the decision-making process. The third-party logistics company chooses its own best decisions based on the postal logistics' decisions. This process constitutes a Stackelberg game, which can be calculated with backward induction method.

The third-party logistics' decision problem is:

$$\max R_2 = \max p_2 x_2 l - \eta_1 x_2 - \eta_2 l - F_2 \quad (6)$$

Since R_2 is the concave function of P_2 , the optimal reaction function can be calculated from form Equation (6):

$$p_2^* = R_2'(p_2) = \frac{(p_1 l^2 + (t_1 + t_2)\delta l + \eta_1 l)(p_1 l x_0 + t_1 \delta x_0 + \lambda(t_1 - t_2))}{(p_1 l + t_1 \delta + p_2 l + t_2 \delta)^2} \quad (7)$$

The decision problem of Postal Logistics is:

$$\max R_1 = \max p_1 x_1 l - \eta_1 x_1 - \eta_2 l - F_1 = \max \left\{ \frac{(p_2 x_0 l + t_2 \delta x_0 - \lambda(t_1 - t_2))(p_1 l - \eta_1)}{(p_1 + p_2)l + (t_1 + t_2)\delta} - \eta_2 l - F_1 \right\} \quad (8)$$

S.t

$$p_2^* = \frac{(p_1 l^2 + (t_1 + t_2)\delta l + \eta_1 l)}{(p_1 l + t_1 \delta + p_2 l + t_2 \delta)^2} \times (p_1 l x_0 + t_1 \delta x_0 + \lambda(t_1 - t_2))$$

$$\left(\frac{\eta_1 x_1 + \eta_2 l + F_1}{x_1 l}, \frac{\eta_1 x_2 + \eta_2 l + F_2}{x_2 l} \right) \in (p_1, p_2) \quad (9)$$

3.2 MUTUAL COOPERATION AND PRICING

Scenario 1. Postal Logistics and third party logistics enterprises have outlets in both the agricultural products suppliers and the demanders, but the road condition has changed. Assuming rural road length l_1 , length of urban highway l_2 , the total length is $l = l_1 + l_2$. Taking into account the urban road traffic is better, the relative complexity of rural road conditions, the road coefficient

$\rho(\rho \neq 1)$ [10] is introduced to facilitate the discussion. Then for the country road whose actual distance is l_1 , the weighted distance is $l_1' = \rho l_1$. Demand of those, who need logistics service is a function on the logistics prices, expressed as $x(p) = b - kp$ [11]. And b is market capacity, k is a factor sensitive to demand and price. As the competition will reduce their income, the two logistics company could co-exist. Considering it is unlikely for the Postal Logistics to be solely responsible for the rural sections, and cooperate with 3PL in roads of towns by each responsible for part of the products, so pricing issues will be discussed below under two cooperation situation respectively.

The two companies each is responsible for a certain volume of agricultural products, with transport distance of the whole journey $l' = \rho l_1 + l_2$. At this time, the Postal Logistics Company's revenue is

$$R_1 = p_1 x_1 l' - v(x_1, l') - F_1$$

The third party logistics company's revenue is

$$R_2 = p_2 x_2 l' - v(x_2, l') - F_2$$

Equivalent behavioural strategy of both companies is to maximize its total revenue, namely:

$$\begin{aligned} \max R_{(1)} = \max & p_1 x_1 l' - F_1 - v(x_1, l') + \\ & p_2 x_2 l' - F_2 - v(x_2, l') = \\ & -kp_1^2 l' + (bl' + k\eta_1)p_1 - kp_2^2 l' + (bl' + k\eta_1)p_2 \\ & - 2\eta_1 b - 2\eta_2 l' - F_1 - F_2. \end{aligned} \quad (10)$$

The graphics of $R_{(1)}$ is a parabola opening down on p_1 and p_2 , so the maximum value can be obtained at the apex.

$$\text{From } \begin{cases} \frac{\partial R_{(1)}}{\partial p_1} = -2kp_1 l' + bl' + k\eta_1 = 0 \\ \frac{\partial R_{(1)}}{\partial p_2} = -2kp_2 l' + bl' + k\eta_1 = 0 \end{cases},$$

we get:

$$p_1 = p_2 = \frac{b(\rho l_1 + l_2) + k\eta_1}{2k(\rho l_1 + l_2)} = \frac{bl' + k\eta_1}{2kl'} \quad (11)$$

Conclusion 3. If the final amount of logistics $x \rightarrow b$, the price elasticity of demand can be is launched reversibly 0. That is to say, due to the rigid demand, logistics demanders will choose the logistics services at all costs.

Proof: From $x(p) = b - kp$, we have $k = (b - x) / p$, if $x \rightarrow b$, then $k \rightarrow 0$, the problem is proved.

Conclusion 4.

$$R_{(1)} = \frac{(bl' + k\eta_1)^2}{2kl'} - 2\eta_1 b - 2\eta_2 l' - F_1 - F_2$$

is the total revenue of logistics service providers. The problem is easy to prove by substituting Equation (11) into (10).

The Postal Logistics Company responsible for rural roads, the transport distance is l_1 . The 3PL Company responsible for urban highway, the transportation distance is l_2 . When the product flows from farmers to urban providers, the Postal Logistics first contact with logistics demanders. At this point it develops a price p to maximize the total revenue, and the whole income is $r = pxl'$.

All the cost is

$$c = v(x, l'_1) + F_1 + v(x, l_2) + F_2.$$

Equivalent behavioural strategy of both companies is to maximize its total revenue, namely:

$$R_{(2)} = pxl' - v(x, l'_1) - F_1 - v(x, l_2) - F_2 = -kl' p^2 + (bl' + 2k\eta_1)p - 2b\eta_1 - \eta_2 l' - \sum F_i \tag{12}$$

From $\frac{\partial R_{(2)}}{\partial p} = -2kl' p + bl' + 2k\eta_1 = 0,$

the best price is

$$p = \frac{2k\eta_1 + bl'}{2kl'} \tag{13}$$

Conclusion 5. The total revenue is

$$R_{(2)} = \frac{(bl' + 2k\eta_1)^2}{4kl'} - 2\eta_1 b - \eta_2 l' - \sum F_i.$$

The problem can be proved by substituting Equation (13) into Equation (12).

Scenario 2. There is no 3PL outlets from agro-food providers to demanders except the postal logistics. Then the Postal Logistics will face two different choices, one is transportation alone, and the other is cooperating with 3PL on downtown roads where there is 3PL outlets. Mark the total distance l , where rural road transport distance l_1 , urban highway transportation distance is l_2 .

1) Postal Logistics Company's network coverage is more comprehensive, with greater choice autonomy. It might be seen as the Postal Logistics commission 3PL to continue the carriage on the town road (regulatory costs neglected). Then the two have formed a secondary principal-agent relationship. As above, logistics demanders' linear inverse demand function is $x(p)$. In this case, the third party logistics company's revenue function is:

$$R_2 = p_2 l_2 x_2 - F_2 - v(x_2, l_2) =$$

$$p_2 l_2 (b - kp_2) - \eta_1 (b - kp_2) - \eta_2 l_2 - F_2,$$

$$S.t \ x_2 \leq x_1 \tag{14}$$

Construct the Lagrangian function:

$$L = p_2 l_2 (b - kp_2) - \eta_1 (b - kp_2) - \eta_2 l_2 - F_2 + \varphi(x_1 - b + kp_2),$$

and solve the problem with Kuhn Tucker conditions to obtain the 3PL's optimal pricing decisions:

$$p_2^* = \frac{\partial R_2}{\partial p_2} = \frac{bl_2 + \eta_1 k}{2kl_2} \tag{15}$$

Due to the volume of Postal Logistics in rural sections l_1 is x_1 , urban roads $x_1 - x_2$, so converted into an equivalent volume level of the entire road is

$$x_0 = \frac{x_1 l'_1 + l_2 (x_1 - x_2)}{l'} = x_1 - \frac{x_2 l_2}{l'}$$

The Postal Logistics Company's revenue function is:

$$R_1 = p_1 l' x_1 - F_1 - v(x_0, l') - p_2 l_2 x_2 = p_1 l' (b - kp_1) - F_1 - \eta_1 x_0 - \eta_2 l' - p_2 l_2 x_2 \tag{16}$$

Eventually we obtain Postal Logistics' optimal pricing:

$$p_1 = \frac{bl' + k\eta_1}{2kl'} \tag{17}$$

2) When postal logistics company selected to ship alone, its revenue function can be expressed as:

$$R'_1 = p_1 x l' - F_1 - v(x, l') = p_1 l' (b - kp_1) - F_1 - \eta_1 (b - kp_1) - \eta_2 l' \tag{18}$$

Ditto, the optimal pricing is:

$$p'_1 = \frac{bl' + k\eta_1}{2kl'} \tag{19}$$

Conclusion 6. When 3PL not cover rural areas, the Postal Logistics' pricing has nothing to do with whether it commissioned 3PL to transport the agro-food. It can be verified from Equation (17) and (19).

Conclusion 7. When the Postal Logistics commissioned 3PL to transport the agro-food, its revenue is less than when carrier alone.

Proof: Substitute the Equations (15) and (17) into (16):

$$R_1 = \frac{b^2 l' - b^2 l_2}{4k} + \frac{2b\eta_1 l_2 - k\eta_1^2}{4l'} + \frac{k\eta_1^2}{4l_2} - F_1 - \frac{\eta_1 b}{2} - \eta_2 l' \tag{20}$$

Substitute the Equation (19) into (18):

$$R'_1 = \frac{b^2 l'}{4k} + \frac{k\eta_1^2}{4l'} - \eta_2 l' - F_1 - \frac{1}{2} \eta_1 b \tag{21}$$

By subtracting Equation (20) from Equation (21) we have:

$$R'_1 - R_1 = \frac{k\eta_1^2}{2l} + \frac{b^2l_2}{4k} - \frac{k\eta_1^2}{4l_2} = \frac{2k^2\eta_1^2l_2 + (b^2l_2^2 - k^2\eta_1^2)l}{4kl_2l} > 0.$$

Therefore, the problem is proved.

4 Numerical example

For a more intuitive understanding of the model above, further analysis is made as the following. Since the revenue under competition must be less than the income in case of mutual cooperation, a brief analysis of the following two cases were mainly on mutual cooperation. Assuming the parameters in the model is $k = 1, l_1 = l_2 = 50, \eta_1 = \eta_2 = 0.5, \rho = 1.2, F_1 = 30, F_2 = 20, b = 100$. By the Mat lab programming, we know the two logistics companies are both priced at 50, and the total revenue is 5.5×10^5 in Scenario 1. The first cooperation mode in Scenario 2, the revenue of postal logistics changes as the price, shown in Figure 2. In the Figure, its highest income is between 2.5×10^5 and 3×10^5 . While selecting carriers alone, the graph of its revenue is a parabola that changes with price as shown in Figure 3, and the highest point value is 2.75×10^5 . Because it is assumed that the haul and traffic have the same coefficient impacting on the cost during numerical example, and the value of total distance and total volume is equal, the result displayed in this case shows that the income is the same whatever Postal Logistics transports alone or cooperate with the third-party logistics.

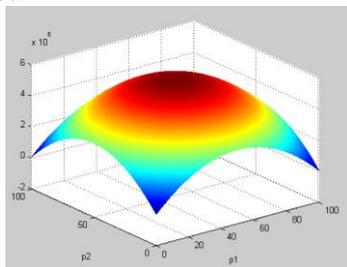


FIGURE 1 The changes of logistics service providers' total revenue with price

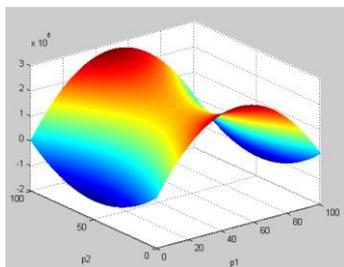


FIGURE 2 The changes of Postal Logistics' total revenue with price in Scenario 2(1)

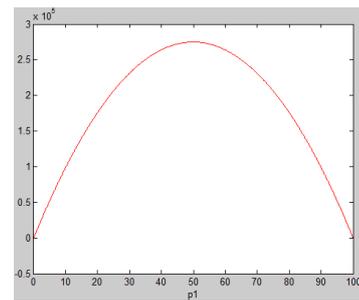


FIGURE 3 The changes of Postal Logistics' total revenue with price in Scenario 2(2)

5 Conclusion

This article gives a detailed analysis of several situations where postal logistics and third party logistics may meet with in the process of agro-food transport from rural to urban areas, and the results are verified by the use of Numerical example. In the case without prejudice to the interests of consumers, cooperation between postal logistics and third party logistics is conducive to reducing costs. However, with the existence of market information asymmetry, the subjects involved in rural logistics usually move toward the direction which is most favourable to their own marketing activities. When logistics service providers take collusion, it is easy to form a monopoly, which not only harms the interests of consumers but to some extent inhibits the market demand. It is not conducive to the healthy development of the logistics industry in rural areas. Therefore, in subsequent studies, it is necessary to take government intervention and activities of other related subjects into account, and further enrich the theory on agricultural logistics and pricing issues in competition as well as cooperation.

Acknowledgments

This project is funded by National Natural Science Foundation of China (71340024), Key Project of College Philosophy and Social Sciences (2014ZDIXM019), Key Project of Humanities and Social Sciences of Jiangsu Province (13EYA003) and Graduate Innovation Fund of Jiangsu Province 'Countermeasures of Fresh Produce's Secure Supply' (284).

References

[1] Liu Z X, Xu Z Y 2003 TPL cooperative game analysis based on the asymmetric information theory *Chinese Journal of Management Science* 11(5) 85-8 (in Chinese)
 [2] Trunick P A 2004 The secret of your 3PL success *Logistics Today* 42(2) 15-21
 [3] Xie T S, Li J 2008 Pricing game analysis for third party logistics services *Journal of Systems Engineering* 23(6) 751-7 (in Chinese)
 [4] Lei L, Wang Q, Fan C 2006 Optional business policies for a supplier-transporter-buyer channel with a price-sensitive demand

The Journal of the operational Research Society 57(3) 281-9 (in Chinese)

[5] Berstein F, Federgruen A 2007 A general equilibrium model for supply chains under price and service competition *Manufacturing and Service Operations Management* 9(3) 242-62

[6] Liu N, Wu Q 2011 Research on coordination strategy of two-level supply chain with the logistic service provider involved *Soft Science* 25(2) 125-9 (in Chinese)

[7] Xiao Y L, Li G C 2012 On cooperation boundary and effect among agricultural products logistics parties based on game theory *China Business and Market* 26(2) 33-8 (in Chinese)

[8] Li L H, Hu Z D 2012 Evolution game model about symbiosis of asymmetric logistics alliance *Journal of Changsha University of Science & Technology (Natural Science)* 9(1) 29-33 (in Chinese)

[9] Guo X L, He S Y 2004 The benefits allocation of alliance under the condition of sole-provider *Science & Technology Progress and Policy* 21(9) 95-6 (in Chinese)

[10] Zheng B, Yang H L, Tang F Z. 2011 Locating optimization and algorithm of rural logistics distribution center in county area *Journal of Dalian Maritime University* 37(1) 95-8 (in Chinese)

[11] Xue S L, Xu Y, Song Y L 2006 Integrative-optimal investigation of pricing and return policies in E-business *Operations Research and Management Science* 15(5) 133-7 (in Chinese)

Authors	
	<p>Jing Xu, born in March, 1990, Xinyang, Henan Province, China</p> <p>Current position, grades: Ph.D Student of School of Management, Jiangsu University, China. University studies: B.Sc. in Management Science and Engineering from Jiangsu University in China. M.Sc. from Jiangsu University in China. Scientific interest: regional logistics system. Publications: 6 papers. Experience: 4 scientific research projects.</p>
	<p>Guanxin Yao, born in September, 1961, Qidong, Jiangsu, China</p> <p>Current position, grades: the Professor of School of Management, Jiangsu University, China. University studies: B.Sc. in Agricultural Mechanization from Jiangsu University in China. M.Sc. from Jiangsu University in China. Scientific interest: regional logistics system. Publications: 50 papers. Experience: teaching experience of 30 years, 10 scientific research projects.</p>

Stakeholder orientation and financial performance of NPOs: development and testing of a mediating model

Yangcheng Hu *

School of Business Administration, Nanchang Institute of Technology, Nanchang, China

Received 5 July 2014, www.tsi.lv

Abstract

Literature examining how stakeholder-oriented strategy helps non-profit organizations (NPOs) cope with environmental uncertainty is limited. Thus, this study builds and tests a mediating model to investigate the relationships among stakeholder orientation, social performance and financial performance of NPOs. Data from 32 NPOs were collected in east China. Multiple regression analysis was employed to examine the relationships among the variables. The results revealed that stakeholder orientation has a positive and significant relationship with both social performance and financial performance. Meanwhile, it was discovered that social performance is significantly related to financial performance. Further, the results provided empirical evidence that the relationship between stakeholder orientation and financial performance was fully mediated by social performance. The limitations and future research directions were also discussed. This paper not only contributes to the theory of stakeholder orientation but also provides practical implications for NPOs on how to improve organizational performance.

Keywords: stakeholder orientation, NPOs, social performance, financial performance

1 Introduction

Over the past few decades non-profit sector has become “the third sector” that is independent of government and enterprise (for-profit sector), playing an important role in defusing “government failure” and “market failure” [1]. China’s non-profit sector has also experienced an unprecedented development and become an indispensable part in building “socialist harmonious society” with its wide range of involvement in social welfare, health care, education, community services, etc. However, non-profit organizations (NPOs) are commonly operating in a more highly competitive environment which is characterized by the increasing demands of stakeholders and the declining grants from governments. China’s NPOs, which are currently undergoing a comprehensive social and economic reform, are no exception to this trend. There has been a great deal of academic and practitioner interests in the concept of stakeholder orientation and its impact on organizational performance. Stakeholder orientation is critical for any organization to succeed in a fast-paced, global and turbulent environment. Thus, how to use stakeholder-oriented strategies to access external resources and to meet stakeholders’ requirements has become a practical problem to be solved for NPOs.

Despite the acknowledgment of the importance of stakeholder-oriented strategy to NPOs, scholars and practitioners are faced with a significant gap in knowledge about this subject, especially in developing countries like China which undergoes comprehensive reform and transition, and where the visibility of NPOs is a fairly new phenomenon. Meanwhile, we have also

observed another two noticeable theoretical gaps in literature. First, the stakeholder orientation literature has focused primarily on enterprises, which produced implications that are not necessarily applicable to NPOs. Second, in the analysis of stakeholder orientation and performance, most scholars have treated organizational performance as a single-dimensional construct, which neglected the multi-dimensional performance characteristics of NPOs. This study attempts to make some contributions by examining the relationships among stakeholder orientation, social performance and financial performance in China’s non-profit context.

The conceptual model is shown in Figure 1. The key interest for this research is whether the relationship between stakeholder orientation and financial performance is direct, indirect or both direct and indirect in the non-profit context. The rest of this paper is organized in the following manner. We first provide a brief review of the relevant literature and develop hypotheses. This is followed by discussions of the research methodology and results of the empirical study. We conclude the paper with the implications of the findings and give some suggestions for future research.

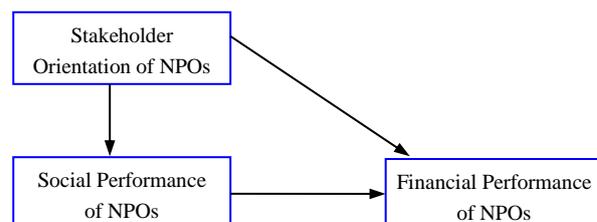


FIGURE 1 Conceptual model

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

2 Hypotheses and conceptual model

2.1 CONCEPTUAL FOUNDATIONS

In the late 1990s, scholars extended the research scope of market orientation to NPOs and conducted fruitful exploration. According to market orientation theory, customer orientation and competitor orientation should be well coordinated. However, market orientation is mainly the comprehensive measurement of two kinds of stakeholders of customers and competitors, and the consideration of other stakeholder groups seems insufficient [2,3]. With the integration of the stakeholder concept, scholars and practitioners began to focus on more types of stakeholders and studies relating to stakeholder orientation have emerged.

American Marketing Association (AMA) proposed a new definition for marketing in 2004. According to AMA, marketing is the activity, and processes for creating, communicating, and delivering values for customers, clients, partners, and society at large. Berman et al. defined stakeholder orientation as how the organization manages different stakeholders through resource allocation decisions [4]. In this vein, Greenley et al. defined stakeholder orientation as the strategic focuses on the interests of various stakeholders and management behaviours to meet different stakeholder's requirements by corporate capabilities and assets [5]. In view of the previous scholars did not pay enough attention to the measurement of stakeholder orientation, Yau et al. developed a stakeholder orientation scale of four dimensions: customer orientation, competitor orientation, shareholder orientation and employee orientation [6].

2.2 HYPOTHESES DEVELOPMENT

Sargeant et al. objected to apply the concept of market orientation to non-profit sector directly and put forward the concept of social orientation, in which the "stakeholder concerns" was treated as a sub-dimension of social orientation [7]. Padanyi and Gainer also argued that single emphasis on customer orientation cannot reflect the characteristics and diversified missions of NPOs [3]. Kelly proposed that NPOs should balance the interests of multiple stakeholders [8]. Hsieh argued that stakeholder orientation is related to performance and verified stakeholder orientation toward different stakeholder impacts performance differently in the non-profit art context [9].

However, currently no literature has provided evidence of this relationship using China's NPOs as samples. Given the multi-dimensional performance characteristics of NPOs, we used both social performance and financial performance as performance indicators. Therefore, the following two hypotheses are advanced:

H1: Stakeholder orientation positively influences social performance in NPOs,

H2: Stakeholder orientation positively influences financial performance in NPOs.

Social performance may affect financial performance by increasing organizational reputation and stakeholder trust. For one thing, positive reputation often leads to good relationships with stakeholders. For another, trust is critical to augment stakeholder satisfaction and interactive behaviour. Customer satisfaction results in customer retention and loyalty, and provides increased financial outcomes [10]. Anderson et al. confirmed that non-financial performance was positively related to economic returns on a sample of 77 enterprises in Sweden [11]. Accordingly, we conclude that, for NPOs, social performance can affect financial performance through social reputation and stakeholder satisfaction. Our discussion leads to the following hypothesis:

H3: Social performance positively influences financial performance of NPOs.

As mentioned earlier, some studies have investigated the relationship between stakeholder orientation and organizational performance while others have tested the inter-relationships among different dimensions of performance such as non-financial performance and financial performance. Since a mediator variable accounts for a significant portion of the relationship between a predictor variable and criterion variable, there is a great need to clarify whether social performance is capable of mediating the relationship between stakeholder orientation and financial performance in the non-profit context. Hence, the following hypothesis is advanced:

H4: Social performance mediates the relationship between stakeholder orientation and financial performance of NPOs.

3. Methods

3.1 SAMPLING AND DATA COLLECTION

This study selects industrial associations and community health service institutes, which act as the representatives of China's "public benefit" and "mutual benefit" NPOs respectfully, as samples for empirical analysis. All samples were taken from east China's Jiangxi province. In addition, organizations which were established less than 2 years were excluded from this study for too young organizations might not have regular stakeholder management activities. Data were collected by distributing questionnaires to the respondents that were required should be high ranking executives. Using convenience sampling method, totally 32 valid questionnaires were recovered, including 11 industrial associations and 21 community health service institutes.

3.2 MEASURES

To ensure the questionnaire items used in previous studies were applicable in China's non-profit sector, first, two-way translation of the items between English and

Chinese was carefully conducted; second, structured interviews with 5 high ranking executives from NPOs were made; third, a pilot test was undertaken. Based on the above three steps we determined the final survey questionnaires. All items were described by seven-point Likert scales ranging from 1 = completely disagree to 7 = completely agree. Following previous researches [6, 9], the present study utilized customer orientation, employee orientation and government orientation to measure stakeholder orientation. A ten-item scale reflects the extent to which the compliance between items and the organization's actuality.

Organization size and age were included as control variables because of their potential impact on organizational performance. Organization size is measured by the number of the organization's full-time employees and organization age is measured by the years of establishment.

3.3 STATISTICAL TECHNIQUES

Reliability analysis was carried out to examine the reliability of measures. Correlation analysis was conducted to find the correlations between variables. Multiple regression analysis was applied to examine the relationships among stakeholder orientation, social performance and financial performance.

4. Results

4.1 RELIABILITIES AND CORRELATIONS

The Cronbach's alpha for stakeholder orientation, social performance and financial performance were 0.925, 0.832 and 0.903 respectively, which exceeded minimum acceptable level of 0.70. The reliability coefficients for the three dimensions of stakeholder orientation ranged from 0.862 to 0.929. Since all the measures have good reliability, the average score of the multi-items for each variable was computed and used in further analysis. Pearson's correlation analysis was used to demonstrate the correlations of control variable, independent variable and dependent variable. As shown in Table 1, stakeholder orientation, social performance and financial performance were significantly and positively correlated as expected.

TABLE 1 Means, standard deviations and correlations

Variable	Correlation matrix		
	1	2	3
1. Stakeholder orientation	0.925	0.434*	0.365*
2. Social performance		0.832	0.704**
3. Financial performance			0.903

Note: N=32, *p<0.05; **p<0.01; Cronbach's alpha on the diagonal

4.2 MULTIPLE REGRESSION ANALYSIS

To test the hypotheses of this study, multiple regression analysis was used. Table 2 illustrated the results of the

regression analysis. Model1a and model 1b are the base models that include the controls. They are both non-significant (p>0.05).

Model 2a demonstrated the direct impact of stakeholder orientation on social performance. It is significant at p<0.05 level (R2=0.228) and explained an additional 21.0% variance over Model1a. Accordingly, this result supported Hypothesis 1 which states stakeholder orientation positively influences social performance of NPOs. Model 2b demonstrated the direct impact of stakeholder orientation on financial performance. It is significant at p<0.05 level (R2=0.187) and explained an additional 16.8% variance over Model 1b. Accordingly, this result supported Hypothesis 2 which states stakeholder orientation positively influences financial performance of NPOs. Model 3 showed the relationship between social performance and financial performance. It is significant at p<0.001 level (R2=0.532) and explained an additional 51.3% variance over Model 1b. Accordingly, this result supported Hypothesis 3 which states that Social performance positively influences financial performance of NPOs.

This study follows Baron and Kenny's [12] procedure to analyse the potential mediating effect of social performance in the relationship between stakeholder orientation and financial performance. See Figure 2 for a simple statistical mediation model.

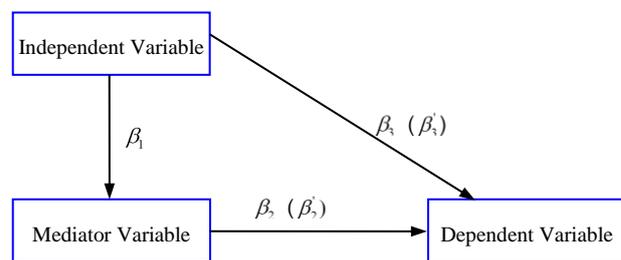


FIGURE 2 Mediation model

Step 1: Use the mediator (M) as the dependent variable in the regression equation and X as an independent variable. In Equation (1), beta_1 should be significant.

$$M = \beta_0 + \beta_1 X + \varepsilon_1 \tag{1}$$

Step 2: Use Y as the dependent variable in the regression equation and M as an independent variable. In Equation (2), beta_2 should be significant.

$$Y = \beta_0' + \beta_2 M + \varepsilon_2 \tag{2}$$

Step 3: Use Y as the dependent variable in the regression equation and X as an independent variable. In Equation (3), beta_3 should be significant.

$$Y = \beta_0'' + \beta_3 X + \varepsilon_3 \tag{3}$$

Step 4: Use Y as the dependent variable in the regression equation and X and M as independent variables simultaneously. In Equation (4), β_2' should be significant.

$$Y = \beta_0'' + \beta_3'X + \beta_2'M + \varepsilon_4. \tag{4}$$

If all these conditions hold and are in the predicted direction, then the effect of X in Equation (4) must be less than in Equation (3). Full mediating effect exists if X has no effect when the mediator is controlled.

According to the above procedure, the first step in this study is to examine the relationship between the independent variable and the mediator. As model 2a showed that stakeholder orientation relates significantly to social performance.

The second step in this study is to examine the effect of the mediator on the dependent variable. As model 3 showed that social performance positively influences financial performance.

The third step in this study is to examine the impact of the independent variable on the dependent variable. As model 2b showed stakeholder orientation positively influences financial performance.

The fourth step in this study is to include the dependent variable and the mediator simultaneously in the model to examine whether the mediator reduces the effect of the independent variable on the dependent variable. As model 4 showed, the relationship between social performance and financial performance is positive and significant, indicating the direct effect of social performance on the dependent variable. In addition, social performance reduced the effect of stakeholder orientation on financial performance to non-significance. Thus social performance played a full mediating role in the relationship between stakeholder orientation and financial performance. Accordingly, the result supported Hypothesis 4 which states Social performance mediates the relationship between stakeholder orientation and financial performance in NPOs.

TABLE 2 Results of regression analysis

Variable	Social Performance		Financial Performance			
	Model1a	Model2a	Model1b	Model2b	Model3	Model4
Control Variables						
Organization Size	0.158	0.104	-0.159	-0.208	-0.274	-0.278
Organization Age	-0.112	-0.120	0.121	0.113	0.202	0.195
Independent Variable						
Stakeholder Orientation		0.462*		0.414*		0.102
Mediator Variable						
Social Performance					0.723***	0.676***
R^2	0.018	0.228	0.019	0.187	0.532	0.540
ΔF	0.250	7.060	0.260	5.389	28.561	19.211

Note: $N=32$, * $p<0.05$, *** $p<0.001$

5. Conclusions and implications

5.1 CONCLUSIONS AND DISCUSSION

Echoing previous research's suggestions [13], this study develops a mediating model to examine the relationships among stakeholder orientation, social performance and financial performance in the context of NPOs.

The findings of this study indicate that stakeholder orientation has significant influences on both social performance and financial performance in NPOs. This result is consistent with the theoretical argument proposed by Maignan et al. [14] in that stakeholder-oriented strategy will contribute to improved non-financial and financial performance. This analysis has also supported previous empirical findings that stakeholder orientation is a critical determinant of performance of non-profit performing arts organizations [9]. Besides, the results reported here suggest that social performance is also a mediating variable, through which stakeholder orientation impacts on financial performance. A recent study performed by Mahmoud and Yusif [15] is further enhanced with our result, as the non-economic performance was seen to fully mediate the relationship between market orientation and economic performance of

NPOs. Thus, this study adds to the current body of knowledge that social performance plays very important role in understanding the relationship between strategic orientation and financial performance.

5.2 IMPLICATIONS

This study makes two major theoretical contributions by providing a better understanding on the relationships among stakeholder orientation, social performance and financial performance. Firstly, while some scholars have studied the relationship between stakeholder orientation and firm performance, research in emerging economies and in non-profit context is largely neglected. This present study, which has been performed on NPOs in China's transitional economy, adds to the theory of stakeholder orientation and NPOs. Secondly, studies on the mediating mechanism in the relationship between stakeholder orientation and financial performance are scarce. This study extends extant literature by examining how stakeholder orientation allows a non-profit organization to improve its social performance through which to improve financial performance.

From the managerial viewpoint, the findings of this study guide the practitioners of NPOs to a better understanding of the essentiality of stakeholder-oriented strategy. With the increasing demands from stakeholders and the growing competition for funds, NPOs are facing more competitive pressures in fulfilling organization's mission and achieving financial sustainability. Stakeholder orientation should be regarded as a key operation concept for China's NPOs to better obtain external resources to meet stakeholder requirements. A non-profit organization hoping to enhance its social and financial performance should pay attention to stakeholder-oriented practices. Specifically, practitioners of NPOs should emphasize the implementation of a set of practices, including assess and meet the beneficiary's demands, stimulate employees' creativity, create a harmonious working environment, and actively undertakes affairs entrusted by local government etc.

In addition, practitioners of NPOs should not take it for granted that efforts to increase stakeholder orientation will automatically enhance financial performance in the non-profit context. Since social performance plays a full mediating role in the relationship between stakeholder orientation and financial performance, practitioners of NPOs should commit to build stakeholder trust and good

social reputation which are critical antecedents of financial sustainability in uncertain environments.

6 Limitations and future research

This research is not without its limitations and should be developed in future research. The first limitation is that our data collection does not include multiple respondents. Since social performance measures in our study involve stakeholder trust and social reputation, future studies should gather information by establishing direct contact with stakeholders. Secondly, as the study employed a cross-sectional research design, a longitudinal research could provide more insight into the relationships investigated in this study. The last limitation is that empirical samples used in this study mainly came from one province in east China and were non-randomly selected, so research data will inevitably have limitations. More types of NPOs from larger scope should be used as samples to further test the views and conclusions presented in this study.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (No. 71162018).

References

- [1] Anheier H K 2009 What kind of nonprofit sector, what kind of society?: comparative policy reflections *The American behavioral scientist* **52**(7) 1082-94
- [2] Duque-Zuluaga L C, Schneider U 2008 Market orientation and organizational performance in the nonprofit context: Exploring both concepts and the relationship between them *Journal of Nonprofit & Public Sector Marketing* **19**(2) 25-47
- [3] Padanyi P, Gainer B 2004 Market orientation in the nonprofit sector: Taking multiple constituencies into consideration *Journal of Marketing Theory and Practice* **12**(2) 43-58
- [4] Berman S, Wicks A, Kotha S, Jones T 1999 Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance *Academy of Management Journal* **42**(5) 488-506
- [5] Greenley G E, Hooley G J, Rudd J M 2005 Market orientation in a multiple stakeholder orientation context: Implications for marketing capabilities and assets *Journal of Business Research* **58**(11) 1483-94
- [6] Yau. O H M, Chow R P M, Sin. L Y M, Tse. A C B, Luk. C L, Lee. J S Y 2007 Developing a scale for stakeholder orientation *European Journal of Marketing* **41**(11/12) 1306-27
- [7] Sargeant A, Foreman S, Liao M N 2002 Operationalizing the marketing concept in the nonprofit sector *Journal of Nonprofit and Public Sector Marketing* **10**(2) 41-65
- [8] Kelly L 2009 Managing stakeholder demands: Balancing responsiveness to clients and funding agents in nonprofit social service organizations *Administration & Society* **41**(2) 158-84
- [9] Hsieh J 2010 Strategic stakeholder orientations and performance consequences-A case of private nonprofit performing arts in the US *International Journal of Nonprofit and Voluntary Sector Marketing* **15**(1) 13-27
- [10] Reinartz W J, Kumar V 2000 On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing *Journal of Marketing* **64**(4) 17-35
- [11] Anderson E W, Cales F, Donald R L 1994 Customer satisfaction, market share and profitability: Findings from Sweden *Journal of Marketing* **8**(3) 53-66
- [12] Baron R M, Kenny D A 1986 The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations *Journal of Personality and Social Psychology* **51**(6) 1173-82
- [13] Ferrell O C, Gonzalez-Padron T L, Hult G T M, Maignan I 2010 From market orientation to stakeholder orientation *Journal of Public Policy & Marketing* **29**(1) 93-6
- [14] Maignan I, Gonzalez-Padron T L, Hult G T M, Ferrell O C 2011 Stakeholder orientation: Development and testing of a framework for socially responsible marketing *Journal of Strategic Marketing* **19**(4) 313-38
- [15] Mahmoud M A, Yusif B 2012 Market orientation, learning orientation, and the performance of nonprofit organisations (NPOs) *International Journal of Productivity and Performance Management* **61**(6) 624-52

Author



Yangcheng Hu, born in June, 1976, Suixi County, Jiangxi Province, China

Current position, grades: the Associate Professor of School of Business Administration, Nanchang Institute of Technology, China.

University studies: B.Sc. in Light Industrial Engineering from Anhui Agricultural University in China. Ph.D. from Zhejiang University in China.

Scientific interest: management decision, non-profit organizations.

Publications: 35 papers.

Experience: teaching experience of 6 years, 5 scientific research projects.

Study on distribution and economic growth based on feder-model: evidence from Fujian province in China

Yong-shi Hu^{1, 2*}, Yue Yu³, Ming-xing Xu³

¹*Department of Traffic and Transportation, Fujian University of Technology, Fuzhou City, Fujian Province, P.R. China, 350108*

²*Theoretical Economics postdoctoral mobile research station, Fujian Normal University, Fuzhou City, Fujian Province, P.R. China, 350108*

³*School of Economics & Management, Fuzhou University, Fuzhou City, Fujian Province, P.R. China, 350108*

Received 6 July 2014, www.tsi.lv

Abstract

The effect of distribution expansion on economic growth is examined as one of the sources of growth. The research available mostly focused on the direct effect of distribution and ignored the externality effect of distribution. The paper tested the direct effect of distribution on economic growth in Fujian province, China with the indicators such as direct contribution and direct contribution rate, and an analytical framework is developed, incorporating the possibility that marginal factor productivities are not equal in the distribution and non-distribution sectors of the economy. Econometric analysis utilizing this framework in the period 1978-2011 indicates that the existence of externality effect generated by distribution sector, which is much larger than the direct effect, significantly affected the economic growth in Fujian province, China. However, the results indicate the marginal productivity of distribution sector is lower than non-distribution sector in Fujian province, which constrains the externality effect of distribution sector. We should take action to improve the level of modernization and informationization of distribution sector and turn distribution sector from labour-intensive to technology-intensive as soon as possible.

Keywords: distribution sector, direct effect, externality effect, feder-model

1 Introduction

The research on the relationship between distribution sector and economic growth is important to clarify the role of distribution sector in regional economy. Researchers have made a thorough inquiry upon this issue from different angles, and have drawn some conclusions such as “distribution industry is a leading industry” [1], “distribution industry is a basic industry” [2], and “distribution industry is a strategic industry” [3], etc. The research above helped to explore the role of distribution sector contributing to the regional economic growth.

Many scholars also did empirical research on distribution and economic growth via mathematical statistics, econometrics and other methods. Qiao Jun estimated the contribution rate with regression analysis and proportion of distribution sector in Chinese national economy from 1986 to 1998, and concluded that the contribution rate of distribution sector in national economy in China was 10.08% from 1986 to 1998. It means that the growth of distribution sector by 1% will bring about the growth of Gross Domestic Products (shorted as GDP) by 0.1088% in China [4]. Zhou Changling, Wen Qi-xiang extended the Solow Model of economy growth and obtained a new function of economy growth. The function illustrates that the growth rate of economy depends on the growth rate of distribution sector when the growth rate of key element,

especially the growth rate of technology is constant [5]. Yang Yi-miao investigated the contribution and its measurement methods of distribution sector to the economic growth from the views of GDP or GNP's growth, employment opportunities increase, national welfare and urban development respectively [6]. Zhao Ping analysed econometrically the direct and indirect contributions of distribution sector to the economic growth using multiple regression analysis and drew three conclusions: firstly, the externality effect of distribution sector to the economic growth was underestimated by using traditional analysis methods. Secondly, distribution sector played a significant role in the macro economy due to its contribution to macro economy, its radiation to regional economy, its control to upstream industry as well as its service capability to the consumption. Thirdly, the important role of distribution sector in the economic growth and the development of society have not been fully revealed since the constraints of distribution system and the block of transmission mechanism distribution [7].

2 The direct effect of distribution sector in Fujian Province, China

Generally, one can measure the contribution of distribution sector made to the growth of regional economy with direct contribution and direct contribution rate. In the paper, the direct contribution of distribution sector is measured by the proportion of distribution sector

* *Corresponding author* e-mail: hys820728@163.com

accounting for GDP. Taking Fujian Province as an example, it can be seen from Table 1 that the proportion of distribution sector in the aggregate output of Fujian province kept increasing in the last 35 years on the whole. Especially, it rose from about 9% in 1978 to 15% in recent years. Distribution sector has been one of the major components of the economy in Fujian Province.

In addition, distribution sector plays a key role in the economic growth in Fujian Province which can be measured by the direct contribution rate. The direct contribution rate can be expressed as

$$\lambda_t = \frac{CI_t - CI_{t-1}}{GDP_t - GDP_{t-1}} \times 100\% , \quad (1)$$

where λ_t – the direct contribution rate, CI_t – the added value of distribution sector in year t , GDP_t – the GDP of Fujian province in year t .

Results adopting Equation (1) are reported in Table 2.

As can be seen from Table 2, the direct contribution rate of distribution sector in Fujian province fluctuated from 14% to 21% on the whole. Although the direct contribution rate became negative in 2005, it resumed high growth in 2006 soon. Since 1978, with the rapid economic development in Fujian Province, the materials exchange between Fujian and the outside world has been increasingly frequently. The distribution sector has become one of the pillar industries in the economic development of Fujian Province gradually.

TABLE 1 Direct effect of distribution sector in Fujian province 1978-2012

Year	the added value of distribution sector (hundred million Yuan)	GDP (hundred million Yuan)	Direct contribution (%)	Direct contribution rate (%)
1978	6.48	66.37	9.76%	—
1979	6.37	74.11	8.60%	-1%
1980	9.11	87.06	10.46%	21.16%
1981	12.5	105.62	11.83%	18.27%
1982	14.46	117.81	12.27%	16.08%
1983	16.57	127.76	12.97%	21.21%
1984	21.3	157.06	13.56%	16.14%
1985	28.17	200.48	14.05%	15.82%
1986	31.51	222.54	14.16%	15.14%
1987	41.85	279.24	14.99%	18.24%
1988	57.98	383.21	15.13%	15.51%
1989	70.37	458.40	15.35%	16.48%
1990	83.93	522.28	16.07%	21.23%
1991	100.98	619.87	16.29%	17.47%
1992	127.98	784.68	16.31%	16.38%
1993	169.75	1114.20	15.24%	12.68%
1994	237.66	1644.39	14.45%	12.81%
1995	296.85	2094.90	14.17%	13.14%
1996	362.61	2484.25	14.60%	16.89%
1997	415.29	2870.90	14.47%	13.62%
1998	667.82	3159.91	21.13%	87.38%
1999	731.24	3414.19	21.42%	24.94%
2000	809.77	3764.54	21.51%	22.41%
2001	858.43	4072.85	21.08%	15.78%
2002	910.97	4467.55	20.39%	13.31%
2003	999.40	4983.67	20.05%	17.13%
2004	1132.76	5763.35	19.65%	17.10%
2005	1039.81	6568.93	15.83%	-11.54%
2006	1204.90	7584.36	15.89%	16.26%
2007	1459.40	9249.13	15.78%	15.29%
2008	1686.09	10823.11	15.58%	14.40%
2009	2030.82	12236.53	16.60%	24.39%
2010	2448.57	14737.12	16.61%	16.71%
2011	2775.49	17560.18	15.81%	11.58%
2012	3097.81	19701.78	15.72%	15.05%

* Note: The distribution sector includes wholesale and retail trade, transportation industry, storage and logistic services, accommodation and catering industry in the paper. Source: Fujian Statistics Yearbook 1979-2012

3 The externality effect of distribution sector in Fujian Province, China

The distribution sector's effect to the economic growth reflects not only in driving the economic growth directly, but also in expanding domestic demand, increasing employment, accelerating the circulation of goods, reducing the operation costs of regional economy, optimizing the structure of the regional economy and improving the efficiency of other industries, etc. [8-9], i.e. distribution sector has externality effect (or makes

indirect contribution) to the economic growth. However, most of studies evaluated the effect of distribution sector to the economic growth from the views of direct contribution and direct contribution rate, while the indirect contribution of distribution sector to the growth of regional economy was ignored. The paper is to analyse the all effect of distribution sector to the regional economic growth (including direct and indirect contributions) comprehensively with the Feder-model proposed by Gershon Feder in 1983 [10].

3.1 FRAMEWORK OF ANALYSIS

The Feder-model was proposed by Feder in 1983, which was used to estimate the effect of exports to the economic growth. Feder divided the national economy into two departments, i.e., export department and non-export department, and deduced a new econometric model to estimate the external effect of the export department to the non-export department based on the production equation of these two departments. Simultaneously, Feder model can be used to estimate the productivity difference between export-oriented sectors and non-export-oriented sectors.

Similarly, one can estimate the contribution of distribution sector made to the regional economic growth with the Feder-model. According to the Feder-model, the paper divides the economic of Fujian province into distribution sector and non-distribution sector, and the production function of those two sectors are as below:

$$D = f(L_C, K_C), \quad (2)$$

$$N = g(L_N, K_N, d), \quad (3)$$

where D are outputs of distribution sectors, N means outputs of non-distribution sectors, L are corresponding sector labour forces stocks, and K stands for the respective sector capital stocks. In Equation (2), it supposes that the outputs of distribution sector will affect the outputs of non-distribution sectors.

In Equation (2) and (3), the factors L and K can be expressed as below respectively:

$$L = L_D + L_N, \quad (4)$$

$$K = K_D + K_N. \quad (5)$$

Suppose that the ratio of marginal factor productivities in the two sectors derivate from unity by a factor δ , i.e.

$$\frac{F_L}{G_L} = \frac{F_K}{G_K} = 1 + \delta, \quad (6)$$

where, the subscripts denote partial derivatives.

In the absence of externalities, and for a given set of prices, a situation where $\delta=0$ means an allocation of resource which maximizes national output. However, the marginal factor productivity of distribution sector is likely to be lower than that of the non-distribution sector (i.e. $\delta < 0$) due to a number of reasons.

Denoting the GDP by Y and since it follows:

$$Y = D + N. \quad (7)$$

For the variable K doesn't exist in the statistical calibre of China, but it's similar to the domestic investment (i.e. I). So, the paper denotes I to replace the variable K . Using Equation (4-6) in Equation (7) yields

$$\frac{dY}{Y} = \alpha \left(\frac{I}{Y} \right) + \beta \left(\frac{dL}{L} \right) + \gamma \left(\frac{dD}{D} \right) \left(\frac{D}{Y} \right), \quad (8)$$

where, dY/Y , dL/L and dD/D labels the growth rate of GDP, labour forces and distribution sector. The variable D/Y represents the proportion of the products of distribution sector in the GDP and the variable I/Y gives the proportion of fixed asset investment in the GDP. In Equation (8), the coefficient γ reflects all effect (including direct and indirect) of distribution sector generated to the economic growth.

In order to estimate the externality effect of the distribution sector and the differences amongst sectorial factor productivities, suppose that the product elasticity of non-distribution sector is constant, i.e.:

$$N = g(L_n, K_n, D) = D^\theta g(L_n, K_n), \quad (9)$$

where θ is a parameter that reflects the externality effect of the distribution sector. One can show

$$\frac{\partial N}{\partial D} = \theta \left(\frac{N}{D} \right). \quad (10)$$

Using Equation (9) and (10), Equation (8) can now be rewritten as:

$$\frac{dY}{Y} = \alpha \left(\frac{I}{Y} \right) + \beta \left(\frac{dL}{L} \right) + \left[\frac{\delta}{1+\delta} + \theta \left(\frac{N}{D} \right) \right] \left(\frac{dC}{C} \right) \left(\frac{C}{N} \right). \quad (11)$$

Therefore, Equation (11) can be rearranged as

$$\frac{dY}{Y} = \alpha \left(\frac{I}{Y} \right) + \beta \left(\frac{dL}{L} \right) + \left[\frac{\delta}{1+\delta} - \theta \right] \left(\frac{dC}{C} \right) \left(\frac{C}{Y} \right) + \theta \left(\frac{dC}{C} \right). \quad (12)$$

Adding a constant term and a random error term to the Equation (8) and Equation (12) respectively, the formulation in Equation (8) and Equation (12) will be the basis of the empirical work in the next section. In Equation (12) the coefficients θ and δ reflect the externality effect of the distribution sector and the derivation of marginal factor productivities from inter-sector.

3.2 EMPIRICAL ANALYSIS OF THE ALL EFFECT

In order to estimate the above equations, the paper collects the relevant data for Fujian Province from 2001 to 2009, as shown in Table 2, where the variable dY/Y is the growth rate of GDP, I/Y means the rate of fixed asset investment, dL/L represents the growth rate of labour force, dC/C is the growth rate of the added value of distribution sector, and C/Y represents the proportion of distribution sector's added value in the GDP. To eliminate the autocorrelation between variables, a lagged variable is introduced into Equation (8). The regression results of Equation (8) are shown in Table 3.

TABLE 2 Relevant data of distribution sector of Fujian province, 2001-2009

Years	dY/Y(%)	I/Y	I/Y (-1)	dL/L	dL/L (-1)	(dC/C)*(C/Y)	dC/C
2001	8.7	0.279	0.276	0.011	0.018	1.29	0.0601
2002	10.2	0.275	0.279	0.02	0.011	1.25	0.0612
2003	11.5	0.303	0.275	0.027	0.020	1.95	0.0971
2004	11.8	0.33	0.303	0.033	0.027	2.62	0.1334
2005	11.6	0.357	0.330	0.03	0.033	-1.3	0.0585
2006	14.8	0.411	0.357	0.043	0.030	2.52	0.0049
2007	15.2	0.467	0.411	0.034	0.043	3.33	0.2112
2008	13.0	0.49	0.467	0.032	0.034	2.42	0.1553
2009	12.3	0.52	0.490	0.043	0.032	0.95	0.0645

TABLE 3 Regression results for distribution industrial effect to the economic growth in Fujian Province, 2001-2009

Dependent Variable: dY/Y
Method: Least Squares
Sample: 2001-2009
Included Observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I/Y	-4.953242	7.371890	-0.671909	0.2385
dL/L	110.7671	52.91133	2.093449	0.1044
dL/L (-1)	114.8827	61.81218	1.858577	0.1366
dC/C*C/Y	0.493903	0.263963	1.871107	0.1347
C	6.660647	1.481671	4.495362	0.0109
R-squared	0.885111	Mean dependent var		12.12222
Adjusted R-squared	0.770222	S.D. dependent var		2.049864
S.E regression	0.982606	Akaike info criterion		3.102965
Sum squared resid	3.862061	Schwarz criterion		3.212534
Log likelihood	-8.963341	F-statistic		7.704046
Durbin-Watson stat	1.933727	Prob (F-statistic)		0.036566

After calculating coefficients of Equation (8) by Table 3, the expression of Equation (8) can be written as below:

$$\frac{dY}{Y} = -4.953 \frac{I}{Y} + 110.497 \frac{dL}{L} + 114.883 \frac{dL}{L} (-1) + 0.494 \frac{dC}{C} \cdot \frac{C}{Y} + 6.661$$

From the regression results, one can see that the adjusted $R^2 = 0.770$ and F -statistic=7.704, which indicate the fitting of data in Equation (8) is excellent and the regression result of Equation (8) is linear significant on the whole. Moreover, all coefficients of dependant variables (except I/Y) are statically significant ($\geq t_{0.05}(5) = 1.476$). Thus, it can be estimated that the contribution of the distribution sector generated to the

economic growth in Fujian is 0.494 of one percentage point (the coefficient γ reflects the all effect of distribution sector generated to the economic growth as explained earlier), which means that 1 percentage growth of distribution sector will drive 0.494 percentage growth of GDP in Fujian in the absence of other factors.

3.3 SPECIFYING THE EXTERNALITY EFFECT

To eliminate the autocorrelation between variables, a lagged variable $I/Y(-1)$ is introduced in Equation (12). The regression results adopting Equation (12) are shown in Table 4.

TABLE 4 Regression results for Fujian Province with specific inter-sectoral externality, 2001-2009

Dependent Variable: dY/Y
Method: Least Squares
Sample: 2001-2009
Included Observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I/Y	65.64538	32.49180	2.020367	0.1366
I/Y(-1)	-62.72183	30.88382	-2.030896	0.1352
dL/L	41.89287	76.81486	0.545375	0.3234
dC/C*C/Y	0.209636	0.317796	1.702656	0.1566
dC/C	2.288329	7.629913	2.307543	0.1713
C	7.470922	1.561390	4.784790	0.0174
R-squared	0.917616	Mean dependent var		12.12222
Adjusted R-squared	0.780309	S.D. dependent var		2.049864
S.E regression	0.960797	Akaike info criterion		2.992613
Sum squared resid	2.769391	Schwarz criterion		3.124096
Log likelihood	-7.466760	F-statistic		6.682948
Durbin-Watson stat	2.314723	Prob (F-statistic)		0.074423

According to Table 4, the expression of Equation (12) is:

$$\frac{dY}{Y} = 66.645 \frac{I}{Y} - 62.722 \frac{I}{Y} (-1) + 41.893 \frac{dL}{L} + 0.210 \frac{dC}{C} \cdot \frac{C}{Y} + 2.288 \frac{dC}{C} + 7.471$$

From the regression results, it can be seen that the adjusted- $R^2 = 0.780$ and F-statistic=6.683, which indicate the fitting of data in Equation (12) is excellent and the regression of Equation (12) is linear significant on the whole. Moreover, all the coefficients of dependant variables (except dL/L) are statically significant ($\geq t_{0.05}(4) = 1.533$). Thus, it can be estimated that the externality effect of distribution sector generated to non-distribution sector in Fujian is 2.288 (the coefficient $\rho = 2.288$ reflects the externality effect of the distribution sector as explained earlier) of one percentage point, which means that 1 percentage growth of distribution sector will drive 2.288 percentage growth of non-distribution sector in Fujian in the absence of other factors, i.e., distribution sector generates externality effect (or indirect contribution) to the growth of non-distribution sector in Fujian Province. Additionally, it can be calculated from the expressions of Equation (8) and (12) that $\delta = -1.359 < 0$. It indicates that the marginal productivity of distribution sector is lower than non-distribution sector, which constrains the externality effect of distribution sector in Fujian province.

4 Concluding remarks

Evidence from Fujian province confirmed that distribution sector strongly correlated to the regional economic growth, i.e., distribution sector directly contributes to the regional economic growth and has significant externality effect to the non-distribution

Reference

- [1] Liu G-G 1999 Converse distribution industry from ending industry to leading industry *Commercial Economics Review* 2 7-8 (in Chinese)
- [2] Huang G-X 2005 Goods distribution industry: the groundwork of economy *Finance & Trade Economics* 4 61-5, 97 (in Chinese)
- [3] Ran J-F, Wen Q-X 2005 Circulation strategic industry theory *Business Economics and Administration* 6 10-5 (in Chinese)
- [4] Qiao J 2000 Empirical analysis and countermeasures research on distribution industry development *Economics Information* 10 18-21 (in Chinese)
- [5] Zhou C-L, Wen Q-X 2003 Circulation speed and economic growth *Modern Economic Science* 25(4) 46-9 (in Chinese)
- [6] Yang Y-M 2006 On the contribution of goods distribution industry *Finance & Trade Economics* 7 16-22 (in Chinese)
- [7] Zhao P 2007 Empirical Study on the influence of distribution industry *Commercial Times* 17 15-7 (in Chinese)
- [8] Ji B-C 2010 On competitiveness and sustainable development of the circulation industry in China *China Business and Market* 24(1) 4-6 (in Chinese)
- [9] Song Z, Wang X-F 2010 Policy study on encouraging consumption of distribution industry *Finance & Trade Economics* 11 77-81 (in Chinese)
- [10] Gershon F 1983 On exports and economic growth *Journal of Development Economics* 12(1-2) 59-73

sector. An efficient and perfect distribution system is significant premises for producer to obtain factors of production, sell products and achieve the value of goods. With the improvement of productivity and the development of market economy, the function of distribution sector will evaluate from weak to strong, from simple to complex due to the demands of the regional economic operation, the externality effect of distribution sector will be more and more indispensable.

However, the externality effect of distribution sector has not fully played due to its lower marginal productivity. One can attributed this phenomenon to the following reasons: firstly, the technical level of distribution sector is low for a long time, which results in the modernization and informationization of distribution sector is incompatible with the industrialization of regional economy. Secondly, the distribution sector is a labour-intensive sector due to its own characteristics, which means it need more input than non-distribution sector to product the same level of output, i.e., the production efficiency of distribution sector is lower natively. So, in order to give full play to the externality effect of distribution sector, we should enhance the level of modernization and informationization of distribution sector and converse distribution sector from labour-intensive to technology-intensive as soon as possible.

Acknowledgements

This work is financial supported by Scientific Research Foundation for Returned Scholars of Fujian University of Technology, grant GY-S12052, which is gratefully acknowledged. The author thanks Prof. Wang Jian and Prof. Li Jianjian for the valuable discussion and recommendation.

Any opinions and views expressed in this paper are those of the authors and do not necessarily reflect those of the sponsoring organization.

Authors	
	<p>Hu Yong-shi, born in April, 1982, Fuzhou City, Fujian Province, China</p> <p>Current position, grades: Teacher of Fujian University of Technology, Fuzhou, China. University studies: M. Sc. of management from Fuzhou University in China. Doctor of management from Fuzhou University in China. Scientific interest: logistics, supply chain management. Experience: 20 scientific research projects.</p>
	<p>Yu Yue, born in August, 1986, Fuzhou City, Fujian Province, China</p> <p>Current position, grades: Ph. D. student of School of Economics & Management, Fuzhou University, Fuzhou, China. University studies: B. Sc. from Nanjing University of Science and Technology in China. M. Sc. of management from Fuzhou University in China. Scientific interest: logistics, supply chain management. Experience: 4 scientific research projects.</p>
	<p>Xu Ming-xing, born in February, 1982, Fuzhou City, Fujian Province, China</p> <p>Current position, grades: Ph. D. student of School of Economics & Management, Fuzhou University, Fuzhou, China; teacher of Fujian Normal University, Fuzhou, China. University studies: B. Sc. of Engineering from Fujian Normal University in China. M. Sc. of Management from Fuzhou University in China. Scientific interest: logistics and supply chain management, e-economics. Experience: a dozen scientific research projects and many academic papers.</p>

Study on the land desertification early-warning system of Xinjiang in China

Yan Zhang*

College of educational science, Xinjiang Normal University, Urumqi, Xinjiang, China, 830056

Received 1 July 2014, www.tsi.lv

Abstract

Land desertification is one of the greatest disasters in the world because of its wide-range influence, long duration, and enormous loss. The desertification early-warning system is very important to decrease the disaster occurrence and the reduction of disaster loss. The grid accumulation desertification early-warning model can provide this. Therefore, Land Desertification Early-warning System of Xinjiang (LDES_XJ) was built in Browser/Server (B/S) structure based on the grid accumulation land desertification early-warning model. The functions include the publishing of land desertification warning information sub-system, the desertification data management, the map browsing and query have been realized in the LDES_XJ. The system is a fundamental platform for the land desertification prevention and control as well as a convenient communication platform for the government, which makes it more convenient for relevant people to obtain the desertification early-warning information and also plays an important role in the desertification prevention and control in Xinjiang, China.

Keywords: desertification prevention, early-warning model, management information system, B/S structure

1 Introduction

Land desertification is one of the largest disasters in the world because of the influence of destruction and desertification prevention and control has become a hot issue and leading edge of the current scientific researches. The preventive measures should be taken to reduce the probability of the disaster and the damage from the disaster. In order to realize the early warning of land desertification, many scholars carried out numerous studies on land desertification warning model. Simple indicator threshold warning was first proposed. For example, Amrita G proposed an indicator of the land desertification warning for bare land [1]. Zhang put forward an indicator mechanism for the desertification disaster warning which has combined with the vegetation, soil, climate and social economy in Hunshadake basin. The comprehensive indicators of early warning were based on the indicator threshold warning with the combination of indicator weight [2]. For instance, Wang considered the indicators included vegetation, soil, and surface conditions, which determined the weight by Delphi Method [3]. Liu determined the indicator weight by AHP and constructed the land desertification early warning model of comprehensive indicators on the grid scale [4]. Chen set up a land desertification ANN early warning model by determining the indicator weight based on the cellular automation under GIS platform with the application [5]. The early warning model of mechanism was constructed considering the mechanism of desertification and features of change. For example, Wang constructed an early warning model of productivity

considering the photosynthesis, temperature, water, soil, and human activities [6]. Li built up a wind-sand expansion warning model based on the gravity, the potential and the expansion model [7]. Li established the grid accumulation desertification early warning model considering the change (fluctuation) and the accumulation of desertification [8]. Along with deep studies, a great development has been made in the land desertification model.

With the development of the land desertification warning model and the computer technology and the geographic information technology (GIS), some progresses have also been made in the land desertification warning model. Wang proposed a framework of information system of the desertification monitoring and early warning; Considering the environment of desertification disaster [9], Ding established a comprehensive desertification environmental database with GIS technology [10]; Jiang and Yang set up a desertification information system based on the desertification census data [11,12]; Wu built up a desertification analysis module based on the land desertification information system [13]; Cai constructed a land desertification grid warning system based on the C/S structure, which combined the GIS technology with Liu's grid dimension comprehensive indicator land desertification warning model [14,15]. With the growing desertification disaster as well as the spread of information network, many people have a higher demand for the desertification disaster warning. It is so important for us to obtain the land desertification warning information as quickly as possible. However, the current

* *Corresponding author* e-mail: zyan1216@163.com

land desertification warning system is a C/S structure system. Few people can access the land desertification warning information. Therefore, it is urgent for us to establish a land desertification early-warning information system with B/S framework.

The land desertification early warning model of grid accumulation was selected as the base framework which adopted Browser/Server (B/S) and the Microsoft IIS6.0 was used as WEB server. The SuperMap iServer Java 6R was chosen as a GIS platform. The VB language was used to code the land desertification warning model. The attribute relational database was constructed by Microsoft SQL Server 2008. It is expected that land desertification information can provide a suitable guidance for the desertification prevention and control and references for social and economic development in Xinjiang.

2 Grid accumulation land desertification early-warning model

The evolution process of land desertification is very slow. Huang studied the desertification evolution of the Maowusu desert and found that the region began to degenerate from the Mid-Tang Dynasty [16]. Dong studied the stratus in Tengger Desert, Taklimakan desert, and Hunshadake Desert, finding that the desertification had started before the quaternary period [17]. It can be known that the desertification takes a long time, usually dozens of years or much longer. Land desertification also has fluctuations during this long

process. Wang pointed out that the desertification would fluctuate with the climate change in North China in recent 50 years and also had an accumulation under the fluctuations [18]. Namely, the desertification development is based on the last state. Aruhan studied the factors of the desertification in Duo-lun County and found the desertification had an accumulation phenomenon [19]. Li proposed a grid accumulation land desertification early-warning model [8]. This model has better spatial scales and temporal scales, which also presents the slowness, change (fluctuation), and accumulation of the land desertification. The model formula is as follows.

$$E_d^n = \frac{\sum_{i=1}^n (U_i - U_0)}{n}, \tag{1}$$

where E_d^n is the desertification early-warning degree at the grid pixel in year n . U_n is the desertification degree at the grid pixel in year n . U_0 is the desertification degree at the grid pixel in the base year, n is year. The assignment of desertification degree between U_0 and U_n is that: non-desertification is 0, mild desertification is 1, moderate desertification is 2, severe desertification is 3, and extremely severe desertification is 4. The warning degree of E_d^n is divided by warning grade, the standards are shown in Table 1.

TABLE 1 The criterion of the desertification early-warning degree

Warning level	Already reducing	Severe reducing	Moderate reducing	Mild reducing	Changeable area	Mild warning	Moderate warning	Severe warning	Extremely Severe warning
Range	[-4,-2)	[-2,-1)	[-1,-0.5)	[-0.5,-0)	0	(0,0.5]	(0.5,1]	(1,2]	(2,4]

While accessing to the information of the land desertification, people also expect to bring back the comprehensive loss from the land desertification. Thus, the loss from the desertification can be determined based on the grid accumulation land desertification warning model when considering the land use type and the land comprehensive ecological value.

$$L^n = (C_c \times E_d^n + M_d \times E_d^n + F_l \times E_d^n + C_s \times E_d^n + W_l \times E_d^n + B_l \times E_d^n) / 4, \tag{2}$$

where L^n is the comprehensive economic loss of the grid pixel which caused by the land desertification in year n ; C_c is the comprehensive ecological value of the construction land; W_l is the comprehensive ecological value of the wetland; B_l is the comprehensive ecological

value of other lands. E_d^n is the desertification warning degree of the grid pixel in year n . L^n can be positive or negative. Positive L^n is the economic loss from land desertification. Negative L^n is the economic benefit when the land desertification improves.

The calculation of comprehensive ecological value for tillage, grass land, forest land, construction land, wetland, and other lands was in reference to Xinjiang Statistical Yearbook and Chinese land ecological system unit area and ecological service value equivalent table summarized by Xie [20,21]. The results are shown in Table 2. Both economic value and ecological value of arable land, grass land, forest land, wetland, and other lands have been considered. However, only ecological value was considered for construction land.

TABLE 2 Total output value of each type land in Xinjiang

	Arable land	Grass land	Forest land	Construction land	Wetland	Other lands
Total value (Ten thousand Yuan/ha)	5.72	5.77	18.02	10.76	51.92	0.34

3 Data source

In order to meet the need of the LDES_XJ, the data include: Xinjiang's land desertification degree data (date from 2000 to 2012); Xinjiang's land using data of year 2000, 2004, and 2008; the annual agricultural output, the animal husbandry output, the forestry output, the secondary industry GDP, and the service industry GDP in Xinjiang Statistical Yearbook from 2000 to 2013. Data involving the land desertification causes include: the annual precipitation, the annual average temperature and humidity, the annual average wind speed, max and min temperature, the annual number of sand storm; the annual water volume data; the economic and the social statistics including population, livestock amount and GDP data from 1949 to the present. Data related to the prevention

and control of desertification mainly comes from the papers.

4 Land desertification early-warning system structure

LDES_XJ must satisfy the querying of the land desertification early-warning information. Meanwhile, a platform for mutual communication is also needed. The functional modules include: the data management, desertification early warning model calculation, the desertification disaster loss calculation, the map browsing and query.

According to the functional analysis of the LDES_XJ, the B/S structure is used for the design of the system. This system is divided into four layers: the application service layer, the middle ware layer, the data layer and the infrastructure layer. The overall framework is shown in Figure 1.

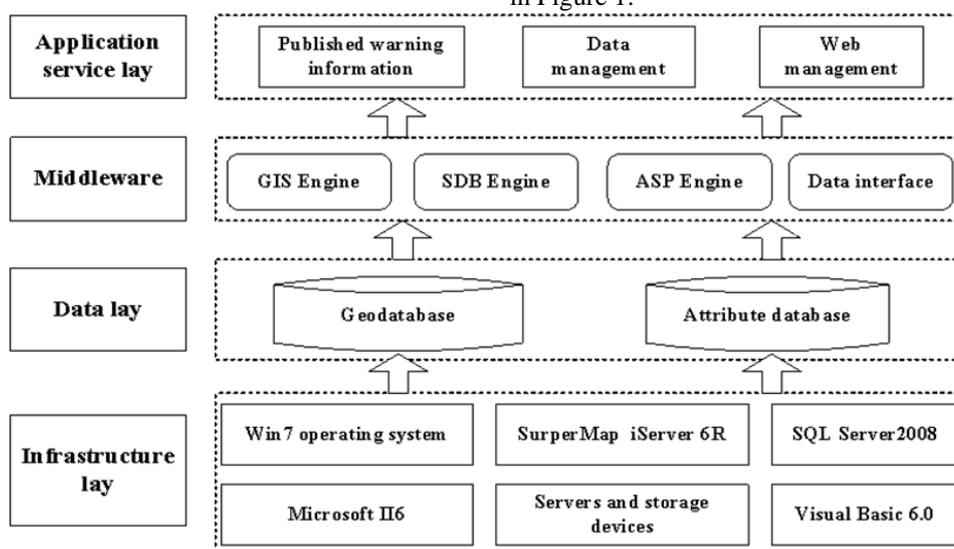


FIGURE 1 The function structure of desertification early-warning system

LDES_XJ involves in a lot of data. According to the attribute of data, the data can be divided into spatial and attribute relational. Spatial data includes vector data and grid data. Super Map Deskpro .NET 6R platform is used to build up the spatial database. While attribute relational data includes social and economic statistical data, weather station data, and hydrology station data for the cause of desertification. In addition, the data of desertification control measures and forum discussion data are also included. Attribute relational database is constructed by Microsoft SQL Server 2008. The spatial and the attribute relational database are connected with key fields [22], forming a large relational database for the LDES_XJ.

5 Application of the land desertification early-warning system of Xinjiang (LDES_XJ)

Microsoft IIS6.0 was used for Web server of the LDES_XJ and SuperMap iServer Java 6R for GIS server and the spatial database. VB language was used to code the land desertification early-warning model. Microsoft SQL Server 2008 was used to construct the attribute relational database. The interfaces of the LDES_XJ server entrance, data management and land desertification early-warning model of the LDES_XJ are displayed in Figures 2-4 (The system is designed in Chinese).



FIGURE 2 The interface of the LDES_XJ server entrance

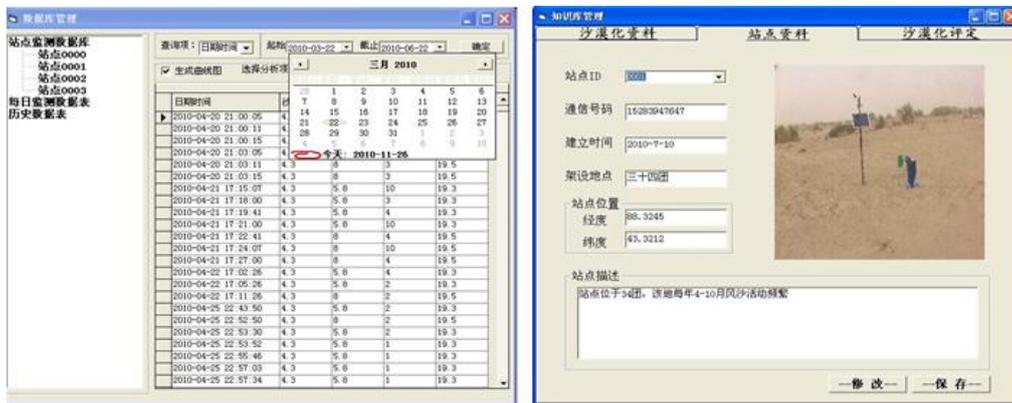


FIGURE 3 Interface of the desertification data management of the LDES_XJ (In Chinese)

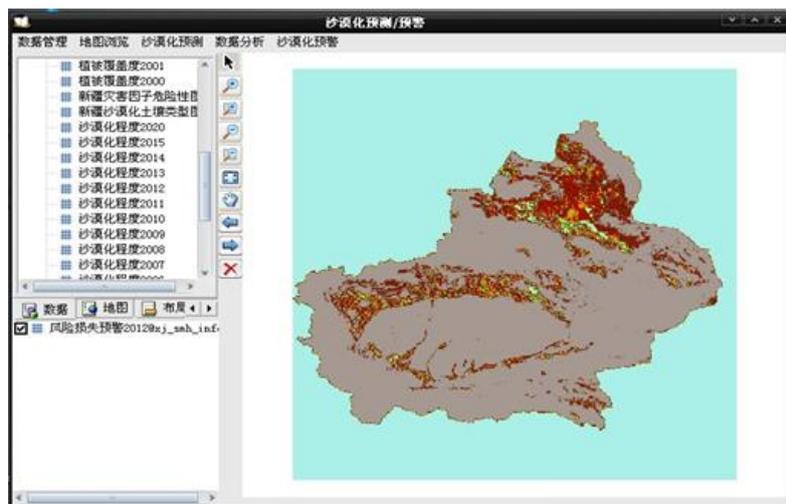


FIGURE 4 Interface of early-warning of the LDES_XJ (In Chinese)

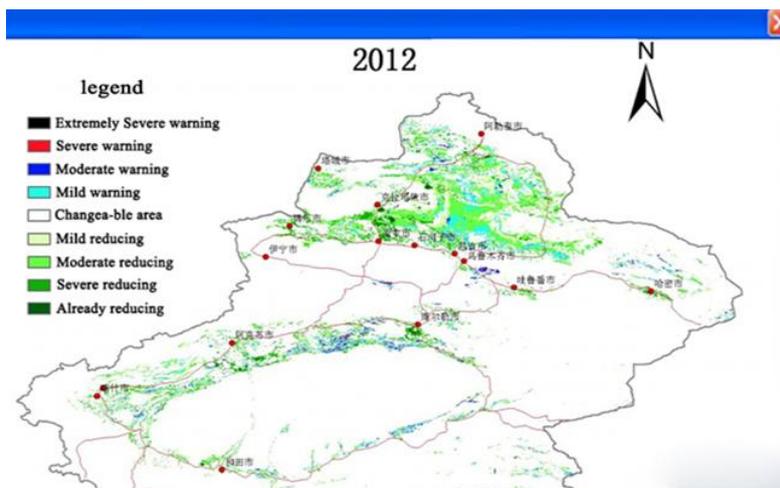


FIGURE 5 The publishing of LDES_XJ

Microsoft Active Server Page (ASP) is used for web publishing [23]. Main sections of the web include: desertification warning, desertification disaster evaluation, desertification cause query, desertification prevention and control, data remote management, desertification forum. The web publishing of the Xinjiang land desertification warning information is shown in Figure 5.

6 Conclusion

The LDES_XJ is very urgent for preventing the desertification in Xinjiang. This paper proposes a land desertification early-warning model of grid accumulation. Microsoft IIS6.0 was used as the WEB server and SuperMap iServer Java 6R as GIS server and the spatial database and VB language for the programming of land desertification warning model. Microsoft SQL Server 2008 was used to construct the attribute relational database and ASP for web publishing. The functions of the web publishing of desertification warning

information, desertification data management, map browsing and query, desertification cause query, desertification control measures query, and online discussion can be realized in the LDES_XJ. It is a fundamental platform for desertification prevention and control in Xinjiang, which enables people to obtain the desertification early-warning information more conveniently. Nowadays, this system plays an important role in the desertification prevention and control in Xinjiang, China.

Acknowledgements

This research is financed by the Natural Science Foundation of China (41301286), the Project of Center for Teacher Education Research in Xinjiang of Research Base of Humanities & Social Sciences in Xinjiang's universities (040512C03), and PhD & Post-doctor Start-up Fund Project of Xinjiang Normal University (XJNUBS1207).

References

- [1] De Soyzaa A G, Whitfordb W G, Herricka J E 1998 Early warning indicators of desertification: examples of tests in the Chihuahuan Desert *Journal of Arid Environments* 39(2) 101-12
- [2] Dong Z, Ding G-d, Ma S-l 2005 Research on indicator system of early-warning of desertification damages in Hunshandake sand *Research of Soil and Water Conservation* 12(6) 79-82 (in Chinese)
- [3] Wang J-h, Liao Y-p, Lin J 2001 Establishment of mathematical warning model on sandy desertification and the warning result of 12 provinces in the north of China *Scientia Silvae Sinicae* 37(1) 58-63 (in Chinese)
- [4] Liu D-l 2010 Research of Sandy Desertification prediction mode based on grid criterion-a case study on the middle of the lower reaches of Tarim River *Xinjiang University Press: Urumqi* 23-40 (in Chinese)
- [5] Chen J-p, Ding H-p, Wang G-w 2004 Desertification evolution modeling through the integration of GIS and cellular automata *Journal of Remote Sensing* 8(3) 254-60 (in Chinese)
- [6] Wang Z-j, Zheng J-l, Wang H-f 2004 Desertization pre-warning model and its application *China Rural Water and Hydropower* 9 4-7 40 (in Chinese)
- [7] Li H-l 2006 Risk Assessment of Sandstorm hazard and security pattern of combat wind erosion-a case study in Daxing country Beijing *Beijing Forestry University Press: Beijing* 20-35 40 (in Chinese)
- [8] Li C-z 2012 Study on the desertification monitoring and early-warning in Xinjiang *Xinjiang University Press: Urumqi* 225-30 40 (in Chinese)
- [9] Wang R-h 2001 Model of desertification disaster warning system for monitoring and evaluation Xinjiang *The western development building a green home academic seminar Sichuan: Chengdu* 2780-5 40 (in Chinese)
- [10] Ding J-l, Tashpolat-T 2001 Study on the framework of information management system of the desertified environment-a case study in Yutian county, Xinjiang *Arid Zone Research* 18(3) 63-6 40 (in Chinese)
- [11] Jiang Y, Li T-b 2002 established desertification land geographic information system of key region China *Journal of Soil and Water Conservation* 16(4) 141-3 40 (in Chinese)
- [12] Yang P, Li S, Ma J-h 2005 Design and application of GIS on desertification in Tibet *Journal of Desert Research* 25(1) 131-5 40 (in Chinese)

- [13] Wu D, Bao Y, Li B-s 2005 The Design and application of land desertification information system in arid and semiarid areas *Geo-information Science* 7(2) 46-9 40 (in Chinese)
- [14] Cai L 2011 Research of Monitoring and Prediction System on Xinjiang's Desertification *Xinjiang University Press: Urumqi* 32-7 40 (in Chinese)
- [15] Meng X-y, Liu Z-h, Du J 2013 Design and application of Tarim desertification monitoring and early warning system *Computer Technology and Development* 23(5) 159-62 (in Chinese)
- [16] Huang Y-z, Wang N-a, He T-h 2009 Process of historical desertification of Muus Desert and relationship between nature and human beings *Scientia Geographica Sinica*, 29(2) 206-11 (in Chinese)
- [17] Dong G-r 2002 Research on desert evolution and climate change of China Ocean press: Beijing 346-50 (in Chinese)
- [18] Wang X-m, Li J-j, Dong G-r 2007 Response of sandy climatic evolution and desertification In recent 50 years in north China *Chinese science bulletin* 52(24) 2882-8 (in Chinese)
- [19] Aruhan, Yang C 2007 Cumulative impacts of driving factors on desertification in Duolun country Inner Mongolia *Journal of Desert Research* 27(6) 936-41 (in Chinese)
- [20] Xie G-d, Cao S-y, Lu C-x 2010 Human's consumption of ecosystem services and ecological debt in China *Journal of Natural Resources* 25(1) 43-51 (in Chinese)
- [21] Xie G-d, Lu C-x, Leng X-f 2003 Ecological assets valuation of the Tibetan Plateau *Journal of Natural Resources* 18(2) 189-96 (in Chinese)
- [22] Kahloula B, Bouamrane K 2013 An Efficient use of a Mapping Knowledge Base *Journal of Digital Information Management* 11(4) 307-13 (in Chinese)
- [23] Li C-z, Zhang Y, Liu Z-h 2014 Optimization of MSW collection routing system to reduce fuel consumption and pollutant emissions *Nature Environment and Pollution Technology* 13(1) 177-84 (in Chinese)

Author



Yan Zhang, born in December, 1978, Urumqi, Xinjiang, China

Current position, grades: the lecturer of College of educational science, Xinjiang normal University, Xinjiang, China.

University studies: B.Sc. in compute technology from Yili normal University of Xinjiang in China. M.Sc. from Zhejiang normal University of Zhejiang in China. D.Sc. from Northeast normal university of jilin in China.

Scientific interest: computer network construction and compute education.

Publications: 10 papers.

Experience: teaching experience of 2 years, 3 scientific research projects.

Study on the location choice of Chinese outward foreign direct investment

Chao Yu^{1*}, Zongshan Pu², Lei Chen²

¹College of Cooperative, Qingdao Agricultural University, Qingdao, Shandong, China, 266109

²Sichuan Normal University Chengdu College, Chengdu, Sichuan, China, 611745

Received 10 June 2014, www.tsi.lv

Abstract

With the steady implementation of the “Go Out” strategy, China has been developing rapidly in the outward foreign direct investment. This paper, based on the 2003-2012 panel data of Chinese outward foreign direct investment, applies systematic GMM model and variable intercept model to study the location choice of Chinese outward foreign direct investment. Research findings show that, the resource endowment and bilateral trade of the host country have significantly positive impacts on Chinese outward foreign direct investment, whereas the market size and technical merit of the host country bring about significant negative impacts. Therefore, domestic corporate outward foreign direct investment should still focus mainly on resource acquisition and trade complementarity rather than acquiring technologies.

Keywords: outward foreign direct investment, location choice, technical merit, GMM model

1 Introduction

Chinese outward foreign direct investment has been developing rapidly in this century. The *2012 Statistical Bulletin of China's Outward Foreign Direct Investment* issued by the Ministry of Commerce shows that, in 2012, Chinese outward foreign direct investment spread over 179 countries (regions) all over the world, and the coverage reached 76.8%. With an investment flow incremental change of 87.8 billion dollars in ten consecutive years and a year-on-year growth of 17.6%, China became one of the world's three biggest outward foreign investor countries. By the end of 2012, Chinese outward foreign direct investment had covered all industrial categories of the national economy and the stock of investment reached 531.94 billion dollars. The cumulative investment stock of the seven industries as rental service, commercial service, finance, mining, wholesale and retail trade and manufacturing accounted for 92.4% of the overall amount Chinese outward foreign direct investment. The sudden rise of China in the outward foreign direct investment with its various characteristics and unique dualistic economy has attracted more and more attention from scholars. With the acceleration of global economic integration, outward foreign direct investment begins to play a more and more significant role in enhancing the national economic development and the international competitiveness of enterprises, and to accelerate the development of a country's outward foreign direct investment has become a focus of the academic world. Location choice is the key to solve the issue. The United Nations Conference on

Trade and Development (UNCTAD) suggested that, for the host country, there are three major influential location factors on the outward foreign direct investment launched by transnational corporations: economic factors (e.g., material availability and cost, labour cost, agglomeration effect, etc.), operation framework of foreign investment and trade (entry and operation conditions, rule stability and transparency, etc.) and corporate operation convenience (preferential policy, operational risks, etc.). Therefore, this paper mainly focuses on these aspects during the discussion.

Scholars at home and abroad apply different methods to study location choice during outward foreign investment and have come to important conclusions. Abroad: Blonigen looked into the U.S. investment in 20 OECD countries and believed that, the market size, market potential, trade costs, population and technology of the host country have a remarkable impact on the inflow of foreign investments [1]. Cheung and Qian analysed the data of Chinese enterprises' investment in 31 countries and came to the conclusion that, the market size, per capita GDP and wage level of the host country have a significant impact on Chinese outward foreign direct investment [2]. Ramasamy et al. used the Poisson regression model and found that, state-owned enterprises tend to invest in countries with abundant resources and close in politics, whereas private enterprises pay greater attention to the development of overseas market [3]. Kolstad and Wiig pointed out that, the institutional quality and resource endowment of the host country have a significant impact on Chinese outward foreign direct investment and enterprises prefer to invest in countries

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

with lower institutional quality but abundant resources [4]. Martínez-Martín used the spatial panel data model to study Spanish transnational corporations' direct investment and concluded that, the market size, trade costs, population, distance off the home country and the technical merit of the host country have a significant impact on foreign capital inflow [5]. In addition, Buckle et al. and Peimin et al. also researched into a country's corporate outward foreign direct investment [6, 7].

In China: He Benfang and Zhang Xiang applied gravity models to conduct empirical study and reached the conclusion that, the trade and labor costs, distance off and other factors of the host country have a significant impact on Chinese outward foreign direct investment [8]. Li Meng and Yu Jinping found that, the market size, resource endowment and trade links of the host country are critically influential on Chinese outward foreign direct investment [9]. Qi Chunling and Zou Chao believed that, Chinese outward foreign direct investment tends to avoid "weak-effect systems" and seek "strong-effect systems". The higher the institutional quality is, the more attractive it is to the Chinese investment [10]. Wang Juan and Fang Liangjing found that, the market size, resource endowment and opening degree of the host country have a significantly positive influence on Chinese outward foreign direct investment, while the impact of political risks and cultural links is not obvious [11]. Song Weijia and Xu Hongwei pointed out that, the resource and technical endowment, infrastructure construction, foreign capital opening degree and bilateral trade of the host country are all very important to the location choice of Chinese outward foreign direct investment [12]. Zong Fangyu et al. looked into the impact of bilateral investment treaty and institutional environment on Chinese outward foreign direct investment, and found that bilateral investment treaty can effectively promote the development of domestic enterprises' outward foreign direct investment [13]. Jiang Guan hong and Jiang Dianchun reported that, Chinese outward foreign direct investment seeks market and resources in developing countries and strategic assets in developed ones [14].

Most of the current researches on the location choice of Chinese outward foreign direct investment are based on the gravity model and the eclectic theory of international production. Conclusions reached upon different variables and models vary and sometimes even poles apart. In fact, Chinese outward foreign direct investments have its own characteristics: on the one hand, it began to develop rapidly only after 2002, and the investment flow before that was small and could hardly bring about any significant economic effect; on the other hand, outward foreign direct investment is a continuous dynamic course; thus, it's more meaningful to use the panel data after 2002 and employ dynamic models to study the location choice of Chinese outward foreign direct investment. This paper uses the 2002-2012 panel data and employs the systematic GMM to conduct empirical study on it.

2 Theoretical analysis

2.1 MARKET SIZE

Location advantage is a critical influential factor on the inflow of foreign capital and market size is an important variable to measure a country's location advantage. Enterprises are mainly driven by the constantly expanding of overseas market in developing their outward foreign direct investment. Dunning's eclectic theory of international production also stresses the effect of market size of the host country in foreign capital inflow as an important influential factor. Generally, the greater the market size and the stronger the economic strength of the host country is, the easier it is for enterprises to use the advantages of the economy of scale and scope to cut down production costs and increase profits, and thus foreign capital inflow is more likely to be attracted; on the contrary, it's harder for host countries with smaller market sizes to attract foreign capital inflows. With the constant incoming of external capitals, the market size of the host country grows and becomes more attractive to external capitals. Therefore, it is hypothesized here that, the market size of the host country is in positive correlation with Chinese outward foreign direct investment. Because GDP (gross domestic product) indicates the overall national output, this paper uses the GDP of the host country as the proxy variable for the market size.

2.2 BILATERAL TRADE

Chinese outward foreign direct investment has been developing rapidly these years and its foreign trade also increases continuously. As a result, it is hypothesized here that, bilateral trade is in positive correlation with Chinese outward foreign direct investment. Since the connection between export and outward foreign direct investment is getting closer, China's export (EXP) to the host country is used here as the proxy variable for the bilateral trade.

2.3 RESOURCE ENDOWMENT

With the further economic development, the imbalance between supply and demand gets more and more severe; the demand of oil, gas and other mineral resources rises increasingly and is strongly dependent on exports. However, in imports from foreign countries, tariffs, transport costs, trade friction and price volatility are inevitable; thus, it's rational to conduct direct investment and production in the host country through outward foreign direct investment, so that the abundant resources of the host country can be used to effectively cut down production costs and improve profits. Recent years, domestic enterprises have accelerated investments mainly targeted at resource exploitation; merger and acquisition, in particular, is developing unexpected speedy. Thus, it is

hypothesized here that, the more abundant the resources of the host country are, Chinese corporate investment is more likely to be attracted. The proportion of the sum of the mineral, metal and fuel exports in its total exports (ORE) of the host country is used here as the proxy variable for the resource endowment of the host country.

2.4 INSTITUTIONAL QUALITY

In general, higher institutional quality of a country (region) contributes to the building of a stable social environment and fairly competitive market environment, reduces information asymmetry and cuts down trade costs and controls investment risks. Conversely, a country (region) with a lower institutional quality may suffer more uncertainty due to its defective rules and inefficient administration. As a matter of fact, the impact of institutional factors on investment is not only derived from the institution of the host country but also from the institution gap between the host country and the investor's home country. With the continuous progress made by China in its legal regulations and government efficiency, the gap between it and the developed countries has been narrowed. Thus, it is hypothesized here that, Chinese outward foreign direct investment is in positive correlation with the institution quality of the host country.

World Governance Index (WGI) is recommended by the World Bank to measure the institutional environment of the host country. WGI includes six aspects, including government efficiency, political stability, rules of law and so on. Here, the index of law of the host country (INST) is used here as the proxy variable for the institutional quality of the host country.

2.5 EXCHANGE RATE

Exchange rate and its fluctuations cause significant impacts on a country's trade balance, capital flows, price level and national income. In general, the continuous appreciation of the home country's currency can reduce the commodity price of the host country, lower the foreign investment costs but also pull down the profit level of the future; on the contrary, the continuous depreciation of home country's currency can improve the profit level of the future while increasing the foreign investment costs. The impact of the appreciation (depreciation) of the home country's currency on enterprises' overseas investment has always been controversial in the academic world. These years, China Yuan (CNY) has been appreciating against major currencies as USA Dollar (USD), and the outward foreign direct investment flow has been increasing rapidly. Therefore, it is hypothesized here that, the appreciation of CNY can significantly promote the development of Chinese outward foreign direct investment. The bilateral exchange rate of China and the host country (EXCH) is here expressed with indirect quotation.

2.6 TECHNICAL MERIT

The international competition is becoming increasingly fierce with the further development of the economy. Investments targeted at technology acquirement have been developing fast. The United States and the European Union have become important investment areas for enterprises from developing countries. According to the *2012 Statistical Bulletin of China's Outward Foreign Direct Investment*, China invested 4.048 billion dollars in U.S. in 2012, and the U.S. had become the second biggest destination of Chinese outward foreign direct investment second only to Hong Kong. With the strengthening of Chinese enterprises' investment in developed countries (regions) these years, China has been more and more obviously motivated by "technology seeking" in its outward foreign direct investment. Thus, it is hypothesized here that host country with higher technical merit is more attractive to Chinese enterprise investment. The export of high and new techniques of the host country (TEC) is here used as the proxy variable for the technical merit of the host country.

2.7 INFRASTRUCTURE CONSTRUCTION

Infrastructure refers to the material facilities that serve the social production and the lives of residents. Since infrastructure construction has the "multiplier effect", that is, it can bring about a total social demand and national income a few times greater than the investment volume. The perfection of a country's (region's) infrastructure construction is critically important to the national economy. Developed transportation, complete information & communication and power system of the host country can effectively cut down the costs of operation and information collection and reduce the information asymmetry. With the further development of economy, infrastructure construction becomes more and more attractive to the foreign capital inflow of a country. Thus, it is hypothesized here that, a host country with better improved infrastructure construction is more attractive to foreign capital inflows. The sum of number of fixed broadband Internet users and telephone circuits owned by every one hundred people of the host country (INT) is here used as the proxy variable for the infrastructure construction of the host country.

2.8 WAGE LEVEL

Labour cost is an important element of enterprise's production cost, and the reduction of production cost is the key for achieving profits and a critical driving force of enterprise's outward foreign investment. The external scale economy believes that, large industrial scale can improve the production efficiency and reduce the production cost if it occurs in an area where a large number of enterprises gather. Enterprises of the home country transfer the manufacturing into other countries

with lower wage levels, and thus to greatly reduce the costs. Meanwhile, since more and more enterprises choose the same destinations for direct investment, they can effectively use the characteristics of external scale economy to have their costs further reduced and achieve higher profits, which is a significant reason why Chinese outward foreign direct investments appear to be in aggregation. Similarly, the eclectic theory of international production believes that, host countries with lower wage levels enable enterprises to achieve higher profits. Therefore, it is hypothesized here that, the wage level of the host country is in positive correlation with Chinese outward foreign direct investments in the country. Since GNI per capita reflects the wage level of the host country as a direct measurer of consumer's average income in that country, it (RGNI) is here used as the proxy variable for the wage level.

In fact, countries with higher technical merits usually have higher wage levels while the ones with low wage levels usually have low technical merit. With the continuous development of Chinese economy and the increasingly improvement of the capital strength and technical power of domestic enterprises, the destination choice of outward foreign investment and the investment flow are changing constantly. This paper provides empirical research findings about the most possible characteristics presented by Chinese enterprises' overseas investments at present.

3 Empirical test

3.1 MODEL SELECTION

This paper studies the influential factors of the host country on Chinese outward foreign direct investment from both dynamic and static angles.

First: Dynamic analysis. Since outward foreign direct investment is a continuous dynamic course, the investment flow of variables lagged by one period is taken as one of the explaining variables during model construction. The model is constructed as below:

$$\begin{aligned} \text{LnFDI}_{it} = & \alpha \text{LnFDI}_{it-1} + \beta_1 \text{LnGDP}_{it} + \beta_2 \text{LnEXP}_{it} + \\ & \beta_3 \text{LnTEC}_{it} + \beta_4 \text{INT}_{it} + \beta_5 \text{ORE}_{it} + \beta_6 \text{RGNI}_{it} + \\ & \beta_7 \text{INST}_{it} + \beta_8 \text{EXCH}_{it} + u_i + v_{it} \end{aligned} \quad (1)$$

where i and t , representing the country and the year respectively, refer to the impact of the host country induced by individual difference; $v_{it} \sim IID(0, \sigma_v^2)$ represents the general error item.

Since the explaining variables of this model are inclusive of lagged terms of explained variables, there is a correlation between the explaining variable and random disturbance term. Big errors will be produced in fixed effect estimation if OLS is employed directly. As a result, this paper applies the systematic GMM method proposed by Arellano and Bover (1995) and Blundell and Bond (1998) to estimate the model. The systematic GMM of

panel data can be further divided into the one-step method and two-step method. The weight matrix used by the two-step method, compared with the one-step estimation, deals with the impact caused by sample heterogeneity more effectively, yet despite that advantage, it can lead to considerably under-evaluated standard error of the estimated parameter. This, the significance judgment of the estimated parameter made by this paper with the two-step method is based on the robustness standard error proposed by Weidmeijer (2005). During the estimation with the systematic GMM method, the serial correlation between the instrumental variable effectiveness and the residual term needs to be tested. In this paper, the instrumental variable effectiveness is examined with the Sargen test, which is based on the number of moment condition divisors, whereas the serial correlation of the residual term is examined with the serial correlation model.

Second: Static analysis. Considering the short time span and large number of variables of data selection, this paper adopts the variable intercept model that contains N individual members. The model is as below:

$$\begin{aligned} \text{LnFDI}_{it} = & \alpha \text{LnFDI}_{it-1} + \beta_1 \text{LnGDP}_{it} + \beta_2 \text{LnEXP}_{it} + \\ & \beta_3 \text{LnTEC}_{it} + \beta_4 \text{INT}_{it} + \beta_5 \text{ORE}_{it} + \beta_6 \text{RGNI}_{it} + \\ & \beta_7 \text{INST}_{it} + \beta_8 \text{EXCH}_{it} + u_i + v_{it} \end{aligned} \quad (2)$$

where i and t represent the country and the year respectively; α_i represents the intercept term; and $v_{it} \sim IID(0, \sigma_v^2)$ represents the general error item. Since different locations have varied advantages, they attract foreign capitals in different ways. Therefore, GLS (cross-section weights) is used to do the estimation.

3.2 DATA DECLARATION

During the selection of the host country, on the one hand, data continuity and availability are taken into consideration; on the other hand, despite the fact that countries as BVI (British Virgin Islands) and the Cayman Islands have great investment flows, they are usually the transfer station of capitals can hardly reflect the ultimate object and motivation of the investment. Therefore, such regions should be removed off the samples. This paper looks into the investment data of 40 countries (regions) where China had invested during 2003-2012. Since data on stocks mainly reflects an accumulative process of the outward foreign direct investment while data on flows can better reflect the development characteristics of a country presented in different periods, this paper takes the foreign direct investment flow (FDI) as the explaining variable. The data is from the annual *Statistical Bulletin of China's Outward Foreign Direct Investment*. Besides, the data of China's outward export comes from the annual *China Statistical Yearbook*. The index of the rule of law of the host country is from the WGI indicators published by the World Bank. The rest data comes from

the World Bank’s WDI data base. Since there is considerable data missing in the 2012 hi-tech product exports of the host country, the 2002-2011 data is used to replace the original considering the data availability.

3.3 EMPIRICAL TEST AND RESULT ANALYSIS

It can be seen from Table 1 that, neither the Sargan test nor AR test is significant, indicating that instrumental variable selection is effective and residual terms are serially uncorrelated. In Model 2, the degree of fitting is also high, and specific to each variable:

First, the operation results of the two models both indicate that, there is a positive correlation between exports, resource endowment of the host country and Chinese outward foreign direct investment. In exports, since product production is divided into several procedures and each of them is usually carried out in separated location. Expansion of any of these production procedures can effectively promote the enlarging of the production scale and the improvement of production efficiency in other procedures. Therefore, the development of foreign trade can effectively promote the outward foreign investment. Meanwhile, with the further development of the foreign trade, information about the more and more extensive and detailed, which enables the market of the host country to be better understood. Thus, enterprises from the home country can make the best of the location advantages of the host country to further scale up their outward foreign investment.

Second, there is a significant negative correlation between the GDP of the host country, technical merit and China’s outward foreign investment. In Model 1, the

infrastructure construction of the host country doesn’t have a significant impact on it; while in Model 2, the negative impact is significant. Thus it can be seen that, in investment destination selection, domestic enterprises are still avoiding the developed countries with strong technical force and perfect infrastructure and paying more attention to the developing countries.

Third, investment lagged by one period has a significant impact on the current period. After the earlier period of outward foreign investment, enterprises from the home country have obtained a better understanding toward the rules of law, customs, product distribution channels, market saturation level, market potential and other information of the host country. Some enterprises even have developed their own regular customer base and sales networks. With the deepening of the understanding toward the market of the host country, domestic enterprises can get more and more detailed information and can make the best of the location advantages of the host country to further scale up the investment. Therefore, wider earlier investment scope and richer gained experience can better promote the development of the direct investment.

Fourth, according to the static analysis, the wage level of the host country has a significant negative impact on a country’s investment, which is consistent with the hypothesis analysis above. However, such impact is not significant according to the dynamic analysis, that is, the influence of the wage level of the host country has no continuity. Besides, the impact of bilateral exchange rate and institutional quality on Chinese outward foreign direct investment is also not significant.

TABLE 1 Operation result of the panel data models

Explaining variable	Model 1		Model 2	
	Coefficient	T-Statistic	Coefficient	T-Statistic
LnFDI(-1)	0.374***	4.315	-	-
LnGDP	-0.274***	-2.629	-0.617***	-9.966
LnEXP	1.002**	5.131	1.697***	23.827
LnTEC	-0.125*	-1.773	-0.135***	-3.200
INT	-0.005	-0.643	-0.011***	-3.245
ORE	0.017***	2.616	0.025***	5.323
INST	0.253	0.736	0.139	0.699
EXCH	0.001	0.510	-0.001	-0.901
LnRGNI	0.188	0.734	-1.929**	-2.186
R ²	-	-	0.90	-
AR(2)_P	-	0.326	-	-
Sargen_P	-	0.421	-	-

4 Conclusion

This paper applies macroeconomic data to study the location choice of Chinese outward foreign direct investment. It turns out that domestic enterprises tend to invest in the countries with abundant resources and in close trading relationship with China and avoid the developed countries with vast market size and advanced technology. However, this characteristic is changing. With the rapid development of domestic economy and

the increasingly improvement of enterprises’ technical merit, especially the constantly deepening impact enforced by the global economic integration, enterprises from the home country have to face more and more fierce international competition wherever their investments locate. The importance of the role played by technology in competition has been proved repeatedly. The competition of the future is the contest of core technologies, and enterprises that master the leading technology would win in the global competition; on the

contrary, enterprises without core technology would be gradually cleaned out even if they have good brand effects, high market shares and other advantages. As a result, the mastery of technology is become more and more important to enterprises.

References

[1] Blonigen B A, Davies R B, Waddel G R, Naughton H T 2007 FDI in space: Spatial autoregressive relationships in foreign direct investment *European Economic Review* 51(5) 1303-25

[2] Buckley P, Clerg L, Cross A, Liu X, Voss H 2007 The determinants of Chinese outward investment *Journal of International Business Studies* 38 499-518

[3] Cheung Y W, Qian X 2009 The empirical of China’s outward direct investment *Pacific Economic Review* 3 312-41

[4] Kolstad I, Wiig A 2010 What determines Chinese outward FDI *Journal of World Business* 47(1) 26-34

[5] Martínez-Martin J 2011 General equilibrium long-Run determinants for Spanish FDI: a spatial panel data approach *Spanish Economic Association* 2(3) 305-33

[6] Peiming W, Joseph D, Donghyun P 2013 Determinants of different modes of FDI: firm-level evidence from Japanese FDI into the US *Open Economies of Review* 24(3) 425-46

[7] Ramasamy B, Yeung M, Sylvie L 2010 Chinese outward foreign direct investment: locating choice and firm ownership *Journal of World Business* 10 1-9

[8] Li M, Yu J 2011 Correlation study on the location advantages of host country and Chinese outward foreign direct investment *World Economy* 6 63-74

[9] Jiang G, Jiang D 2012 Location choice of China’s outward foreign investment: panel data verification based on investment attraction model *World Economy* 9 21-40

[10] Wang J, Fang L 2011 Influential factors on the location choice of Chinese outward foreign direct investment *Social Scientist* 173(9) 79-83

[11] Qi C, Zou C 2013 Institutional quality and institutional distance of the host country and Chinese outward foreign direct investment *Location Contemporary Finance & Economics* 7 100-108

[12] Song W, Xu W 2012 Study of influential factors on the location choice of outward foreign direct investment *Finance and Economics Research* 10 44-50

[13] Zhang H, Wang J 2009 Location factors of host country and relation study on China’s OFDI-the empirical evidence based on quantile regression *China Industrial Economics* 255(6) 151-160

[14] Zong F, Lu J, Wu C 2012 Bilateral investment treaty, institutional quality and location choice of corporate outward foreign direct investment *Economic Research* 5 71-82

Acknowledgements

This study is supported by the key school-level projects of Qingdao Agricultural University (6613Q05).

Authors	
	<p>Chao Yu, born in August, 1980, Qingdao City, P.R. China</p> <p>Current position, grades: Qingdao Agricultural University, Qingdao, China. University studies: Dr. Degree in Management Science and Engineering. Scientific interest: Econometrics, International Finance. Publications: 10.</p>
	<p>Zongshan Pu, born in October, 1981, Gansu Province, P.R. China</p> <p>Current position, grades: the lecturer of Sichuan Normal University Chengdu College, Sichuan, China. University studies: Master degree in Administrative Management at Southwest Jiaotong University in 2011. Scientific interest: Econometrics, Management. Publications: 5.</p>
	<p>Lei Chen, born in March, 1981, Chongqing City, P.R.China</p> <p>Current position, grades: the Senior Accountant at Sichuan Normal University Chengdu College, Sichuan, China. Lecturer of The Engineering & technical College at Chengdu University since 2010, Sichuan, China. University studies: Master degree in Accounting at Southwest Jiaotong University in 2010. Scientific interest: Accounting, Auditing, Econometrics, Management. Publications: 7.</p>

The collaborative optimization of uncertain supply chain network under multi-generation co-existence

Zhang Lei*

School of Business Administration, Zhejiang University of Finance & Economics, Hangzhou, 310018, China

Received 1 August 2014, www.tsi.lv

Abstract

This paper studies the collaborative optimization of multi-generation co-existence supply chain network of single manufacturer, multi-distribution centre and market. Firstly, we considered the effects of prices, time and substitutability on the market demand of both new and old products and the whole production and marketing decision, and then subdivide the products demand of each generation. Secondly, considering that as the lifecycle stage of each generation differs, there exist differences in demand characteristics, which make the structure of supply chain network obviously different, we built WCVaR risk optimization model of production-distribution network under co-existence, which subjects to scattering distribution constraints and solve the model by Lingo11.0 and then made simulation analysis. Lastly, we verified the validity of the model with the optimization results of numerical simulation.

Keywords: multi-generation co-existence, supply chain network, collaborative optimization, WCVaR

1 Introduction

With the development of science and technology, globalizing market competition and complicated and variable demand, the lifecycle of innovative perishable goods like mobile phones, computers, toys, etc. has been shorter and shorter, and the short lifecycle of perishable goods quickens the replacement of products, so the situation of co-existence in market of multi-generation products is normal to see. About the researches on multi-generation co-existence, domestic and foreign scholars have obtained abundant achievements, and their research contents mainly focus on the market prediction of new products, the decision of new products entering a market, price decision, production decision and products innovation and benefit coordination, (Kurawarwala, 1998; Hu Zhineng, 2013) etc. However, the structural design of supply chain under multi-generation co-existence hasn't received due attention. The design optimization of supply chain network is one of the most important strategic problems in enterprises and related research achievements which mainly focus on the design, analysis, coordination, management, optimization, and recombinant of supply chain has been quite abundant, involving production and purchase game model, production-plan model, inventory model, distribution system model, logistics distribution model, network optimization model, facility location model, etc. (Carlos J Vidal, 1997; ChenJian, 2001; Marc Goetschalckx, 2002; Haralambos Sarimveis, 2008; Melo M T, 2009).

However, the existing achievements on network design optimization cannot be completely applied to the

supply chain network of multi-generation co-existence. What leads to this is that the existing researches on supply chain mainly aim at single product or unrelated multiple products, ignoring the diversity of periodical characteristics of products. Generally, in order to maintain a sound operational situation, enterprises need to conduct continuous researches to develop new products, the most satisfying picture to see is when the decline stage of the first generation products is still on the go, the second generation products have entered the growth stage of market and the third generation products have accessed to the introductory stage namely to start conceiving the latest generation products. Not only can it give full play to investment efficiency of the first generation, but also the following new products can occupy the market in succession, which can meet the demand of market and make the sales revenue and profit of enterprises grow stably as well. Some scholars held the idea that the measures taken by the supply chain should be different as the lifecycle stage of products differ, thereby, when facing the periodical demand characteristics of the lifecycle stage of the perishable goods are in the highly dynamic change, the operational strategies of the single supply chain are not compatible with the market competition any more, as a result, different operational strategies should be integrated and reconstructed fast and dynamically according to the change of periodical characteristics. However, the researches on single-product or the unrelated multiple products in most existing documents are based on the static demand characteristics while in the circumstance of multi-generation co-existence, the new and the old are

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

closely associated with each other. Meanwhile, there exists differences in demand characteristics because the lifecycle stage of each generation differs, which makes their structures of supply chain network obviously different and there is much remarkable difference with the co-existence of single product and multiple independent products. So the network optimization of double channel in co-existence should be the collaborative optimization with different supply chain patterns rather than the independent optimization with single pattern.

In view of this, this paper introduces WCVaR on the basis of the researches that have done by other scholars to measure the risk of production-distribution network which is with risk preference in the co-existence of new and old products. By making risk assessment, predicting the exposure degree of risk and the loss magnitude, we can take risk aversion measures in advance. This paper builds network optimization model which aims to minimize the risk of production-distribution network in the co-existence of new and old products and meets certain service level. The model takes the situation in which the uncertain demand information is discrete points into consideration in order to explore the production-distribution strategy with minimal risk under the worst situational network.

2 Problems description and explanation

2.1 PROBLEMS DESCRIPTION

When the perishable goods of new generation hit the market, firstly, their quality, function, and appearance design are always of higher level compared to the old products. More often than not the new products can take the place of the old but the old can't in turn, in other words, the demand is of unidirectional substitution. Secondly, in the aspect of pricing, the price of the new products is much higher as they appear on the market. Some consumers will be immediately in demand of the new products, but some consumers will hesitate because of their higher sensitivity to price than to function and they won't afford to buy before the price is goes down. So the demand elasticity of consumers in hesitation is large. For the old products, we can retain the customers and hold the market share by reducing the price and taking promotion measures, and we must take the sales potential of old products into full consideration when pricing the new products. Also, the price reduction of old products should decrease the customer loss of new products as much as possible so that the enterprises can gain more profits. Therefore, in the circumstance of co-existence, the prices of both new and old products have great influence on the demand. Moreover, with more market competitors, the competition is increasingly fierce. Only by making quick response to market can the enterprises satisfy the consumers' individualized demand, improve consumer satisfaction and then attract and retain

the consumers and finally win the market. So the response time would also affect the demand of new and old products.

In conclusion, the actual market demand of new and old products could be quite complicated because of the unidirectional substitution of new products, the interaction between the prices of new and old products and the effects of response time. In addition, the lifecycle stages of new and old products differ, so do their production-distribution patterns, modes of transportation and inventory locations, which makes their network design more complicated. But in reality, it's quite difficult to obtain the initial distribution information of demand through the historical data because of the specificity of products and sometimes the data we got are just some scattering points, which, as a result, greatly restrict the solutions to the existing models. So, in the circumstance of co-existence of multi-generations, in order to realize the risk minimization of the whole network with certain service level, enterprises need to exercise the optimal design of production-distribution network of two generations in the case of high uncertainty and complication of demand, and simultaneously to determine the price, lead time, production quantity, locations of distribution centres, integrated optimization of storage and the transportation problems.

2.2 MODEL ASSUMPTIONS AND PARAMETERS

1) Assumptions are made as follows in order to simplify the complication of models.

Assumption 1. The perishable goods of two generations are of the same brands, the new generation is in growth stage while the old generation is in mature stage.

Assumption 2. The demand variables in each market are mutually independent, the initial demand subjects to the discrete distribution of point set only.

Assumption 3. The actual demand in each market is negatively linear with the lead time and price.

Assumption 4. The demand of products is in unidirectional substitution, and the substitution rate is negatively linear with the price gap between the new and old products.

Assumption 5. The price is identical in the same market stage.

Assumption 6. The order cycles of both new and old products are identical and known, and both of them adopt the (t,s) ordering policy.

Assumption 7. The raw material suppliers and manufacturers adopt JIT distribution policy, and the manufacturers do not have raw material in stock.

Assumption 8. Not taking the restriction of production capacity and transport capacity of the factory into consideration.

Assumption 9. To get closer to reality and keep the generality, the transport of both new and old products adopt the PTP&HUB distribution modes.

2) Parameter determination.

Superscripts and subscripts:

i is the sales market;

j is the distribution centre;

k is the classifications of products;

s is the sample size.

Time parameters:

g_l is the unit production time of new semi-finished products;

l_l is the unit reprocessing time of new semi-finished products;

t_k is the production time of unit product k ;

t_o is the time interval for ordering;

$tm_{(*)}$ is the transport time between facilities.

Cost parameters:

v_k is the variable production cost of unit product k ;

q_l is the unit variable production cost of new semi-finished product;

e_l is the unit variable processing cost of new semi-finished product;

r_i is the regular freight from factory to market i ;

v_{ki} is the unit variable freight of product k from factory to market i ;

r_j is the regular freight from factory to distribution centre j ;

v_{kj} is the unit variable freight of product k from market to distribution center j ;

r_{ji} is the regular freight from distribution center j to market i ;

v_{kji} is the unit variable freight of product k from distribution center j to market i ;

cs_k is the unit shortage cost of product k .

Related parameters about price demand:

a_k is the price elasticity of demand of product k ;

b_k is the time elasticity of demand of product k ;

c is the conversion rate of demand the old products to the new products;

Δpc is the price gap between new and old products when $c=0$;

D_{ki}^{0s} is the initial demand of product k in market i ;

D_{ki}^s is the core demand of product k in market i ;

D_{ki}^s is the actual demand of product k in market i ;

p_s is the appearance probability of the s th sample.

0-1 variables:

Y_i is a 0-1 variable for whether to transport from factory to market i ;

Z_j is a 0-1 variable for whether to transport from factory to distribution center j ;

F_{ji} is a 0-1 variable for whether to transport from distribution center j to market i .

Decision variables:

pc_k is the price of product k ;

t_{ik} is the due date for product k in market i ;

x_k is the production of product k ;

m_1 is the production of new semi-finished products;

w_1 is the reprocessing amount of new semi-finished products;

ζ_{ki} is the shortage amount of product k in market i ;

y_{ki} is the traffic volume for product k from factory to market i ;

z_{kj} is the traffic volume for product k from factory to distribution centre j ;

f_{kji} is the traffic volume for product k from distribution center to market.

3 Model building and solving

3.1 MODEL BUILDING

As the lifecycle stage of both new and old generation differs, there exist differences in production-distribution patterns, storage locations and response time, which concretely represented in the differences of loss function composition, time constraint, flow conservation and demand relationship. For new products in growth stage, we adopt the production mode of ordering-assembling of which the loss function mainly consider the income loss, production and processing cost of semi-finished goods, freight and shortage loss. While for the old products, we produce according the inventory, and it's loss function mainly includes the income loss, production cost, freight and the shortage loss. In this paper, we build WCVaR risk optimization model of the whole production-distribution network in the circumstance of co-existence, which subjects to the box discrete distribution. On the basis of providing the customers of new and old products with satisfying service level, the objective function is to minimize the worst-case conditional value-at-risk of the whole production-distribution network to the least. The WCVaR risk optimization model of production-distribution network in co-existence is built as follows.

$$MinT = \sum_{k=1}^2 \theta_k + \sum_{k=1}^2 v_k x_k + q_1 m_1 + e_1 w_1 + \sum_{i=1}^I \left(r_i Y_i + \sum_{k=1}^2 v_{ki} y_{ki} \right) + \sum_{j=1}^J \left(r_j Z_j + \sum_{k=1}^2 v_{kj} z_{kj} \right) +, \tag{1}$$

$$\sum_{j=1}^J \sum_{i=1}^I \left(r_{ji} F_{ji} + \sum_{k=1}^2 v_{kji} f_{kji} \right) + \sum_{k=1}^2 \sum_{i=1}^I cs_k \zeta_{ki}$$

$$s.t. \alpha_k + \frac{1}{1-\beta} \pi^T u_{ks} \leq \theta_k, k = 1, 2, \tag{2}$$

$$u_{ks} \geq -\sum_{i=1}^I pc_k D_{ki}^s - \alpha_k, \forall k, s, \tag{3}$$

$$u_{ks} \geq 0, \forall k, s, \tag{4}$$

$$x_1 + w_1 = \sum_{i=1}^I y_{1i} + \sum_{j=1}^J z_{1j}, \tag{5}$$

$$x_2 = \sum_{i=1}^I y_{2i} + \sum_{j=1}^J z_{2j}, \tag{6}$$

$$w_1 = m_1, \tag{7}$$

$$z_{kj} = \sum_{i=1}^I f_{ijk}, \forall k, j, \tag{8}$$

$$E\{D_{ki}^s\} - \zeta_{ki} \leq y_{ki} + \sum_{j=1}^J f_{ijk} \leq \max\{D_{ki}^s\} - \zeta_{ki}, \forall k, i, \tag{9}$$

$$\sum_{i=1}^I \zeta_{ki} \leq \max\{\sum_{i=1}^I D_{ki}^s\} - x_k - w_k, \tag{10}$$

$$g_1 m_1 + t_{i1} \leq to, \forall i, \tag{11}$$

$$t_1 x_1 + l_1 w_1 + Y_i t_{i1} \leq t_{i1}, \forall i, \tag{12}$$

$$t_1 x_1 + l_1 w_1 + Z_j t_{1j} + F_{ji} t_{ji} \leq t_{i1}, \forall i, \tag{13}$$

$$t_2 x_2 + Z_j t_{2j} + t_{i2} \leq to, \forall j, i, \tag{14}$$

$$F_{ji} t_{ji} \leq t_{i2}, \forall j, i, \tag{15}$$

$$t_2 x_2 + Y_i t_{i2} \leq to, \forall i, \tag{17}$$

$$D_{ki}^s = D_{ki}^{0s} - a_k p c_k - b_k t_{ik}, \forall i, k, \tag{18}$$

$$D_{i1}^s = D_{i1}^s + c D_{i1}^s, \forall i, \tag{19}$$

$$D_{2i}^s = D_{2i}^s - c D_{2i}^s, \forall i, \tag{20}$$

$$c = 1 + (p c_2 - p c_1) / \Delta p c, \tag{21}$$

$$x_k, m_1, w_1, y_{ki}, z_{kj}, f_{kji} \geq 0. \tag{22}$$

The Equation (1) represents the minimal worst-case conditional value-at-risk of production - distribution network in the co-existence of two generations, mainly considering the income loss, production, assembling, transport and shortage cost. Equation (2) represents the income risk in the worst case, Equations (3) and (4) represent the discrete constraints; Equation (5) represents the conservation of production and transport volume of the new products; Equation (6) represents the conservation of production and transport volume of the old products; Equation (7) represents that the production of new semi-finished products is equal to the processing capacity; Equation (8) represents the conservation of the entering and exiting amount in distribution centres; Equation (9) represents the conservation of demand and traffic volume of two generations; Equation (10) represents the constraint of shortage amount; Equations (11) to (13) express the composition of actual response and constraints of new products; Equations (14) to (17) express the composition of actual response and constraints of old products; Equation (18) represents the price and time have effects on the demand; Equations (19) and (20) express the substitution rate has effect on

actual demand; Equation (21) represents the relationship between substitution rate and price gap of two generations; Equation (22) is to make constraints that each variable isn't less than zero.

3.2 MODEL TRANSFORMATION

In Equation (2), π^T represents discrete points, Zhu and Fukushima (2009) defined the discrete points of box set. When π subjects to box set distribution, then the probabilities of discrete points satisfy:

$$\pi \in P_D^B = \{\pi : \pi = \pi^0 + \eta, e^T \eta = 0, \underline{\eta} \leq \eta \leq \bar{\eta}\}. \tag{23}$$

In Equation (23), π^0 is a known distribution, η is a constant, $e^T \eta$ is to guarantee that π is a probability distribution, e is a unit vector, then P_D^B is called as box discrete distribution.

So, the constraint (2) can be equivalently transformed to:

$$\alpha_k + \frac{1}{1-\beta} \sum_{s=1}^S p_s u_{ks} + \frac{1}{1-\beta} \sum_{s=1}^S (\eta_{ks} \xi_{ks} + \eta_{-ks} \omega_{ks}) \leq \theta_k, \tag{24}$$

$$k = 1, 2$$

$$e_{ks} z z_k + \xi_{ks} + \omega_{ks} = u_{ks}, \forall k, s, \tag{25}$$

$$\xi_{ks} \geq 0, \omega_{ks} \leq 0, \forall k, s. \tag{26}$$

Now, the model has transformed to mixed integer programming model which can be solved with some related methods and software.

3.3 MODEL SOLVING

The mixed integer programming model which can be solved with branch and bound method, cutting plane approach and Lagrangian algorithm has the features of multi variables, many constrained formulas with respond time composition and constraints, and moreover great amount of data of samples need to be processed in the solving process. Compared to the general models of logistics network design, this model is much more complicated so that it's difficult to obtain the accurate solution with common methods. For optimization model solving, Lingo is considered as a professional software package which is a perfect choice to solve combinational optimization with the characteristics of efficient running and easy operation. This paper will solve the model by programming with Lingo11.0.

4 Numerical simulation

A certain manufacturer of perishable products with high-tech plans to construct a production-distribution integration network of co-existence of new and old products, and now there exists a manufactory, three markets and three distribution centres. For the new

products in growth stage, the unit production cost is 700 yuan, the production efficiency is 0.012 hour per piece. For semi-finished products, the variable production cost is 450 yuan per piece, the production efficiency is 0.008 hour per piece, the variable reprocessing cost is 230 yuan per piece and the reprocessing efficiency is 0.006 hour per piece; While for the old products in mature stage, the unit production cost is 430 yuan, the production efficiency is 0.01 hour per piece. The freight from manufactory to each market is respectively 12.5 yuan, 9.5 yuan and 10 yuan. The freight from manufactory to each distribution centre is respectively 3.6 yuan, 3 yuan and

4.2 yuan. The transport time from manufactory to each market is respectively 4 hours, 3 hours and 6 hours, while the transport time from market to each distribution centre is respectively 2.5 hours, 4 hours and 3.5 hours. The delivery lead time for new products is from 30 to 50 hours in each market and the price range is [1500,2000]; While the delivery lead time for old products is from 15 to 25 hours and the price range is [1000,1200]; $\Delta pc = 1000$ and the interval time of ordering is stipulated to be 200 hours; $a_1 = 0.01$, $a_2 = 0.03$, $b_1 = 2.5$, $b_2 = 1$, data of other related parameters is shown in Table 1 and Table 2.

TABLE 1 The unit variable transport cost from distribution centre to market (yuan per piece)

distribution centre \ market	I_1	I_2	I_3
J_1	7	6	6.5
J_2	9	4	5
J_3	5.5	2.5	8

TABLE 2 The transport time from distribution centre to market (hour)

distribution centre \ market	I_1	I_2	I_3
J_1	3	5	4
J_2	2	8	6
J_3	4.5	7.5	3

4.1 SIMULATION RESULTS

Take 500 points as the historical data sample of initial demand with Monte Carlo Method., and the points correspond with D_{ki}^s in Equation (3), namely $s=500$; let $\beta_1 = \beta_2 = 0.95$; $0.0001 \leq \eta \leq 0.0008$, solving the model

by programming with lingo 11.0, it takes 39 min 19 s to get the global optimal solution, and the calculation results are as follows: the target risk value $T = -3019568$ yuan, and the optimal operation strategy of production-distribution network is shown Table 3 and Table 4.

TABLE 3 The optimal solution of the price of both new and old products and lead time

pc_1	pc_2	t_{11}	t_{12}	t_{21}	t_{22}	t_{31}	t_{32}
1832	1200	30	15	30	15	30	15

TABLE 4 The optimal solution of the production of new and old products

x_2	xp_1	xrp_1
2953	1972	1972

TABLE 5 The logistics distribution of both new and old products and the optimal solution of 0-1 transport variables

z_{12}	z_{13}	z_{22}	z_{23}	f_{123}	f_{223}	f_{131}	f_{231}
614	1358	1092	1861	614	1092	878	973
f_{132}	f_{232}	Z_2	Z_3	F_{23}	F_{31}	F_{32}	else
480	888	1	1	1	1	1	0

As is shown in Table 3, when the new and old products coexist, in order to obtain the maximal profit, the price of old products is 1200 yuan, which is the maximum of permitted price, but its price is still far lower than the 1832 yuan of new products. To attract consumers as many as possible with lower price and then increase the sales volume, we don't price the new products at their highest, but considering the huge sales potential of old products, the price of new products shouldn't be too low for fear that the consumers of old products would turn to the new. Only in this way can the enterprises get more

profits. From Table 4, we can see that the optimal production of old products is 2953 pieces which is far more than the 1972 pieces of new products, what leads to this is that the old products are in mature stage and its demand is maximal and stable while the new products are in the grow and the demand is still increasing, from which we also can conclude that the old products are in great demand at present. Meanwhile, the new products are produced only by ordering-assembling mode, because the category of products in growth is numerous and the customers have high personalized requirement, the

production without delay costs high and it's difficult to satisfy the changeable demand in short time. By adopting ordering-assembling, not only can the scale production for semi-finished goods reduce the cost and improve the accuracy of prediction, but also can shorten the response time and assemble the products with diversity to adapt to the changeable demand. Furthermore, the optimal lead time of both new and old products is the minimum, the reason is that whatever stage they are in, growth stage or mature stage, the response time is the key factor to improve the service level, to enhance the competition ability and to win the market.

From Table 5, we can know that both new and old products which are distributed through J2 and J3 choose the HUB only as the centralized distribution, this is because that the manufacturer is far away from the markets. In spite of the guaranteed timeliness of nonstop transportation, the freight is quite high when directly distributed from manufacturer to 3 markets, which is respectively 12.5 yuan, 9.5 yuan and 10 yuan per piece while the highest and the lowest unit freight from distribution centers to market is just respectively 12.2 yuan and 7 yuan, so we don't adopt the PTP direct distribution. Similarly, the unit transportation cost from distribution centers J1 to 3 markets respectively reaches to 23.1 yuan, 22.1 yuan and 22.5 yuan per piece, which makes the total freight much higher than the freight from J1 and J2, while the freight to 3 markets in the optimal scheme is respectively 20, 17, 17.5 yuan per piece, so we don't choose J1. For old products, it's likely to reduce the logistics cost with the scale advantages of centralized distribution and the new products can be assembled in distribution centres as well in order to make response to the market demand in short time. On the other hand, the storage and transportation of both new and old products can be integrated by choosing the same distribution centres and distribution modes, cutting down the cost of storage, transportation and management to the maximal extend. From the logistics assignment of each node, we can see that the model takes the choice of routes and logistics assignment into consideration from the optimal view of the whole production-distribution network.

In conclusion, the strategies of production and sale in present market list as follows: the old products are still in great demand, so we adopt the low price strategy and maximal price level to stabilize the market and increase

the sales; The new products should be priced rationally to find a balance between its own profit and the profit reduction of old products so as to realize the maximal profit of the two products. As for the production, the short term demand of old products is steady, so we should take the effects of price, time and demand substitution into full consideration according to the existing demand information, and then arrange the production reasonably; While the new products should also attach much importance to the demand of old products and pay due attention to the effects of its own price, time and substitution rate, and then price and make production plan for semi-finished goods reasonably. In terms of the choices of distribution modes and distribution centres, both new and old products should follow the principle that cost is to keep optimal in a certain period of time, thus reducing the transportation cost to the maximal extend. In terms of the logistics assignment, the scale benefit brought by the integrated transportation of the two products should receive due attention to minimize the cost of logistics and management. So, effective and reasonable production-distribution network can provide powerful guidance for decision making in the market where the new and old products coexist, and the correct decision not only can effectively weight various costs, reasonably arrange the logistics assignment for the two products, minimize the risk of supply chain and win much more profits, but also can incent various strategies to optimize and improve constantly.

4.2 SENSITIVITY ANALYSIS

By analysing the sensitivity of some parameters in model can the decision maker choose appropriate optimization strategy of production-distribution network according to the actual situation.

To analyse the effects of the risk preference of decision maker on the WCVaR risk value of the whole production-distribution network, now analyse the sensitivity respectively for the confidence levels of both new and old products (β_1 and β_2). Taking the parameters in examples as reference value, we scale down the ratio between β_1 , β_2 and the reference value by 5% and then explore the relationship between the WCVaR risk value and the lessoned ratio. The result is shown in Figure 1.

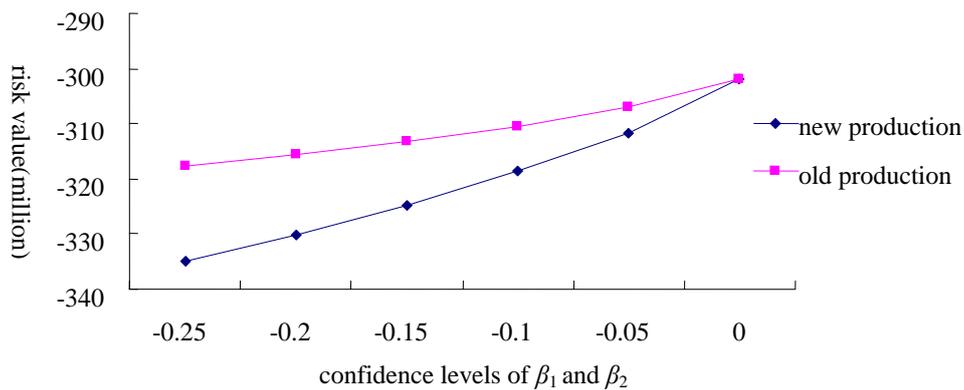


FIGURE 1 Relationship between the change of β_1 and β_2 and target value

As is shown in Figure 1, when other conditions are fixed, respectively change the confidence levels of both new and old products (β_1 and β_2), the target value is much more sensitive to the confidence level of new products than to that of the old products, which indicates that the decision maker's attitude to the risk of new products can arouse greater fluctuation for the whole production-distribution network, reflecting the new products market has stronger sensitivity. So, the decision maker could take

different measures in different markets according to the actual operation to reduce the risk loss.

The alteration of confidence level β_1 and β_2 do not have effects on the production, distribution modes, option of distribution centres and low assignment, but it does affect the price strategies of both new and old products. The Table 6 shows the optimal price of the two products at different confidence levels.

TABLE 6 The optimal price of new and old products at different confidence levels

Confidence level	β_1	0.95			0.90			0.85		
		β_2	0.95	0.90	0.85	0.95	0.90	0.85	0.95	0.90
New product		1832	1847	1857	1839	1854	1864	1842	1855	1865
Old product		1200	1200	1200	1200	1200	1200	1200	1200	1200

From Table 6, we can see that however the confidence level β_1 and β_2 changes, the price of old products remains at 1200, which means that the decision maker's risk attitude can't affect the price of old products. When the values of β_1 and β_2 shrink simultaneously, the price of old products rises, which indicates that when the risk preference decreases, the decision maker tends to adopt conservative marketing strategy, namely to sustain the market share to the most at low risk level. While when β_1 is fixed and only β_2 is reduced, the price of new products also increase, the width of which, however, is over the effects of its own confidence level. This is because the risk value is much more sensitive to the confidence level of new products than to that of the old products, the risk reduction brought by the reduction of unit β_2 is less than what brought by

unit β_1 , which, as a result, makes the integral risk level increase relatively and finally price high for new products so as to maintain the high profits. So, in the market where the new and old products coexist, the decision maker's risk preference to different products market would affect mutually and the decision should be made after integral trade off and consideration according the actual situation. Under the box discrete distribution, the disturbance value of probability distribution of the random variable D_{ki}^s is between $\bar{\eta}_{ks}$ and $\underline{\eta}_{ks}$. In order to acquire the effects of demand fluctuation extend on the WCVaR risk value of network, now we analyze how the WCVaR risk value of production-distribution network changes under different disturbance. As shown in Figure 2.

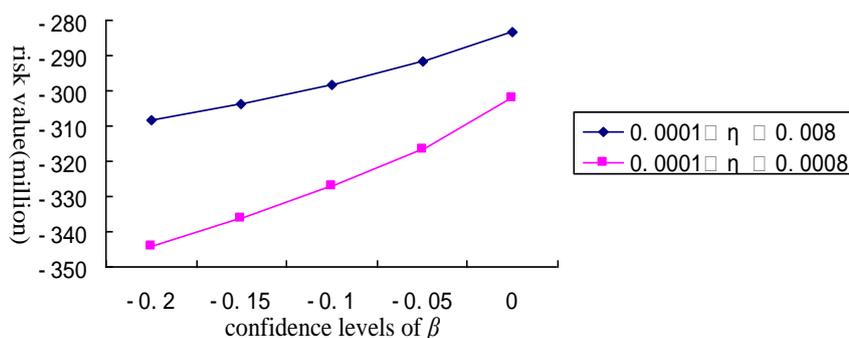


FIGURE 2 The relationship between disturbance value and target value at different confidence levels

From Figure 2, the corresponding decision value and WCVaR risk values would change as the disturbance value varies. At the same confidence level, the WCVaR risk value would increase as the disturbance value of demand probability distribution increases, which indicates that the larger fluctuation extends, the higher network risk. Furthermore, when widening the variation range of disturbance value, the WCVaR risk value tends to be stable. It means that the model is with good robustness in the disturbance which is brought by uncertain random distribution. Thus, it can be seen that the WCVaR method can effectively measure the risk of production-distribution network when the demand of products market is with different fluctuation extend, and the larger the fluctuation extends, the more robust the model is. It provides beneficial guidance, reference and help for the decision makers of supply chain to analyse, measure and manage the actual risk of production-distribution network.

5 Conclusion

What attracts customers to perishable products most is the uniqueness in technology, performance and value. The uniqueness of products is always perishable because once a product is in good graces of customers, the

competitors would produce similar goods with strong substitutability as soon as possible, which force the enterprise to utilize newer technology and update the products constantly to ensure the stable income in the long term. According to the co-existence of multi-generations, this paper builds risk optimization model which considers the substitution of new products to old products, the effects of price and lead time on demand and generate discrete data sample of demand with Monte Carlo Method. We explore the optimal strategy which meets the minimal WCVaR risk value of the whole production-distribution network, namely simultaneously determine the production scale of both new and old products, price, lead time, locations of distribution centres and the optimal transport route which are verified through numerical simulation.

Acknowledgments

This work was supported by GuangXi Education Department Foundation of China (No. 2013YB146), National Social Science Foundation of China (No. 12CGL045), Fujian Provincial Natural Science Foundation of China (No. 2011J01381) and China Postdoctoral Science Foundation (No. 2013M531541).

References

- [1] Kurawarwala A A, Matsuo H 1998 Product growth models for medium-term forecasting of short life cycle products *Technological Forecasting and Social Change* **57**(3) 169-96
- [2] Hu Z, Pei Y 2013 Impact of product sampling on the diffusion of multi-generation products based on the pricing strategy *Journal of Systems Engineering* **28**(3) 316-26 (in Chinese)
- [3] Vidal C J, Goetschalckx M 1997 Strategic production-distribution models: A critical review with emphasis on global supply chain models *European Journal of Operational Research* **98**(1) 1-18
- [4] Goetschalckx M, Vidal C J, Dogan K 2002 Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms *European Journal of Operational Research* **143**(1) 1-18
- [5] Sarimveis H, Patrinos P, Tarantilis C D, Kiranoudis C T 2008 Dynamic modeling and control of supply chain systems: A review *Computers & Operations Research* **35**(11) 3530 -61
- [6] Melo MT, Nickel S, Saldanha-da-Gama F 2009 Facility location and supply chain management-A review *European Journal of Operational Research* **196**(2) 401-12
- [7] Aitken J, Childerhouse P, Towill D 2003 The impact of product life cycle on supply chain strategy *International Journal of Production Economics* **85**(2) 127-40
- [8] Zhu S, Fukushima M 2009 Worst-case conditional value-at-risk with application to robust portfolio management *Operations Research* **57**(5) 1155-68

Author



Zhang Lei, born in January, 1978, Hangzhou County, Zhejiang Province, China

Current position, grades: the Associate Professor of School of Business Administration, Zhejiang University of Finance & Economics, China.

University studies: D.Sc.Tech. from Southwest Jiaotong University in China.

Scientific interest: supply chain management.

Publications: 20 papers

Experience: teaching experience of 4 years, 2 scientific research projects.

On dynamic iterative algorithm and the loss of newsvendor problem

Hao-ran Shi^{1, 2}, Xiaoqing Liu^{1*}, Yao Yang¹

¹School of Energy and Environment, Xihua University, 610039, Chengdu, P.R. China

²School of Transportation and Logistics, Southwest Jiaotong University, 610031, Chengdu, P.R. China

Received 1 July 2014, www.tsi.lv

Abstract

Based on the thinking and method of dynamic programming, this paper calculates the expected profit of every selling cycle of newsvendor by historical data and calculates the expected profit of the second selling cycle by that of the first selling cycle. We will get the optimal purchasing quantity as the expected profit begins to fall. Besides, this paper also discusses the stock loss and gets the optimal purchasing quantity by empirical examples.

Keywords: newsvendor problem, marginal income, dynamic iterative algorithm, optimal purchasing quantity

1 Introduction

Stocking is important for production and sales. It is divided into random stock model and definite stock model. Newsvendor problem is a typical random stock model. Traditional algorithms for the newsvendor model are common in textbooks and journals [1-5]. Many researchers have probed into this problem. Literature [6-10] discuss object function, strategic variation and model parameters respectively. Other researches [11-13] study the problem from time restriction and financial benefits. And more apply it to purchasing and inventory management of the supply chain [14-16] rather than focus on algorithms. This paper will adopt the principle and idea of dynamic programming, does some calculation and works out the optimal formula for newsvendor problem.

2 Traditional dynamic iterative algorithm

Traditional newsvendor problem model is shown in Literature [14]. The newsvendor gets the newspaper wholesale at 0.3 Yuan from the news agency and sells it at 0.45 Yuan. If the newspaper cannot be sold out, he will lose 0.3 Yuan. If out-of-stock, then the expected loss is 0.15 Yuan. Historical sales record is as Table 1, then how many pieces of newspapers should the newsvendor purchase to reduce the loss to the minimum?

TABLE 1 Historical sales record of the newspapers

Demand (pieces)	120	130	140	150	160
Demand probability P(D)	0.15	0.2	0.3	0.25	0.1

Solution: If calculated by traditional algorithm, we can get:

If purchasing 120 pieces, the average loss will be:

$$(120 - 120) \times 0 \times 0.15 + (130 - 120) \times 0.15 \times 0.2 + (140 - 120) \times 0.15 \times 0.3 + (150 - 120) \times 0.15 \times 0.25 + (160 - 120) \times 0.15 \times 0.1 = 0 + 1.5 \times 0.2 + 3 \times 0.3 + 6 \times 0.1$$

If purchasing 130 pieces, the average loss will be:

$$(130 - 120) \times 0.3 \times 0.15 + (130 - 130) \times 0 \times 0.2 + (140 - 130) \times 0.15 \times 0.3 + (150 - 130) \times 0.15 \times 0.25 + (160 - 130) \times 0.15 \times 0.1 = 3 + 0 + 1.5 \times 0.3 + 3 \times 0.25 + 4.5 \times 0.1 = 2.1$$

We can get the average loss for other purchases, as is shown in Table 2.

TABLE 2 Loss for different numbers of newspapers' sales amount

D	120	130	140	150	160	Average loss
P(D)	0.15	0.2	0.3	0.25	0.1	
F						
Q						
120	0	1.5	3	4.5	6	2.925
130	3	0	1.5	3	4.5	2.1
140	6	3	0	1.5	3	2.175
150	9	6	3	0	1.5	3.6
160	12	9	6	3	0	6.15

From Table 2, we can see that when the purchasing amount equals to the demand, the failure cost is 0. As a result, the loss in Table 2 can be divided into two types, one is the backlog of loss, on the left of 0; the other is out-of-stock loss, on the right of 0. The backlog of loss is real while out-of-stock loss talks about the probability. In the traditional algorithm, two losses weight the same and it is necessary to reduce their total amount to the

* Corresponding author e-mail xiaoqingliou1999@163.com

minimum. The result is shown in Table 3, the same as in Table 2. We also get the curve, as in Figure 1.

TABLE 3 Contrast of two losses

D	120	130	140	150	160			
P(D)	0.15	0.2	0.3	0.25	0.1	backlog of loss	out-of-stock loss	Average loss
120	0	1.5	3	4.5	6	0	2.925	2.925
130	3	0	1.5	3	4.5	0.45	1.65	2.1
140	6	3	0	1.5	3	1.5	0.675	2.175
150	9	6	3	0	1.5	3.45	0.15	3.6
160	12	9	6	3	0	6.15	0	6.15

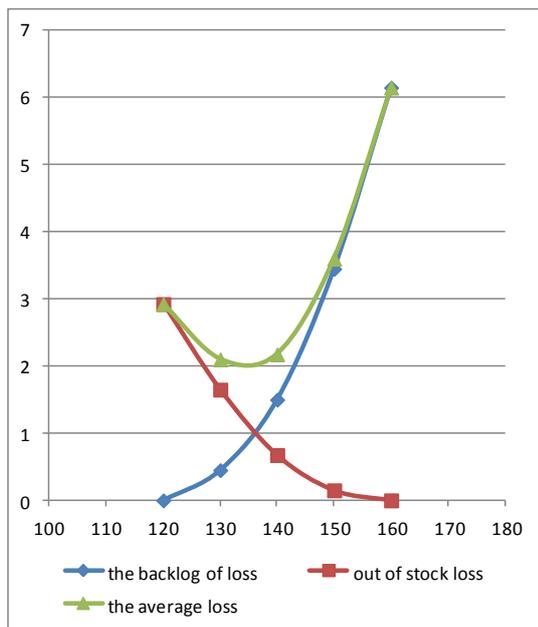


FIGURE 1 Failure cost for the newsvendor problem

If calculated by dynamic iterative algorithm, first we need to get the cumulative frequency for every sale amount, as in Table 3.

TABLE 4 Sales probability and cumulative frequency.

Demand (D)	120	130	140	150	160
Demand Probability P(D)	0.15	0.2	0.3	0.25	0.1
Cumulative frequency	1	0.85	0.65	0.35	0.1

Cumulative frequency means the frequency of the least amount of sales. For example, if the cumulative frequency for 120 pieces is 1, it means that 120 pieces of newspapers can be for 100% sure. In the same point, the cumulative frequency for 140 pieces is 0.65, indicating

TABLE 5 The process of iterate algorithm for the newsvendor problem

D	120	130	140	150	160			
P(D)	0.15	0.2	0.3	0.25	0.1	Selling profit	Increasing the unit backlog of loss	Expected profit
120	18	0	0	0	0	18	0	18
130	18	1.275 - 0.45	0	0	0	19.275	0.45	18.825
140	18	1.275 - 0.45	0.975 - 1.05	0	0	19.8	1.05	18.75
150	18	1.275 - 0.45	0.975 - 1.05	0.525 - 1.95	0	19.275	1.95	17.325
160	18	1.275 - 0.45	0.975 - 1.05	0.525 - 1.95	0.15 - 2.7	17.475	2.7	14.775

that 65% of 140 can be sold. It consists of 0.1 (cumulative frequency for 160 pieces), 0.25 (cumulative frequency for 150) and 0.3 (cumulative frequency for 140). And $1 - 0.65 = 0.35$, indicating that 0.35% of 140 pieces cannot be sold, either 120 can be sold with the possibility of 0.15 or 130 can be sold with the possibility of 0.2.

Then we are getting the expected profit.

The purchasing of 120 pieces consists of two parts: the profit and the loss. Then there is:
 $E[C(120)] = 120 \times 0.15 \times 1 = 18$.

The purchasing of 130 pieces also consists of two parts: one is the expected profit for selling 120 pieces, the other is the expected profit higher than that of the 120. The above formula has got the profit for 120 pieces. As for the sales amount that is more than 120, for every step strength of 10, the probability of selling them is 85%. Then, the expected profit for purchasing 130 pieces is:
 $E[C(130)] = E[C(120)] + 10 \times 0.15 \times 0.85 - 10 \times 0.3 \times 0.15 = 18 + (1.275 - 0.45) = 18 + 0.825 = 18.825$

In the same way, we can also get the expected profit for other purchasing amount.
 For 140 pieces:

$$E[C(140)] = E[C(130)] + 10 \times 0.15 \times 0.65 - 10 \times 0.3 \times 0.35 = 18.825 + (0.975 - 1.05) = 18.825 - 0.075 = 18.75$$

For 150 pieces:

$$E[C(150)] = E[C(140)] + 10 \times 0.15 \times 0.35 - 10 \times 0.3 \times 0.65 = 18.75 + (0.525 - 1.95) = 18.75 - 1.425 = 17.325$$

For 160 pieces:

$$E[C(160)] = E[C(150)] + 10 \times 0.15 \times 0.1 - 10 \times 0.3 \times 0.9 = 17.325 + (0.15 - 2.7) = 14.775$$

For the dynamic iterative algorithm, the optimal purchasing is also 130 pieces. In fact, we can stop by 140 pieces because increasing every 10 pieces, selling profit is smaller than the backlog of loss, which means the expected profit will fall. With the increase of the pieces, selling profit is smaller and smaller than the backlog of loss, meaning that the expected profit will also be small. That is the advantage of the dynamic iterative algorithm, in which the calculation stops once the sales profit is smaller than the backlog of loss. Compared to the traditional algorithm, the dynamic iterative algorithm reduces some steps with the same results and is able to deduce the optimal purchasing quantity.

The result is shown in Table 5 and Figure 2.

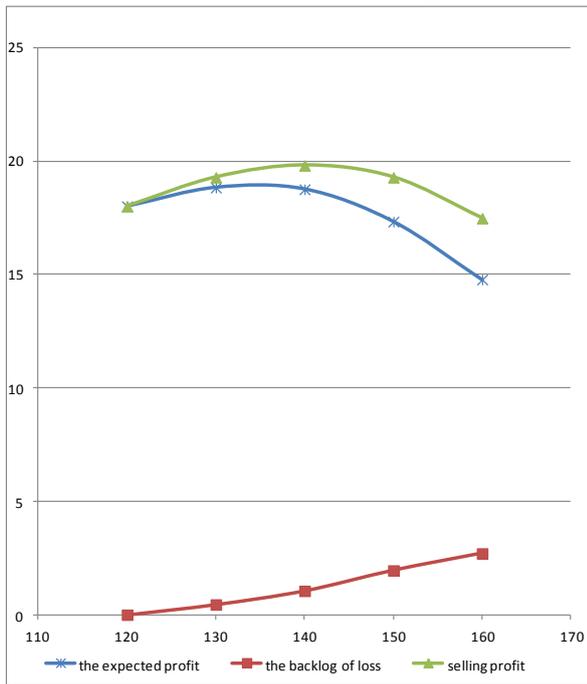


FIGURE 2 Iterative algorithm for the newsvendor problem

General deduction: Suppose the minimum sales amount is a , the maximum sales amount is b , $n \in [a, b]$ and the distributed probability $P(n)$ is known. Selling one piece of newspaper, the newspaper boy can get k Yuan, and loses h Yuan for every cumulative piece. d refers to the step length between No. x and No. $x + 1$ or No. $x - a$. It doesn't matter whether d is the average or not, or to say, it doesn't matter whether d is the well-distributed step length or not $\sum_{n=a}^b P(n) = 1$ the maximum expected profit is:

$$E(C(n)) = \begin{cases} n \times k, & n = a \\ E(C(n-1)) + dk \sum_{n=a}^n P(n) - dh \left(1 - \sum_{n=a}^n P(n) \right), & n > a \end{cases} \quad (1)$$

Equation (1) is the traditional iterative algorithm for the maximum expected profit.

From the Equation (1), we can see $n \times k$ is above 0, which means the minimum sales profit. Generally speaking, $d \times k \times \sum_{n=a}^n P(n) - d \times h \left(1 - \sum_{n=a}^n P(n) \right)$ increases from the second and begins to fall later until it is below 0. Thus, the maximum expected profit is found.

When $d \times k \times \sum_{n=a}^n P(n) - d \times h \left(1 - \sum_{n=a}^n P(n) \right) < 0$, the expected profit will fall. So, when there is the first

$d \times k \times \sum_{n=a}^n P(n) - d \times h \left(1 - \sum_{n=a}^n P(n) \right) < 0$, its previous value will be the maximum expected profit. That is to say, when

$d \times k \times \sum_{n=a}^n P(n) - d \times h \left(1 - \sum_{n=a}^n P(n) \right) < 0$ is the last value above 0, it is the maximum profit.

The maximum profit n will get if we calculate:

$$d \times k \times \sum_{n=a}^n P(n) - d \times h \left(1 - \sum_{n=a}^n P(n) \right) < 0$$

and

$$d \times k \times \sum_{n=a}^n P(n) - d \times h \left(1 - \sum_{n=a}^n P(n) \right) > 0.$$

When there is the last:

$$d \times k \times \sum_{n=a}^n P(n) - d \times h \left(1 - \sum_{n=a}^n P(n) \right) \geq 0,$$

the above formula will become:

$$k \times \sum_{n=a}^n P(n) - h \left(1 - \sum_{n=a}^n P(n) \right) \geq 0$$

or

$$k \times \sum_{n=a}^n P(n) \geq h \left(1 - \sum_{n=a}^n P(n) \right) \geq 0$$

as d is the step length and $d > 0$.

Obviously, $k, h \sum_{n=a}^n P(n), 1 - \sum_{n=a}^n P(n)$ are above 0, thus

there is:

$$\frac{\sum_{n=a}^n P(n)}{1 - \sum_{n=a}^n P(n)} \geq \frac{h}{k} \quad (2)$$

When there is:

$$d \times k \times \sum_{n=a}^{n+1} P(n+1) - d \times h \left(1 - \sum_{n=a}^{n+1} P(n+1) \right) \leq 0,$$

for the first time, the above formula will be:

as d is the step length and $d > 0$. Obviously,

$k, h \sum_{n=a}^n P(n), 1 - \sum_{n=a}^n P(n)$ are above 0, thus there is:

$$\frac{\sum_{n=a}^{n+1} P(n)}{1 - \sum_{n=a}^{n+1} P(n+1)} \leq \frac{h}{k} \tag{3}$$

From Equations (2) and (3), we can get the relationship between the optimal purchasing quantity n and $\frac{h}{k}$:

$$\frac{\sum_{n=a}^{n+1} P(n)}{1 - \sum_{n=a}^{n+1} P(n+1)} \leq \frac{h}{k} \leq \frac{\sum_{n=a}^n P(n)}{1 - \sum_{n=a}^n P(n)} \tag{4}$$

From Equation (4), we can deduce the traditional optimal purchasing quantity for the newsvendor problem:

$$\sum_{n=a}^{n+1} P(n+1) \leq \frac{k}{h+k} \leq \sum_{n=a}^n P(n) \tag{5}$$

Substituting Equation (5) into the previous equation, we can get:

$$\sum_{n=a}^{n+1} P(n+1) \leq \frac{0.15}{0.45} \leq \sum_{n=a}^n P(n),$$

$$\sum_{n=a}^{n+1} P(n+1) \leq 0.33 \leq \sum_{n=a}^n P(n).$$

Then when the cumulative frequency is 0.33, the left side is the maximum profit, the right side is the second maximum profit with the corresponding frequency of 0.15 and 0.35 respectively. The cumulative frequency for the maximum profit is 0.35, indicating that 130 pieces is the optimal purchasing quantity.

3 Newsvendor problem with inventory loss

Many goods such as fresh agricultural products, food and perishable goods are experiencing losses either in the stock or in the process of sales. Thus, loss needs to be taken into consideration for the optimal purchasing quantity or the stock quantity. The formula is deduced as follows:

Suppose selling every piece of good can make k Yuan, and loses h Yuan if it is stocked, the cost is c Yuan, the loss rate of the stock or in the process of sales is d . the probability P of the everyday sales amount r is known.

Then what is the purchasing amount for the maximum expected profit or the minimum loss?

Suppose the everyday sales amount is r , the probability $P(r)$ is already known - $\sum_{r=0}^{\infty} P(r) = 1$. The purchasing quantity is Q .

When the supply is bigger than the demand ($r \leq Q$), there will be the loss for stocked newspapers and the expected loss is:

$$\sum_{r=0}^Q h(Q-r)P(r).$$

When the supply is smaller than the demand ($r > Q$), there will be the loss for out-of-stock newspapers and the expected loss is:

$$\sum_{r=Q+1}^{\infty} k(r-Q)P(r).$$

Thus, when the purchasing amount is Q , the expected loss is:

$$\sum_{r=0}^Q h(Q-r)P(r)$$

$$C(Q) = \sum_{r=0}^Q h(Q-r)P(r) + \sum_{r=Q+1}^{\infty} k(r-Q)P(r) + Q \times d \times c$$

We will then calculate the minimum Q for $C(Q)$. $C(Q)$ should meet the demand of the followings:

Condition (1): $C(Q) \leq C(Q+1)$;

Condition (2): $C(Q) \leq C(Q-1)$.

From condition (1), we can deduce:

TABLE 6 Demand and probability for a single season

Demand	100	200	300	400	500	600	700
Probability	0.17	0.2	0.25	0.12	0.1	0.08	0.08
Cumulative rate	1	0.83	0.63	0.38	0.26	0.16	0.08

$$Q \times d \times c + \sum_{r=0}^Q h(Q-r)P(r) + \sum_{r=Q+1}^{\infty} k(r-Q)P(r) \leq$$

$$(Q+1) \times d \times c + \sum_{r=0}^{Q+1} h(Q+1-r)P(r) + \sum_{r=Q+2}^{\infty} k(r-Q-1)P(r) -$$

$$d \times c + h \left\{ \sum_{r=0}^Q (Q-r)P(r) - \sum_{r=0}^{Q+1} k(Q-1-r)P(r) \right\} +$$

$$k \left\{ \sum_{r=Q+1}^{\infty} (r-Q)P(r) - \sum_{r=Q+2}^{\infty} (r-Q-1)P(r) \right\} \leq 0$$

and get:

$$-d \times c - h \sum_{r=0}^Q P(r) + k \sum_{r=Q+1}^{\infty} P(r) \leq 0$$

there is:

$$-d \times c - h \sum_{r=0}^Q P(r) + k \sum_{r=Q+1}^{\infty} P(r) + k \sum_{r=0}^Q P(r) \leq k \sum_{r=0}^Q P(r) - d \times c + k \leq (h+k) \sum_{r=0}^Q P(r).$$

We can deduce:

$$\sum_{n=0}^Q P(r) \geq \frac{k}{h+k} - \frac{d \times c}{h+k} \tag{6}$$

From condition (2), we can also deduce:

$$\sum_{n=0}^{Q+1} P(r) \leq \frac{k}{h+k} - \frac{d \times c}{h+k} \tag{7}$$

Here we get the optimal purchasing quantity for the inventory loss d :

$$\sum_{n=0}^{Q+1} P(r) \leq \frac{k}{h+k} - \frac{d \times c}{h+k} \leq \sum_{r=0}^Q P(r) \tag{8}$$

From the above formula, the optimal purchasing quantity with inventory loss is larger than that without inventory loss. The increase amount depends on the sum of value and the loss rate and the ratio of the backlog of loss to out-of-stock loss. The bigger the ratio, the more the optimal purchasing amount will add. But at the same time, the expected profit will substantially decrease because the added purchasing amount is offset by the loss. Thus, it is important to reduce the loss rate.

4 Applications

The following example is derived from Reference [5]. Let's calculate it by dynamic iterative algorithm.

Part A sold in one store is 18 Yuan per each with the cost of 10 Yuan. The storage cost of every part for a single quarter is 4 Yuan. The probability distribution of demand for the first quarter (3 months) is shown in the following Table 6. We are calculating the optimal purchasing quantity (Stocked parts are not depreciated and can be sold at the original price).

$$E(100) = 100 \times 8 - 100 \times 2 = 600,$$

$$E(200) = (100) + 100 \times 8 \times 0.83 - 100 \times 0.17 \times 4 - 100 \times 0.83 \times 2 = 600 + 430 = 1030,$$

$$E(300) = (200) + 100 \times 8 \times 0.63 - 100 \times 0.37 \times 4 - 100 \times 0.63 \times 2 = 1030 + 230 = 1260,$$

$$E(400) = (300) + 100 \times 8 \times 0.38 - 100 \times 0.62 \times 4 - 100 \times 0.38 \times 2 = 1260 - 20 = 1240.$$

At this time, there is no need to continue the calculation because the profit will fall. When the purchasing quantity is 300 pieces, the expected optimal profit is 1260 Yuan.

Substitute Equation (5) into the previous Equation, we can get:

$$\sum_{n=a}^{n+1} P(n+1) \leq \frac{6}{6+4} \leq \sum_{n=a}^n P(n).$$

Notice: k is the profit, h is the backlog of loss (here refers to as storage cost). The profit for every piece is $k = 18 - 10 - 2 = 6$, and the storage cost h is 4 Yuan (it will be sold next month). Thus there is:

$$\sum_{n=a}^{n+1} P(n+1) \leq 0.6 \leq \sum_{n=a}^n P(n).$$

The optimal purchasing quantity is 300 pieces. The result is equal to the result for using Equation (1).

If the rate of loss is 0.1, substitute it into Equation (8):

$$\sum_{n=0}^{Q+1} P(r) \leq \frac{6}{6+4} - \frac{0.1 \times 10}{6+4} \leq \sum_{r=0}^Q P(r),$$

$$\sum_{n=0}^{Q+1} P(r) \leq 0.5 \leq \sum_{r=0}^Q P(r)$$

The optimal purchasing quantity is 400 pieces.

5 Conclusions

The dynamic iterative algorithm takes the advantage of the previous calculation and reduces some steps. Once the sales income of added sales is less than the backlog of loss, the calculation can be stopped and the optimal value is found. That is, when the marginal income in below 0, the corresponding purchasing quantity will be the optimal purchasing quantity. Here the marginal income refers to the total of sales income and the backlog of loss brought by added purchasing quantity. The step length d is a random number. It is well adaptive.

Acknowledgments

This work was supported by the Scientific Research Fund of Sichuan Provincial Education Department (Grant No.13204684), Key Scientific Research Foundation of Xihua University (Grant No. Z1120412) and Academic Cultivation Project of Key Laboratory of Fluid and Power Machinery Engineering, Xihua University (Grant No. SBZDPY-11-10).

References

- [1] Xuemin, Mou L, Xu Q 2008 Research Progress and Summary of newsboy model, *Statistics and Decision* **26**(9) 11-4 (in Chinese)
- [2] Li H, Ding X 2006 Solving Newsboy Problem by using Computer Simulation based down-hill Algorithm *Microcomputer Information* **22** 146-7 (in Chinese)
- [3] Li H 2008 Newsboy Problem Solved by Computer Analogue Method *Journal of Anhui Radio&TV University* **3** 126-8 (in Chinese)
- [4] Qian S 2002 Operational Research Beijing: Tsinghua University Press 369-70 (in Chinese)
- [5] Rui S, Ceng Z An Analysis of the Issue of the Per-Duration Need for a Discrete-Time Stochastic Storage Problem, *Operations Research and Management Science* **17**(1) 79-83 (in Chinese)
- [6] Lau H-S, Lau A H-L 1995 The multi-product multi-constraint newsboy problem: Applications, formulation and solution *Journal of Operations Management* **13** 153-62 (in Chinese)
- [7] Pearson M A 2000 The Incorporation of Target Performance Measures and Constrained Optimisation in the Newsboy Problem. *The Journal of the Operational Research Society* **51**(6) 744-54
- [8] Dana J D., Petruzzi N C 2001 Note: The Newsvendor Model with Endogenous Demand *Management Science* **47**(11) 1488-97
- [9] Dekker R, Frenk J B G, Kleijn M J, Kok A G 2000 On the newsboy model with a cutoff transaction size *IIE Transactions* **32** 461-9
- [10] Shore H 2004 General Solution for the Newsboy Model with Random Order Size and Possibly a Cutoff Transaction Size *Journal of the Operational Research Society* **55**(11) 1218-28
- [11] Su X, Lin Z, Yang L 2004 A newsboy model with budget cost constraints *Journal of Jilin University (Science Edition)* **42**(3) 371-4 (in Chinese)
- [12] Li M, Wang K, Fang F 2003 The Time-Based Newsboy Problem: Application, Formulation and Solution. *Systems Engineering – Theory Methodology Applications* **12**(2) 146-52
- [13] Cai Q, Lu Q, Zhu D The Newsboy Problem Model with Changing Forecast Accurate *Forecast Ing* **22**(5) 42-6 (in Chinese)
- [14] Chen L, Wu J, Zhang Y 2000 Tutorial model and method of operational research Beijing: Tsinghua University Press (in Chinese)
- [15] Zhou Y, Ying R, Chen X, Wang Z 2013 Two – product newsboy problem based on prospect theory *Journal of Management Sciences in China* **16**(11) 17-29 (in Chinese)
- [16] Liu Y, Zhang J, Wang L 2013 Optimal joint pricing and ordering decisions in newsvendor model with two demand cases *Control and Decision* **28**(9) 1419-22 (in Chinese)

Authors	
	<p>Hao ran Shi, born in November, 1973, Dazhou country, Sichuan Province, P.R. China</p> <p>Current position, grades: Master degree at Xihua University. University studies: logistics engineering. Professional interests: operational research.</p>
	<p>Xiao qing Liu, born in October, 1980, Anshun country, Guizhou province, P.R. China</p> <p>Current position, grades: Ph.D at Xihua University. University studies: environmental engineering. Professional interests: Environmental impact assessment.</p>
	<p>Yao Yang, born in May, 1968, Huzhou country, Zhejiang province, P.R. China</p> <p>Current position, grades: M.E at Xihua University. University studies: hydraulic and hydro-power engineering. Professional interests: hydropower station pressure pipe test.</p>

A market segmentation model of enterprise marketing based on an improved grey correlation analysis methods

Mei Yang*

Chongqing College of Electronic Engineering, Chongqing, China

Received 12 June 2014, www.tsi.lv

Abstract

In order to make out quick and effective enterprise sales strategy, this paper studies market segmentation for key brand products and proposes a market segmentation model based on an improved grey correlation analysis methods. Key indicators about market segmentation are analysed first. These indicators are then subject to standardization to be in the unified measurement. On this basis, classical grey correlation analysis methods get improved and Hamming distance is adopted to cope with fuzzy information. Grey correlation coefficient and grey correlation degree of market segmentation are acquired. AHP is introduced to assign weight to indicators of market segmentation in order to get weighed comprehensive grey correlation degree. Finally, test is given to the effectiveness, scientific nature and feasibility of the model and algorithm through case study of a brand product.

Keywords: market segmentation, marketing strategy, grey correlation analysis, correlation degree, model

1 Introduction

Before launching a new product to the market, enterprises usually need to be familiar with the market segmentation and plan for it. Market segmentation serves to lowering the marketing cost, yielding more profits and increasing the competitiveness, which is significant to the operation of enterprises [1-4].

However, factors that influence marketing effect and promotion prospect should be taken into consideration when doing market segmentation, including market factor, economic factor, development factor, competitive factor, etc. Besides, market position, promotion distribution and promotion areas should also have a place in the analysis. Thus, market segmentation is a complicated systematic analytical process. It involves with multiple indicators for evaluation to get effective and clear market segmentation.

Currently, there are many ways for the multi-attribute system decision analysis [5-8], such as Analytical Hierarchy Process, Delphi method, fuzzy comprehensive evaluation method, expert score method, etc. These methods are adapted to different cases with practical results. However, subjectivity, uncertain information, complicated calculating process, little reliability stand in the way of scientific and effective decision analysis. Thus, this paper proposes a market segmentation model based on an improved grey correlation analysis method.

2 Evaluation index system of market segmentation

2.1 THE SELECTION OF INDICATORS

Marketing factors are multi-attribute, uncertain and complicated. Market information differs from each other. Thus, evaluation indicators are selected in a scientific way from multiple perspectives to ensure that market segmentation is accurate, effective and objective. Based on previous researches, this paper makes out the following rules for the selection of indicators.

Objectivity: The characteristics of the enterprise and real market operation should be taken into consideration when selecting indicators. Subjectivity should be avoided.

Scientific feature: Indicators should be selected from a scientific perspective. These indicators should have scientific inner and extension meanings to be objective and reasonable.

Dynamic feature: Marketing is dynamic. As market information changes, the marketing environment changes along with it. Therefore, indicators should reflect the dynamic information of market to ensure the reliability of market segmentation.

Integral feature: Market segmentation is a complicated system decision analysis process. So, indicators should be correlated within the system. An overlap or lack should be avoided.

Based on these principles, this paper selects market share, marketing profit rate, sales growth rate, customer satisfaction and marketing investment cost as evaluation indicators.

* *Corresponding author* e-mail: meizi11106@163.com

2.2 WEIGHT OF EVALUATION INDICATORS OF MARKET SEGMENTATION

Evaluation indicators contribute differently to market segment. Thus, they should be given different weight. Expert score method, fuzzy comprehensive analysis method, grey correlation method, entropy weight method and AHP are existing ways to address weight distribution. To avoid subjectivity and ensure the reasonability and reliability of weight distribution, this paper uses AHP to give weight to indicators of market segmentation. Through expert opinions and surveys, here comes the relative importance matrix A between every two indicators:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}_{n \times n} \quad (1)$$

The maximum characteristics value λ A is obtained. Calculate the coincidence indicator CI :

$$CI = \frac{\lambda A_{\max} - n}{n - 1} \quad (2)$$

After referring to relevant table, the average random coincidence indicator RI can be obtained. Calculate the coincidence ratio CR of evaluation indicator:

$$CR = \frac{CI}{RI} \quad (3)$$

If evaluation indicator fits the threshold of coincidence, the weight of the evaluation indicator is:

$$w_i = \frac{\sum_{j=1}^n a_{ij}}{\sum_{i=1}^n \sum_{k=1}^n a_{ik}} \quad (4)$$

3 Market segmentation model based on grey correlation decision analysis model

3.1 STANDARDIZATION OF EVALUATION INDICATORS OF MARKET SEGMENTATION

Market profit rate, market share, sales growth rate and customer satisfaction are indicators of profit type while marketing investment cost is indicator of cost type. In these indicators, accurate data are available through survey and statistical analysis. For some indicators, there is only fuzzy statistical information. So, to unify the evaluation standard, indicators should be subject to standardization to eradicate difference of scale and type.

When the indicator is of profit type, the value of segment market U_i about evaluation indicator j is $v_{ij} = [v_{ij}^a, v_{ij}^b]$. The standardized value \tilde{v}_{ij} is:

$$\tilde{v}_{ij} = \left[\frac{v_{ij}^a - \min_{1 \leq i \leq m} v_{ij}}{\max_{1 \leq i \leq m} v_{ij} - \min_{1 \leq i \leq m} v_{ij}}, \frac{v_{ij}^b - \min_{1 \leq i \leq m} v_{ij}}{\max_{1 \leq i \leq m} v_{ij} - \min_{1 \leq i \leq m} v_{ij}} \right] \quad (5)$$

In particular, when $v_{ij}^a = v_{ij}^b$, i.e. when v_{ij} is an accurate value, there is:

$$\frac{v_{ij}^a - \min_{1 \leq i \leq m} v_{ij}}{\max_{1 \leq i \leq m} v_{ij} - \min_{1 \leq i \leq m} v_{ij}} = \frac{v_{ij}^b - \min_{1 \leq i \leq m} v_{ij}}{\max_{1 \leq i \leq m} v_{ij} - \min_{1 \leq i \leq m} v_{ij}}$$

When the indicator is of cost type, the value of segment market U_i about evaluation indicator j is $v_{ij} = [v_{ij}^a, v_{ij}^b]$. The standardized value \tilde{v}_{ij} is:

$$\tilde{v}_{ij} = \left[\frac{\min_{1 \leq i \leq m} v_{ij} - v_{ij}^a}{\max_{1 \leq i \leq m} v_{ij} - \min_{1 \leq i \leq m} v_{ij}}, \frac{\min_{1 \leq i \leq m} v_{ij} - v_{ij}^b}{\max_{1 \leq i \leq m} v_{ij} - \min_{1 \leq i \leq m} v_{ij}} \right] \quad (6)$$

In particular, when $v_{ij}^a = v_{ij}^b$, that is, when v_{ij} is an accurate value, there is:

$$\frac{\min_{1 \leq i \leq m} v_{ij} - v_{ij}^a}{\max_{1 \leq i \leq m} v_{ij} - \min_{1 \leq i \leq m} v_{ij}} = \frac{\min_{1 \leq i \leq m} v_{ij} - v_{ij}^b}{\max_{1 \leq i \leq m} v_{ij} - \min_{1 \leq i \leq m} v_{ij}}$$

3.2 GREY CORRELATION DECISION ANALYSIS OF MARKET SEGMENTATION

With uncertain data, only when potential market information and statistical data are referred to in the decision analysis of market segmentation scheme, can the process is objective and reasonable. However, for fuzzy information with range, traditional grey correlation decision analysis has certain limitations. Thus, this paper integrates calculation equation of fuzzy information distance into grey correlation decision analysis.

Suppose $V_1 = [v_1^a, v_1^b]$ and $V_2 = [v_2^a, v_2^b]$ are fuzzy interval number with uncertain information. In particular, $v_1^a \leq v_1^b, v_2^a \leq v_2^b$, there is:

$$d_{V_1, V_2}^T = \frac{\left[|v_1^a - v_2^a|^T + |v_1^b - v_2^b|^T \right]^{\frac{1}{T}}}{\sqrt[T]{2}} \quad (7)$$

It refers to distance between V_1 and V_2 .

In particular, when $T = 1$, $d_{V_1, V_2}^T = d_{V_1, V_2}^1$ is the Hamming distance between V_1 and V_2 . There is:

$$d_{V_1, V_2}^1 = \frac{\left[|v_1^a - v_2^a|^T + |v_1^b - v_2^b|^T \right]^{\frac{1}{T}}}{\sqrt[T]{2}} \quad (8)$$

When $T = 2$, $d_{V_1, V_2}^T = d_{V_1, V_2}^2$ is the Euclidean distance between V_1 and V_2 . There is:

$$d_{V_1, V_2}^T = \sqrt{\frac{|v_1^a - v_2^a|^2 + |v_1^b - v_2^b|^2}{2}} \tag{9}$$

Suppose the optimal value of market segmentation about evaluation indicator j is $u_j = [u_j^a, u_j^b]$. The grey correlation coefficient between segmented market U_i about indicator j and the optimal value $u_j = [u_j^a, u_j^b]$:

$$\delta_{ij} = \frac{\min_i \min_j |d_{v_{ij}, u_j}^T| + \rho \max_i \max_j |d_{v_{ij}, u_j}^T|}{|d_{v_{ij}, u_j}^T| + \beta \max_i \max_j |d_{v_{ij}, u_j}^T|} \tag{10}$$

where ρ refers to discrimination coefficient of grey correlation analysis model, $\rho \in 0, 1$, usually $\rho = 0.5$.

Given that the indicators have different weight, the weighed grey correlation degree σ_i between segmented market U_i and the optimal value $u_j = [u_j^a, u_j^b]$ is:

$$\sigma_i = \sum_{j=1}^n w_j * \delta_{ij} \tag{11}$$

After the consultation with management team and other experts, we set up the threshold σ_0 of market segmentation. When there is:

$$\exists \forall \sigma_i \geq \sigma_0 \tag{12}$$

It means that the segmented market U_i has a feasibility and prospect for investment.

For s segmented market that fits the threshold σ_0 , if there is:

$$U_{\max} = \max U_1, U_2, \dots, U_s = U_k \tag{13}$$

It means that the segmented market U_i has the most investment value and brightest prospect. Thus, it can be a key focus.

3.3 MARKET SEGMENTATION MODEL AND ALGORITHM BASED ON GREY CORRELATION DECISION ANALYSIS METHOD

The algorithm can be described as follows:

Step 1: According to selection principles of indicators in Chapter 1.1, the ideas from management team and relevant experts are collection for the selection of indicators.

Step 2: Based on Equations (1)-(4) in subsection 1.2, distribute the weight to these indicators.

Step 3: Standardize these indicators according to Equations (5) and (6).

Step 4: Analyse the distance between different evaluation indicators and the optimal value of indicators according to Equations (7)-(9).

Step 5: Acquire the weighed grey correlation degree between different indicators and the optimal value of indicators according to Equations (10) and (11).

Step 6: Acquire the hierarchy of segmented market that fits the threshold according to Equations (12) and (13) that pave the way for the development of key market.

4 Case study and test

Test is given to the effectiveness, scientific nature and feasibility of the model and algorithm through case study of a brand product. Market share, marketing profit rate, sales growth rate, customer satisfaction and marketing investment cost as evaluation indicators. The relative importance matrix A is obtained through AHP method:

$$A = \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{3} & 5 & 5 \\ 3 & 1 & 3 & 7 & 9 \\ 3 & \frac{1}{3} & 1 & 5 & 9 \\ \frac{1}{5} & \frac{1}{7} & \frac{1}{5} & 1 & 1 \\ \frac{1}{5} & \frac{1}{9} & \frac{1}{9} & 1 & 1 \end{pmatrix}_{5 \times 5}$$

In the relative importance matrix, $CI = 0.075 \leq 1$ all fitting the coincidence. Then the weighed of evaluation indicator is $W = 0.169, 0.460, 0.284, 0.047, 0.040$.

According to surveys of marketing performance and analysis on market information, market is divided into high-end market, middle-end market and low-end market. The evaluation result of each market is shown in Table 1.

TABLE 1 Evaluation result of segmented market of a home appliance product

	High-end market U_1	Middle-end market U_2	Low-end market U_3
Market share	11.50	22.30	15.60
Marketing profit rate	15.28-15.78	17.65-18.25	7.32-7.84
Sales growth rate	8.40-8.60	6.05-6.35	10.06-10.28
Customer satisfaction	0.90	0.95	0.90
Marketing investment cost	1260.00	980.00	840.00

The evaluation result after standardization is shown in Table 2.

TABLE 2 Evaluation result of segmented market after standardization

	High-end market U_1	Middle-end market U_2	Low-end market U_3
Market share	0.000	1.000	0.380
Marketing profit rate	0.728-0.774	0.945-1.000	0.000-0.048
Sales growth rate	0.556-0.603	0.000-0.071	0.948-1.000
Customer satisfaction	0.000	1.000	0.000
Marketing investment cost	0.000	0.667	1.000

The optimal value sequence of different evaluation indicators is:

$$u = 1.000, 0.945, 1.000, 0.948, 1.000, 1.000, 1.000.$$

The distance matrix B between different segmented market about indicator and the optimal value is:

$$B = \begin{pmatrix} 1.000 & 0.000 & 0.620 \\ 0.222 & 0.000 & 0.949 \\ 0.395 & 0.938 & 0.000 \\ 1.000 & 0.000 & 1.000 \\ 1.000 & 0.333 & 0.000 \end{pmatrix}_{5 \times 3}.$$

The grey correlation coefficient matrix C between different segmented market about indicator and the optimal value is:

$$C = \begin{pmatrix} 0.333 & 1.000 & 0.446 \\ 0.693 & 1.000 & 0.345 \\ 0.558 & 0.348 & 1.000 \\ 0.333 & 1.000 & 0.300 \\ 0.333 & 0.600 & 0.000 \end{pmatrix}_{5 \times 3}.$$

Therefore, the weighted grey correlation sequence is calculated as $\sigma = 0.563, 0.799, 0.574$. After consultation with management team and experts, we set the threshold as $\sigma_0 = 0.60$. In segmented market for this

new type of home appliance, middle-end market is the best choice, followed by low-end market. The development value of high-end market is the lowest. Thus, middle-end market can be a target to shot. In this market, this enterprise can lower the sales cost, yield profit and increase competitiveness.

5 Conclusions

This paper studies market segmentation for key brand products and proposes a market segmentation model based on an improved grey correlation analysis methods. In this model, selection principles of evaluation indicators of market segmentation are first proposed. Weighed distribution is adopted based on AHP method to make sure the indicators are objective, scientific and reliable. These indicators are then subject to standardization to be in the unified measurement. In the process of grey correlation analysis, fuzzy distance calculation is introduced to deal with fuzzy and uncertain information. Weighed comprehensive grey correlation degree is then available, making the segmentation more reliable. Finally, test is given to the effectiveness, scientific nature and feasibility of the model and algorithm through case study. The model and algorithm provides scientific evidence for the enterprise to do market segmentation.

References

- [1] Xu J, Ren Yu-long 2003 Application of Extension Method in Choosing Target Market of Electricity Marketing *Journal of Chongqing University* **26**(11) 142-145+154 (in Chinese)
- [2] Wang J, Wang N 2010 Telecom Market Segmentation Research based on extension evaluation theory *Journal of Beijing University of Post and Telecommunication (Social science edition)* **12**(6) 60-5 (in Chinese)
- [3] Ran J, Shao P 2008 To study on customers segmenting strategy of personal mobile communication based on customer value evaluation *Value Engineering* (10) 56-9
- [4] Liao Y 2008 Study on market segmentation based on customers' satisfaction trap *Statistics and Information Forum* **23**(5) 5-10
- [5] Chen T, Feng X, Liang B, Sun W 2013 Grey Fuzzy Evaluation of CBM Development and Area Selection *Journal of Northeastern University(Natural Science)* **34**(3) 429-33 (in Chinese)
- [6] Ma D, Liu X, Wang S 2010 Comprehensive evaluation of high-speed railway track based on analytic hierarchy process *Subgrade Engineering* **151**(6) 6-8
- [7] Wan Shu-ping 2009 New method for interval multi-attribute group decision making with incomplete attribute weights *Computer Integrated Manufacturing Systems* **15**(4) 104-9
- [8] Xiang Y, Zou P, Ma L 2013 Multi-attribute Group Decision-making Based on Rapid Self-adjusting Delphi Method with Biased Information Sampling *Journal of Beijing University of Technology* **39**(8) 1172-8 (in Chinese)
- [9] Wu J, Wu X, Chen Y 2010 Validation of simulation models based on improved grey relational analysis *Systems Engineering and Electronics* **32**(8) 1677-9
- [10] Wang T, Yang A, Bu L 2013 Mechanism scheme design based on multi-attribute extension grey relevant optimized decision-making model *Systems Engineering – Theory & Practice* **33**(9) 2321-9
- [11] Liu S, Cai H, Yang Y 2013 Advance in grey incidence analysis modelling *Systems Engineering – Theory & Practice* **33**(8) 2041-6
- [12] Liu S F, Dang Y G, Fang Z G 2010 Grey system theory and application *Beijing: Science Press (in Chinese)*
- [13] Shi H X, Liu S F, Fang Z G, Zhang A 2008 The model of grey periodic incidence and their rehabilitation *Chinese Journal of Management Science* **16**(3) 131-6 (in Chinese)

Author



Mei Yang, born in October, 1982, Chongqing, China

Current position, grades: lecturer at Chongqing College of Electronic Engineering.

University studies: Master Degree in Probability and Mathematical Statistics at Chongqing University in 2010.

Scientific interest: mathematical modelling.

Study on enterprise extension marketing model based on extension engineering methods

Tian Jia*

Business Administration Department, Henan Polytechnic, Zhengzhou 450046, P.R. China

Received 1 July 2014, www.tsi.lv

Abstract

This paper studies the marketing strategy of enterprises and proposes an enterprise extension marketing model based on extension engineering methods under market economy. The matter-element model for enterprise marketing is established on extension theory, and the matter-element features, values and discourse domain are studied to produce more marketing strategies. These strategies are subject to extension priority-degree evaluation for convergence analysis. Compatibility analysis is also carried out. Empirical studies prove the model to be feasible and effective.

Keywords: extension marketing, extension engineering, marketing strategy, market, model

1 Introduction

In the era of knowledge economy and with the rapid development of information technology and computer science, it is significant to make out marketing strategy and launch into the market so as to increase competitiveness, gain market share and grow sustainably [1-3]. Therefore, many researchers, both home and abroad, have studied this issue from different perspectives with various methods, achieving remarkable results [4-7].

However, in the process of planning marketing strategy and its implementation, many factors need to be taken into consideration, such as market factor, enterprise factor, human factor, etc. The influence of these factors can be undermined if their coordination and conflict can be addressed. Extenics is a useful tool to turn to as it is an intelligent design subject that copes with conflict and incompatibility.

Currently, some researchers have integrated extenics into the analysis of marketing strategy. But this is far from enough. Based on previous researches, this paper proposes an enterprise extension marketing model based on extension engineering methods, which shall be a reference and guidance for enterprise marketing practice.

2 Extension model of enterprise marketing strategy

2.1 MATTER-ELEMENT MODEL OF ENTERPRISE MARKETING STRATEGY

Extenics is an intelligent design subject proposed by Chinese scholar Professor Cai Wen. It conducts an extension analysis for design tasks and copes with conflicts by a combination of qualitative and quantitative

method. In this way, extension engineering methods are given birth. Matter-element as the basic unit to describe design task, uses a sequence set with three elements $R_{market} = O_{market}, C, V$ to describe the enterprise marketing strategy that has one-dimension. O_{market} refers to the name of the strategy, C refers to characteristics of the strategy and V refers to value of the characteristics.

In real terms, marketing strategy is restricted by many factors. Therefore, there is a necessity to take these conflict and incompatible factors together with marketing evaluation indicators as characteristics of matter-element. Then, the matter-element of marketing strategy with n characteristics is expressed by:

$$R_{market} = O_{market}, C, V = \begin{bmatrix} O_{market} & c_1 & v_1 \\ & c_2 & v_2 \\ & \vdots & \vdots \\ & c_n & v_n \end{bmatrix}. \quad (1)$$

In the expression, $c_i, 1 \leq i \leq n$ refers to the broad sense of characteristics of marketing strategy O_{market} . $v_i, 1 \leq i \leq n$ refers to the value of characteristics c_i .

2.2 EXTENSION ANALYSIS OF ENTERPRISE MARKETING STRATEGY

The planning and extension of marketing strategy are in fact an extension analysis of the matter-element model. Extension transformation enriches such analysis. Based on the constructed matter-element model, discourse domain, characteristics of marketing strategy and values of the characteristics are three approaches to do the extension analysis in order to get more valuable strategies

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

and models. The extension model can be achieved in the following ways.

1) Transform the value of the characteristics in matter-element model R to achieve the extension transformation of marketing demand and the value of marketing capability indicators, there is:

$$\begin{aligned} R_{TV} &= R^*, k_V, k_V^* \mid R^* \in R \rightarrow R_{iTV}, \\ k_V &= k \vee R, \\ k_V^* &= k^* \vee R \rightarrow R_{iTV}, \end{aligned} \tag{2}$$

where R_{TV} refers to extension transformation of the value of the characteristics in matter-element model R . k_V and k_V^* refers to correlation function before and after extension transformation respectively. $k_V^* = k^* \vee R \rightarrow R_{iTV} \geq 0$ refers to effective extension transformation of the value of the characteristics in model R .

2) Transform the characteristics in matter-element model R to achieve the extension transformation of factors influencing marketing and marketing capability indicators, there is:

$$\begin{aligned} R_{TC} &= R^*, k_C, k_C^* \mid R^* \in R \rightarrow R_{iTC}, \\ k_C &= k \wedge R, \\ k_C^* &= k^* \wedge R \rightarrow R_{iTC}, \end{aligned} \tag{3}$$

where R_{TC} refers to extension transformation of the characteristic in matter-element model R . k_C and k_C^* refers to correlation function before and after extension transformation respectively. In particular, $k_C^* = k^* \wedge R \rightarrow R_{iTC} \geq 0$ refers to effective extension transformation of the characteristics in model R .

3) Transform the discourse domain in matter-element model R to achieve the extension transformation of factors influencing marketing and marketing force indicators, there is:

$$\begin{aligned} R_{TO} &= R^*, k_O, k_O^* \mid R^* \in R \rightarrow R_{iTO}, \\ k_O &= k \circ R, \\ k_O^* &= k^* \circ R \rightarrow R_{iTO}, \end{aligned} \tag{4}$$

where R_{TO} refers to extension transformation of the discourse domain in matter-element model R . k_O and k_O^* refers to correlation function before and after extension transformation respectively. In particular, $k_O^* = k^* \circ R \rightarrow R_{iTO} \geq 0$ refers to effective

extension transformation of the discourse domain in model R .

2.3 CONVERGENCE ANALYSIS OF ENTERPRISE MARKETING STRATEGY

Based on the extension analysis, many marketing strategies that meet the enterprise's demand are available at hand. Thus, convergence analysis is necessary for the study. Suppose there are m marketing strategies by extension analysis. Here can acquire q convergence indicators after the consultation with design experts. The correlation function k_i^j of marketing strategy i about convergence indicator j can be expressed in the following forms.

If the strategy target is expressed by range U_{0j} , the extension correlation function k_i^j is:

$$k_i^j = \rho(v_i^j, U_{0j}) / |U_{0j}|, \tag{5}$$

where $\rho(v_i^j, U_{0j})$ refers to the extension distance of marketing strategy i about convergence indicator j . The detailed calculation is referred to Equation (10) and (11).

If the strategy target is expressed by range set U_{0j} and U_j , the extension correlation function k_i^j is:

$$k_i^j = \rho(v_i^j, U_{0j}) / (\rho(v_i^j, U_j) - \rho(v_i^j, U_{0j})). \tag{6}$$

Extension correlation function needs to be normalized to get $k_i^{\otimes j}$ so that the indicators are unified:

$$k_i^{\otimes j} = \begin{cases} k_i^j / \max_{v_i^j \in U_{0j}} k_i^j, & k_i^j > 0 \\ 0, & k_i^j = 0. \\ k_i^j / \max_{v_i^j \in U_{0j}} |k_i^j|, & k_i^j < 0 \end{cases} \tag{7}$$

Given different weight w_j of convergence indicators, the extension priority-degree ψ_i of marketing strategy i is:

$$\psi_i = w_1, w_2, \dots, w_q * \begin{bmatrix} k_i^{\otimes 1} \\ k_i^{\otimes 2} \\ \vdots \\ k_i^{\otimes q} \end{bmatrix} = \sum_{j=1}^q w_j * k_i^{\otimes j} \tag{8}$$

Analyse all extension priority-degree ψ_i . When ψ_i meets the requirement of the given threshold, it means that this marketing strategy is feasible and effective.

3 Compatibility analysis of enterprise marketing strategy

Enterprise marketing strategy subject to convergence analysis may not accord with the anticipated marketing target as some indicators can conflict with each other and be incompatible. Thus, compatibility analysis is needed so as to be closer to the anticipated marketing target for the purpose of increasing marketing ability and competitiveness.

The general model of compatibility analysis is expressed as:

$$P = G \otimes L, \tag{9}$$

where G refers to the strategy target and L refers to factors and conditions that restrict the implementation of the strategy.

In extension theory, there are three ways to conduct the compatibility analysis. First, with the marketing target unchanged, transform the restriction conditions to solve the problem; second, when the restriction conditions fail to change, adjust the target to solve the problem; third, adjust both the target and the restriction conditions. Compatibility function is introduced to measure the effectiveness of these three ways.

Suppose the value of the marketing target of matter-element model is V_i , the restricted value is $V_0 = [v_0^{\min}, v_0^{\max}]$. If V_i is the accurate value, then the extension distance between the strategy target value and the restricted value V_0 is:

$$\rho V_i, V_0 = \left| V_i - \frac{v_0^{\max} + v_0^{\min}}{2} \right| - \frac{v_0^{\max} - v_0^{\min}}{2}. \tag{10}$$

If V_i is fuzzy value with uncertain information and $V_i = [v_i^{\min}, v_i^{\max}]$, the extension distance between the strategy target value and the restricted value V_0 is:

$$\rho V_i, V_0 = \frac{\rho v_i^{\min}, V_0 + \rho v_i^{\max}, V_0}{2} = \frac{\left| v_i^{\min} - \frac{v_0^{\max} + v_0^{\min}}{2} \right| + \left| v_i^{\max} - \frac{v_0^{\max} + v_0^{\min}}{2} \right| - v_0^{\max} - v_0^{\min}}{2}. \tag{11}$$

If the marketing strategy target and factors and restriction conditions that influence the implementation of the strategy are not compatible and of single characteristics, which means only one condition fails to meet the strategy target, then the compatibility function $K P = G \otimes L$ is expressed as:

$$K P = G \otimes L = -\rho V_i, V_0. \tag{12}$$

If the marketing strategy target and factors and restriction conditions that influence the implementation

of the strategy are not compatible and of multiple characteristics, which means many conditions fail to meet the strategy target, and when every characteristics fits the compatibility, the compatibility function $K P = G \otimes L$ is expressed as:

$$K P = G \otimes L = \min_{1 \leq i \leq m} -\rho V_i, V_0 \mid \min_{1 \leq i \leq m} -\rho V_i, V_0 \wedge \left(\bigwedge_{i=1}^m -\rho V_i, V_0 \geq 0 \right) \tag{13}$$

If the marketing strategy target and factors and restriction conditions that influence the implementation of the strategy are not compatible and of multiple characteristics, which means many conditions fail to meet the strategy target, and when one characteristics fits the compatibility, the compatibility function $K P = G \otimes L$ is expressed as:

$$K P = G \otimes L = \max_{1 \leq i \leq m} -\rho V_i, V_0 \mid \max_{1 \leq i \leq m} -\rho V_i, V_0 \wedge \left(\bigvee_{i=1}^m -\rho V_i, V_0 \geq 0 \right) \tag{14}$$

According to the compatibility analysis, when $K P = G \otimes L > 0$, the marketing strategy target and factors and restriction conditions that influence the implementation of the strategy are compatible. When $K P = G \otimes L < 0$, the marketing strategy target and factors and restriction conditions that influence the implementation of the strategy are incompatible. When $K P = G \otimes L = 0$, they are in between.

A marketing strategy that fits the compatibility is labelled as feasible and effective. One that doesn't fit the compatibility is infectious in one way or another and thus needs improvement to better serve the market.

4 Case study and analysis

The marketing strategy of a new product of a home appliance enterprise is subject to case study that further illustrates the extension marketing model based on extension engineering methods. This product is a high-end product. Market survey and marketing ideas of the management team are taken into consideration. Based on Equations (1)-(3), launch area, market target and launch layer are studied. Three extension marketing plans are shown in Table 1.

TABLE 1 Extension marketing plan of new products of a home appliance enterprise

Marketing Plan	Profit Margin	Market Share	Sales Growth	Production Quota
A	13.5	15.0	8.5	68.5
B	16.8	12.0	7.2	65.0
C	15.0	14.0	6.0	66.5

Based on expert views and the marketing ability of the enterprise, the anticipated marketing profit rate is 15-20 with the market share registering 13-18. The sales growth rate stands at 7-10 and the production quota is 60-70. According to the convergence analysis expression, the extension distance (Table 2) and the extension correlation function (Table 3) of different marketing strategy plans are acquired.

TABLE 2 Extension distance of marketing plan

Marketing Plan	Extension Distance			
	Profit Margin	Market Share	Sales Growth	Production Quota
A	1.50	-2.50	-1.50	-1.50
B	-1.80	0.50	-0.20	-5.00
C	0.00	-1.50	1.00	-3.50

TABLE 3 Extension correlation function of marketing plan

Marketing Plan	Extension Correlation Function			
	Profit Margin	Market Share	Sales Growth	Production Quota
A	1.000	-1.000	-1.000	-0.300
B	-1.000	1.000	-0.133	-1.000
C	0.000	0.600	1.000	-0.700

Given the weight of matter-element characteristics $W = 0.20, 0.30, 0.30, 0.20$, the priority-degree sequence is $\psi = -0.460, -0.139, 0.340$. Suppose the threshold of priority-degree is $\psi_0 = 0$, from the sequence it is clear that plan A and B are feasible. This company values on marketing development and promotion of new products. So the market share and sales growth rate are conditions

References

- [1] Liu Y 2009 Power Market Segmentation Priority-degree Evaluation and Differentiated Marketing *Management World* **2009**(6) 85-9
- [2] Hou K, Chen Y, Gong L 2012 Marketing management system based on customer credit evaluation model *Computer Engineering and Design* **33**(10) 3979-83
- [3] Rashad Y, Tshelo N, Uarba M M 2011 Project management and project integration management in relationship with service marketing *International Conference on Management and Service Science* Wuhan China *IEEE Computer Society Press* 2011 1-4
- [4] Wang H, Liu Y 2010 Target Market Choice for Residential Marketing by Priority-degree Evaluating *Journal of Xian Science and Technology University (Social Science)* **20**(5) 18-22
- [5] Song C, Zhao L 2011 Construction and Analysis of Marketing Evaluation Model *Commercial Times* **2011**(24) 31-2
- [6] Cai J 2010 A Study on Marketing of Household Appliances *Tianjin University* Doctoral dissertation 2010 1-11
- [7] Jiang Y, Shang J, Liu Y 2010 Maximizing customer satisfaction through an online recommendation system: A novel associative classification model *Decision Support Systems* **48**(3) 470-9
- [8] Wang T, Yang A, Zhong S 2014 Products Extension Adaptive Design Based on Case Reuse *International Journal of Control and Automation* **7**(1) 295-306
- [9] Zhao Y W, Zhang G X 2012 A New Integrated Design Method Based On Fuzzy Matter-Element Optimization *Journal of Materials Processing Technology* **129**(1-3) 612-8
- [10] Wang T, Yang A, Bu L 2013 Mechanism scheme design based on multi-attribute extension gray relevant optimized decision-making model *Systems Engineering – Theory & Practice* **33**(9) 2321-9.
- [11] Cai W, Yang C 2013 Fundamental Theory and Method of Extenics *Chinese Science Bulletin* **58**(13) 1190-9.
- [12] Yang C, Cai W 2007 Extension Engineering *Beijing: Science Press* 162-9

for compatibility analysis. Profit rate and production quota are secondary factors.

From the compatibility analysis model, the compatibility function of plan A is:

$$K_{A-P=G\otimes L} = 1.000 | \min 1.000, 1.000 \wedge -\rho V_2, V_0 \geq 0 \wedge -\rho V_3, V_0 \geq 0 > 0.$$

This strategy does fit the anticipated target. The compatibility function of plan B is:

$$K_{B-P=G\otimes L} = -1.000 | \min -1.000, 0.139 \wedge -\rho V_2, V_0 \geq 0 \wedge -\rho V_3, V_0 < 0 < 0.$$

It doesn't fit the anticipated target and the matter-element characteristics needs to be subject to extension transformation so better serve the marketing demand.

5 Conclusions

This paper studies the marketing strategy of enterprises and proposes an enterprise extension marketing model based on extension engineering methods. These strategies are subject to convergence analysis and compatibility analysis. It provides guidance to make out a marketing strategy in a smart and quick way. From empirical studies, it is clear that this extension marketing model can lead the convergence analysis and compatibility analysis so as to cope with conflicts and conduct effective evaluation. It serves to scientific management and planning of enterprises and helps them to increase competitiveness and grow sustainably.

Author



Tian Jia, born in August, 1978, Zhengzhou, Henan, China

Current position, grades: Lecturer since 2010.

University studies: Master's Degree in Business Administration from Zhongnan University of Economics and Law in 2009.

Scientific interest: Business administration, entrepreneurship education of college students.

Publications: 3 papers.

Experience: teacher at Henan Polytechnic since July 2004.

An improved Grey prediction model and the application in college sports information management system

Ziya Wang^{1*}, Ran Li²

¹Department of Physical Education, Anhui Jianzhu University, Hefei, Anhui, China

²School of Management, Anhui Jianzhu University, Hefei, Anhui, China

Received 1 March 2014, www.tsi.lv

Abstract

The sports elective course is an important means for college students to exercise their physical education and quality education. Because the students can choose the elective courses voluntarily online, teachers and colleges cannot learn the number of the students who choose the sports elective courses before the deadline. In this paper, we propose an improved GM (1, 1) prediction model to forecast the number of the students who take part in the sports elective courses and apply this model in the college sports information management system. Firstly, this paper puts forward the Grey prediction model with time parameter. Then, this paper studies the constructing mechanism and the modelling characteristics of this model. At last, we use the improved GM (1, 1) model to forecast the number of students who select the college sports elective courses in college sports information management system. High precision of the fitting and forecasting are obtained in the experiment while the result verifies the validity of the model.

Keywords: Grey precision model, college sports information management system, sports elective course

1 Introduction

In recent years, the development of the quality education is fast and powerful in China. In order to enrich the extracurricular life of college students, many universities set up the sports elective course. The main aim of the sports elective course is to improve the physical quality. In addition, the main means of physical education curriculum is the physical exercise. Sports elective course exercises the students comprehensively. At the same time, sports elective course can enhance the physical fitness of students, promotes the students' health and improves the physical quality of college students through the rational exercise. The sports elective course is one of the important ways to implement the quality education and cultivate an all-round development of the talents.

With the development of the computer technology, the college information management system becomes more and more popular. In the meantime, students select the course by college information management system online. Because the college students can choose the elective courses voluntarily and they can also deselect the courses online before the deadline, teachers could not ensure the number of students who take part in the sports elective course. It makes a lot of trouble for the institutes and teachers arrange the curriculum planning. To solve this problem, we propose an improved GM (1,1) and apply in the college sports information management system to predict the number of the students who participate in the sports elective course.

At present, there are little literatures to forecast the number of students who participant in the sports elective course. In this paper, we use the improved Grey forecasting model to predict the number of students who select the sports elective course. The Grey prediction theory is an important part in the Grey system theory which is established by the scholars of China. It becomes a significant research branch in the prediction theory. According to the accumulation of the sequence, the Grey prediction theory excavates the inherent regularity of the data sequence to reveal the future development trend. The Grey GM(1,1) model is the core model of the Grey prediction. It has a higher precision for fitting and forecasting the combined series which have the Grey exponential rule. Therefore, GM(1,1) model has an extensive research background.

Grey model is an important branch of Grey system theory [1-3], since it was pioneered by professor Deng in 1982. In order to enhance the model precision, scholars have been researching new modelling technology in practice. The results demonstrate that the selection of background value, the manner of accumulated generating and the mode of original data are the main factors which influence the precision of Grey model. Thus, some methods for improving model precision are presented correspondingly. Li and Dai [4] improved the model predictive precision by modifying the initial value. Tseng et al. [5] and Wang et al. [6] proposed a hybrid grey model to forecast series with seasonality, applying the ratio-to-moving-average method in order to calculate the seasonal indexes and remove the seasonal factor. For

* Corresponding author e-mail: 76926762@qq.com

oscillatory sequence, Qian and Dang [7] used accelerating translation transformation and improved smoothness by weighted mean generating and modelling finally. Zeng and Xiao [8] improved original data smoothness with the power function. Dang et al. [9, 10] constructed a new class of weakening buffer operator and strengthening buffer operator by using the buffer operator axiom for shock disturbed data sequence. Papers [11, 12] also researched accumulated generating. Rao et al. [13, 14] established hopping model and phase model and studied the matrix form for solving model parameters. Peng [15] established generalized unequal interval Grey forecasting model, and presented the matrix form for solving model parameters. For background value-building, Tan [16] improved the background value for equal interval sequence and unequal interval sequence; Literature [17] thought that the actual systems are usually fractional order. Literatures [18] studied the stability of fractional order system and its controllability and objectivity. The Grey forecasting model has been proposed as a promising alternative to time-series forecasting [19, 20].

On the basis of the existing researches, this paper put forward a Grey prediction model with time parameter. According to the Grey prediction theory, we structure a new Grey prediction model which is suitable for time sequence. Then, we study the constructing mechanism and the model characters. At final, we apply this model to forecasting the number of college students who select the college sports elective courses. When teachers' entry the college sports information management system, they can select the number of the students who participant in the sports elective courses each year as the samples. As the results, they can get the predicting outcomes. The numerical experiment shows that this method has high accuracy. The structure of this paper is as follows. The first part is introduction. The second part is the generation of the operator. The third part is the $GM(1,1)$ model. The fourth part is the construction of the Grey $GM(1,1,t^\alpha)$ model with time power form. The fifth part is the characteristic of the $GM(1,1,t^\alpha)$ model. The sixth part is the computer simulation and the last part is the conclusion.

2 The generation of the operator

2.1 ACCUMULATED GENERATING OPERATOR AND REGRESSIVE GENERATING OPERATOR

We suppose that the original sequence is:

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\},$$

where, n is the dimensionality, $X^{(0)}$ is the Grey sequence. $n \geq 1, x^{(0)}(i) \geq 0$

We define the new sequence

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}. \tag{1}$$

where $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$, $X^{(1)}$ is said to be the one order accumulated generating operation series of $X^{(0)}$, namely $X^{(1)} = AGOX^{(0)}$.

If we accumulate the Equation (1) for r orders, we can get:

$$x^{(r)}(k) = \sum_{i=1}^k x^{(r-1)}(i). \tag{2}$$

If $x^{(1)} = x^{(1)}(k) - x^{(1)}(k-1)$, $X^{(1)}$ is said to be the one order regressive generating operation series of $X^{(0)}$, namely $X^{(1)} = IAGOX^{(0)}$

If we accumulate the Equation (1) for r orders, we can get $x^{(r-1)} = x^{(r)}(k) - x^{(r)}(k-1)$.

2.2 INVERSE ACCUMULATED GENERATING OPERATOR

We suppose that $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ is n -dimensions nonnegative descending sequence:

$$x^{(0)}(i) \geq x^{(0)}(i+1), i = 1, 2, \dots, n-1. \tag{3}$$

We define the new sequence:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\},$$

where $x^{(1)}(k) = \sum_{i=n}^k x^{(0)}(i)$.

It is the reverse accumulation generation sequence:

$$x^{(0)} = x^{(1)}(k) - x^{(1)}(k+1). \tag{4}$$

We can learn that $x^{(1)}$ is a monotonous drop sequence. When $x^{(0)}$ is a nonnegative monotonic function, $x^{(1)}$ have a tendency of monotonic decrease after reverse accumulation.

2.3 INDEX ACCUMULATED GENERATING OPERATOR

We suppose that the original sequence is:

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}, \tag{5}$$

where, n is the dimensionality, $X^{(0)}$ is the Grey sequence. $n \geq 1, x^{(0)}(i) \geq 0$.

We define $x^{(0)}(k) = be^{a(k-1)}, k = 1, 2, \dots, n$. We call this type as homogeneous discrete exponential function and we define:

$$x^{(0)}(k) = Be^{a(k-1)} + C, k = 1, 2, \dots, n, a, b, B, C \in R. \tag{6}$$

We can know that:

$$-a = \ln\left(1 - \frac{b}{B}\right) \tag{7}$$

and:

$$B = \frac{be^a}{e^a - 1}, C = \frac{b}{1 - e^a}. \tag{8}$$

If $x^{(1)}(1) = x^{(0)}(1) = b$ and:

$$\frac{b}{1 - e^a} = \frac{u}{a}. \tag{9}$$

We can get:

$$x^{(1)}(k) = \left(x^{(0)}(1) - \frac{u}{a}\right)e^{a(k-1)} + \frac{u}{a}, \tag{10}$$

$$x^{(1)}(t) = \left(x^{(0)}(1) - \frac{u}{a}\right)e^{a(t-t_0)} + \frac{u}{a}. \tag{11}$$

3 GM(1,1) model

We introduce the most classic Grey model $GM(1,1)$. The first 1 means that there is one variable in this model and the other one means that we use first order differential equation.

The mean ideal of the $GM(1,1)$ is as follows. We set

$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ is the original sequence and get:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}, \tag{12}$$

where $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$. We call the a first order linear ordinary differential equations is the albino differential equation of $GM(1,1)$:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b. \tag{13}$$

The differential form is:

$$x^{(0)}(k) + az^{(1)}(k) = b, \tag{14}$$

where, a is called the development coefficient which represents the development state of the prediction value. b is called the Grey action quantity and it represents change contained in the data. a, b is the first order parameter bag of $GM(1,1)$ model: $[a, b]^T = (B^T B)^{-1} B^T Y_n$, where:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix},$$

$$Y_n = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T. \tag{15}$$

Background value is:

$$z^{(1)}(k+1) = \frac{1}{2}[x^{(1)}(k) + x^{(1)}(k+1)], k = 1, 2, \dots, n-1.$$

The discrete solution of $GM(1,1)$ is:

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a}) \cdot e^{-ak} + \frac{b}{a}. \tag{16}$$

The reducing value is:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^{-a})(x^{(0)}(1) - \frac{b}{a}) \cdot e^{-ak}, k = 1, 2, \dots, n. \tag{17}$$

We suppose that $\bar{x}^{(r)}(k)$ is the fitted value of the $x^{(r)}(k)$ and we define $q^{(0)}(k) = x^{(0)}(k) - \bar{x}^{(0)}(k)$ as the residual of $x^{(0)}$ at time k .

$$S_1^2 = \frac{1}{n} \sum_{i=1}^n (x^{(0)}(i) - \bar{u}_1)^2$$

$$S_2^2 = \frac{1}{n} \sum_{j=1}^n (\bar{x}^{(0)}(i) - \bar{u}_2)^2. \tag{18}$$

\bar{u}_1 is the mean value of the original sequence and \bar{u}_2 is the mean value of the predicted sequence. The ratio is $C = \frac{S_2}{S_1}$ and the probability is:

$$P = P\{|q^{(0)}(k) - \bar{q}|\} < 0.6945S_1\},$$

where \bar{q} is the mean value of the residual error sequence. We define the prediction accuracy of good, qualified, just the mark and unqualified according to P and C . The value of P and C are shown as Table.1

TABLE 1 Residual inspection standard of $GM(1,1)$

The level of prediction accuracy	P	C
good	>0.95	<0.35
qualified	>0.8	<0.5
just the mark	>0.7	<0.65
unqualified	≤0.7	≥0.65

4 The construction of the Grey GM(1,1,t^α) model with time power form

Definition 1: We assume $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$. We call $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ as the first-order accumulated generating sequence (1-AGO) of $X^{(0)}$, where $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$, we call $Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$ as the proximate mean generation sequence of $X^{(1)}$:

$$Z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k-1)), k = 2, 3, \dots, n \quad (19)$$

Definition 2: Assuming $X^{(0)}, X^{(1)}, Z^{(1)}$ as the **Definition 1**, we call $x^{(0)}(k) + az^{(1)}(k) = bk^\alpha + c$ as the basic form of GM(1,1,t^α), where α is a nonnegative constant. We call $\frac{dx^{(1)}}{dt} + ax^{(1)} = bt^\alpha + c$ as the winterization equation of GM(1,1,t^α).

Theorem 1: Assuming $X^{(0)}$ a non-negative sequence:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

where $x^{(0)}(k) \geq 0, k = 1, 2, \dots, n$. $X^{(1)}$ is 1-AGO of $X^{(0)}$. $Z^{(1)}$ is the proximate mean generation sequence of $X^{(1)}$. If $\hat{\gamma} = [a, b, c]^T$ is the parameter list and:

$$Y = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix}, B = \begin{pmatrix} -z^{(1)}(2) & 2^\alpha & 1 \\ -z^{(1)}(3) & 3^\alpha & 1 \\ \vdots & \vdots & 1 \\ -z^{(1)}(n) & n^\alpha & 1 \end{pmatrix} \quad (20)$$

The least square parameter estimation of $x^{(0)}(k) + az^{(1)}(k) = bk^\alpha + c$ about GM(1,1,t^α) model satisfies $\hat{\gamma} = (B^T B)^{-1} B^T Y$.

Proof: taking data to the GM(1,1,t^α) model $x^{(0)}(k) + az^{(1)}(k) = bk^\alpha + c$, we can get:

$$\begin{aligned} x^{(0)}(2) + az^{(1)}(2) &= 2^\alpha b + c, \\ x^{(0)}(3) + az^{(1)}(3) &= 3^\alpha b + c, \\ &\vdots \\ x^{(0)}(n) + az^{(1)}(n) &= n^\alpha b + c \end{aligned}$$

$Y = B\hat{a}$ is a set of estimates for a, b and c . Substituting $-az^{(1)}(k) + bk^\alpha + c$ with $x^{(0)}(k), k = 2, 3, \dots, n$, we can get the error sequence $\varepsilon = Y - B\hat{a}$. Assuming:

$$\begin{aligned} s &= \varepsilon \cdot \varepsilon^T = Y - B\hat{a}^T(Y - B\hat{a}) = \\ &\sum_{k=2}^n (x^{(0)}(k) + az^{(1)}(k) - bk^\alpha - c)^2 \end{aligned}$$

The a, b, c which make the s minimum is as follows:

$$\frac{\partial s}{\partial a} = 2 \sum_{k=2}^n (x^{(0)}(k) + az^{(1)}(k) - bk^\alpha - c)z^{(1)}(k) = 0,$$

$$\frac{\partial s}{\partial b} = -2 \sum_{k=2}^n (x^{(0)}(k) + az^{(1)}(k) - bk^\alpha - c)k^\alpha = 0,$$

$$\frac{\partial s}{\partial c} = -2 \sum_{k=2}^n (x^{(0)}(k) + az^{(1)}(k) - bk^\alpha - c) = 0.$$

From the above equations, we can get a, b and c . By $Y = B\hat{a}$, we get:

$$B^T B \hat{\gamma} = B^T Y, \hat{\gamma} = (B^T B)^{-1} B^T Y \quad (21)$$

From above, we can obtain:

$$\hat{\gamma} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = (B^T B)^{-1} B^T Y \quad (22)$$

Theorem 2: We assume $\hat{\gamma} = [a, b, c]^T = (B^T B)^{-1} B^T Y$ for B and Y which referred in **Theorem 1**. Therefore, the time response function of the winterization equation

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = bt^\alpha + c$$

is as follows:

$$x^{(1)}(t) = be^{-at} \int e^{at} t^\alpha dt + \frac{c}{a}$$

In the model GM(1,1,t^α), the time response sequence of $x^{(0)}(k) + az^{(1)}(k) = bk^\alpha + c$ can be get discretely by the time response function of the winterization equation.

5 The characteristic of the GM(1,1,t^α) model

Theorem 3: When $\alpha = 0$, the model GM(1,1,t^α) changes to $x^{(0)}(k) + az^{(1)}(k) = bk^0 + c = b_0$. That is, GM(1,1,t^α) degrades to the GM(1,1) model. If B and Y are referred as the **Theorem 1** $\hat{\gamma} = [a, b, c]^T = (B^T B)^{-1} B^T Y$.

We can get the time response function of the winterization function $\frac{dx^{(1)}}{dt} + ax^{(1)} = b_0$ is as follows:

$$x^{(1)}(t) = \left(x^{(1)}(1) - \frac{b_0}{a} \right) e^{-a(t-1)} + \frac{b_0}{a}$$

1) The time response sequence of the GM(1,1) model $x^{(0)}(k) + az^{(1)}(k) = b_0$ is as follows:

$$\hat{x}^{(1)}(k+1) = \left(x^{(1)}(1) - \frac{b_0}{a} \right) e^{-ak} + \frac{b_0}{a}, k = 0, 1, 2, \dots, n$$

2) The reduction value:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^{-a}) \left(x^{(0)}(1) - \frac{b_0}{a} \right) e^{-ak}, k = 1, 2, \dots, n.$$

Property 1: From the **Theorem 3**, we can know that the $GM(1,1,t^\alpha)$ model adapts the sequence modelling, which has the approximate non-homogeneous exponential rule $x(t) \approx ce^{at}$ when $\alpha = 0$.

Theorem 4: When $\alpha = 1$, the $GM(1,1,t^\alpha)$ model changes to $x^{(0)}(k) + az^{(1)}(k) = bk + c$. That is, $GM(1,1,t^\alpha)$ changes to $GM(1,1,t)$. If B and Y are referred as the **Theorem 1**, $\hat{\gamma} = [a, b, c]^T = (B^T B)^{-1} B^T Y$.

1) The time response function of the winterization function $\frac{dx^{(1)}}{dt} + ax^{(1)} = bt + c$ is as follows:

$$x^{(1)}(t) = \left(x^{(1)}(1) - \frac{b}{a} - \frac{ac-b}{a^2} \right) e^{-a(t-1)} + \frac{b}{a}t + \frac{ac-b}{a^2}.$$

2) The time response sequence of the $GM(1,1,t)$ model $x^{(0)}(k) + az^{(1)}(k) = bk + c$ is as follows:

$$\hat{x}^{(1)}(k+1) = \left(x^{(1)}(1) - \frac{b}{a} - \frac{ac-b}{a^2} \right) e^{-ak} + \frac{b}{a}k + \frac{ac-b}{a^2}, k = 0, 1, 2, \dots, n.$$

3) The reduction value:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^{-a}) \left(x^{(1)}(1) - \frac{b}{a} - \frac{ac-b}{a^2} \right) e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n.$$

Property 2: From the theorem 4, we can know the $GM(1,1,t)$ model adapts the sequence modelling which has the approximate non-homogeneous exponential rule $x(t) \approx ce^{at} + bd$ when $\alpha = 1$.

Theorem 5: When $\alpha = 2$, the $GM(1,1,t^\alpha)$ model changes to $x^{(0)}(k) + az^{(1)}(k) = bk^2 + c$. That is, $GM(1,1,t^\alpha)$ changes to $GM(1,1,t^2)$. If B and Y are referred as the **Theorem 1**, $\hat{\gamma} = [a, b, c]^T = (B^T B)^{-1} B^T Y$.

1) The time response function of the winterization function $\frac{dx^{(1)}}{dt} + ax^{(1)} = bt^2 + c$ is as follows:

$$x^{(1)}(t) = \left(x^{(1)}(1) - \frac{a^2b + a^2c - 2ab + 2b}{a^3} \right) e^{-a(t-1)} + \frac{b}{a}t^2 - \frac{2b}{a^2}t + \frac{2b + a^2c}{a^3}.$$

2) The time response sequence of the $GM(1,1,t^2)$ model $x^{(0)}(k) + az^{(1)}(k) = bk^2 + c$ is as follows:

$$x^{(1)}(t) = \left(x^{(1)}(1) - \frac{a^2b + a^2c - 2ab + 2b}{a^3} \right) e^{-ak} + \frac{b}{a}k^2 - \frac{2b}{a^2}k + \frac{2b + a^2c}{a^3}, k = 0, 1, 2, \dots, n.$$

3) The reduction value:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^{-a}) \left(x^{(1)}(1) - \frac{a^2b + a^2c - 2ab + 2b}{a^3} \right) e^{-ak} + \frac{2b}{a}k - \frac{(a+2)b}{a^2}, k = 1, 2, \dots, n.$$

Property 3: From the **Theorem 5**, we can know, the $GM(1,1,t^2)$ model adapts the sequence modelling, which has the approximate non-homogeneous exponential rule when $\alpha = 2$, $x(t) \approx ce^{at} + bt + d$. When α gets other value, we can research the time response sequence and the related properties of the $GM(1,1,t^\alpha)$ according to the specific value. In the practical application, when selecting the Grey model $GM(1,1,t^\alpha)$ with time power item, we can eliminate the parameters a, b and c in the model by using the Grey derivative information coverage principle. Then, we can get the expression of the original data series about the parameter α . Therefore, we ensure the value of α and give the optimization steps of α by using the intelligent algorithm.

6 Computer Simulations

When teacher entry the college sports information management system, they choose the numbers of the students who select the sports elective courses for each year as the samples. Then, they fit the samples. If the fitting precision is high, we predict the number of participants. The flow is as follows (Figure 1).

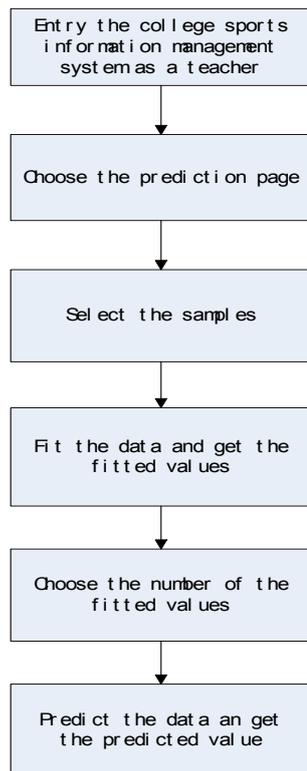


FIGURE 1 The Prediction Process of $GM(1,1,t^\alpha)$ Model

We choose the numbers of the students who take part in the sports elective course for ten years from a college. We fit the data of first 7 years as the samples set and get the fitted values. The other data of next 3 years are compared with the predicted values.

We fit the data and get the fitted values. Then we calculate the error of the two columns of the values. The results are as shown in Table 2.

TABLE 2 The error of the actual values and the fitted values

N7	Year	Actual values	Fitted values	Error
1	2001	413	413	0.00%
2	2002	495	479	-3.23%
3	2003	560	540	3.59%
4	2004	532	538	1.21%
5	2005	581	584	0.516%
6	2006	627	635	1.2%
7	2007	663	661	-0.302%

Form Table 2, we can find that the actual values and the fitted values are almost the same. The fitting effect is

References

[1] Deng J L 2002 Grey Theory Foundation, *Huazhong University of Technology Press* Wuhan China
 [2] Deng J L 2002 Grey Forecasting and Decision *Huazhong University of Technology Press* Wuhan China
 [3] Liu S F, Lin Y 2006 Grey Information Theory and practical Applications *Springer-Verlag* London
 [4] Li J F, Dai W Z 2005 Research on the ameliorating $GM(1,1)$ Model and its application in the power quantity modeling of Shanghai city *Syst Eng Theor Pract* 3 140-5

good. From the 7 samples, we see that the most error is 3.59% and the error becomes smaller as the number of the samples increases. The fitting of the data achieves a good result. So, we can make prediction next.

As the good result of fitting data, we predict the values for next 3 samples by $GM(1,1,t^\alpha)$ model. Then, we calculate the error between the actual values and the predicted values. The results are shown in Table 3.

TABLE 3 The error of the actual values and the fitted values

N3	Year	Actual values	Predicted values	Error
1	2008	742	748	0.808%
2	2009	859	864	0.58%
3	2010	916	927	0.10%

From Table 3 we can see that the actual values and the predicted values are exactly similar. This means that the prediction has obtained the good effect. The $GM(1,1,t^\alpha)$ model applying to the aerobics performance prediction is feasible and effective.

7 Conclusions

The computer technology and the network technology become more and more popular. These technologies are applied to many fields, including the college sports information management system. Colleges introduce this system to manage the information and the students better. Through college sports information management system, it is convenient for students to select or deselect the courses before deadline, especially the sports elective course.

The main objective of the sports elective course is to improve the physical quality while its primary means is the physical exercise. Due to the uncertainty of enrolment, the institutes and the teachers could not arrange the rational curriculum planning. In order to forecast the number of the elective courses, we do below work:

- 1) we propose the Grey prediction model with time parameter;
- 2) we also study the constructing mechanism and the model characters;
- 3) we apply this model in college information management system. The experimental results show that this method has high accuracy. It also has a broad application and a practical background.

[5] Tseng F M, Yu H C, Tzeng G H 2001 Applied hybrid grey model to forecast seasonal time series *Technol Forecast Soc Change* 67(2-3) 91-302
 [6] Wang Q J, Liao X H, Zhou Y H, Zou Z R, Zhu J J, Peng Y 2005 Hybrid grey model to forecast monitoring series with seasonality *Central South Univ Technol* 12(5) 623-7
 [7] Qian W Y, Dang Y G 2009 $GM(1,1)$ model based on oscillation sequences *Syst Eng Theor Pract* 29(3) 149-54

- [8] Zeng X Y, Xiao X P 2009 Study on generalization for $GM(1,1)$ model and its application *Control Decision* **24**(7) 92–105
- [9] Dang Y G, Liu S F, Liu B, Tang X W 2004 Study on the buffer weakening operator *Chin J Manage Sci* **12**(2) 108–11
- [10] Dang Y G, Liu B, Guan Y Q 2005 On the strengthening buffer operators *Control Decision* **20**(12) 1332–36
- [11] Xiao X P, Deng J L 2001 Novel results on AGO space *J Grey Syst* **13**(4) 325–30
- [12] Song Z M, Xiao X P, Deng J L 2002 The character of opposite direction AGO and its class ratio *J Grey Syst* **14**(9) 9–14
- [13] Rao C, Xiao X P, Peng J 2006 A $GM(1,1)$ control model with pure generalized AGO based on matrix analysis *Proceedings of the 6th World Congress on, Intelligent Control and Automation (WCICA06)* **1** 574–7
- [14] Rao C, Xiao X P, Peng J 2006 A new $GM(1,1)$ model for prediction modeling of step series *Proceedings of the First International Conference on Complex Systems and Applications (Published in Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms* 522–6
- [15] Peng K K 2011 Research on Grey modeling and prediction of ultimate bearing capacity of single pile *Wuhan University of Technology* Wuhan China
- [16] Tan G J 2000 The structure method and application of background value in Grey system $GM(1,1)$ model (II) *Syst Eng Theor Pract* **20**(5) 125–8
- [17] Torvik P J, Bagley R L 1984 On the appearance of the fractional derivative in the behavior of real material *J Appl Mech Trans ASMF* **1984** **51**(2) 294–8
- [18] Matignon D 1996 Stability results for fractional differential equations with applications to control processing *Computational Engineering in Systems and Application Multi conference Lille IMACS IEEE-SMC* **2** 963–8
- [19] Deng J L 1989 Control problems of grey systems *Syst Contr Lett* **1**(1) 288–94
- [20] Deng J L 1989 Introduction to grey system theory *Grey Syst* **1**(1) 1–24

Authors



Ziya Wang, born in September, 1979, Anhui, China

Current position, grades: Master degree, lecturer.

University studies: Department of Physical Education of Anhui Jianzhu University.

Scientific interest: Sports training and sports economics management.

Publications: 4.

Experiences: teaching in the Department of Physical Education of Anhui Jianzhu University since 2006.



Ran Li, born in October, 1978, Anhui, China

Current position, grades: studying for doctoral degree, associate professor.

University studies: A Doctor of Business Administration of Hefei University of Technology, 2012.

Scientific interest: environmental economics, supply chain and financial management.

Publications: 5.

Experiences: teaching in School of management of Anhui Jianzhu University since 2005.

An improved method of controlling bullwhip effect and the analysis of the bullwhip effect

Ran Li^{1, 2*}

¹*School of Management, Hefei University of Technology, Hefei, Anhui, China*

²*School of Management, Anhui Jianzhu University, Hefei, Anhui, China*

Received 1 April 2014, www.tsi.lv

Abstract

The bullwhip effect is an important parameter to measure whether the logistics management is good or not. The bullwhip effect affects the production, inventory, transport efficiency in logistics management seriously. In this paper, we establish the structural model of supply chain with multi distribution centre and apply control method to inhibition the bullwhip effect. We analyse the control mechanism of the bullwhip effect and present the control arithmetic to control the bullwhip effect. All the processes are under the circumstance that the demand is worst according to the control theory. At last we processed a stochastic control simulation experiment to control the bullwhip effect. The result shows that the bullwhip effect is inhibited and the bullwhip effect is reduced and stable. The first part of this paper is the related problem description. The second part is basic model and quantitative description of bullwhip effect. The third is control method. The last part is a simulation example.

Keywords: bullwhip effect, multi distribution centre, logistics management

1 Introduction

The bullwhip effect is a phenomenon of demand fluctuations transfer increase in supply chain. This phenomenon means that the orders retailers delivery to manufacture is different to the actually orders in logistics management. This distortion spreads to the upstream in an enlarged form. The consequences of the bullwhip effect are self-evident to the enterprises. So, bullwhip effect is very serious in logistics management. Manufacturers pay excess production cost of raw materials, raw material shortages, manufacturing overtime payment and the high level of inventory due to the poor demand forecasting. These will lead to the extra storage costs, backlog of funds, low efficiency of the transportation process and the extra transportation costs. All of these will cause the enormous economic losses in logistics management.

The first person recognized the bullwhip effect is Forrester. He pointed out that the changes manufacturer perceive is far exceeds the customers perceive through a series of case studies. And he also noticed this effect in the supply chain of each class will be amplified [1]. Many scholars also pointed out the bullwhip effect exist in many industries through the numerical analysis of actual data from the economic angle. SHU Liang you and Yanfeng Ouyang thought that the bullwhip effect can be restrained when the supply chain members share the demand information and the larger the range of sharing demand information is, the bigger the function restraining the bullwhip effect is [2, 3]. V. Gaur and Krane S D believed that the order quantity has a tendency to increase

with moving to the upstream in the supply chain [4,5]. Yanfeng Ouyang researched the stability of bullwhip in system and characterization of the bullwhip effect in linear, time-invariant supply chains. [6, 7]. Senge and Steman observed management behaviour of beer distribution game in a wider view under the same conditions. They found a small sale volatility of retailer can be amplified to make orders or yield change greatly of each member through each link in the supply chain [8, 9]. Towill confirmed that the inventory management has a effect on information distorted in supply chain [10]. Sterman J D first used the (s, S) ordering strategy to prove the existence of bullwhip effect [11]. Sucky, E studied the effect of bullwhip. He thought the effect of bullwhip is overrated [12]. Lee considered the $AR(p)$ model and $ARIMA(0,1,1)$ model. He studied the bullwhip effect caused by fluctuations in prices and supplier out of stock [13]. Xu and Dong considered the retailers and suppliers are in the same $AR(1)$ model and predicted the demand. He got the result that the RMSEP of supplier is high than that of retailer after applying VMI [14]. Chen, Drezner, Ryan and David confirmed the impact of demand forecasting on the bullwhip effect. They not only proved the existence of bullwhip effect theoretically, but also quantized the variability each stage increase in the supply chain [15]. Blackburn, Kahn and Rinks are some of the earliest scholars studied the influence of bullwhip effect on enterprise economic profit. They reviewed the USA manufacturing industry and USA economic history to research the bullwhip effect in the view of system dynamics [16-18]. Metters quantified the bullwhip effect in the supply chain with the heuristic algorithm. He get

* *Corresponding author* e-mail: ranran19780212@126.com

the result: economic profit of enterprises enhance the rate of up to 30%. He also found the promoting effect has a close relationship with scope of business enterprise and the enterprise cost structure [19]. Bottain studied on the influence of technology for radio frequency identification (RFID) and electronic product code (EPC Network) network technology in the operation of Italy FMCG supply chain system. They found that these advanced technologies can improve the supply chain system by visualization and reduce the safety stock level of the enterprise. These technologies weaken the bullwhip effect to a great extent and promote the profits of fast moving consumer goods supply chain through above principles [20]. Giuliano, Fernanda and Comenges studied a single-product serial supply chain. They considered a control parameter can switch the chain from a series of filters to a series of amplifiers in the bullwhip effect and analyse how the optimal values of the parameters change when discontinuities are in an order policy [21]. Chandra and Grabis quantified the bullwhip effect in the case of serially correlated external demand if autoregressive models are applied to obtain multiple steps demand forecasts. They find the MRP can reduce magnitude of the bullwhip effect while providing the inventory performance comparable to that of a traditional order-up approach [22]. Clark and Hammond discussed the relationship between BPR process of the food industry in USA and channel performance. They thought that VMI has achieved more satisfactory success than pure EDI in the food industry through empirical analysis [23]. Anupindi and Bassok applied the contract model to reduce the bullwhip effect [24, 25]. With the development of economy and the intensification of market competition, the research and control of bullwhip effect have become indispensable part of business management in supply chain.

In this paper, we established a supply chain model with multi distribution centres and introduce a demand disturbance. Then, we will improve the quantization method of the bullwhip effect and apply the robustness to control the bullwhip effect. And, we propose a control theory method of bullwhip effect to make the bullwhip effect minimum in supply chain. At last, we simulate the improved method through numerical analysis and validate the bullwhip is reduced and stable in logistics management.

2 The model of supply chain

Now, we begin to discuss the improved supply chain structure model of distribution centre. At first, we make assumptions in our model:

Assumption 1: Demand per period at each retailer location is an independent and equivalently distributed random variable.

Assumption 2: A periodic review procedure is used. In each period, the following sequence of events happens at each stocking location: order, delivery and sale.

Assumption 3: Orders are delivered after a constant lead time.

Assumption 4: Excess demand is backlogged at all levels.

Assumption 5: Fixed plus variable ordering costs are happened at the retailer level and the variable costs are charged at the distribution centre only.

Assumption 6: Holding costs and shortage costs are charged against expected (end of period) inventory levels.

Assumption 7: Pipeline holding costs are paid by the receiving location.

Assumption 8: All the costs are stationary.

In assumption 1, the stationary of demand means that the source of demands serial correlation is only the retailers' ordering decisions. Regular review and backlogging are common practices in many industries. We assume that the ordering cost structure charges fixed replenishment costs to retailers and not to the wholesaler. This condition could occur when wholesalers have adopted effective order filling technology and/or where the manufacturers absorb the cost of filling orders. In the latter case, it could be argued that these costs are passed on to the wholesaler and are reflected in the unit cost (and hence in the holding cost). We have observed these planning when wholesalers have long term contracts with their suppliers.

Next, we consider the nominal system of the supply chain. i.e.:

$$x_{1,k+1} = x_{1,k} + u_{1,k} - d_{1,k}, \tag{1}$$

$$x_{2,k+1} = x_{2,k} - Lu_{1,k} + u_{2,k}. \tag{2}$$

Equation (1) is supply chain upstream part inventory dynamic equation for the market customer layers. $x_{1,k}$ is the order inventory about a customer supply chain upstream portion. It is an n -dimensional column vector. $u_{1,k}$ is customer order quantity. It is also an n -dimensional column vector. d_1 is deterministic demand. It is an n -dimensional vector. Equation (2) is the upstream part inventory dynamic equation for distribution centre. $x_{2,k}$ is order inventory about distribution centre. It is an m dimensional vector. $Lu_{1,k}$ is customer order aggregation amount about distribution centre. Matrix L translates n -dimensional vector aggregation about customer order into m -dimensional demand vector about distribution. Among them, L is a matrix for m row n column. That is:

$$\begin{matrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2n} \\ \dots & \dots & \dots & \dots \\ \lambda_{m1} & \lambda_{m2} & \dots & \lambda_{mn} \end{matrix}, \tag{3}$$

where $\lambda_{ij} \geq 0$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$. $\sum_i \lambda_{ij} = 1$, $j = 1, 2, \dots, n$. Actually, row vector $(\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{in})$ of the

middle aggregation matrix L is a weighted coefficient vector. It dedicates that the i intermediate rally point of the supply chain is the order allocation proportion for n customer.

Equations (1) and (2) can also be written the form of a matrix, that is:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} I & 0 \\ -L & I \end{bmatrix} \begin{bmatrix} u_{1,k} \\ u_{2,k} \end{bmatrix} + \begin{bmatrix} -d_{1,k} \\ 0 \end{bmatrix}. \tag{4}$$

The nominal system about matrix form of supply chain is as following:

$$x_{k+1} = x_k + Bu_k + d_k, \tag{5}$$

where, $B = \begin{bmatrix} I & 0 \\ -L & I \end{bmatrix}, d_k = \begin{bmatrix} -d_{1,k} \\ 0 \end{bmatrix}.$

When the supply chain is subjected of disturbance about the ended uncertain demand, we write the terminal disturbance as Fw_k . When the disturbance transfers to the front supply chain distribution centre and manufacturer (Figure 1), it forms the bullwhip effect:

$$x_{k+1}^f = x_k^f + Bu_k^f + d_k + Fw_k. \tag{6}$$

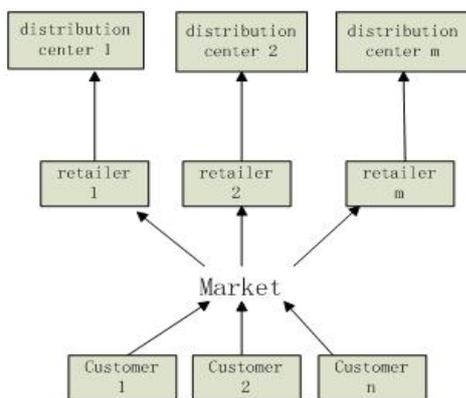


FIGURE 1 Multi distribution centre flow chart

Where $F = diag(F_1, 0)$, F_1 is n -dimensional vector. F is $(n+m)$ dimensional vector. We can describe the ended uncertain demand of the supply chain as following:

$$\xi_1 = d_1 + F_1w, \tag{7}$$

where ξ_1 is n -dimensional demand vector of the supply chain. d_1 is n -dimensional vector of the certain demand. w is n -dimensional vector of uncertain disturbance. Because of the system formula of supply chain (6) is subjected of the ended demand, the uncertain environment will effect inventory variables (state variables), and order variables(control variables).Then, inventory variables and order variables become x_k^f, u_k^f .

Now, we study the deviation between inventory and order in supply chain system further. That is:

$$\dot{x}_k = x_k^f - x_k, \tag{8}$$

$$\dot{u}_k = u_k^f - u_k. \tag{9}$$

The deviation system of multiple distribution centres in supply chain is as following:

$$\dot{x}_{k+1} = \dot{x}_k + B\dot{u}_k + Fw_k. \tag{10}$$

For the quantitative description of the bullwhip effect we mostly adopt the variance form to descriptive the bullwhip effect quantitatively in supply chain. The description has influence on the quantitative analysis of the bullwhip effect but the description is not convenient for a complex structure, such as the multi-distribution centre in supply chain. It is more difficult to study the dynamic control of the bullwhip effect further. In this thesis, we adopt the deviation description of the bullwhip effect. The concept of the bullwhip effect is that the enhancement effect when the ended demand fluctuation is forward in the supply chain. In this thesis, the object is multiple distribution centre model. The bullwhip effect is the enhanced process that the lower demand fluctuation causes the higher demand fluctuation. We adopt to compare the fluctuation of the front inventory and order with the following demand fluctuation in order to describe the bullwhip effect. That is:

$$r_{1,k}^2 = \frac{\dot{x}_{1,k}^T Q_1 \dot{x}_{1,k} + \dot{u}^T T^{1,k} \dot{u}_{1,k}}{w_k^T w_k}, \tag{11}$$

$$r_{2,k}^2 = \frac{\dot{x}_{2,k}^T Q_2 \dot{x}_{2,k} + u_{2,k}^T \dot{u}_{2,k}}{w_k^T w_k}, \tag{12}$$

where Q is positive semi definite matrix $Q = diag(Q_1, Q_2)$, Q_1 and Q_2 are also positive semi definite matrix. $r_{1,k}$ describes the bullwhip effect of the demand fluctuation about customer. $r_{2,k}$ describes the bullwhip effect of the ended demand fluctuation about distribution centre. In this way, the bullwhip effect can be described by more general parameters, such as $r_{1,k}$ and $r_{2,k}$ in supply chain .The bigger the value of $r_{1,k}$ and $r_{2,k}$, the stronger the bullwhip effect. On the contrary, the smaller the value of $r_{2,k}$, the weaker the bullwhip effect.

3 The H_∞ control of the bullwhip effect

3.1 THE CONTROL MECHANISM OF THE BULLWHIP EFFECT

Bullwhip effect is a high risk exists in the marketing, it is a The result of the game about demand forecast revisions, order quantity decision, price fluctuation and so on between vendors and suppliers which increase the supplier's production, supply, inventory management and marketing instability. The reason of the bullwhip effect is that when supply chain information transfer from the final clients to the original suppliers the information distorted and gradually enlarged, the demand information

to appear more and more large fluctuations because the share could not be effectively to realize the information.

For the deviation formula in the supply chain, the parameters $r_{1,k}$ and $r_{2,k}$ and Equation (11) in the bullwhip effect describe the process that the ended demand fluctuation causes the front fluctuation about inventory and order in the supply chain .It is a question how to select a \dot{u}_k to weak the bullwhip effect as possible, especially the ended demand fluctuation, that is the worst disturbance conditions. We Select control \dot{u}_k in order to reduce the bullwhip effect to the lowest degree in supply chain .This is a H_∞ control question by using the words of analysis of system control theory. The essence of the question is that when the disturbance w is big, that is the bullwhip effect, we make J had a smallest value by selecting \dot{u}_k , or:

$$\min_{\dot{u}_k} \max_{w_k} J = \frac{1}{2} \sum_{k=0}^N (\dot{x}_k^T Q \dot{x}_k + \dot{u}_k^T \dot{u}_k - \beta^2 w^T w) , \quad (13)$$

where $Q = Q_x^T Q_x = \text{diag}(Q_1, Q_2) = \text{diag}(Q_{x_1}^T Q_{x_1}, Q_{x_2}^T Q_{x_2})$ is positive semi definite matrix. β is weighted factor which is about the disturbance w .

3.2 THE CONTROL ALGORITHM OF H_∞

Its significance is clear for modern logistics that the supply chain Equations (10) and (13) describes the control problem of H_∞ in the bullwhip effect. That is how to consider the worst situation in the condition of uncertain circumstance. The management strategy of supply chain is to make the inventory status and order control deviation minimize, that is make the bullwhip effect minimize.

$$\dot{x}_k - > 0, \dot{u}_k - > 0 . \quad (14)$$

For the Equations (10) and (13), we can get the following result easily. That is, if and only if

$$I - \beta^2 F^T S_{k+1} F > 0, 0 \leq k \leq N . \quad (15)$$

This question has the only saddle point solution:

$$\dot{u}_k = -B^T S_{k+1} [I + (BB^T - \beta^{-2} FF^T) S_{k+1}]^{-1} \dot{x}_k , \quad (16)$$

$$w_k = \beta^{-2} F^T S_{k+1} [I + (BB^T - \beta^{-2} FF^T) S_{k+1}]^{-1} \dot{x}_k , \quad (17)$$

where S_k is fit for the formula of *Riccati*:

$$S_k = Q + S_{k+1} [I + (BB^T - \beta^{-2} FF^T) S_{k+1}]^{-1} , S_n = 0 . \quad (18)$$

If (I, B) is positive definite and (I, Q_x) is measured, when $k \rightarrow \infty$, then $S_k \rightarrow S > 0$, it is said that there are feedback control \dot{u}_k in the asymptotically stable system. Obviously, Equations (10) and (13) meet the condition that the (I, B) is positive definite and (I, Q_x) is

measured. So the supply chain system has asymptotically stable solution of H_∞ .We only need to solve the solution of Equation (18).

In simulation, for Equation (18), when $k \rightarrow \infty$, after sufficiently many iterations calculation, if $\|S_{k+1} - S_k\| \rightarrow 0$, we can regard that we have got the stable solution of *Riccati*. Then, we can get the stable solution about u .

Therefore, the inventory quantity and order quantity in supply chain system are:

$$x_k^f = \dot{x}_k + x_k , \quad (19)$$

$$u_k^f = \dot{u}_k + u_k , \quad (20)$$

where we can set the inventory status x_k and the quantity control u_k by the plan which is formulated by supply chain management.

4 Numerical analysis

We fist get the change of parameters of bullwhip effect in supply chain through H_∞ control. We assume that there is a large-scale supply chain. In addition, in this supply chain, there are ten customer groups and five distribution centres $n = 10, m = 5$. At the same time, we hypothesize that $F = [I_{10}, 0_5]$ and the average price of the products $p = 10$.

The initial condition of the inventory deviation is:

$$x_1^T = (0.45, 0.11, 0.09, 0.32, 0.17, -0.05, 0.02, -0.09, 0.30, 0.35, 0.28, -0.03, 0, 11, 0.17, 0.26),$$

(Unit of measurement: thousand)

Customer distribution aggregation layer matrix:

$$L = \begin{bmatrix} 0.2 & 0.1 & 0.4 & 0.1 & 0 & 0.3 & 0.2 & 0.1 & 0 & 0.3 \\ 0.2 & 0.4 & 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0.5 & 0.2 & 0.1 \\ 0.3 & 0.1 & 0.2 & 0.1 & 0.4 & 0.1 & 0.2 & 0.1 & 0.4 & 0.2 \\ 0 & 0.1 & 0.1 & 0.4 & 0.3 & 0.1 & 0 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.3 & 0.1 & 0.1 & 0.1 & 0.3 & 0.6 & 0.1 & 0.3 & 0.2 \end{bmatrix}$$

The demand and the inventory in supply chain are:

$$u^T = (1.01, 1.07, 1.19, 0.96, 1.03, 1.12, 0.87, 1.22, 1.18, 1.33, 1.12, 0.91, 1.22, 1.21, 1.13),$$

(Unit of measurement: thousand).

$$x^T = (1.10, 1.09, 1.21, 0.97, 1.08, 1.17, 0.91, 1.32, 1.21, 1.35, 1.14, 0.95, 1.32, 1.28, 1.16),$$

(Unit of measurement: thousand).

k is limited time and we assume that $k = 10$.

When $k \rightarrow \infty$, here is the gain matrix:

$$K = \begin{pmatrix} K_1 & K_2 & K_3 \\ K_4 & K_5 & K_6 \\ K_7 & K_8 & K_9 \end{pmatrix}$$

$$K_1 = \begin{pmatrix} 0.0629 & -0.0716 & 0.0412 & -0.0020 & 0.0502 \\ 0.0811 & -0.0156 & -0.0936 & -0.0108 & -0.0489 \\ -0.0746 & 0.0831 & -0.0446 & 0.0292 & 0.0011 \\ 0.0826 & 0.0584 & -0.0907 & 0.0418 & 0.0398 \\ 0.0264 & 0.0918 & -0.0807 & 0.0509 & 0.0781 \end{pmatrix}$$

$$K_2 = \begin{pmatrix} -0.0300 & -0.0848 & 0.0203 & -0.0786 & -0.0729 \\ -0.0606 & -0.0892 & -0.0474 & 0.0923 & 0.07385 \\ -0.0497 & 0.0061 & 0.0308 & 0.0990 & 0.0159 \\ 0.02320 & 0.0558 & 0.0378 & 0.0549 & 0.0099 \\ -0.0053 & 0.0868 & 0.0492 & 0.0634 & -0.0710 \end{pmatrix}$$

$$K_3 = \begin{pmatrix} -0.0165 & -0.0737 & 0.0094 & -0.0387 & -0.0584 \\ -0.0900 & 0.0884 & -0.0407 & 0.00170 & 0.0397 \\ 0.0805 & 0.0912 & 0.0489 & 0.0021 & -0.0058 \\ -0.0018 & 0.0150 & -0.0622 & 0.0635 & -0.0539 \\ 0.0889 & -0.0880 & 0.0373 & 0.0589 & 0.0688 \end{pmatrix}$$

$$K_4 = \begin{pmatrix} -0.0804 & 0.0311 & 0.0642 & -0.0447 & 0.0918 \\ -0.0443 & -0.0928 & 0.0389 & 0.0359 & 0.0094 \\ 0.0093 & 0.0698 & -0.0368 & 0.0310 & -0.0722 \\ 0.0915 & 0.0867 & 0.0900 & -0.0674 & -0.0701 \\ 0.0927 & 0.0357 & -0.0931 & -0.0762 & -0.0485 \end{pmatrix}$$

$$K_5 = \begin{pmatrix} -0.0296 & -0.0740 & -0.0098 & 0.0737 & 0.0706 \\ 0.0661 & 0.0137 & -0.0832 & -0.0831 & 0.0244 \\ 0.0170 & -0.0061 & -0.0542 & -0.0200 & -0.0298 \\ 0.0099 & -0.0976 & 0.0826 & -0.0480 & 0.0026 \\ 0.0834 & -0.0325 & -0.0695 & 0.0600 & -0.0196 \end{pmatrix}$$

$$K_6 = \begin{pmatrix} -0.0021 & -0.0530 & -0.0633 & 0.0288 & -0.0610 \\ -0.0324 & -0.0293 & -0.0263 & -0.0242 & -0.0548 \\ 0.0800 & 0.0642 & 0.0251 & 0.0623 & -0.0658 \\ -0.0261 & -0.0969 & 0.0560 & 0.0065 & -0.0544 \\ -0.0777 & -0.0914 & -0.0837 & -0.0298 & -0.0128 \end{pmatrix}$$

$$K_7 = \begin{pmatrix} -0.0684 & 0.0515 & -0.0122 & -0.0003 & 0.0681 \\ 0.0941 & 0.0486 & -0.0236 & 0.0919 & 0.0491 \\ 0.0914 & -0.0215 & 0.0531 & -0.0319 & 0.0628 \\ -0.0029 & 0.0310 & 0.0590 & 0.0170 & -0.0513 \\ 0.0600 & -0.0657 & -0.0626 & -0.0552 & 0.0858 \end{pmatrix}$$

$$K_8 = \begin{pmatrix} -0.0428 & -0.0675 & 0.06516 & -0.0137 & -0.0848 \\ 0.05144 & 0.05885 & 0.0076 & 0.0821 & -0.0520 \\ 0.0507 & -0.0377 & 0.0992 & -0.0636 & -0.0753 \\ 0.0239 & 0.0057 & -0.0843 & -0.0472 & -0.0632 \\ 0.0135 & -0.0668 & -0.0114 & -0.0708 & -0.0520 \end{pmatrix}$$

$$K_9 = \begin{pmatrix} 0.0560 & -0.0662 & 0.0858 & 0.0878 & -0.0377 \\ -0.0220 & 0.046 & 0.0551 & 0.0751 & 0.0846 \\ -0.0516 & 0.0298 & -0.0026 & 0.0100 & -0.0139 \\ -0.0192 & 0.0295 & -0.0128 & 0.0244 & -0.0630 \\ -0.0807 & -0.0098 & -0.0106 & 0.0174 & 0.0809 \end{pmatrix}$$

We can get the Figure 2: the change curve of r_1 and r_2 . This picture shows the change of the parameters of the bullwhip effect.

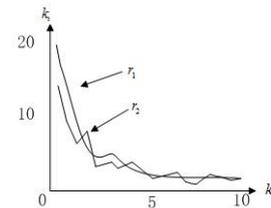


FIGURE 2 Change curve of r_1 and r_2

We can see that the bullwhip effect parameter $r_{1,k}$ and $r_{2,k}$ in supply chain decreased through H_∞ control. This is means that this method reduced the bullwhip effect when demand disturbance is maximal at the terminal of the supply chain.

The bullwhip effect will lead the financial loss in logistics management. So, we discuss the economic loss. We just show the economic loss of the first and second middle aggregation points to the first five customers demand as the bullwhip effect. p_{1k} means the economic loss of middle aggregation point at k and p_{2k} means the economic loss of distributor at k .

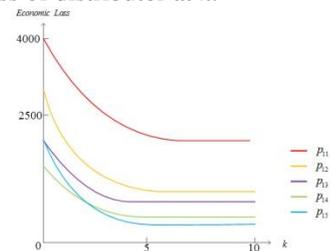


FIGURE 3 The economic loss of the middle aggregation point at k

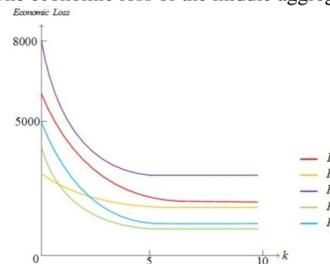


FIGURE 4 The economic loss of the distributor at k

From Figures 3 and 4, we can see clearly that the economic loss of middle aggregation point and distributor reduce obviously and tend to be stable through the H_∞ control. The numerical analysis shows that the bullwhip effect is well suppressed ($r_{1,k}$ and $r_{2,k}$ diminish) and the economic loss reduces greatly through the H_∞ control.

5 Conclusions

Bullwhip effect, or demand information distortion, has been a subject of both theoretical and empirical studies in the operations management literature. In this paper, we present a hierarchical model framework for the analysis of the bullwhip effect of inventories in multi-echelon distribution supply chains. The work we have done is as follows:

1) We establish a hierarchical model framework of multi-echelon distributions.

2) We describe the bullwhip effect in a quantitative method.

3) We analyse this quantitative description of bullwhip effect through the H_∞ control. Through the numerical analysis, we can see that the bullwhip effect is restrained and weakened effectively. This means that the bullwhips effect is controlled in logistics management.

Acknowledgments

This paper belongs to the project of “the research of multi-source information service system under the cloud computing environment” supported by Chinese National Natural Science Foundation (No. 71131002) and the project of “the research of Evidence reasoning theory and system under two-dimensional Semantic” supported by Chinese National Natural Science Foundation (No. 71071048).

References

- [1] Forrester J 1961 *Industrial Dynamic* New York: MIT press and Wiley&Sons Inc
- [2] Shu L, Yan Q 2005 Analysis on Restraining Bullwhip Effect by Different Demand Information Sharing *Journal of Jiaozuo Institute of Technology* 24(5) 405-9
- [3] Ouyang Y 2007 The effect of information sharing on supply chain stability and the bullwhip effect *European Journal of Operational Research* 182 1107-21
- [4] Gaur V, Giloni A, Seshadri S 2005 Information sharing in a supply chain under ARMA demand *Management Science* 51(6) 961-9
- [5] Krane S D, Braun S N 1991 Production smoothing evidence from physical product data *J Political Economy* 99(3) 558-81
- [6] Ouyang Y 2005 System-level stability and optimality of decentralized supply chains *PhD Dissertation University of California Berkeley USA*
- [7] Ouyang Y, Daganzo C F 2006 Characterization of the bullwhip effect in linear, time-invariant supply chains: Some formulae and tests *Management Science* 52(10) 1544-56
- [8] Senge P M 1990 *The Fifth Discipline* New York: Doubleday USA
- [9] Senge P M, Sterman J D System thinking and organizational learning: Acting locally and thinking globally in the organization of the future *European Journal of Operational Research* 59(3) 137-45
- [10] Towill R D 1996 Industrial dynamics modeling of supply chains *International Journal of Physical Distribution&Logistics Management* 26(2) 23-41
- [11] Sterman J D 1995 The beer distribution game In Heicie J, Meile L, eds. *Games and Exercises for Operations Management* New Jersey: Prentice Hall 101-12
- [12] Sucky E 2009 The bullwhip effect in supply chains – An overestimated problem? *IJPE* 118(1) 311-22
- [13] Lee H, Padmanabhan V, Whang S 1997 Information distortion in a supply chain: The bullwhip effect *Management Science* 43 546-58
- [14] Xu K, Dong Y 1998 Vendor Managed Inventory: Towards a Better Coordination of Information in a Supply Chain *ICM98_Shang-hai*
- [15] Chen F, Drezner Z, Ryan J K, Simchi-Levi D 2000 Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information *Management Science* 46(3) 436-43
- [16] Blackburn J D 1991 Time-based competition: The next battleground in American manufacturing *Homewood: Business One Irwin*
- [17] Osczelkan E C, Cakany D M 2009 Reverse bullwhip effect in pricing *European Journal of Operational Research* 192(1) 302-12
- [18] Rinks D B 2002 System dynamics in supply chains *Proceedings of the 2002 Euroma Conference* Copenhagen 443-57
- [19] Chen L, Lee H L 2009 Information sharing and order variability control under a generalized demand model *Management Science* 55(5) 781-97
- [20] Bottani E, Montanari R, Volpi A 2010 The impact of RFID and EPC network on the bullwhip effect in the Italian FMCG supply chain *International Journal of Production Economics* 124(2) 426-32
- [21] Caloiero G, Strozzi F, Comenges J Z 2008 A supply chain as a series of filters or amplifiers of the bullwhip effect *International Journal of Production Economics* 114(2) 631-45
- [22] Chandra C, Grabis J 2005 Application of multi-steps forecasting for restraining the bullwhip effect and improving inventory performance under autoregressive demand *European Journal of Operational Research* 166(2) 337-50
- [23] Clark T, Hammond J 1997 Reengineering channel reordering processes to improve total supply chain performance *Production and Operation Management* 6(3) 248-65
- [24] Anupindi R 1993 Supply Management Under Uncertainty *Graduate School of Industrial Administration* Carnegie Mellon University
- [25] Bassok Y, Bixby A, Srinivasan R, Wiesel H Z 1997 Design of component supply contract with commitment revision flexibility *IBM J Res Develop* 41(6) 693-702

Author



Ran Li, born in October, 1978, Anhui, China

Current position, grades: studying for doctoral degree, associate professor.

University studies: A Doctor of Business Administration of Hefei University of Technology, 2012.

Scientific interest: Environmental Economics, Supply Chain and Financial Management.

Publications: 5.

Experiences: teaching in School of management of Anhui Jianzhu University since 2005.

An extension evaluation model of the operation state of aero engine

Jian Chu^{1*}, Guoyu Wang², Shan Xu³

¹Joint Laboratory of Information Sensing & Intelligent Control, Tianjin University of Technology and Education, Tianjin, 300222, China

²School of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin, 300222, China

³Tianjin Bohai Vocational Technical College, No11 Juliansan Roda Hexi District Tianjin, 300221, China

Received 12 June 2014, www.tsi.lv

Abstract

Traditional fault diagnosis with single parameter fails to evaluate the operation state of aero engine. This paper analyzes the extension evaluation based on supervision information of engine's performance and works out an evaluation system as well as an extension evaluation model and algorithm based on incidence function in extension theory. Through the model, parameters under the operation state of aero engine are studied. The incidence function in extension theory between classic domain and section domain of all attributes and parameters is established in the corresponding evaluation system. This makes it possible to acquire the state level of aero engine according to incidence in extension theory.

Keywords: aero engine, operation state, incidence function, evaluation, model

1 Introduction

Monitoring the operation state of aero engine is important to prolong its lifespan and ensure a safety flight. With the development of science and technology, modern engine is more and more complicated and can be affected by many factors, some of which are even unknown. Parameters in current monitoring method also fail to reveal the overall performance of the engine. Many remain untestable as is limited to conditions. Without enough data, it becomes a technical difficulty to evaluate the performance of the engine [1-4].

When the engine breaks down, it is significant to obtain relevant information on the operation state of aero engine for fault diagnosis that serves to the maintenance and prediction of the engine. Currently, there are three ways of analytical methods: (1) System operation state analysis based on signal treatment [5, 6]; (2) System operation state analysis based on analytic model [7, 8]; (3) System operation state analysis based on knowledge diagnosis [9, 10].

However, these methods are fuzzy and uncertain that cannot provide accurate diagnosis for the operation state when there are only small samples and little information.

Therefore, this paper constructs an extension evaluation model following the incidence function in extension theory as well as the algorithm. It hopes to advance the fault diagnosis and prediction by finding out possible fault information and learning about the state level to increase the efficiency and accuracy of the

diagnosis. This is significant in terms of theory and engineering.

2 Establishment of the evaluation system for the operation state of aero engine

The aero engine controlling system is becoming complicated and bigger in size with the advancement of automation. The fault of the system will bring huge losses, which makes it necessary to improve its reliability and safety as well as the diagnosis technique. However, there are many factors that count. It is hard to monitor the reliability with only one source of information. Thus, various kinds of monitoring methods and flexible and sensitive parameters are needed to realize the judgment on the overall performance of aero engine. The operation state analysis relies on sensor technique, testing technique, computer technology, display technique and artificial intelligence analysis. It also requires the support of large amount of maintenance tests. Fault diagnosis, determination of the state level and the prediction of reliability and accuracy are the key. Therefore, based on previous researches and after technical communication with experts, this paper categories parameters of operation state into low/high pressure rotor speed performance, low/high pressure turbine guide vane angle adjustment performance, turbine exhaust control performance, vibration performance of nozzle, oil performance pressure, slip property and cycle life performance as a part of the evaluation system, as is shown in Figure 1.

* *Corresponding author* e-mail: chujian6@126.com

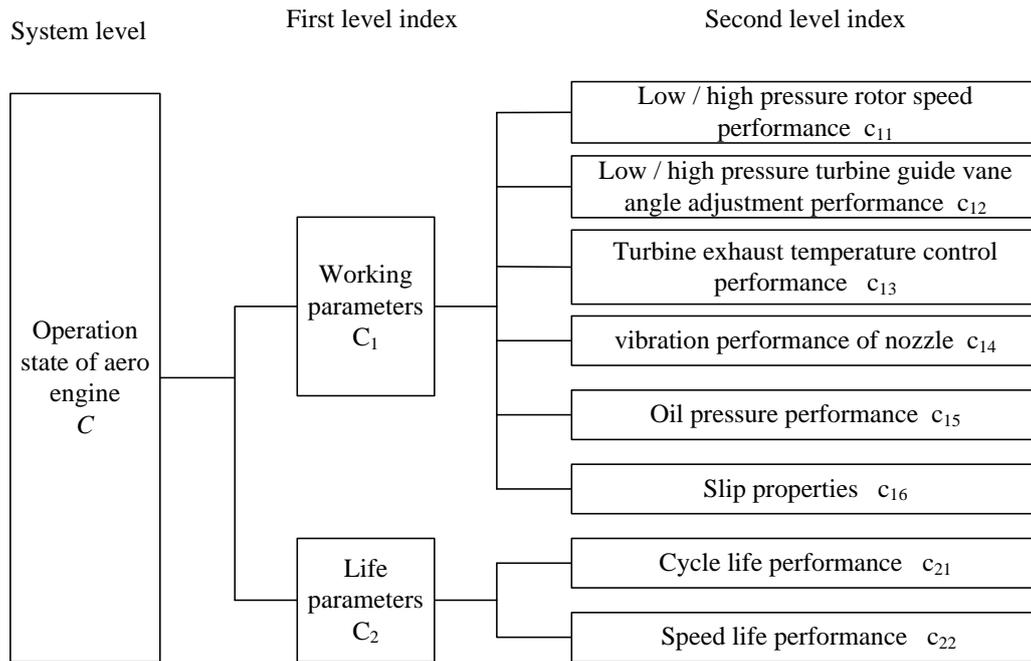


FIGURE 1 The evaluation system for the operation state of aero engine

The extension analysis of the operation state takes the advantage of all state information and existing knowledge to get a whole process of operation state and fault state evaluation.

3 Extension evaluation model and algorithm of the operation state of aero engine

3.1 DETERMINATION OF PERFORMANCE FACTOR SET OF THE OPERATION STATE

According to Figure1, the performance factor set has several levels of indexes. The system level shows the overall performance of the engine and is expressed by $C = \{C_1, C_2\}$ and fits $C_1 \cap C_2 = O$, which means that parameters of a subordinate level do not mix. The first level indexes are C_1 and C_2 , with $C_1 = \{c_{11}, c_{12}, c_{13}, c_{14}, c_{15}, c_{16}\}$ and $C_2 = \{c_{21}, c_{22}\}$. The second level shows the specific performance of the engine.

3.2 STANDARDIZATION OF INDEXES OF THE OPERATION STATE

Some of the parameters can be measured up accurately while others cannot. Some have a positive effect on the operation state while others may have a negative effect. Therefore, parameters of different types should be normalized and standardized for better analysis.

For those fuzzy or uncertain parameters, fuzzy evaluation is given to the degree of membership. Specific standards are shown in Table 1.

If parameters are measurable, when parameter $v^i(c)$ is the positive index, the standardized index will be:

$$\tilde{v}^i(c) = [\tilde{v}_1^i(c), \tilde{v}_2^i(c)] = \left(\frac{v_1^i(c)}{v_2^{ik} | \max_{1 \leq j \leq m} (v_2^{ij})}, \frac{v_2^i(c)}{v_1^{ik} | \max_{1 \leq j \leq m} (v_1^{ij})} \right), \quad (1)$$

TABLE 1 The degree of membership of fuzzy evaluation of parameters

The degree of membership	Performance level
0.9	Best state
0.7	Good state
0.5	Medium state
0.3	Poor state
0.1	Worst state
0.8,0.6,0.4,0.2	State in between

When $v^i(c)$ is the negative index, the standardized index will be:

$$\tilde{v}^i(c) = [\tilde{v}_1^i(c), \tilde{v}_2^i(c)] = \left(\frac{v_1^{ik} | \min_{1 \leq j \leq m} (v_1^{ij})}{v_2^i(c)}, \frac{v_1^{ik} | \min_{1 \leq j \leq m} (v_1^{ij})}{v_1^i(c)} \right), \quad (2)$$

$v_2^{ik} | \max_{1 \leq j \leq m} (v_2^{ij})$ refers to the maximum value of $v^i(c)$ in the intervals of classic domain. $v_1^{ik} | \min_{1 \leq j \leq m} (v_1^{ij})$ refers to the minimum value of $v^i(c)$ in the intervals of classic domain. After standardization, the index fits $0 \leq \tilde{v}_1^i(c) \leq 1$ and $0 \leq \tilde{v}_2^i(c) \leq 1$. And indexes of classic domain and section domain fit [0,1]. Therefore, the state value, classic domain and section domain of all indexes are standardized and the difference is eliminated.

3.3 WEIGHT OF INDEXES OF THE OPERATION STATE

Weight refers to the priority of parameters. It is important to the analysis of the operation state. The more accurate, scientific and objective the weights are the closer to the reality and more convincible the result will be. Gray related analysis makes up the weakness of traditional mathematical calculation. It can address the system which has small samples, little information and regularities. It is close to qualitative analysis. Gray related analysis depends on similarity to judge how closely parameters are related. A closer curve means highly related.

Apply those parameters to the analysis. Suppose there is a pair parameter sequence:

$$X_i = (x_i(1), x_i(2), x_i(3), \dots, x_i(k)), k=1,2,3,\dots,n.$$

Select the ideal parameter sequence: $X_0 = (x_0(1), x_0(2), x_0(3), \dots, x_0(k)), k=1,2,3,\dots,n$. And it shall fit:

$$x_0(k) = \max_{1 \leq i \leq n} x_i(k), k=1,2,3,\dots,n.$$

Compare the tested result with the standard result. And calculate the incidence coefficient $\varphi_i(k)$ at an index k , and the expression is:

$$\varphi_i(k) = \frac{\min_i \min_k |x_0(k) - X_i(k)| + \beta \max_i \max_k |x_0(k) - X_i(k)|}{|x_0(k) - X_i(k)| + \beta \max_i \max_k |x_0(k) - X_i(k)|} \quad (3)$$

$$\beta \in (0,1), k=1,2,\dots,n$$

Calculate all incidence coefficients, normalize them and get the weight of each index.

3.4 CALCULATION BASED ON INCIDENCE FUCTION IN EXTENSION THEORY

Incidence in extension theory is constructed [13-15]. Suppose the classic domain of index i is $V_{cla}^i(c) = [v_{cla-lef}^i(c), v_{cla-rig}^i(c)]$, $v_{cla-lef}^i(c) \leq v_{cla-rig}^i(c)$ and its section domain is $V_{sec}^i(c) = [v_{sec-lef}^i(c), v_{sec-rig}^i(c)]$, $v_{sec-lef}^i(c) \leq v_{sec-rig}^i(c)$. When the characteristic value $v^i(c)$ of index i is accurate, the extension distance ρ_i^{cla} between index i and classic domain $V_{cla}^i(c)$ is:

$$K_i = \begin{cases} -\rho_i^{cla} / |v_{cla-lef}^i(c) - v_{cla-rig}^i(c)| & v^i(c) \in [v_{cla-lef}^i(c), v_{cla-rig}^i(c)] \\ \rho_i^{cla} / (\rho_i^{sec} - \rho_i^{cla}) & v^i(c) \notin [v_{cla-lef}^i(c), v_{cla-rig}^i(c)] \end{cases} \quad (8)$$

$$\rho_i^{cla} = \left| v^i(c) - \frac{v_{cla-lef}^i(c) + v_{cla-rig}^i(c)}{2} \right| - \frac{v_{cla-rig}^i(c) - v_{cla-lef}^i(c)}{2} \quad (4)$$

When the characteristic value $v^i(c)$ of index i is accurate, the extension distance ρ_i^{sec} between index i and section domain $V_{sec}^i(c)$ is:

$$\rho_i^{sec} = \left| v^i(c) - \frac{v_{sec-lef}^i(c) + v_{sec-rig}^i(c)}{2} \right| - \frac{v_{sec-rig}^i(c) - v_{sec-lef}^i(c)}{2} \quad (5)$$

When the characteristic value $v^i(c)$ of index i is fuzzy, that is when $v^i(c) = [v_{lef}^i(c), v_{rig}^i(c)]$, the extension distance ρ_i^{cla} between index i and section domain $V_{cla}^i(c)$ is:

$$\rho_i^{cla} = \left(\frac{1}{2} \left| v_{lef}^i(c) - \frac{v_{cla-lef}^i(c) + v_{cla-rig}^i(c)}{2} \right| \right) + \left(\frac{1}{2} \left| v_{rig}^i(c) - \frac{v_{cla-lef}^i(c) + v_{cla-rig}^i(c)}{2} \right| \right) \quad (6)$$

When the characteristic value $v^i(c)$ of index i is fuzzy, that is when $v^i(c) = [v_{lef}^i(c), v_{rig}^i(c)]$, the extension distance ρ_i^{sec} between index i and section domain $V_{sec}^i(c)$ is:

$$\rho_i^{sec} = \left(\frac{1}{2} \left| v_{lef}^i(c) - \frac{v_{sec-lef}^i(c) + v_{sec-rig}^i(c)}{2} \right| \right) + \left(\frac{1}{2} \left| v_{rig}^i(c) - \frac{v_{sec-lef}^i(c) + v_{sec-rig}^i(c)}{2} \right| \right) \quad (7)$$

When the extension distances are acquired, then there comes the incidence function between classic domain $V_{cla}^i(c)$ and index i is:

Suppose the weight of index i is w_i and it fits $\sum_{i=1}^n w_i = 1$, then the incidence in extension theory Φ_i between index i and classic domain $V_{cla}^i(c)$ is:

$$\Phi_i = \sum_{i=1}^n (w_i K_i). \tag{9}$$

According to close principle, if there is:

$$\Phi_0 = \max_{1 \leq t \leq m} (\Phi_t) = \max(\Phi_1, \Phi_2, \dots, \Phi_m) = \Phi_t, 1 \leq t \leq m. \tag{10}$$

Then the state attribute belongs to the classic $V_{cla}^t(c)$, which means the fault state of the operation state is in the state t . Further maintenance and repair of aero engine can base on the predicted state.

3.5 EXTENSION EVALUATION MODEL AND ALGORITHM OF THE OPERATION STATE

The algorithm can be described as the follows:

Step 1: Monitor the aero engine that needs the extension evaluation analysis based on data extraction and produce standard state samples of data monitoring;

Step 2: Consult relevant experts and engineers, acquire parameters of the operation state and construct an evaluation system based on the operation state analysis;

Step 3: Under the evaluation system, acquire the classic domain and the section domain of indexes based on design knowledge and experience;

Step 4: Acquire the standardized monitoring data, classic domain and the section domain based on

Equations (1) and (2) and get the after-standardization data;

Step 5: Acquire weights of indexes under the evaluation system based on Equation (3);

Step 6: Calculate the extension distance between index and its corresponding classic domain based on Equations (4) and (6);

Step 7: Calculate the extension distance between index and its corresponding section domain based on Equations (5) and (7);

Step 8: Acquire the incidence function in extensive theory between index and classic domain based on Equation (8);

Step 9: Acquire the incidence in extensive theory between index and classic domain based on Equation (9);

Step 10: Based on the incidence in extensive theory and the close principle, acquire the classic domain of the operation state of aero engine and determine the state level of the aero engine to be monitored and prepare for further maintenance.

4 Empirical tests

This paper analyses and explains the extension evaluation model with the example of an aero engine in normal state. It leaves the interval to collect parameter information and consult with maintenance staff or experts to get the fuzzy judgment on relevance parameters under the evaluation system (Refer to Figure 1). This paper uses “Best, Good, Normal and Poor” to evaluate. The monitoring data after standardization is shown in Table 2 while the classic domain and the section domain is shown in Table 3.

TABLE 2 Monitoring data after standardization

Parameters	Weight	Indexes	Weight	Standard value
Working parameters	0.585	Low/high pressure rotor speed performance	0.35	0.80-0.90
		Low/high pressure turbine guide vane angle adjustment performance	0.12	0.70-0.80
		Turbine exhaust control performance	0.16	0.50-0.60
		Vibration performance of nozzle	0.23	0.60-0.70
		Oil performance pressure	0.06	0.80-0.90
Life parameters	0.415	Slip property	0.08	0.70-0.80
		Cycle life performance	0.45	0.60-0.70
		Speed life performance	0.55	0.40-0.50

TABLE 3 Classic domain and section domain of indexes after standardization

Indexes	Classic domain				Section domain
	Best	Good	Normal	Poor	
Low/high pressure rotor speed performance	0.80-0.90	0.70-0.80	0.50-0.70	0.10-0.50	0.1-0.9
Low/high pressure turbine guide vane angle adjustment performance	0.70-0.90	0.60-0.70	0.40-0.60	0.10-0.40	0.1-0.9
Turbine exhaust control performance	0.80-0.90	0.70-0.80	0.50-0.70	0.10-0.50	0.1-0.9
Vibration performance of nozzle	0.80-0.90	0.60-0.80	0.50-0.60	0.10-0.50	0.1-0.9
Oil performance pressure	0.80-0.90	0.70-0.80	0.50-0.70	0.10-0.50	0.1-0.9
Slip property	0.80-0.90	0.60-0.80	0.40-0.60	0.10-0.40	0.1-0.9
Cycle life performance	0.70-0.90	0.60-0.70	0.50-0.60	0.10-0.50	0.1-0.9
Speed life performance	0.80-0.90	0.70-0.80	0.50-0.70	0.10-0.50	0.1-0.9

The extension space between index, classic domain and section domain and incidence coefficient are acquired based on the extension evaluation model and algorithm in

Section 3.5. The specific values are shown in Tables 4 and 5.

TABLE 4 The extension space between index, classic domain and section domain

Indexes	The extension space between index and classic domain				The extension space between index and section domain
	Best	Good	Normal	Poor	
Low/high pressure rotor speed performance	0.00	0.05	0.15	0.35	-0.05
Low/high pressure turbine guide vane angle adjustment performance	-0.05	0.05	0.15	0.35	-0.15
Turbine exhaust control performance	0.25	0.15	-0.05	0.05	-0.35
Vibration performance of nozzle	0.15	-0.05	0.05	0.15	-0.25
Oil performance pressure	0.00	0.05	0.15	0.35	-0.05
Slip property	0.05	-0.05	0.15	0.35	-0.15
Cycle life performance	0.05	0.00	0.05	0.15	-0.15
Speed life performance	0.35	0.25	0.05	-0.05	-0.35

TABLE 5 Coefficient parameter between index and classic domain

Indexes	Coefficient parameter between index and classic domain			
	Best	Good	Normal	Poor
Low/high pressure rotor speed performance	0.000	-0.500	-0.750	-0.875
Low/high pressure turbine guide vane angle adjustment performance	0.250	-0.250	-0.500	-0.700
Turbine exhaust control performance	-0.417	-0.300	0.250	-0.250
Vibration performance of nozzle	-0.375	0.250	-0.167	-0.375
Oil performance pressure	0.000	-0.500	-0.750	-0.875
Slip property	-0.250	0.250	-0.500	-0.700
Cycle life performance	-0.250	0.000	-0.250	-0.500
Speed life performance	-0.500	-0.417	-0.125	0.125

The Incidence in extension theory of index can be acquired with the weight taken into consideration, that is, $\Phi_0 = \max(-0.035, -0.221, -0.313, -0.452) = \Phi_1$. It is clear that this engine works in a good state and asks no maintenance at present.

5 Conclusions

This paper analyses the extension evaluation based on supervision information of engine's performance and works out an evaluation system as well as an extension

evaluation model and algorithm based on incidence function in extension theory. Through the model, parameters under the operation state of aero engine are studied. And the incidence function in extension theory between the classic domain and the section domain of all attributes and parameters is established in the corresponding evaluation system. This makes it possible to acquire the state level of aero engine according to incidence in extension theory and support the computed-based monitoring and prediction of the operation state.

Reference

[1] Yang E, Zhang Z, Liu G, Cui D 1999 Real-time system for condition monitoring of propulsion system using BP-ART hybrid neural networks *Journal of Propulsion Technology* 20(6) 10-5

[2] Huang X 2004 Sensor fault diagnosis and reconstruction of engine control system based on autoassociative neural network *Chinese Journal of Aeronautics* 2004 17(1) 23-7

[3] Xu Q, Shi J 2004 Aeroengine fault diagnosis based on support vector machine *Journal of Aerospace Power* 20(2) 298-302

[4] Li H, Fei Y 2007 Discussion about techniques and development for fault diagnosis of aeroengine *Aviation Maintenance and Engineering* 5 41-2

[5] Chen M Z, Zhou D H, Liu G R 2005 A new particle predictor for fault prediction of nonlinear time-varying systems *Developments in chemical Engineering and Mineral Processing* 13(3) 379-88

[6] Shao J, Xu M, Wang R 2009 Bayesian Model-based Fault Diagnosis for the Rotor *Aircraft Engineering and Aerospace Technology* 81(1) 19-24

[7] Luo J H, Pattipati K R, Qiao L, Chigusa S 2007 *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews* 37(6) 1163-73

[8] Ge M, Xu Y S, Du R X 2008 *IEEE Transactions on Automation Science and Engineering* 5(1) 127-39

[9] Wu J D, Liu C H 2008 Investigation of engine fault diagnosis using discrete wavelet transform and neural network *Expert Systems with Applications* 35(3) 1200-13

[10] Pashayev A M, Askerov D D, Sadiqov R A, Abdullayev P S 2004 Fuzzy-neural approach for aircraft gas turbine engines diagnosing AIAA First Intelligent Systems Technical Conference September 2004 1-15

[11] Wei G The method of gray related analysis to multiple attribute decision making problem with interval numbers *Journal of Systems Engineering and Electronics* 2006 28(9) 1358-9

[12] Dang Y, Liu S, Liu B, Yu Y 2004 Study on incidence decision making of multi-attribute interval number *Journal of Nanjing University of Aeronautics and Astronautics* 36(3) 403-6

[13] Zhao Y W, Zhang G X A 2012 New Integrated Design Method Based On Fuzzy Matter-Element Optimization *Journal of Materials Processing Technology* 129(1-3) 612-8

[14] Wang T, Yang A, Zhong S, Zhao Z 2014 Extension Adaptive Design Model Of Scheme Design For Complex Mechanical Products *Tehnički vjesnik/Technical Gazette* 21(1) 123-33

[15] Yang C, Cai W 2007 *Extension engineering Beijing: Science Press*

Authors	
	<p>Jian Chu, born on April 4, 1961, Tianjin, China</p> <p>Current position, grades: professor. University studies: Master of Engineering. Experience: Engaged in automation specialty, building automation, electrical controls and sensors work of teaching since 1983.</p>
	<p>Guoyu Wang, born in June, 1987</p> <p>Current position, grades: postgraduate at Tianjin Vocational and Technical University.</p>
	<p>Shan Xu, born in 1981</p> <p>Current position, grades: lecturer. University studies: Master degree at Hebei University, majoring in technology of power system and its automation. Experience: Teaching in Department of Electrical Engineering of Tianjin Bohai Vocatioanl and Technical College for 9 years.</p>

The bionic design, virtual prototype modelling and motion simulation of biped robot with heterogeneous legs

Hualong Xie¹, Nan He¹, Fei Li^{2*}

¹School of Mechanical Engineering & Automation, Northeastern University, Shenyang, Liaoning, 110819, P.R. China

²School of Information Science & Engineering, Shenyang University of Technology, Shenyang, Liaoning, 110870, P.R. China

Received 1 March 2014, www.tsi.lv

Abstract

The research of humanoid robot and intelligent prosthesis is integrated and a new-style humanoid robot named biped robot with heterogeneous legs (BRHL) is proposed to provide an ideal test-bed for intelligent bionic leg (IBL). The research background and concept of BRHL are introduced. The existed problems of common humanoid robot leg are analysed in detail. Based on bionics, the joint structure and driving scheme of artificial leg with pneumatic muscle actuator and bionic leg with magneto-rheological damper are designed. Using Pro/E, the virtual prototypes of artificial leg, bionic leg and BRHL are established. The bionic characteristics of artificial leg and bionic leg are analysed and the motion simulation of BRHL using ADAMS is done. The simulation indicates that BRHL can simulate amputees with IBL well and is an ideal test-bed for IBL.

Keywords: biped robot with heterogeneous legs, intelligent bionic leg, artificial leg, pneumatic muscle actuator, motion simulation

1 Introduction

In the process of human social development, it has led to a lot of people of physical disabilities as a result of natural disasters, accidental injury, traffic accident and war, and many other factors. The 2nd national sampling investigation of the disabled indicates that it has at least 24120 thousand people with physical disabilities in china and the lower limb amputees are about 440 thousands [1]. In the US, around 1.6 million people live with limb loss. About 97% of all vascular limb loss is lower-limb amputations, of which 25.8% are above-knee amputations [2]. The main research challenges in the design of transfemoral prostheses are the efficiency with respect to the metabolic/external energy consumption and the adaptability to various walking conditions [3].

Intelligent bionic legs (IBL) controlled by a micro processing unit (MPU) is an advanced intelligent prosthesis (IP) [4, 5]. Thanks to precise MPU control, amputees with IP can change their gaits according to their needs. The prosthesis gait of amputees used in everyday life could include running, walking on slope, riding a bicycle, etc. To guarantee the IBL control performance, repetitive walking test of amputees with IBL is necessary. However, IBL test will exhaust the amputees. It is not only costly but also painful to human subjects. It may even lead to accidental harm to amputees.

Currently, a leg simulator is generally used in IBL tests, as shown in Figure 1. This leg simulator can produce human hip motion and drive IBL to test swing and stance performance. But it cannot be used to test the

walking stability, dual-leg coordination and IBL's gait tracking to the healthy leg of amputees.

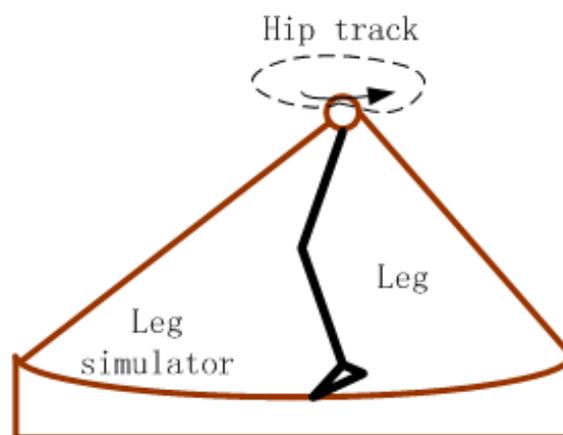


FIGURE 1 Sample leg simulator

The study of humanoid robot can help human understand own walking mechanism and use these characteristics for serving itself. At present, some devices, such as endoskeleton and exoskeleton prosthesis, have adopted certain technology of the humanoid robots:

- 1) The dynamic walking study of humanoid robots can help us design reasonable prosthesis mechanism which would make leg disabled people liberate from wheelchair or help hand disabled people restore life self-care ability.
 - 2) The limbs coordinated motion study of humanoid robots can help us develop the exoskeleton prosthesis which would increase the strength and speed of human.
- For example, the BLEEX robot leg belongs to

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

exoskeleton prosthesis and has important military significance.

The ultimate research goal of humanoid robots is to achieve a full range of anthropomorphism. In order to provide an idea test-bed for IBL, the research of humanoid robot and intelligent prosthesis is integrated and a new-style humanoid robot named biped robot with heterogeneous legs (BRHL) is proposed by the robotics group at Northeastern University, China [6]. A BRHL model is shown in Figure 2.

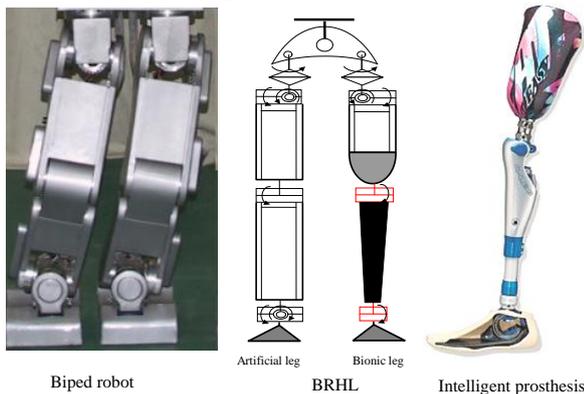


FIGURE 2 A BRHL model

BRHL is composed of two legs. One is artificial leg (AL) and the other is above-knee bionic leg (BL). AL has 6 degrees of freedom (3 DOF hip, 1 DOF knee and 2 DOF ankle). BL includes an artificial leg hip joint, intelligent bionic leg and flexible prosthetic foot. The artificial leg is corresponding to amputee's healthy leg and bionic leg corresponding to intelligent prosthesis used by amputees. So BRHL can simulate amputees' walking with IBL well and is an ideal test-bed for intelligent prosthesis. Since BRHL mainly orients area of rehabilitation medicine, its design focuses on personification. In the paper, the bionic design, virtual prototype modeling and motion simulation of a simplified BRHL which can walk forward are done.

2 Bionic analysis of common humanoid robot leg

The bionic design of knee joint is the key problem of BRHL development. Its design mainly refers the structure of common humanoid robot leg. Though the humanoid robots have been further studied and human basic movements, such as stable walking, up and down stairs, turn and dance movements have been realized, many problems still exists in the developed of humanoid robots:

- 1) Its walking gait is unnatural and has a larger difference from that of human leg.
- 2) Limited to own mechanism structure, the impact between feet and ground is large and walking speed is slow.
- 3) The energy consumption and rigidity of drive are larger and go against the realization of joint flexible movement.

The bones of human leg are the result of natural selection and evolution for a long time and are most

suitable for human walking. At present, knee joints of humanoid robot most adopt uniaxial axis mechanism and motor drive which have essential difference with joint structure, motion mechanism and driving mode of human leg. It results in the difference of leg gait between human and humanoid robot. The main differences can be summarized in the following three aspects.

2.1 KNEE JOINT STRUCTURE

The human knee joint is mainly composed of the femur, tibia and patella. The contact surface between femoral bottom and tibial top is irregular. During flexion and extension activity, the knee joint is driven by expansion motion of medial and lateral muscles. There are both rolling and sliding between the two contact surfaces. The outstanding feature of knee joint is that its instantaneous centre of rotation (ICR) is not fixed and similar a "J" curve, as shown in Figure 3 [7]. Thus the leg has alterable thigh and calf length and high obstacle negotiation performance. At the same time, the change of knee joint ICR and leg length can adjust the torque applied by the ground reaction force on joint and reduce the needed muscle extension force of hip joint, which can improve the walking stability and efficiency of the legs.

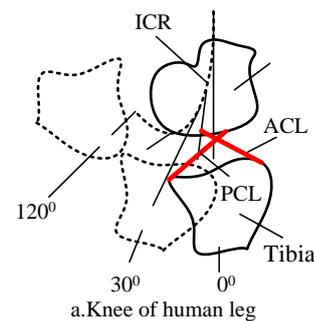


FIGURE 3 ICR of human knee joint

Now most joints of the humanoid robot and artificial limb adopt uniaxial knee mechanism which is shown in Figure 3. Its ICR is fixed and has obvious difference from that of human knee joint. To maintain stability during support phase, the robot generally keep legs bend when stand and walks slowly which would cause walking gait is unnatural.

2.2 MECHANISM RIGID-FLEXIBLE CHARACTERISTIC

The human leg is a flexible system as an organic whole. Depending on foot arch and ligaments adhered, human leg can generate flexibility, reduce impact and collision between feet and ground when walking, protect joints and increase ability of bearing human weigh and pressure. The flexibility of human feet can store energy and contribute to realize running and jumping. The knee joint of human has ligaments and meniscus mechanism, which can play a buffer action and protect joints from injuries in the process of human movement, as shown in Figure 4.

In contrast, the most legs of common humanoid robot are connected by rigid bars and its joints are driven by motors. Thus common humanoid robot would have large rigidity and poor flexibility which is easy to cause unstable walking especially under uneven road. To solve above problems, the feet of common humanoid robot are usually designed as rectangular plates to increase contact area with ground. At the same time, the speed when foot contacts ground is limited and buffer devices are installed to reduce impact and collision between foot and ground when walking. Although above methods can reduce the impact and collision to insure robot walking steadily, but cause its gait unnatural.

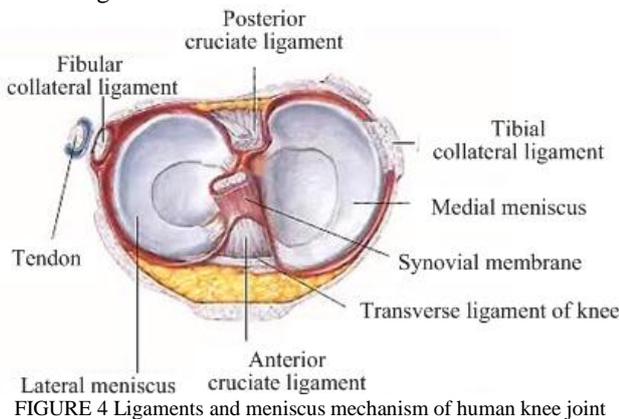


FIGURE 4 Ligaments and meniscus mechanism of human knee joint

2.3 JOINT ACTUATOR

When human walks in normal gait, muscles connected with ligaments on both sides of joints would accept neural signal and its expansion motion would drive joints to rotate and slide. The muscles and ligaments with good flexibility can effectively buffer impact and collision of ground and the walking is steady. Currently, most joints of legs are driven by electric or hydraulic actuator, etc. The actuator of common humanoid robot has obvious difference with that of human. Humanoid robot joints can be distinguished as three categories, including active, semi-active and passive, by the types of actuators that were used in the according mechanism.

1) Active joint: Joint is always driven by electromotor, pneumatic or hydraulic device and can swing optionally if actuator is allowable. Its energy consuming is much larger.

2) Passive joint: Joint without actuator basically consumes no energy and has ability of storing some energy by mechanism design. It can walk only depending on its gravity. Its walking gait is natural, but road condition is limited.

3) Semi-active joint: Joint is always driven by intelligent damper, such as electro-rheological or magneto-rheological damper, etc. It can provide large damping force, but active force is less.

3 Bionic design and modelling of artificial leg

3.1 MECHANISM DESIGN OF KNEE JOINT

Currently, there are some kinds of knee joint configurations, mainly including uniaxial and multiaxial knee joint. Since artificial leg simulates human healthy leg to generate natural gait for intelligent bionic leg, its design emphasizes particularly on humanoid performance. To achieve that, 4-bar multiaxial knee mechanism which is shown in Figure 5 is adopted. In order to ensure that the artificial leg can keep balance in support phase, limit stop is added to avoid the shank excessive extension. All joints adopt bearing support and each rod is connected by shaft bossing. At the same time, in order to guarantee walking stability, 4-bar mechanism with both sides is adopted. The parameter values of 4-bar knee mechanism obtained by mechanism optimization are shown in Table 1 [8].

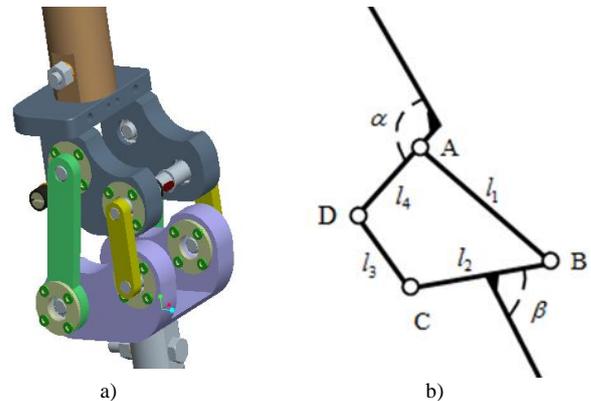


FIGURE 5 4-bar multiaxial knee joint: a) Virtual prototype, b) Schematic diagram

TABLE 1 Values of 4-bar knee parameters

l_1	l_2	l_3	l_4	α	β
0.089m	0.079m	0.04m	0.059m	2.49rad	1.57rad

The calculation formula of ICR coordinates can be written as:

$$\begin{cases} x_{ICR} = \frac{y_C - y_B + ax_B - bx_C}{a - b} \\ y_{ICR} = \frac{x_C - x_B + a^{-1}y_C - b^{-1}y_B}{a^{-1} - b^{-1}} \end{cases} \quad (1)$$

$$a = \frac{y_A - y_B}{x_A - x_B}, b = \frac{y_C - y_D}{x_C - x_D}$$

The ICR of 4-bar multiaxial knee joint is calculated, as shown in Figure 6.

Compared with uniaxial knee mechanism used by common humanoid robot, it has many advantages as follows:

1) ICR of 4-bar multiaxial knee joint is not fixed and similar to "J" curve. It can effectively simulate normal movement of human knee joint.

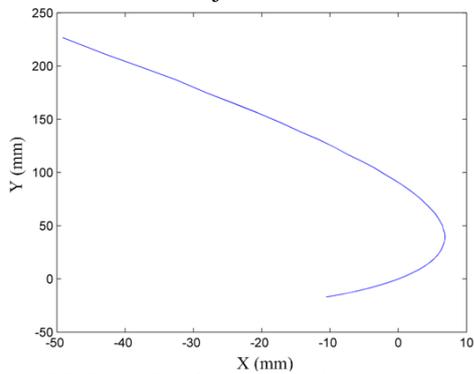


FIGURE 6 ICR of 4-bar multiaxial knee joint.

2) During swing phase, it can effectively reduce the length of thigh and shank under the same swing angle which is shown in Figure 7. Thus it has good obstacle-surmounting performance.

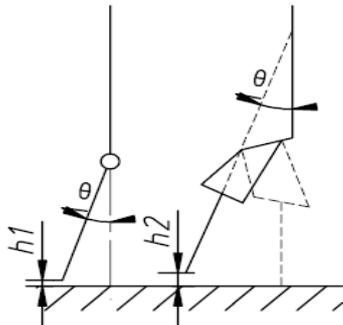


FIGURE 7 Contrast of obstacle-surmounting performance for uniaxial and multiaxial knee joint.

3) During mid-swing phase or sitting down, its ICR would down to the normal position and effectively improve sitting posture.

4) It can effectively improve the stability of support phase which is shown in Figure 8.

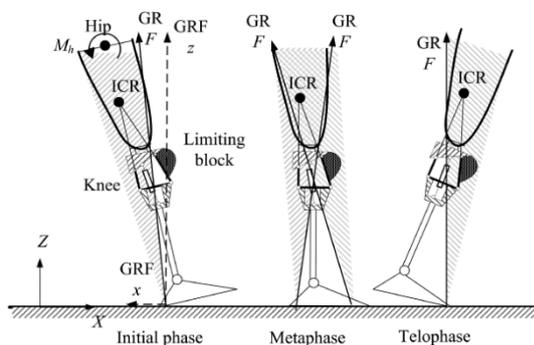


FIGURE 8 Stability of 4-bar multiaxial knee joint during support phase

3.2 DRIVING SCHEME DESIGN OF KNEE JOINT

The human knee joint is driven by antagonistic muscles which not only can provide driving force for precise position control, but also have good flexibility to absorb vibration and buffer impact. Pneumatic muscle actuator

(PMA) has advantages of light weight, simple structure, large output force, good flexibility and similar force-length characteristics to human muscle, etc. In the paper, PMA is selected as actuator of knee joint.

The knee joint with 4-bar mechanism has four drive shafts. The control torque analysis of each drive shaft is shown in Figure 9. The needed control torque of shaft A is minimum and shaft D takes second place. From the view of energy saving, shaft A or D should be selected. Restricted by its own structure of 4-bar mechanism, the space of axis A is less and goes against to install sprocket drive system. Shaft D is selected as drive shaft which can not only meet the requirements of actual mechanical structure, but also can reduce energy consumption.

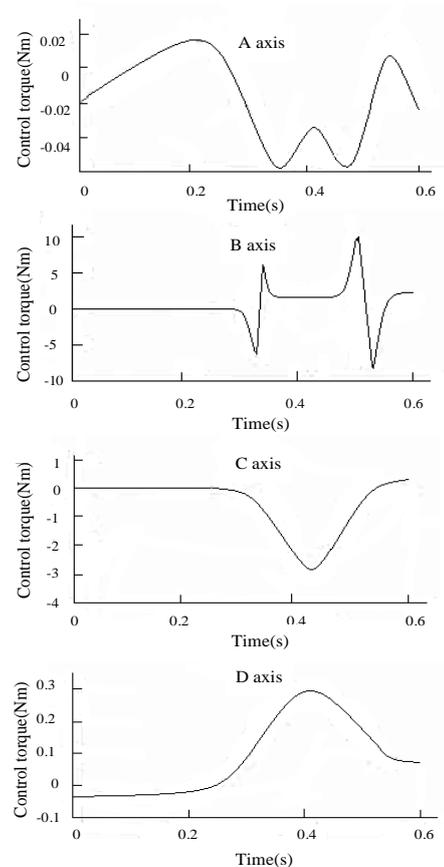


FIGURE 9 Control torque analysis of each drive shaft

Two sets of placement schemes for PMA are designed. The first scheme with single drive is shown in Figure 10, that is to say, through a shaft to drive the movement of knee joint with two sides. The second scheme with separating drive is shown in Figure 11, that is to say, two parallel 4-bar linkages are driven respectively with two drive shafts.

Compared with the first scheme, the second scheme require rotation angle of two axes synchronous and consistent. Thus its control system is complicated. The first scheme, which has simple structure and can simplify control is adopted.

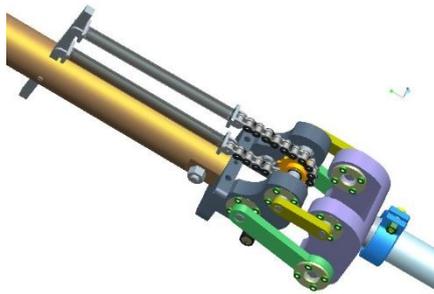


FIGURE 10 The first placement scheme of PMA

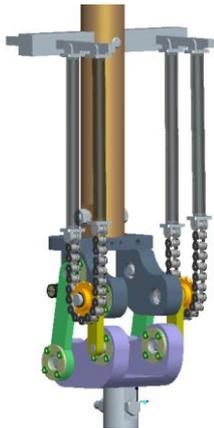


FIGURE 11 The second placement scheme of PMA

The shaft D is driven by a couple of pneumatic artificial muscles through sprocket installed and provides a driving force for 4-bar knee joint. Ends of the shaft are supported by bearings. The sprocket drive is adopted because it has advantages of high transmission accuracy, simple structure, strong bearing capacity, easy maintenance and long service life, etc. The sprocket is connected with drive shaft by key. In order to prevent axial movement of sprocket, sprocket is fixed by shaft shoulder and clasp on both sides separately.

3.3 MODEL SELECTION OF PNEUMATIC ARTIFICIAL MUSCLE

The model selection of pneumatic artificial muscle is mainly determined by two parameters: diameter and length. The working principle of knee joint driven by pneumatic artificial muscle is shown in Figure 12. The load which is connected with pneumatic artificial muscle by sprocket and chain can be considered as a uniform and thin rod with length l and quality m . The rotation angle of knee joint is described using θ and its initial value is zero. The initial length and diameter of pneumatic artificial muscle are described by L_0 and D_0 .

According to moment balance, the following relationships can be obtained.

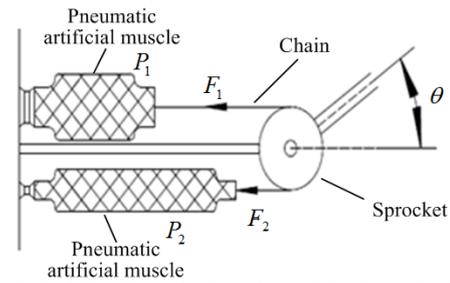


FIGURE 12 Working principle of knee joint driven by PMA

$$T = mg \frac{l}{2} \sin \theta, \tag{2}$$

$$T = (F_1 - F_2) \frac{D_0}{2}, \tag{3}$$

Then the tension difference can be derived by simultaneous Equations (2) and (3).

$$F_1 - F_2 = \frac{mgl}{D_0} \sin \theta, \tag{4}$$

With reference to the human body parameters [9], shank length is $0.48m$, shank weight is $3.225kg$, foot weight is $1.125kg$, and flexion-extension angle range of knee joint is $0^\circ \sim 135^\circ$. The sprocket diameter is $0.368m$. When θ is equal to 90° , the tension difference is maximum. Input above parameters values into Equation (4), the tension difference can be calculated as $685.6N$.

In order to determine the length of pneumatic artificial muscle, the contraction length ΔL of pneumatic artificial muscle when human knee bends to the maximum angle should be calculated first. The ΔL can be calculated as follows:

$$\Delta L = \pi D_0 \frac{\theta}{360^\circ}, \tag{5}$$

Input $\theta = 135^\circ$ and $D_0 = 36.8mm$ into Equation (5), the ΔL can be calculated as $43.3mm$. Since the maximum shrinkage rate of pneumatic artificial muscle produced by FESTO is about 25%, the needed length of pneumatic artificial muscle is about $200mm$.

The characteristic curve of FESTO-MAS-20-200N-AA-MC-O-ER-EG adopted in the paper is shown in Figure 13. In the picture, "1", "2", "3" and "4" denote maximum output force, maximum working pressure, maximum shrinkage rate and maximum preload respectively.

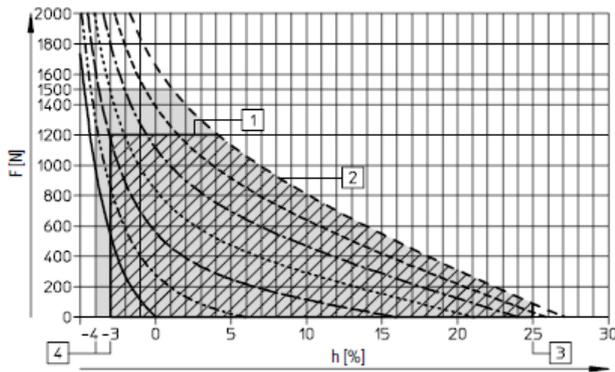


FIGURE 13 The characteristic curves of FESTO-MAS-20-200N-AA-MC-O-ER-EG

3.4 VIRTUAL PROTOTYPE OF ARTIFICIAL LEG

The virtual prototype of artificial leg established using software Pro/E is shown in Figure 14. The hip joint of artificial leg is driven by dc servo motor with harmonic gear reducer. Its advantage is that hip joint has simple transmission mechanism and high transmission efficiency. The ankle joint of artificial leg is driven by dc servo motor with harmonic gear reducer and spur gear transmission. The horizontal structure of ankle joint is more compact and transmission accuracy is better.

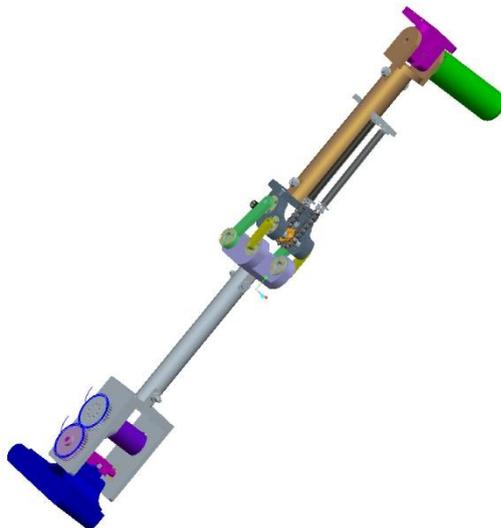


FIGURE 14 Virtual prototype of artificial leg

4 Bionic design and analysis of bionic leg

4.1 BIONIC DESIGN OF BIONIC LEG

The virtual prototype of bionic leg established using software Pro/E is shown in Figure 15. The knee joint of bionic leg is also adopts the same 4-bar multiaxial knee mechanism. Compared with artificial leg, the knee joint of bionic leg is semi-controlled by magneto-rheological damper to adjust rotation performance. MR damper has a series of advantage, such as simple structure, small volume, smart response, low energy demand and large

resistance etc. Prosthetic foot is fixed to BL, so its ankle joint has no degree of freedom.

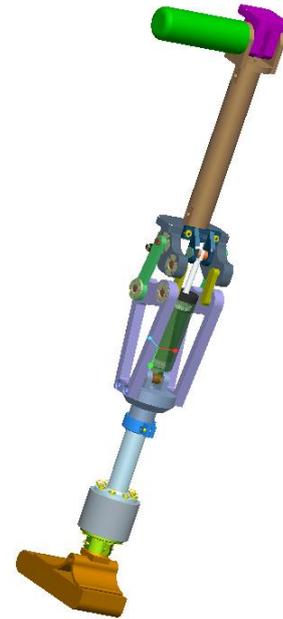


FIGURE 15 Virtual prototype of bionic leg

4.2 PERFORMANCE ANALYSIS OF BIONIC LEG

The parameters which describe stability of bionic leg are shown in Figure 16.

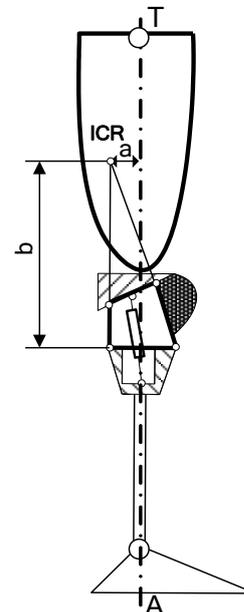


FIGURE 16 The stability parameters of bionic leg.

Here "a" denotes the distance between knee joint ICR and load line at full extension position. "b" denotes the distance between knee joint ICR and hinge point of down bar and shank along load line at full extension position. The static stability of bionic leg is described using γ which denotes bending angle of knee joint when "a" is equal to zero. It can be obtained from Figure 15,

$a = 35.28\text{mm}$, $b = 293.71\text{mm}$, $\gamma = 4.74^\circ$. The stance phase stability of 4-bar mechanism is very good.

The dynamic stability of bionic leg is determined by the relative position of knee joint ICR and load line. According to flexibility of swing phase, the prosthetic knee joint of four bar mechanism can be distinguished as three categories, including absolute stable, coordinated and flexible. Absolute stable 4-bar knee joint refers that its ICR always after load line, whether the heel touches or tiptoe leaves the ground. Thus the knee bend is difficult and gait is unnatural during transition from support phase to swing phase. But it has strong stability and is suitable for patients with weak muscle strength of residual limb. Coordinated 4-bar knee joint refers that its ICR locates stable coordination area and has higher position. It has relatively good stability and doesn't need big hip joint torque to ensure ICR in stable coordination area. So it is suitable for patients who need not only better stability, but also certain flexibility under weak muscle strength of residual limb. The ICR of flexible 4-bar knee joint is also locates stable coordination area. But it has less distance and higher position relatively. So its stability is less than that of coordinated 4-bar knee joint and needed hip joint torque is higher. But it has good flexibility and can easily realize knee bend during support telophase. It is suitable for patients with strong strength of residual limb and sport prosthesis.

For 4-bar knee joint designed in the paper, the relationship between ICR and load line from initial phase to telophase of support is shown in Figure 8. It can be seen that it belongs to flexible 4-bar knee joint.

5 Motion simulation of BRHL

The software of ADAMS integrates with the latest theory results of multi-body dynamics, a variety of convenient modelling tools, efficient solver, powerful post-processing modules and visual interfaces. It can automatically establish and solve equations of system model which is almost impossible to be achieved manually. So ADAMS can be used for system simulation of particularly complex robot system. In the paper, the virtual prototype of BRHL is established using Pro/E and imported into ADAMS. The virtual prototype of simplified BRHL is shown in Figure 17. The auxiliary vehicle is used to keep walking stability of BRHL.

Based on virtual prototypes above, the motion simulation of BRHL is done by adding joint angle curves which are previous planned. The continuously walking simulation of BRHL under flat environment is shown in Figure 18. The simulation indicates that BRHL can simulate amputees with intelligent prosthesis well and is an ideal test-bed for IBL.

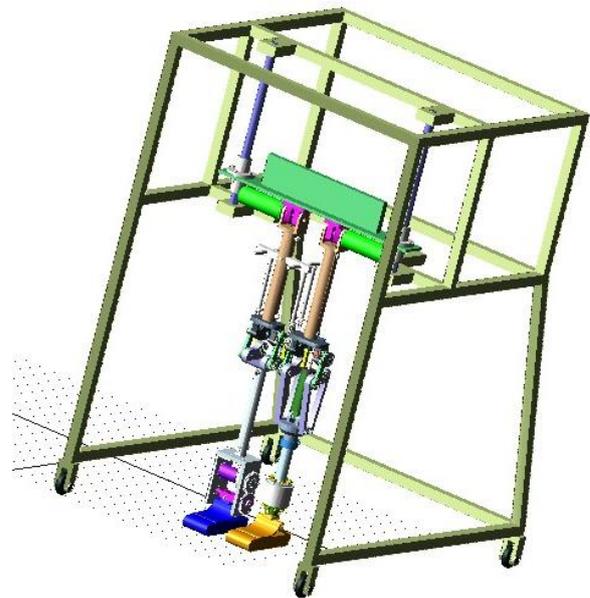


FIGURE 17 Virtual prototype of simplified BRHL

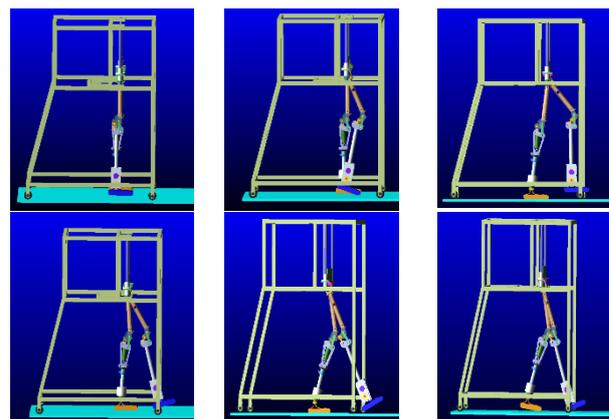


FIGURE 18 Continuously walking simulation of BRHL

6 Conclusions

The ICR of 4-bar multiaxial knee joint is similar to "J" curve and is consistent with that of human leg. The artificial leg driven by PMA can simulate the control mode of human knee joint by antagonistic muscles. IBL semi-controlled by magneto-rheological damper which consume less energy has good stability and flexibility. BRHL has good humanoid characteristics and is an ideal test-bed for IBL.

Acknowledgments

Financial supports from National Natural Science Foundation of China (ID 51105070), the Fundamental Research Funds for the Central Universities (ID N120403002), Scientific Study Project of Liaoning Province Education Department (ID L2013048) are highly appreciated.

References

- [1] Gong S Y, Yang P, Song L and Chen L L 2011 *Journal of Hebei University of Technology* **40**(2) 6-9
- [2] Ziegler G K, MacKenzie E J, Ephraim P L, Trivison T G, Brookmeyer R 2008 *Archives of Physical Medicine and Rehabilitation* **89**(3) 422-9
- [3] Hoover C D, Fulk G D, Fite K B 2012 *ASME DC Journal of Medical Devices* **6**(1) 011005-1
- [4] Obe S Z, Sykes A, Lang S, Cullington I 2005 *Robotica* **23**(3) 337-344
- [5] Shandiz M A, Farahmand F, Osman N A, Zohoor H 2013 *International Journal of Advanced Robotic Systems* **10** 1-10
- [6] Xie H L, Wang B R, Cong D H, Xu X H 2006 *Journal of System Simulation* **18**(11) 3234-3237
- [7] Poliakov O M, Lazarev V B, Chepenyuk O O 2013 *Journal of Rehabilitation Robotics* **1**(2) 109-23
- [8] Xie H L, Xu Z W, Shi J S, Xu X H 2008 *Journal of Northeastern University (Natural Science)* **29**(10) 1470-1473
- [9] Wang B R 2005 *Research and Development of Biped Robot with Heterogeneous Legs* Cambridge Univ. Press: Shenyang, chapter 2

Authors



Hualong Xie

Current position, grades: Associate professor at Northeastern University since 2010.

University studies: bachelor degree in Mechanical Electronic Engineering, Master degree in Mechanical Design and Theory, PhD degree in Control Theory and Control Engineering from Northeastern University, China, in 2000, 2003 and 2006, respectively.

Scientific interest: robot, intelligent control, intelligent bionic leg and biomechanics.



Nan He

Current position, grades: postgraduate at School of Mechanical Engineering & Automation in Northeastern University, China.

University studies: Bachelor degree in Mechanical Design, Manufacturing and Automation from Liaoning University of Technology.

Scientific interest: robot and intelligent control.



Fei Li

Current position, grades: Lecturer in the School of Information Science & Engineering at Shenyang University of Technology, China.

University studies: bachelor degree in measurement and control technology and instrumentation, master degree in control theory and control engineering from Shenyang Institute of Chemical Technology, China, in 2001 and 2004, respectively. The Ph.D. degree in pattern recognition and intelligent system from Northeastern University, China, in 2007.

Scientific interest: robotics, control theory and computer vision.

Influence of liquid physical properties on liquid film flow characteristics of uneven wall

Mei Liu, Songling Wang, Zhengren Wu*

School of energy power and mechanical engineering, North China Electric Power University Baoding, China, 071003

Received 1 August 2014, www.tsi.lv

Abstract

Two-dimensional model of inclined uneven wall was established based on VOF method to numerically simulate the flow characteristics of liquid film. The impact of physical property on the flow field is studied. Water, acetone and ethyl alcohol were selected as the medium. The results indicate that liquid film thickness increases with liquid viscosity, while the phase difference between free surface and the uneven wall has no change. The continuous uniform film is easy to form while considering the surface tension. Furthermore, the phase difference and liquid film thickness both increase when taking the surface tension into account.

Keywords: uneven wall, liquid film, flow characteristics, physical properties, VOF method

1 Introduction

Liquid film flow is a common phenomenon in nature. It plays a pivotal role in the equipment of traditional industries and high-tech areas due to the advantages of high heat and mass transfer coefficient. The most important factors in the design of these devices are the thickness and the flow rate of the liquid film, which are the critical parameters affect the heat and mass transfer characteristics. Therefore, the investigation of the flow and heat transfer characteristics of liquid film has become an increasingly important research topic.

Over decades, plenty of theoretical and experimental research in the field of film has been done. Alexandre [1] studied the stability of liquid film flowing down an inclined wavy plane based on the finite element method. Ye [2] discussed the influence of interfacial shear stress and Reynolds number on the linear stability of liquid film surface wave on inclined wall based on boundary layer model. Hu [3] simulated a uniformly heated film flowing down an inclined plate by using Lagrangian finite element method, and the saturated periodic wave, quasiperiodic wave, multi-peaked wave and solitary hump wave were received. Pak [4] analysed the influences of the electric field and the wavy structures by using the weighted-residual integral boundary-layer model. On the other side, some scholars have experimentally measured the liquid film surface wave velocity, wavelength, frequency and other parameters and studied the evolution of film and film-wave instability development by high-speed digital camera technology [5], high-frequency capacitance-type water film

thickness [6], three-dimensional laser Doppler velocimetry (LDA) [7, 8].

As the development of computer technology and calculation methods, the adoption of numerical simulation method to investigate liquid film flow characteristics and heat and mass transfer performance has become a hot topic of intensive interests. Beata [9] established a two-phase counter current structured packing model to determine the effect of liquid and gas flow rates and physical properties of the flowing liquids on the interfacial area. Liu [10] computationally simulated the flow field of falling film down a vertical plate at high Reynolds number by using volume of fluid (VOF) method. Gu [11] also used the VOF method to investigate the gas-liquid two-phase liquid film flow and to study influence of the structure and liquid-phase flow rate. Sun [12] used the Fluent software for the purpose of consideration of corrugated plate film.

This thesis uses CFD software Fluent and based on the VOF method to simulate the liquid film flow on inclined uneven wall. The impact of physical properties on liquid film formation and development process is investigated to provide theoretical basis for the utility, optimization and development of liquid film.

2 Models and boundary conditions

2.1 PHYSICAL MODEL

The physical model in this paper is shown in Figure 1 and physical properties of each kind of liquid can be seen in Table 1.

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

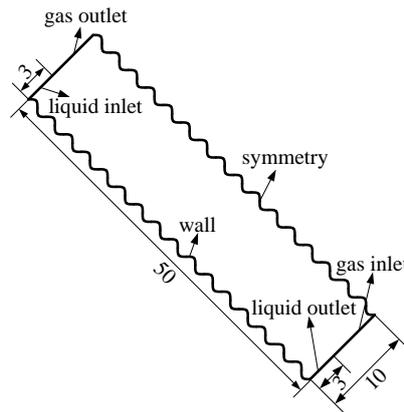


FIGURE 1 Physical model and boundary conditions

TABLE 1 Physical properties of liquid phase

fluid	ρ_L [kg/m ³]	μ_L [Pa·s]	σ [N·m ⁻¹]	wavy wall θ_w [°]
water	998.2	0.001003	0.0722	57
acetone	791.0	0.00033	0.0229	40
ethyl-alcohol	790.0	0.00120	0.0214	40

2.2 GOVERNING EQUATIONS

Liquid film on uneven wall is the gas-liquid two-phase counter current process and VOF method is used to track the free surface.

(1) Continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{u} = 0. \tag{1}$$

(2) Momentum equation:

$$\frac{\partial}{\partial t} \rho \mathbf{u} + \nabla \cdot \rho \mathbf{u} \mathbf{u} = -\nabla p + \rho \mathbf{g} + \mathbf{F}_s + \nabla \cdot [\mu \nabla \mathbf{u} + \nabla \mathbf{u}^T] \tag{2}$$

(3) Volume fraction continuity equation:

$$\frac{\partial \alpha_q}{\partial t} + \mathbf{u} \cdot \nabla \alpha_q = 0. \tag{3}$$

Here, ρ and μ respectively represents density and dynamic viscosity in each cell, \mathbf{u} is velocity vector, \mathbf{g} is gravitational acceleration, p is the pressure, t is time, \mathbf{F}_s is the body force due to surface tension, and α_q is the volume fraction of the q^{th} phase. The volume fraction of each phase satisfies:

$$\sum_{q=1}^n \alpha_q = 1. \tag{4}$$

Physical properties in governing equations of two-phase flow such as the density and viscosity can be described as:

$$\rho = \alpha_L \rho_L + 1 - \alpha_L \rho_G, \tag{5}$$

$$\mu = \alpha_L \mu_L + 1 - \alpha_L \mu_G. \tag{6}$$

In this paper, the surface tension model in simulation is the CSF (Continuum Surface Force) model, which the surface tension is attached to the VOF calculations as the source term \mathbf{F}_s , it can be expressed as:

$$\mathbf{F}_{VOL} = \sigma_{ij} \frac{\rho \kappa_i \nabla \alpha_i}{\frac{1}{2} (\rho_i + \rho_j)}, i, j = L, G, \tag{7}$$

where σ_{ij} is the surface tension coefficient, κ_i represents the interface curvature and it is the divergence of the unit normal vector on free surface:

$$\kappa = \nabla \cdot \hat{n}. \tag{8}$$

The force due to surface tension between the liquid and gas phase is called wall adhesion. We can incorporate it into the calculation of surface tension, and the normal vector at the wall is:

$$\hat{n} = \hat{n}_w \cos \theta_w + \hat{t}_w \sin \theta_w, \tag{9}$$

where the contact angle θ_w is the angle between the wall and the interface tangent, \hat{n}_w and \hat{t}_w is the unit normal vector and unit tangent vector at the phase interface.

2.3 BOUNDARY CONDITIONS

Assume that at the initial time the domain is full of stationary gas, that is, $t=0$, $\alpha_L=0$, $\alpha_G=1$. The wall is no slip and adiabatic, the gas and liquid inlets both are velocity inlet boundary conditions and the outlets are pressure outlet, the outlet pressure is atmospheric pressure, and the edge opposite to the wall is set to symmetric boundary.

3 Results and Analysis

We mainly study the influence of viscosity and surface tension on liquid film flow characteristics.

3.1 INFLUENCE OF VISCOSITY

From Table 1 we know that there is less difference of the density and surface tension between acetone and ethyl alcohol, but the viscosity difference is large. Figure 2 and Figure 3 respectively shows the acetone and ethyl alcohol film formation process on inclined corrugated wall *a* at the inlet velocity of 0.05 m/s. The larger viscosity ethyl alcohol flows longer than acetone during the same time due to the viscosity retardation. Therefore, the larger fluid viscosity, the shorter distance of film flows.

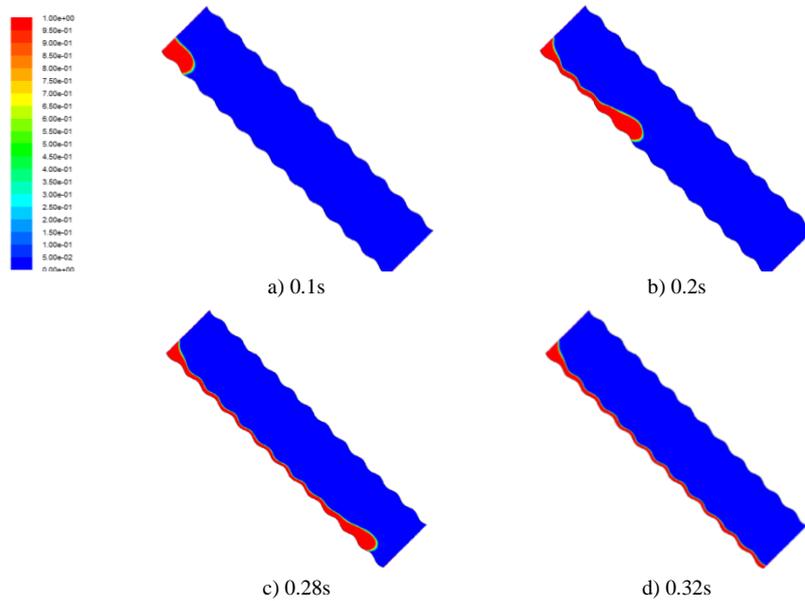


FIGURE 2 Film of acetone flow on corrugated wall *a*

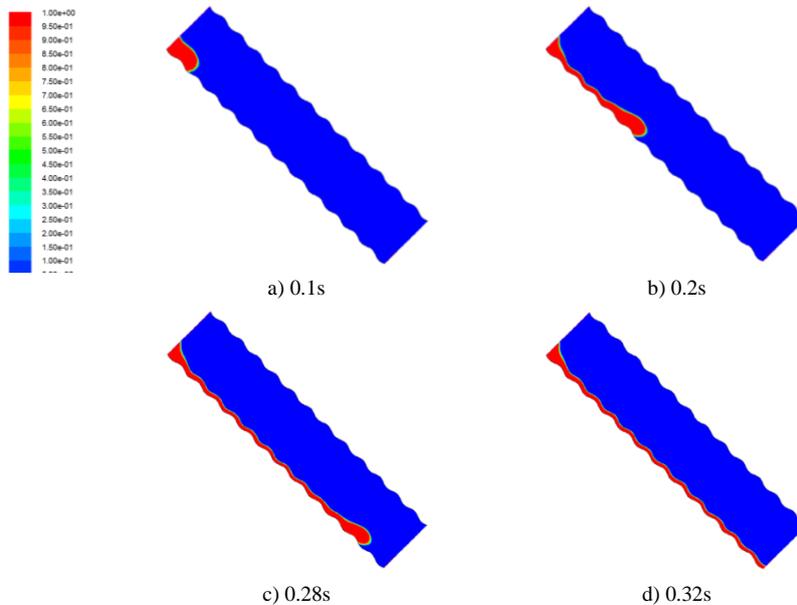


FIGURE 3 Film of ethyl alcohol flow on corrugated wall *a*

The stable velocity vector on corrugated wall *a* of the two kinds fluid are presented in Figure 4. We can see from the figure that the overall speed of the film decreases when the fluid viscosity increases. That is due to the fluid friction loss increases with viscosity, thus the overall film speed reduces. Figure 5 shows the free surface of acetone and ethyl alcohol film at the inlet velocity of 0.05m/s. It indicates that the phase of

the free surface film hardly changes while the viscosity increases, but the film thickness increases. It is probably due to the larger viscosity fluid can stay longer on the wall surface. Therefore, the fluid film thickness increases with the viscosity, while the phase difference between the free surface and the wall does not change.

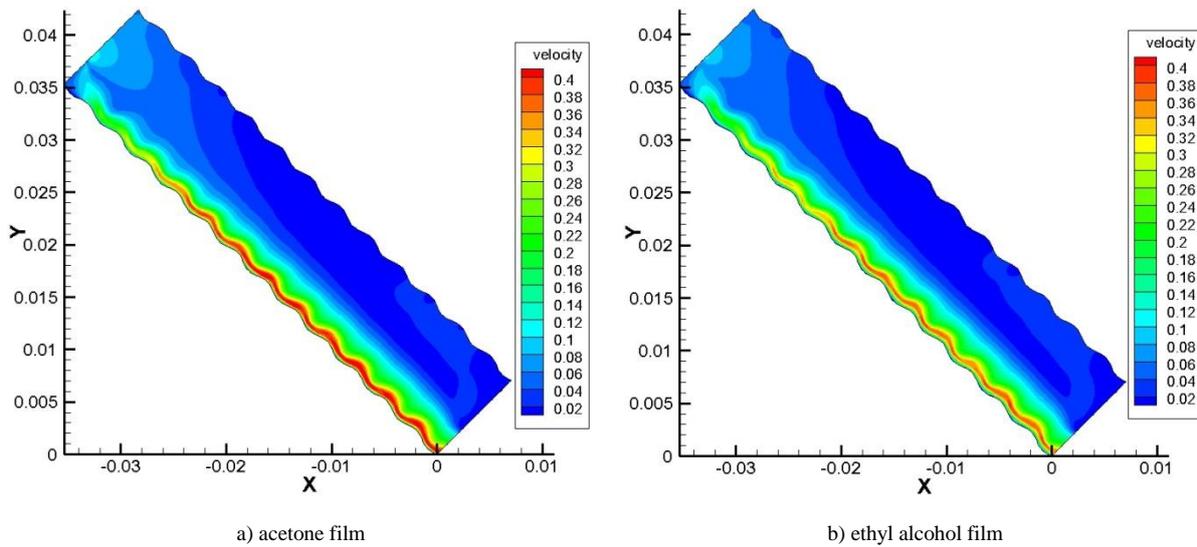


FIGURE 4 Velocity vector of acetone and ethyl alcohol film on corrugated wall *a*

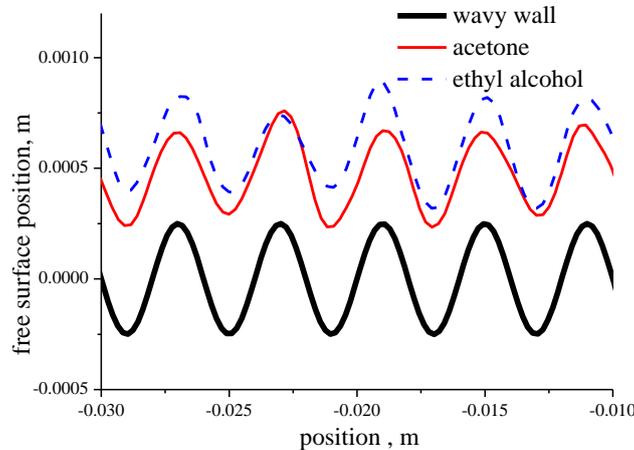


FIGURE 5 Influence of viscosity on free surface

3.2 INFLUENCE OF SURFACE TENSION

Figure 6 shows the distribution of the liquid film on the wavy wall *a* with and without surface tension while the inlet-velocity is 0.1 m/s. It is clear that a uniform film could not form without considering the surface tension.

The gas mixed between the wall and film would reduce the contact area, thereby the heat transfer efficiency reduces. While considering the surface tension, although there would be phase difference, the continuous and uniform film still forms.

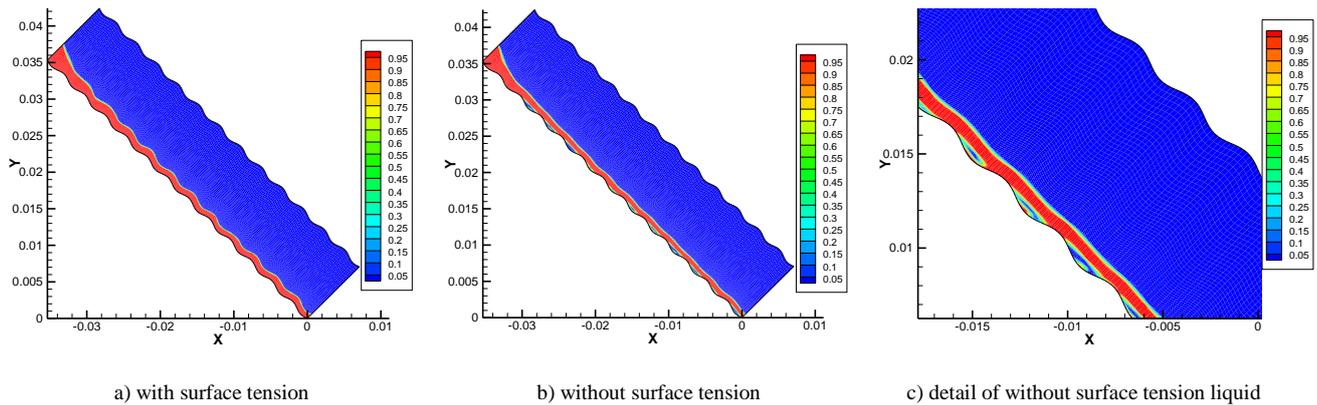


FIGURE 6 Influence of surface tension on liquid film distribution

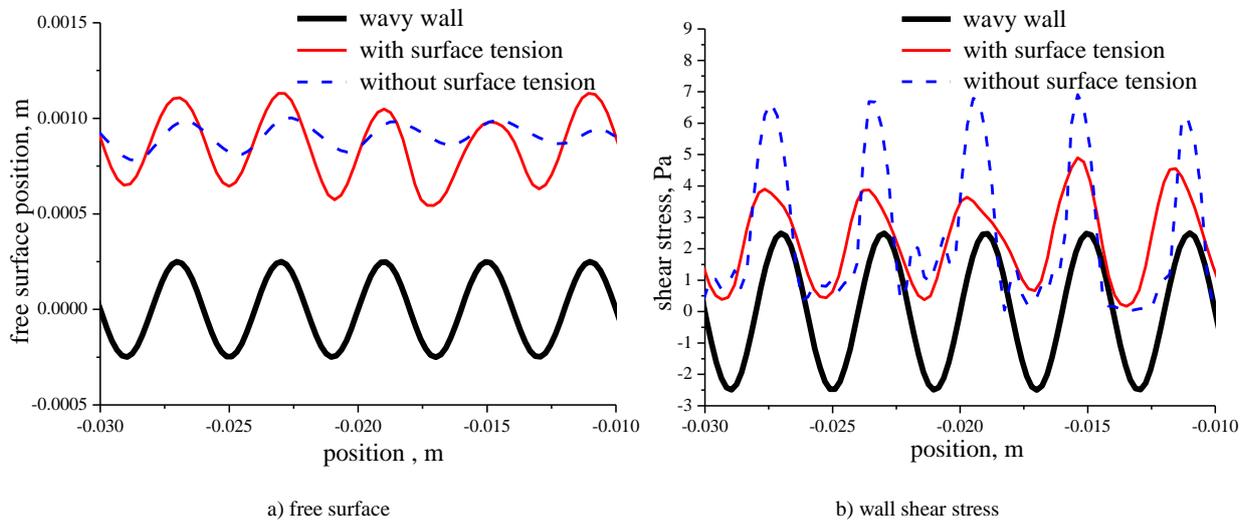


FIGURE 7 Influence of surface tension

Figure 7 respectively shows the free surface and the wall shear stress under the two different conditions. Obviously, the phase difference reduces and the liquid film almost immediately changes follow the wall while considering the surface tension. The liquid film thickness somewhat decreases while compared with the condition of considering surface tension. The change of the phase and film thickness would cause changes in wall shear stress, as Figure 7(b) shows. While considering the surface tension, the position of the maximum shear stress moves from the wave crest to the wave trough, and the amplitude of shear stress reduce compared with the condition of no-considering surface tension.

The above results indicate that the surface tension plays a very important role while handling the problem of a thin liquid film, and its variation would even change the structure of the whole flow field. Therefore, the surface tension must be considered.

4 Conclusions

This paper numerically simulates liquid film flow on uneven wall and investigates the impact of fluid properties on liquid film flow characteristics. The simulation results show that: The film thickness increases and overall film velocity reduces with the viscosity of fluid, but it does not affect the phase difference. It is not easy to form a uniform film without surface tension, and wall shear stress and the phase difference will change.

Acknowledgment

The project was supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (20110036110009).

References

- [1] Ern A, Joubaud R, Lelièvre T 2011 Numerical study of a thin liquid film flowing down an inclined wavy plane *Phys D: Nonlinear Phenom* 1714-23
- [2] Ye X-M, Zhang Y, Wang S-L 2007 Investigation on Stability of Sheared Wavy Liquid Films *Proceedings of the CSEE* 103-6
- [3] Hu J, Hu G-H, Sun D-J 2004 Direct Simulation of a Uniformly Heated Film Flowing Down an Inclined Plate *Chinese Journal of Computational Physics* 415-20 (in Chinese)
- [4] Pak M, Hu G-H 2011 Influences of a uniform electric field on the stability of film flows along an inclined wavy wall *Chinese Journal of Hydrodynamics* 368-76 (in Chinese)
- [5] Lu C, Duan R, Jiang S 2008 Experimental study of flow instabilities of falling films *J. Tsinghua Univ (Sci & Tech)* 107-9 (in Chinese)
- [6] Song G J, Hu P, Wei S, et.al 2012 Experimental Study on Surface-Wave Evolution Behavior of Water Film Falling Down Vertical Plate *Atomic Energy Science and Technology* 804-8
- [7] Liu C, Cheng J, Zhang W 2007 LDV Measurements of Liquid Velocity Profiles of Bubbly Flow in Structured Packing *Chemical Industry and Engineering Progress* 92-6
- [8] Yu L, Zeng A, Yu G 2006 The Study on the Liquid Velocity Fluctuation and the Mass Transfer in the Gas-Liquid Falling Film *Journal of Chemical Engineering of Chinese Universities* 696-701 (in Chinese)
- [9] Szulcowska B, Zbicinski I, Gorak A 2003 Liquid Flow on Structured Packing CFD Simulation and Experimental Study *Chemical Engineering and Technology* 580-4
- [10] LIU Y, TONH Y, RenHaigang 2010 Numerical Investigation on Flow characteristics of Falling Water Film Down a Vertical Plate at High Reynolds Number *Infrared Technology* 567-71
- [11] GU F, 2004 CFD Simulations of the Local-Flow and Mass-Transfer in the Structured Packing Tianjin: *Tianjin University (in Chinese)*
- [12] SUN F 2012 Numerical simulation on liquid flow characteristics of the wavy plate *Modern Chemical Industry* 90-3 (in Chinese)

Authors	
	<p>Mei Liu</p> <p>Current position, grades: Lecturer at the North China Electric Power University.</p> <p>University studies: Master of management at the North China Electric Power University 2002-2005, Doctor student at the North China Electric Power University, Thermal energy engineering, 2012.</p> <p>Scientific interest: fluid mechanics, the technology economy and management, energy management and control, low carbon economy.</p>
	<p>Songling Wang</p> <p>Current position, grades: Professor at the Energy & Power Engineering School, North China Electric Power University.</p> <p>Scientific interest: fluid dynamics.</p> <p>Publications: more than 200 scientific papers and 3 books.</p>
	<p>Zhengren Wu</p> <p>Current position, grades: lecturer at the Energy & Power Engineering School, North China Electric Power University.</p> <p>Scientific interest: The characteristic analysis of the film on the different wall. The research on the drag reduction characteristics of the different structure.</p> <p>Publications: 20.</p>

Multifractal analysis on gene and PPI networks

Danling Wang¹, Yanfei Wang^{2*}

¹*School of Mathematics and Physics, University of Science and Technology, Beijing, China*

²*School of Sciences China Agricultural University Tsinghuadonglu 17, 100083, Beijing, China*

Received 1 August 2014, www.tsi.lv

Abstract

Multifractal analysis is a useful way to systematically describe the spatial heterogeneity of both theoretical and experimental fractal patterns. In this paper, we introduce a new box-covering algorithm to compute the generalized fractal dimensions of complex networks. We apply our method on networks built on disease-related gene microarray data and PPI networks. For each microarray data, we compare the difference of multifractal behaviour between gene networks that reconstructed from patients and normal micorarrays. The result suggests that multifractality exists in all the gene networks we generated and the differences in the shape of the D_q curves are obvious for all microarray data sets. Meanwhile, multifractal analysis could provide a potentially useful tool for gene clustering and identification between healthy people and patients. For the analysis of PPI networks, the results support that the algorithm is a suitable and effective tool to perform multifractal analysis of complex networks, and this method can be a useful tool to cluster and classify real PPI networks of organisms.

Keywords: multifractal analysis, self-similarity, gene networks, PPI networks

1 Introduction

Complex networks have been studied extensively due to their relevance to many real-world systems such as the World Wide Web, the internet, energy landscapes, and biological and social systems. After analysing a variety of real complex networks, Song et al. [1] found that they consist of self-repeating patterns on all length scales, i.e., they have *self-similar structures*. In order to unfold the self-similar property of complex networks, Song et al. [1] calculated their fractal dimension, a known useful characteristic of complex fractal sets [2-4], and found that the box counting method is a proper tool for further investigations of network properties. The tools of fractal analysis provide a global description of the heterogeneity of an object. However, this approach is not adequate when the object may exhibit a multifractal behaviour. Multifractal analysis is a useful way to systematically characterize the spatial heterogeneity of both theoretical and experimental fractal patterns. It was initially proposed to treat turbulence data, and has recently been applied successfully in many different fields including time series analysis [5], financial modelling [6], biological systems [768] and geophysical systems [9].

In recent years, bioinformatics has become a more and more notable research field since it allows biologists to make full use of the advances in computer science and computational statistics in analysing the data of an organism at the genomic, transcriptomic and proteomic levels [10]. DNA technology, i.e. microarray of large sets of nucleotide sequences, is a modern tool that is used to obtain information about expression levels of thousands

of genes simultaneously. The gene networks built based on microarray data become a popular research field.

In this paper, we aim to compare the difference of multifractal behaviours between gene networks that reconstructed from patients and normal people microarrays and some PPI networks. However, work in such high dimensional and large data is extremely difficult. So for gene microarrays, firstly, we apply *Fuzzy Membership test* (FM-test) [11] to get the most important genes that are related with the disease; then we construct networks based on the microarray data of the selected genes by calculating the correlation coefficient. Next we apply the modified fixed-size box-covering method on them to detect their multifractal behaviours. For PPI networks, we firstly use cytoscape to we need to find the largest connected part of each data set, and then we adopt our method to check multifractal characteristics in different organisms. Secondly, we also we randomly chose several sub-networks from different parts of the human PPI network with the same nodes and compare multifractal characteristic between them.

2 Methods

The most common algorithms of traditional multifractal analysis are the fixed-size box-counting algorithms [12]. For a given measure μ with support E in a metric space, we consider the partition sum:

$$Z_\varepsilon(q) = \sum_{\mu(B) \neq 0} [\mu(B)]^q, q \in \mathbb{R},$$

* *Corresponding author* e-mail: yfmu@sina.com

where the sum is evaluated over all different nonempty boxes B of a given size ε in a grid covering of the support T . The exponent $\tau(q)$ is defined by:

$$\tau(q) = \lim_{\varepsilon \rightarrow 0} \frac{\ln Z_\varepsilon(q)}{\ln \varepsilon}$$

and the generalized fractal dimensions of the measure are defined as:

$$D(q) = \tau(q) / (q - 1), \text{ for } q \neq 1,$$

and:

$$D(q) = \lim_{\varepsilon \rightarrow 0} \frac{\ln Z_{1,\varepsilon}}{\ln \varepsilon}, \text{ for } q = 1,$$

where $Z_{1,\varepsilon} = \sum_{\mu(B) \neq 0} \mu(B) \ln \mu(B)$. The generalized fractal dimensions are numerically estimated through a linear regression of $(\ln Z_\varepsilon(q)) / (q - 1)$ against $\ln \varepsilon$ for $q \neq 1$, and similarly through a linear regression of $Z_{1,\varepsilon}$ against $\ln \varepsilon$ for $q = 1$. The $D(q)$ corresponding to negative values of q deal with the structure and the properties of the regions where the measure value is small.

Our group proposed a new box-covering algorithm to compute the generalized fractal dimensions of network [13]. For a network, we denote the matrix of shortest path lengths by $B = (b_{ij})_{N \times N}$, where b_{ij} is the length of the shortest path between nodes i and j . Then we use $B = (b_{ij})_{N \times N}$ as input data for multifractal analysis based on our fixed-size box counting algorithm as follows:

- a) Initially, all the nodes in the network are marked as uncovered and no node has been chosen as a seed or centre of a box.
- b) Set $t=1, 2, \dots, T$ appropriately. Group t nodes into T different ordered random sequences. More specifically, in each sequence, nodes which will be chosen as seed or centre of a box are randomly arrayed.

Remark: T is the number of random sequences and is also the value over which we take the average of the partition sum $\bar{Z}_r(q)$. In this study, we set $T=1000$ for all the networks in order to compare them.

- c) Set the size of the box in the range $r \in [1, d]$, where d is the diameter of the network.

Remark: When $r=1$, the nodes covered within the same box must be connected to each other directly. When $r=d$, the entire networks could be covered in only one box no matter which node was chosen as the centre of the box.

- d) For each centre of a box, search all neighbours within distance r and cover all nodes which are found but have not been covered yet.
- e) If no newly covered nodes have been found, then this box is discarded.

- f) For the nonempty boxes B , we define their measure as $\mu(B) = N_B / N$, where N_B is the number of nodes covered by the box B , and N is the number of nodes of the entire network.

- g) Repeat (d) until all nodes are assigned to their respective boxes.

- h) Repeat (c) and (d) for all the random sequences, and take the average of the partition sums $\bar{Z}_r(q) = (\sum^t Z_r(q)) / T$, and then $\bar{Z}_r(q)$ for linear regression.

Linear regression is an essential step to get the appropriate range of $r \in [r_{\min}, r_{\max}]$ and to get the generalized fractal dimensions D_q . In our approach, we run the linear regression of $[\ln \bar{Z}_r(q)] / (q - 1)$ against $\ln(r/d)$ for $q \neq 1$, and similarly the linear regression of $\bar{Z}_{1,r}$ against $\ln(r/d)$ for $q=1$, where $\bar{Z}_{1,r} = \sum_{\mu(B) \neq 0} \mu(B) \ln \mu(B)$ and d is the diameter of the network.

3 Results and discussion

3.1 MULTIFRACTAL ANALYSIS OF GENE NETWORKS

Four different gene microarray data sets are used in our work: Colorectal cancer gene microarray data [16]; Type II diabetes gene microarray data [17]; Type I diabetes gene microarray data [18] and Lung cancer gene microarray data [11]. There are two parts in each data, the first part consists of genes expression values from patients or drugs sensitive people; the second part consists of genes expression values from healthy donators or patients after medication or treatment. For each original data, we firstly use FM-test [11] to select around 2000 genes which are most possibly related with disease, then build patients gene networks and normal people gene networks respectively. For colorectal data, CP is the patient gene networks, HP is healthy people gene network; for type II diabetes data, IR is insulin resistant people network and IS is insulin sensitive people network; for type I diabetes data, TP is patients network and TM is patients after medication network; for lung cancer data, LP is patients network and LN is normal people network. We analysed multifractal behaviour of two networks for each microarray data. All the networks are full connected.

We calculated the D_q spectra for these gene networks of different datasets and then summarize the numerical results in Table 1 including the number of nodes (N), number of the threshold t , maximum value of D_q , limit of D_q , and ΔD_q . Figures 1 and 2 show that the generalized fractal dimension D_q of these gene networks are decreasing functions of q and multifractality exists in these networks. From the table and figures we see, multifractal characteristic exists in all the gene networks

we analysed. Meanwhile, the D_q curves of gene networks from normal people are mostly higher than the ones from

patients, especially for the first three microarray data.

TABLE 1 Numerical results on gene networks

Networks	N	t	Max D_q	Lim D_q	ΔD_q
CP	2000	0.72	3.44	1.20	2.24
CH	2000	0.85	3.22	2.23	0.99
IR	2304	0.95	2.34	1.47	0.87
IS	2299	0.95	2.14	1.08	1.06
TP	2000	0.72	3.39	1.3	2.05
TH	2000	0.81	3.36	2.20	1.16
LC	2000	0.97	2.36	1.91	0.45
LN	2000	0.96	2.43	1.90	0.53

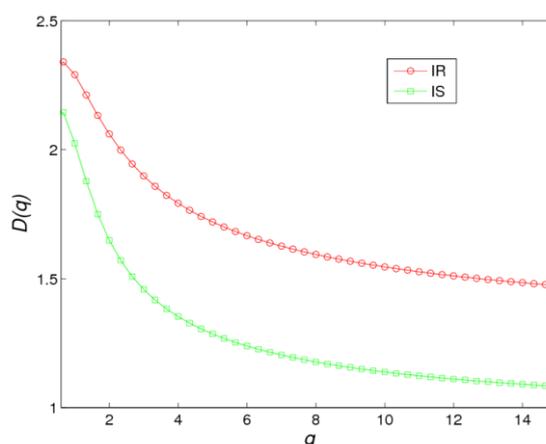
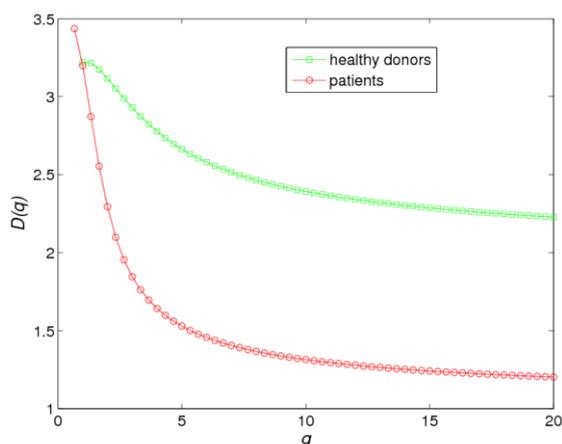


Figure 1 a) Colorectal cancer microarray data, b) Type II diabetes microarray data

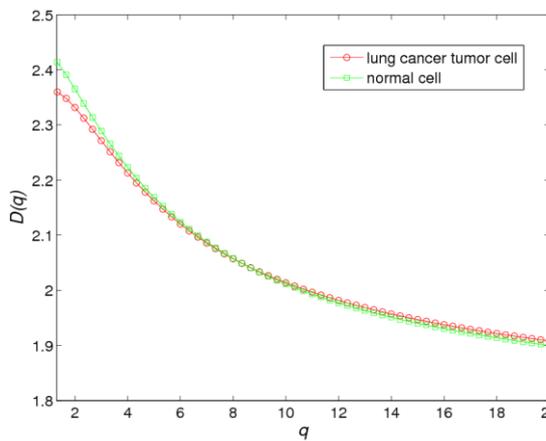
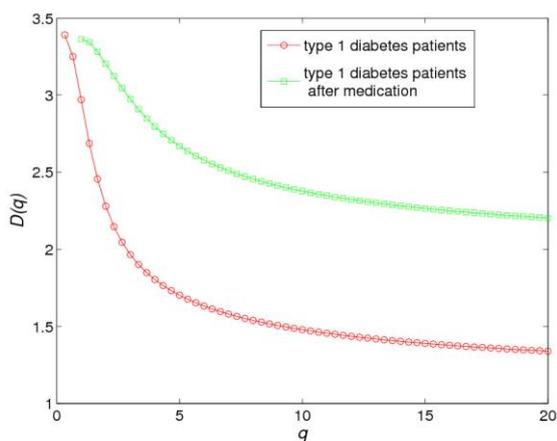


Figure 2 a) Type I diabetes microarray data; b) lung cancer microarray data

3.2 MULTIFRACTAL ANALYSIS OF PPI NETWORKS

The protein-protein interaction data we used here are mainly downloaded from two databases: the PPI networks of *Drosophila melanogaster* (fruit fly), *C.elegans*, *Arabidopsis thaliana* (a type of plant) are downloaded from BioGRID. The PPI networks of *S. cerevisiae* (baker's yeast), *E.coli* and *H.pylori* are downloaded from DIP [19]. We also use the same human PPI network data as in Lee and Jung [20].

Our fractal and multifractal analyses are based on connected networks which do not have separated parts or isolated nodes. In order to apply them to protein-protein interaction networks, some preparation is needed in advance. Firstly, we need to find the largest connected part of each data set. For this purpose many tools and methods could be used. In our study, we adopt Cytoscape [21] which is an open bioinformatics software platform for visualizing molecular interaction networks and analysing network graphs of any kind involving nodes and edges. In using Cytoscape, we could get the largest

connected part of each interacting PPI data set and this connected part is the network on which fractal and multifractal analyses are performed.

We calculated the D_q spectra for seven PPI networks of different organisms and summarize the corresponding

numerical results in Table 2 including the number or nodes (N), number of edges (E), diameter of the network (d), maximum value of D_q , limit of D_q , and ΔD_q . These results show multifractality exists in PPI networks.

TABLE 2 Numerical results of Protein-protein interaction networks

Networks	N	E	d	Max D_q	Lim D_q	ΔD_q
Human	8934	41341	14	4.89	2.65	2.24
D.melanogaster	7476	26534	11	4.84	2.87	1.97
S.cerevisiae	4976	21875	10	4.62	2.48	2.14
E.coli	2516	11465	12	4.15	2.10	2.05
H.pylori	686	1351	9	3.47	1.91	1.56
Arabidopsis thaliana	1298	2767	25	2.51	1.62	0.88
C.elegans	3343	6437	13	4.47	1.49	2.98

TABLE 3 Numerical results of sub-networks of Human PPI

Networks	N	E	d	Max D_q	Lim D_q	ΔD_q
Subnetwork of Human PPI	3505	4651	24	3.65	1.99	1.66
Subnetwork of Human PPI	3505	5262	27	2.97	2.83	0.14
Subnetwork of Human PPI	3505	5353	22	3.95	2.19	1.76
Subnetwork of Human PPI	3505	7055	15	4.22	2.28	1.94
Subnetwork of Human PPI	3505	7509	15	3.55	2.94	0.61
Subnetwork of Human PPI	3505	8750	16	3.81	2.59	1.22
Subnetwork of Human PPI	3505	10652	10	4.02	2.47	1.55

From Figure 3a we could see that all PPI networks are multifractal and there are two clear groupings of organisms based on the peak values of their D_q curves. The first group includes human, Drosophila melanogaster, S.cerevisiae, and C.elegans. The second group just includes two bacteria E.coli and H. pylori. We could also see that the PPI networks of the seven organisms have similar shape for the D_q curves. They reach their peak values around $q = 2$, then decrease sharply as $q > 2$ and finally reach their limit value when $q > 10$. So we can take $\lim D_q = D(20)$ and use $\Delta D_q = \max D_q - \lim D_q$ to verify how the D_q function changes along each curve.

Then we randomly chose several sub-networks from different parts of the human PPI network. These sub-networks all contain 3505 nodes and different numbers of edges. Since these sub-networks are chosen randomly,

overlapping between them is allowed. Then we calculated the D_q spectra for sub-networks of human protein-protein interaction network [20] and summarize the corresponding numerical results in Table 3 including the number or nodes (N), number of edges (E), diameter of the network (d), maximum value of D_q , limit of D_q , and ΔD_q . These results show multifractality exists in PPI networks.

From Figure 3b we could see that not all parts of a PPI network have the same multifractal behaviour. More specifically, among these sub-networks, the ΔD_q values vary from one to another which means that the edge distribution of some parts of a network is symmetric while that of the other parts may not be. This may help to understand the diversity and complexity of protein-protein interactions.

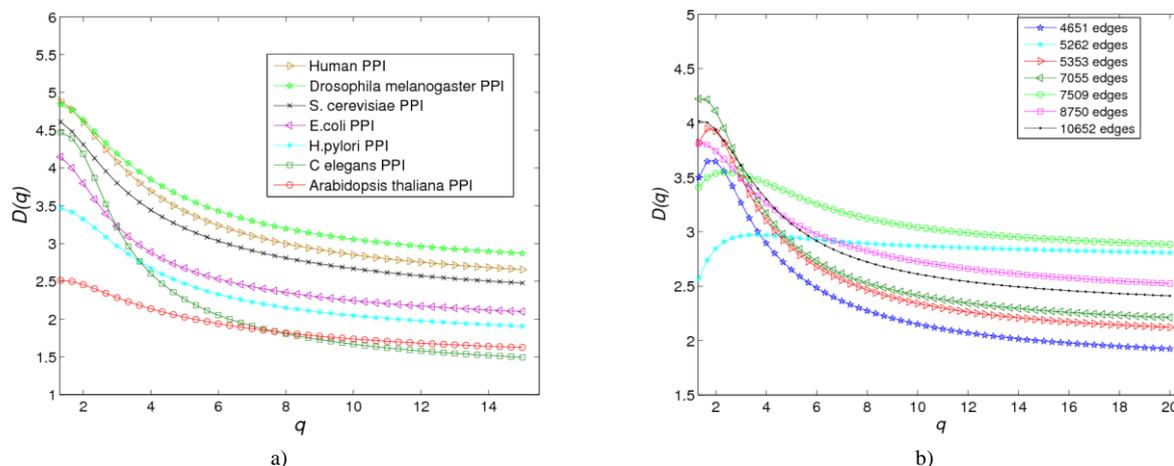


Figure 3 (a) The D_q curves for PPI networks; (b) The D_q curves for sub-networks of human PPI networks

4 Conclusions

A modified algorithm for analysing the multifractal behaviours of complex networks is introduced in this paper. We apply this modified fixed-size box-covering method on gene networks reconstructed from patients and normal gene microarrays. Firstly, we use the fuzzy membership test to get the most important genes that related with the disease; then we construct networks based on the microarray data of the selected genes by calculated the correlation coefficient. From the results we see, multifractality exists in all the gene networks we generated and the difference in the shape of the D_q curves

are obvious for these microarray datasets. Thus multifractal analysis could provide a potentially useful tool for gene clustering and identification between healthy people and patients. We also apply our method on some PPI networks, these results support that multifractal analysis can be a useful tool to cluster and classify real networks such as the PPI networks of organisms.

Acknowledgment

This work is supported by Chinese Universities Scientific Fund No. 2013XJ010 and the Fundamental Research Funds for the Central Universities No. FRF-TP-13-020A.

References

- [1] Song C, Havlin S, Makse H A 2005 Self-similarity of complex networks *Nature London* **433** 392-5
- [2] Mandelbrot B B 1983 *The Fractal Geometry of Nature Academic Press* New York
- [3] Feder J 1988 *Fractals Plenum Press* New York
- [4] Falconer K 1997 *Techniques in Fractal Geometry Wiley* New York
- [5] Cansessa E 2000 Multifractality in time series *J Phys A: Math Gen* **33** 3637-51
- [6] Anh V V, Tieng Q M, Tse Y K 2000 Cointegration of stochastic multifractals with application to foreign exchange rates *Int Trans Opera Res* **7** 349-63
- [7] Yu Z G, Anh V, Lau K S 2001 Multifractal characterisation of length sequences of coding and noncoding segments in a complete genome *Physica A* **301** 351-61
- [8] Yu Z G, Anh V, Lau K S 2001 Measure representation and multifractal analysis of complete genome *Phys Rev E* **64** 31903
- [9] Kantelhardt J W, Koscielny-Bunde E, Rybski D, Braun P, Bunde A, Havlin S 2006 Long-term persistence and multifractality of precipitation and river runoff records *J Geophys Res* **111** D01106
- [10] Keedwell E, Narayanan 2005 Introduction to Artificial Intelligence and Computer Science, in *Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems John Wiley & Sons Ltd* Chichester UK
- [11] Liang LR, Wang S X, Lu Y, Mandal V, Patacsil D, Kumar D (2006) FM-test: a fuzzy-set-theory-based approach to differential expression data analysis *BMC Bioinformatics* **7** (S4)
- [12] Halsey T C, Jensen MH, Kadanoff LP, Procaccia I, Shraiman B I 1986 Fractal measures and their singularities: the characterization of strange sets *Phys Rev A* **33** 1141-51
- [13] Wang D L, Yu Z G, Anh V 2012 Multifractal analysis of complex networks *Chin Phys B* **21** 080504
- [14] Werhli A V, Grzegorzczak M, Husmeier D 2006 Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian models *Bioinformatics* **22** 2523-31
- [15] Borate B R, Chesler E J, Langston M A, Saxton A M, Voy B H 2009 Comparison of threshold selection methods for micorarray gene co-expression matrices *BMC Research Notes* **2** 240
- [16] Collado M, Garcia V, Garcia J M, Alonso O, Lombardia L, Diaz-Uriarte R, Fernandez L A, Zaballos A, Bonilla F, Serrano M 2007 Genomic profiling of circulating plasma RNA for the analysis of cancer *Clin. Chem.* **53**(10) 1860-3
- [17] Yang X, Pratley R E, Tokraks S, Bogardu C, Permana P A 2007 Microarray profiling of skeletal muscle tissues from equally obese, non-diabet insulin-sensitive and insulin-resistant Pima Indians *Diabetologia* **45** 1584-93
- [18] van Oostrom O, de Kleijn D P V, Fledderus J O, Pescatori M, Stubbs A, Tuinenburg A, Lim S K, Verhaar M C 2009 Folic acid supplementation normalizes the endothelial progenitor cell transcriptome of patients with type I diabetes: a case-control pilot study *Cardiovascular Diabetology* **8** 47
- [19] DIP: <http://dip.doe-mbi.ucla.edu/>
- [20] Lee C Y, Jung S 2006 Statistical self-similar properties of complex networks *Phys Rev E* **73**(6) 066102
- [21] Cytoscape software: <http://cytoscapeweb.cytoscape.org/>

Authors



Danling Wang, born in 1982, Hebei Province, China

Current position, grades: lecturer at the University of Science and Technology Beijing.
University studies: PhD at Queensland University of Technology (2008-2011).
Scientific interest: fractal and multifractal analysis, chaos systems and dynamics.
Publications: 8



Weiwei Zhu, born in September, 1988, Qingdao, China

Current position, grades: lecturer at China Agricultural University.
University studies: PhD at Queensland University of Technology (2008-2011).
Scientific interest: fuzzy sets theory, data mining, machine learning.
Publications: 9

Calculation of microstress in machining distortion of titanium alloy monolithic component based on x-ray diffraction experiment

Yong Yang^{*}, Weiwei Zhu

School of Mechanical Engineering, Qingdao Technological University, Qingdao, China

Received 1 March 2014, www.tsi.lv

Abstract

Machining distortion of titanium alloy monolithic component is closely related to the internal stress of material. In this paper, the various factors causing width effects of X-ray diffraction line were analysed, and the effect of stacking fault on diffraction spectrum was excluded according to the result of TEM experiment. The true diffraction spectrum, which can reflect inside information of Ti6Al4V titanium alloy, was determined using MDI JADE peak shape analysis method. Further, the calculation model of microstress was constructed. Combining with the result of X-ray diffraction experiment, microstress in machining distortion of titanium alloy monolithic component was calculated. This work establishes the foundation for investigating the mechanism of machining distortion of titanium alloy monolithic component.

Keywords: titanium alloy, machining distortion, microstress, diffraction line width effect

1 Introduction

Titanium alloys, specifically Ti6Al4V, are used widely in aerospace industry, which offer favourable mechanical characteristics such as high strength-to-weight ratio, toughness, superb corrosion resistance and biocompatibility [1, 2]. But titanium alloys are also difficult-to-machine materials with considerable manufacturing problems [3]. The distortion of titanium alloy monolithic component due to CNC machining is one of the most striking process problems that exist in the manufacturing process of aerospace parts [4, 5], which greatly impacts the production quality and efficiency and also leads to great economic losses.

During machining process, cutting force, heat, clamping force and initial residual stress in blank cause machining distortion together. While initial residual stress and machining stress are the root reasons leading to machining distortion. So, study on the internal stress is helpful to reveal the machining distortion mechanism of titanium alloy monolithic component [6-8].

According to stress classification method presented by Macherauch, the internal stress can be divided into three categories [9]:

1) the first class internal stress, that is called macro residual stress, contains large numbers of small grains. When the balance between the first class internal stress is broken, the macro dimension change must be caused;

2) the second class internal stress, that is called microstress. When the balance between the second class

internal stress is broken, the macro dimension change can also be caused;

3) the third class internal stress. When the balance between the third class internal stress is broken, the macro dimension change will not be caused.

At present, the researches of macro residual stress are relatively complete, and the measurement methods are also more complete [10-12]. But relatively few researches of microstress have been reported. In this paper, to obtain a more general understanding on machining distortion, the microstress was studied deeply. The titanium alloy Ti6Al4V monolithic component was taken as research object, and the calculation model of microstress was constructed.

2 Calculation model of microstress

When the internal stress exists in the polycrystalline material, the strain must also exist internal strain corresponding, which causes the local distortion and results in changes of its internal structure. These changes are reflected in the diffraction spectrum line, and microstress can be measured by analysing these diffraction informations [13, 14].

By measuring the diffraction line half height to width, and other factors to cause diffraction line widths after separation, it can measure the size of the micro internal stress, which is the theoretical basis of the micro internal stress measurement.

Assuming that diffraction line width is only micro an effect factors, then, assumes that a Crystal surface spacing for d_0 , due to micro internal stress of role,

^{*} *Corresponding author's* e-mail: yangyong913@gmail.com

makes the Crystal surface of surface spacing and d_0 has some deviated, assuming that d_+ and d_- respectively with the material diffraction line width half tall in the diffraction Angle 2θ and 2θ , the mean value of the microstrain average is:

$$\varepsilon_{average} = \left(\frac{\Delta d}{d} \right)_{average} \quad (1)$$

and due to $\Delta 2\theta = 2\theta_+ - 2\theta_- = 2\theta_0 - 2\theta_-$, there is:

$$\beta_{hkl} = 4\Delta\theta, \quad (2)$$

where β_{hkl} is the diffraction linear half maximum strength space corresponding full Angle width.

A ray diffraction profile and maximum intensity corresponding to the full width of premises (width). By using the equation:

$$\varepsilon_{\phi\psi} = \frac{d_{\phi\psi} - d_0}{d_0} = -(\theta_{\phi\psi} - \theta_0) \cot \theta_0. \quad (3)$$

Using relationship with $\frac{\Delta d}{d} = -\cot \theta \Delta \theta$, there is:

$$\left(\frac{\Delta d}{d} \right)_{average} = \varepsilon_{average} = \frac{\beta_{hkl}}{4} \cot \theta_{hkl}. \quad (4)$$

$$\beta_{hkl} = 4\varepsilon_{average} \tan \theta_{hkl}. \quad (5)$$

In the above formula, β_{hkl} is radians, if converted into degrees, there is :

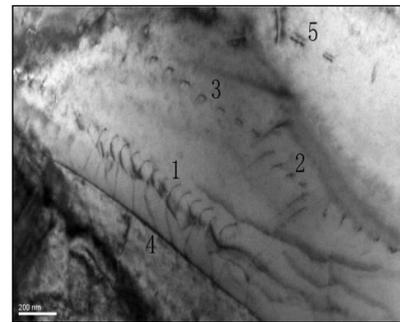
$$\sigma_{average} = E\varepsilon_{average} = E \frac{\pi\beta_{hkl} \cot \theta_{hkl}}{180^\circ \times 4}.$$

The above equation is the calculation model of microstress. But in actual measurement process, instrument etc. system factors, and the crystal block size, stacking fault microscopic internal stress and other materials internal organizational form factors will affect the linear diffraction profiles, the need for specific materials, determine the impact of factors to determine algorithm.

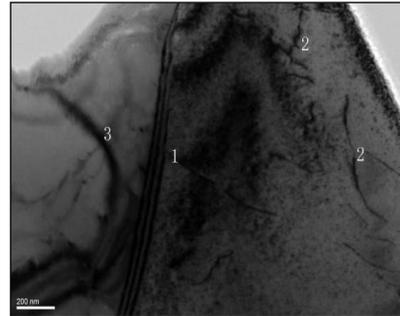
3 Analysis of effect factors of x-ray diffraction line width

3.1 STACKING FAULT OBSERVATION

Experimental material was Ti6Al4V titanium alloy, and the hardness was about 34 HRC. Mechanical thinning (plane grinding), dimpling and ion thinning were performed to achieve the desired TEM sample. The CM200FEG high resolution transmission electron microscopy was used and the microstructure image is shown in Figure 1.



a) Microstructure image along [1213] direction



b) Microstructure image along [001] direction

FIGURE 1 Titanium alloy monolithic component after process two-phase microstructure image

Combining with high resolution electron microscopy image, and Figure 1a identification of 1, 2, 3 place are different forms of dislocation plug product list, 4 place is grain boundary, 5 in dislocation. Figure 1b place 1 for grain boundary, 3 place such as to pour stripe, 2 place two positions for edge dislocation. After observation and analysis, it can be concluded that: Ti6Al4V titanium alloy material cutting internal dislocation motion violent, as dislocation slip one of the main form of the edge dislocation massive exist, in grain boundary place appeared obvious phenomenon of dislocations, and stacking fault there are not obvious. So for diffraction line width influence factors of stacking fault factors, in view of the Ti6Al4V material does not exist, in the follow-up of the solution can be ignored.

3.2 WIDTH EFFECT OF X-RAY DIFFRACTION LINE

For polycrystalline materials, due to the anisotropy, grain size is not the same as. By Edwald graphic method [15], when the crystal block size is relatively large, and each wafer in a crystal plane hkl the corresponding reciprocal approximation for a geometric point, cognate surface hkl in the numerous crystal block corresponding reciprocal points form a thickness reciprocal ball, as in shown in Figure 2a. And the characteristic of the interference function of analysis, diffractive domain (interference function is not zero area, can be understood as the diffraction ball thickness) of the shape and size of the crystal shape and size of a reciprocal relationship, i.e. material central

crystal block size is smaller, diffractive domain is bigger, this resulted in slightly deviated from Prague corner of direction that also exists diffraction, causing diffraction width, as shown in Figure 2b.

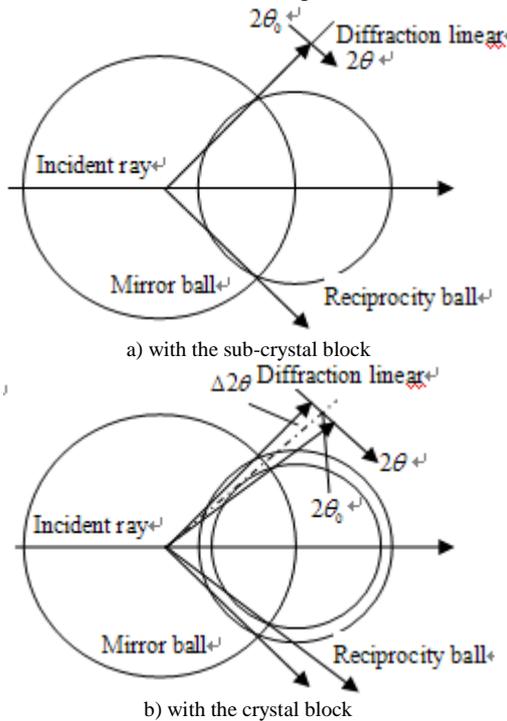


FIGURE 2 Diffraction cone schematic diagram

Therefore, the existence of the sub-crystal block for diffraction line width to influence is very serious. Assumption of X-ray diffraction spectrum line width effect separately from the crystal block size factors, well-known German chemist Debye and his graduate student Scherrer is deduced the Scherrer formula can be used for calculation of the crystal block width effect, and the single crystal and polycrystalline are applicable.

$$D_{hkl} = \frac{0.89\lambda}{\beta_{hkl} \cos \theta_{hkl}}, \tag{7}$$

where $D_{hkl} = Nd_{hkl}$ is the crystal block size, half tall wide β_{hkl} units for radian.

Based on above research, can determines sub-crystal block size and micro distortion both is effect physical width of main factors, Asia Crystal block size, and micro distortion of judgment method, is through observation the different diffraction orders of the diffraction lines of physical width β , if $\beta \cos \theta$ for constant, is description wide of by Asia Crystal block wide of effect caused, if beta cot θ for constant, is description wide of is by micro distortion caused, if both are not constant is description this two kind of factors common exists, affecting the physical width of the linear .

4 Calculation of microstress

4.1 X-RAY DIFFRACTION EXPERIMENT

In accordance with the principles of calibration samples location and orientation of the selected measurement points, residual stress test use X-ray stress Analyser XSTRESS3000, this instrument use solid linear detector technology, X-rays can be directly converted into electric signal. In order to improve the measurement precision, the larger collimator, Psi (ψ) and Phi (ϕ) swing and swing method were used.

The four measurement points are designed and shown in Figure 3. By measuring the stress free Ti sample of powder, measurement can be carried out after instrument calibration and the measuring process is shown in Figure 4.



FIGURE 3 Measurement point design



FIGURE 4 Measuring process

4.2 CALCULATION OF MICROSTRESS

The XSTRESS3000 measurement data is converted to MDI JADE procedures to be identified format. Fitting work make up by phase retrieval, deduction of back strength, deduction K_{a2} , deduction of equipment of spectral line widths, as well as for smoothing.

MDI JADE software in the analysis of map is automatically deducted from the instrument width effect with map smooth, get the measurement point of the real diffraction spectra (physical width of spectral lines), is shown in Figure 5.

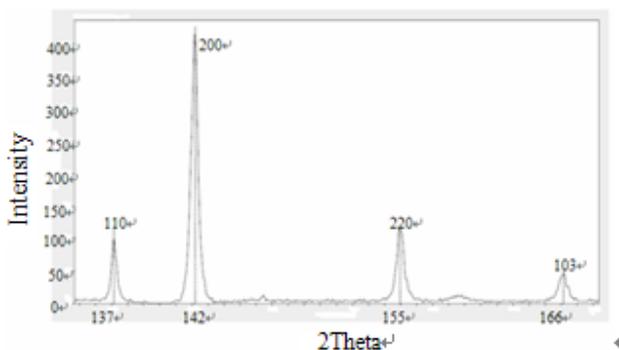


FIGURE 5 Measure point diffraction line (physical width)

For the cognate crystal surfaces 110 and 220, the diffraction peak shape similarity, the half high width β in numerical value is approximately the same, and the 2θ angle is quite different, so the crystal block size and microscopic distortion two factors are present. The line width is not only simple mechanical superposition, but a convolution relation, JADE procedures in the convolution process regard the diffraction line approximation as a specific function to solve. Approximate function method is the crystal block width function and microscopic distortion line width function as a known function, such as the Gauss function, Cauchy function or Cauchy square functions etc.

Due to the two kind of line width effect is not a simple mechanical superposition, but they form convolution, so by a line cannot completed. Need to measure two or more than two diffraction peak half high width β , due to the crystal block size and crystal indices related, so the diffraction plane selection to choose the same direction, for example (110) and (220), get of each spectral line width β and physical diffraction angle θ . At this time regard $\frac{\sin \theta}{\lambda}$ as X-

axis, identify $\frac{\beta \cos \theta}{\lambda}$ and $\frac{\sin \theta}{\lambda}$ linear relation, with

the method of the least squares fit a straight line, the slope of the straight line for micro strain two times, straight line in the ordinate of intercept for crystal block size reciprocal, thus determining the crystal block size D and microscopic distortion ϵ value.

Using MDI JADE program for data operation processing is as follows:

- 1) Measurement of specimen on the reverse side one points more than two diffraction peaks, for the same direction secondary diffraction;
- 2) Reading in data in JADE, for phase retrieval, deduct back bottom strength, deduction and map smooth, full spectrum fitting; into the JADE, for phase retrieval, background subtraction, deduction, as Atlas strength smooth, full spectrum fitting;
- 3) Checking instrument half tall wide compensate curve is correct, select menu "Report - Size&Strain Plot" instruction, open the calculation dialog;
- 4) Selecting size/strain, adjust D value to 1.5. If the

shape is more closer to the Gauss function, it is set to two, if the shape is more closer to the Cauchy function, take D=1. The value of D size influence single value of the experimental results, for sample regularity there are no influence;

- 5) Saving the image. Saving calculation results export to text format, calculation results is shown in Figure 6.

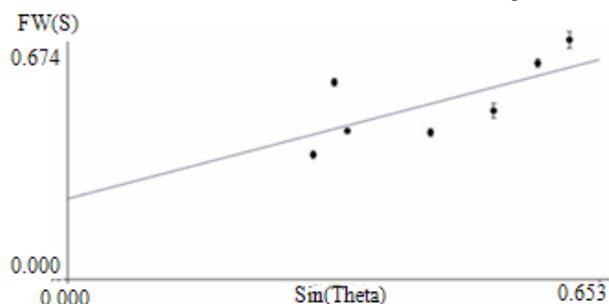


FIGURE 6 Calculation results

In the Figure 6, FW (S) represents the physical width. Strain represents micro strain, as the ratio of train value with the plane spacing, Strain is 0.2461%, the strain value was 0.01638, R represents error. The method of approximate function is the calculation of average grain size, because different crystal face size is not the same, the calculation results represent the direction of each diffraction the size of the grain size. Calculation results of microstress are shown in Table 1.

TABLE 1 Calculation results of microstress

Time node	Measurement points	ϵ (nm/nm)	σ (MPa)
Machining finished	No.1	0.018306926	211.81
	No.2	0.019924986	230.53
	No.3	0.01614891	186.84
	No.4	0.015250658	176.45
40 days passed	No.1	0.009100099	105.29
	No.2	0.011571049	133.88
	No.3	0.00934011	108.07
	No.4	0.010033111	116.08
80 days passed	No.1	0.012416515	143.66
	No.2	0.014536224	168.18
	No.3	0.016056433	185.77
	No.4	0.012811376	148.23

5 Conclusions

- 1) The calculation model of microstress in machining distortion of titanium alloy monolithic component was proposed. The premise of calculation is that the diffraction spectrum line must be determined first, and the calculation is a solving process using convolution relationship.
- 2) The various factors causing width effects of X-ray diffraction line were analysed comprehensively, and the effect of stacking fault on diffraction spectrum was excluded according to the result of TEM experiment. The true diffraction spectrum, which can reflect inside information of Ti6Al4V titanium alloy, was determined using MDI JADE peak shape analysis method.
- 3) According to calculation model of microstress, combining with the result of X-ray diffraction experiment, the microstress in machining distortion of

titanium alloy monolithic component were calculated.

4) This work establishes the foundation for investigating the mechanism of machining distortion of titanium alloy monolithic component.

References

- [1] Megson T H G 2012 Aircraft Structures for Engineering Students Leeds University Press UK Chapter 8
- [2] Durul U, Tugrul O 2011 *International Journal of Machine Tools and Manufacture* **51**(3) 250-80
- [3] Armendia M, Garay A, Iriarte L M 2010 *Journal of Materials Processing Technology* **210**(2) 197-203
- [4] Izamshah R, Mo J P T, Ding S 2011 *Journal of Engineering Manufacture* **203**(9) 687-98
- [5] Guo H, Zuo D W, Wu H B, Xua F, Tong G Q 2009 *Materials Science and Engineering* **499**(2) 230-3
- [6] Jitender K and Rai P X 2008 *International Journal of Machine Tools & Manufacture* **48**(6) 629-43
- [7] Yang Y, Wang Y L and Li C H 2011 *Advanced Science Letters* **4**(5) 1-5 (in Chinese)
- [8] Liu G 2009 *Journal of Materials Processing Technology* **209**(6) 2788-93
- [9] Merodio J, Ray W and Javier R 2013 *International Journal of Non-Linear Mechanics* **56** 43-9
- [10] Hu Y X and Ramana V 2012 *Surface and Coatings Technology* **206**(15) 3374-85
- [11] André Mézin 2006 *Surface and Coatings Technology* **200**(18-19) 5259-67
- [12] Böhm J, Wanner A, Kampmann R 2003 *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **200** 315-322
- [13] Kaouache B, Labat S and Thomas O, Maitrejean S, Carreau V 2008 *Microelectronic Engineering* **85**(10) 2175-8
- [14] Zhong Z Q and Zhang Z J 2006 *Metallurgical analysis* **10** 18-19
- [15] Jiang C H and Yang C Z 2010 *X-ray diffraction technique and its application* East China University Of Science And Technology Press: Shanghai chapter 7

Acknowledgments

The authors gratefully acknowledge the support of the National Natural Science Foundation of China (Grant No. 51105216) and the Shangdong Province Excellent Young and Middle-Aged Scientists Research Awards Fund (Grant No. BS2011ZZ006).

Authors	
	<p>Yong Yang, born in June, 1976, Qingdao, China</p> <p>Current position, grades: Doctor of Mechanical Engineering, Associate professor in Qingdao Technological University. University studies: Mechanical Engineering in Wuhan University of Science and Technology. Scientific interest: advanced manufacturing technology. Publications: 3 Patents, 28 Papers</p>
	<p>Weiwei Zhu, born in September, 1988, Qingdao, China</p> <p>Current position, grades: MSc student of Mechanical Engineering. University studies: Mechanical Engineering in Qingdao Technological University. Scientific interest: machining distortion controlling. Publications: 2 Papers.</p>

Research and implement of CATIA parametric modelling-based cutter information integration in VERICUT

Xianguang Kong^{*}, Yihui Li, Lei Yin, Xiaowen Wang

School of Electromechanical Engineering Xidian University, Xian, Shaanxi, 710071, China

Received 1 August 2014, www.tsi.lv

Abstract

In order to achieve the integration of the VERICUT tool library and tool management software, this paper studies the VERICUT tool integration technology based on CATIA parametric modelling. As the tool modelling ability is limited, VERICUT can neither achieve complex tool modelling nor generate the tool model by parameterization. Therefore, the secondary development based on VERICUT is unable to meet the integration requirements of tool information. This paper proposes a new integration approach, which transforms the integration of VERICUT and tool management software into VERICUT and CATIA tool information. Meanwhile, a feasible integrated development process is put forward. First, the tool model is parametrically driven by CATIA. Then, the integration of the CATIA tool library and tool management software is realized based on CAA. Finally, the transformation of CATIA and VERICUT tool model is accomplished, which indirectly realizes the integration of the VERICUT tool library and tool management software. During the NC machining simulation, VERICUT inherits the tool information and model generated by CATIA NC programming, which can ensure consistency with the tool information in tool management software. This paper solves the problem of integration of tool management software and VERICUT in the context of a digital manufacturing project; successful application of the proposed approach has greatly improved the efficiency of NC programming.

Keywords: VERICUT, CATIA, tool management software, parametric modelling, TLS, CAA

1 Introduction

VERICUT, an NC machining simulation software pushed out by CGTECH, can simultaneously simulate the tool path and the machine tool, to detect potential problems in machine processing. In order to realize the simulation of NC machining processes, VERICUT first establishes the geometric and kinematic model of the machine tool, then builds other manufacturing resources, such as the geometric model of cutters, workpieces and fixtures, and specifies the tool path or the NC program and sets appropriate parameters. This is followed by the simulation and optimization of the machine process. Meanwhile, widely applied in enterprises, tool management software can effectively manage information about tool components, tool diagrams and tool suppliers. However, each of the two kinds of software has its own tool library. They manage and use the tool information independently, therefore resulting in the problem of "Information Island", which prevents workers from dynamically calling the tool information from the software. Workers have to adjust and maintain tool information, which leads to a big workload and is error prone [1].

In order to allow VERICUT to dynamically call the tool information from tool management software in the NC machining verification, there is a need for a secondary development based on VERICUT. However, VERICUT only provides a standard method to shape

models because of limitations of its tool modelling function. That is, it designs tool profiles with a symmetrical rotary centre and composed of lines and arcs by profile definition, and cannot build complicated tool models. The most important issue is the fact it does not have the function of parameterized modelling. The above problems result in a VERICUT tool model that cannot vary with the tool attribute value. Therefore, integration of tool information cannot be achieved by a secondary development based on VERICUT.

To this end, this paper proposes a new integrated approach. When using VERICUT for NC machining simulation, we take advantage of the 3D parametric modelling function of CATIA to establish the tool models in NC processing, and generate VERICUT tool models with CATIA as the parametric design platform. Thus the consistency of tool information in the VERICUT NC machining verification process and tool management software is achieved. This paper presents a case study based on a digital manufacturing project, and solves the problem of integration of tool management software and VERICUT. Through the integration of the VERICUT tool library, NC workers can directly inherit tools from the CATIA tool library, and are able to master the conditions of existing tools, thereby improving the efficiency of NC programming.

^{*} *Corresponding author* e-mail: kongxg@vip.sina.com

2 VERICUT tool integration approach based on CATIA parametric modelling

CATIA NC programming and VERICUT NC machining verification are two closely linked aspects. VERICUT uses the NC code generated by CATIA for NC machining verification. Therefore, the problem of tool information consistency in VERICUT and tool management software is solved when the following is ensured; the tools selected in CATIA NC programming are in the tool management software, and VERICUT NC machining verification only needs to inherit the tool information and models in CATIA NC programming [2].

The VERICUT tool library contains information of the cutting part, cutter bar and clamping part, which is stored in the tool library files in TLS format, and is used after calling the tool library, which is already set to edit. When using VERICUT to do NC machining simulation, the first step is calling the NC program in the VERICUT circumstance, and then defining the tool list to build the

mapping relationship of the tool number assigned in G-code and the main tool library file [3]. The NC code is generated by CATIA and the tool's ID number in the NC code is what is recorded in the tool library. We need to extract the tool's ID number and assign the tool's attribute value to the VERICUT tool library. This can indirectly realize the integration of information from the tool management software and VERICUT.

Different tools with different machining processes in NC machining result in different specifications of tool models, with different tool selections, generated by VERICUT tool TLS files. As the tools only have attribute value in tool management software, and VERICUT tool model information is stored in the tool library files in TLS format, we can build a tool model in CATIA and then generate the TLS files [4]. The integration scheme of tool management software and VERICUT is shown in Figure 1. The integration is of of CATIA and VERICUT, which realizes the integration of tool management software and VERICUT.

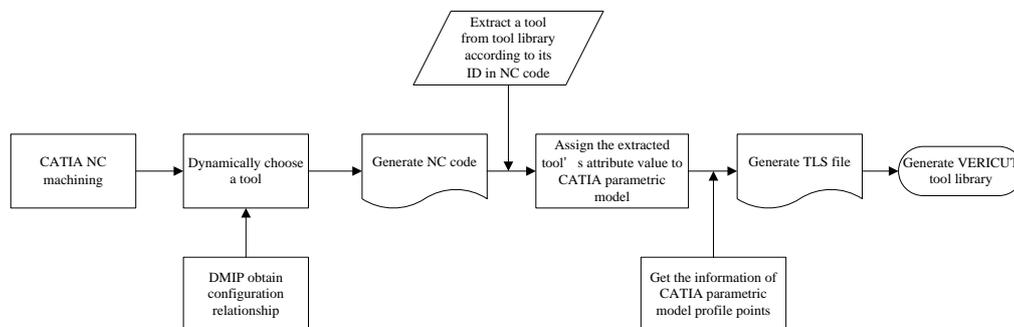


FIGURE 1 The integration scheme of tool management software and VERICUT

3 Tool integration development process

3.1 BUILDING THE PARAMETRIC TOOL MODEL BASED ON CATIA

There are two methods for parametric modelling in CATIA. One is the program driving method, which uses advanced object-oriented language such as VB, VC, C++, etc. to drive CATIA by editing command to draw the parametric model; the other is the dimension driving method. The former can directly call the API object to draw graphics. As the graphics are completely generated by the program, there is a large amount of code, and programming is difficult [6]. The program has to be run every time a part is generated, resulting in a low running speed. However, the latter can modify the parameters to change the model. The code of this program is simple and has a high running speed. Considering the running speed and the low difficulty of tool modelling, this paper uses the dimension driving method to build the tool model [7, 8].

The parametric design method can generally be divided into the following four steps:

1. Extract geometric feature parameters and do customized naming;

2. Establish the corresponding mechanism of geometrical quantity numerical information and parameters, as well as the corresponding relations among the parameters;

3. Edit parameters to modify geometric entities indirectly;

4. Introduce the parameter verification, supervise the design process and test whether the design meets the requirements.

The operating process is as follows; Open CATIA, and enter the Part Design module. Click $f(x)$ and establish the tool parameters, number the tool contour points, revolve the tool contour line and generate the model. Click $f(x)$ and choose the parameter of the Part Number, click "add formula", choose "Part Number" in the parameters on the feature tree, then click OK. The establishment of the tool parametric model is then completed [8].

3.2 PARAMETRICALLY DRIVE AND GENERATE CATIA TOOL MODEL

As the powerful engineering software, CATIA has high flexibility. Users can conduct various developments with different methods according to their needs. Specifically,

there are two ways of secondary development of CATIA: Using micro and Component Application Architecture (CAA). As a platform for Dassault Systemes products extension and customers customizing and developing, CAA uses object-oriented programming (OOP) language, which has become the mainstream of software development and design. Using COM technology and OLE technology, the development of CAA can be regarded as a combination and expansion of its component objects. As a type of software architecture, COM has better module independence and scalability, which makes the programming of CAA much easier and more standardized, with more concise code [9].

In view of this function, there is a need to call the tools in the tool library of the tool management software, and dynamically assign the tool's attribute value to the tool model in the CATIA NC machining module. Therefore, we adopt the CATIA tool integration method based on CAA to achieve the generation of the CATIA tool model through dynamic data calls and parametric drive. The integration scheme is shown in Figure 2.

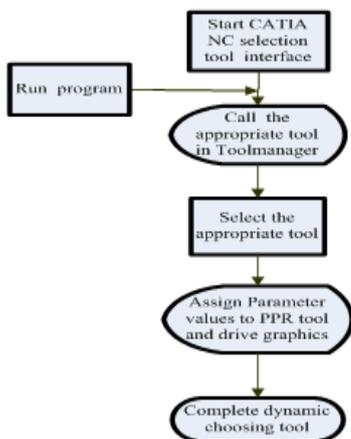


FIGURE 2 Tool library integrated solution

3.3 REALIZING THE TRANSFORMATION BETWEEN CATIA AND VERICUT TOOL MODELS

The integration problem of CATIA and VERICUT to be solved is how the contour point coordinates of the tool model in CATIA can be transformed into that in VERICUT, that is, how to generate the *.TLS file[10]. As part of the VERICUT tool library file, the TLS file contains cutter and toolholder description data used for cutting simulation. The TLS file can be expressed by the following example:

```

CGTECH Tool Library File
Version-5.4
TOOLID "1" {
  UNITS MILLIMETER
  TOOLTYPE MILLING
  STACK YES
}
  
```

Kong Xianguang, Li Yihui, Yin Lei, Wang Xiaowen

```

CUTTER {ASSEMBLY {SOR {PTS { (0,0) (0.75,0) }
  ARC (0.75,0.25,0.25) CCW
  PTS {(1,0.25) (1,0.625) (0.5,0.625) (0.5,1.125)
    (0,1.125) }
  COLOR 6 }}}}}
  
```

This indicates that a tool, with its ID as 1, has a profile consisting of straight lines and arcs. The endpoints of the straight lines are (0, 0), (0.75, 0), (1, 0.25), (1, 0.625), (0.5,0.625), (0.5, 1.125), (0, 1.125), the centre of the arc is (0.75, 0.25), the start point is (0.75, 0), end point is (1, 0.25), the radius is 0.25, and the direction is counter-clockwise. The unit is mm [12].

This shows that the TLS file generates the tool model using information about tool contour points. Therefore, reading the coordinates of the points is an important step.

The development process is shown in Figure 3:

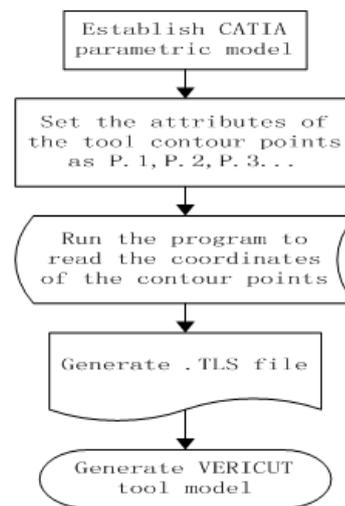


FIGURE 3 The integration process of VERICUT and CATIA

4 Development and validation of tool integration module

Using Windows Server 2000, Windows 2000 Professional/XP/NT operating system, SQL Server 2000 database management system and Delphi 7, this paper achieves the development of a tool information integration module based on VERICUT. The realization process is shown in Figure 4. After debugging and verification, the platform is officially used in enterprises, and solves the problem of "Tool Information Island" between CATIA, tool management software and VERICUT [11].

The module based on VERICUT has two major functions:

1. Transformation of the tool in NC files into a CATIA physical model: assignment of the attribute value of the tool in the tool management software to the CATIA parametric model;

2. Direct generation of the TLS file tool shape in VERICUT: transformation of the contour points coordinates of the CATIA physical model into a .tls file; finding it in the tool management in VERICUT, and

opening it to generate the tool shape after saving the TLS file in the designated location.

Based on the above two types of integration, this paper achieves the integration of tool management software and VERICUT, which enables the

transformation of NC code to TLS file, so that tool information generated by VERICUT is consistent with that used in the previous step. The generating speed of the VERICUT tool is also increased, and the consistency of the tool information is ensured.

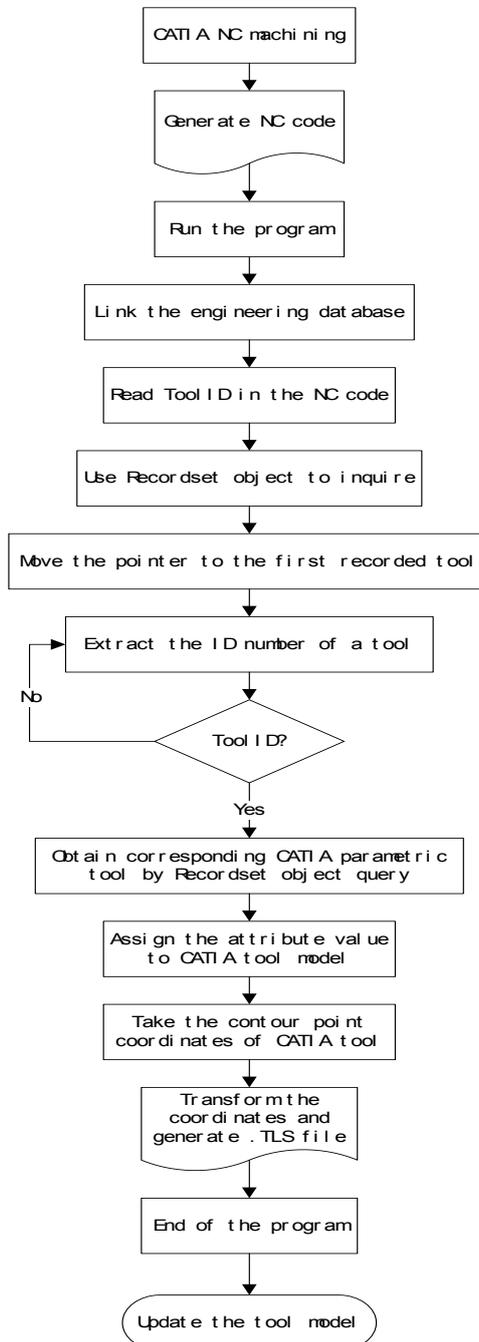


FIGURE 4 The development process of tool integration model based on VERICUT

5 Conclusions

Based on a digital manufacturing project, this paper studies the VERICUT tool library integration technology, analyses the structure and establishment theory of the VERICUT tool library and the tool integration method of VERICUT. The usage of

parametric modelling, based on CATIA is proposed to realize VERICUT tool information integration. Through the tool information integration of CATIA and tool management software, as well as CATIA and VERICUT, the integration of VERICUT and tool management software is achieved. In addition, combined with the practical situation, we developed a software

module that realizes the integration of the VERICUT tool library. The successful integration of the VERICUT tool library and tool management software plays the following role:

1) Greatly reduced the quantity of work associated with setting the tool parameters, and improved the efficiency of NC programming by directly inheriting the tool information from the NC code generated by CATIA during the NC programming.

2) Through the integration of the VERICUT tool library, the selected tool in the NC verification is related

to the production field tool, which ensures the suitability of the NC program.

3) Improved the degree of automation and integration of the digital manufacturing system.

Acknowledgements

This work was supported by “the Fundamental Research Funds for the Central Universities”.

This work was supported by Shaanxi Provincial scientific and technological research projects under Grant Nos. DF0102130401.

References

- [1] Delacour J, Cuinier J-L 2004 Presentation of the first PLM integrated Optical Simulation Software for the Design and Engineering of Optical Systems *Proceedings of SPIE – The International Society for Optical Engineering* **5249** 42-53
- [2] Xie L H, Shen Y H 2005 NC machining of CATIA V5 *Beijing: Tsinghua University Press*
- [3] Xu P C 2007 Road of digital manufacturing system construction *Aviation manufacturing technology* 68-71
- [4] Dong Y X, Xi P 2006 The secondary development based on CATIA interface *Aviation manufacturing technology* 83-6
- [5] Dong Y X, Xi P 2005 NC coordinate simulation system of secondary develop five coordinate machining center based on CATIA *Mechanical engineer* 41-3
- [6] Cheng M, Deng S F, Zhu R 2006 CAD/CAE integration based on CATIA platform *Computer aided design and graphics journal* **18** 1078-82
- [7] Deng D M, Zhou L S, Chen G 2007 Based on CATIA component library build library tool design and implementation *South China university of science and technology journal* **35** 138-42
- [8] Dassault Systems (2001) *CAA V5 For CATIA Foundations* 18-156
- [9] Wang S, Cheng N F, Zhu Y M 2007 Entity modeling methods of function gradient material based on CATIA platform *China mechanical engineering* **18** 454-6
- [10] Yang L H, Zhang H M 2001 Research and implementation of information integration technology of CATIA product based on COM component *Computer engineering and application* 132-4
- [11] Guo Z, Zang S L, Wei G J 2006 *CATIA V5 secondary development technology and its application in design of stamping dies*. Die and Mould industry **32** 1-4
- [12] Li W, Zhang Y Y 2001 Development and application of conversion software of NC program based on CATIA *Aeronautical Manufacture Technology* 55-7

Authors



Xianguang Kong

Current position, grades: Associate professor at Xidian University, College of Mechanical & Electrical Engineering.
University studies: Ph.D in Mechanical Manufacturing and Automation (2005, Northwestern Polytechnical University).
Scientific interest: advanced hybrid manufacturing technology, CAD/CAPP/CAM/CIMS.



Yihui Li

Current position, grades: Doctor of Engineering, Director of Application of information technology Research Office, Xi'an High Voltage Apparatus Research Institute Co., Ltd., Xi Dian Group, China.
University studies: Ph.D on Mechanical Manufacturing and Automation (2009, Northwestern Polytechnical University).
Scientific interest: enterprise informationization strategy, enterprise resource planning, it planning, digitized manufacturing technology.
Publications: 3 patents, 10 publications.



Lei Yin

Current position, grades: Associate Professor at Xidian University.
University studies: Ph.D. in Northwestern Polytechnical University (2000–2005); Postdoctoral in Xidian University(2007-2010).
Scientific interest: system modeling and simulation.
Publications: 4 patents, 16 publications.



Xiaowen Wang

Current position, grades: Postgraduate at Xidian University, College of Mechanical & Electrical Engineering.

Bilinear model for ontology mapping

Jian-Zhang Wu^{1, 2*}, Yu Xiao³, Wei Gao⁴

¹School of Computer Science & Engineer, Southeast University, Nanjing 210096, China

²Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing 210096, China

³School of Continuing Education, Southeast University, Nanjing 210096, China

⁴School of Information Science and Technology, Yunnan Normal University, Kunming, 650500, China

Received 15 July 2014, www.tsi.lv

Abstract

As a model of concept representation, ontology has widely applied to various disciplines. Ontology mapping is used to create the link between different ontologies. In this paper, we present a new ontology mapping algorithm by virtue of bilinear model. The linear mapping pair is given by the iterative procedure. Two strategies are manifested to obtain the finally ontology mapping. The simulation experimental results show that the proposed new technologies have high accuracy and efficiency on ontology mapping in certain applications.

Keywords: ontology, ontology mapping, linear mapping, bilinear model, dimensionality reduction

1 Introduction

As a knowledge representation and conceptual shared model, ontology has been applied in image retrieval, knowledge management and information retrieval search extension. Acting as an effective concept semantic model, ontology is also employed in disciplines beyond computer science, such as social science (for instance, see [1]), biology science [2] and geography science [3].

The ontology model is actually a graph $G=(V,E)$, each vertex v in an ontology graph G represents a concept and each edge $e=v_i v_j$ on an ontology graph G represents a relationship between concepts v_i and v_j . The aim of ontology mapping is to bridge the link between two or more ontologies. Let G_1 and G_2 be two ontology graphs corresponding to ontology O_1 and O_2 respectively. For each $v \in G_1$, find a set $S_v \subseteq V(G_2)$ where the concepts corresponding to vertices in S_v are semantically close to the concept corresponding to v . One method to get such mapping is, for each $v \in G_1$, to compute the similarity $S(v, v_j)$ where $v_j \in V(G_2)$ and to choose a parameter $0 < M < 1$. Then S_v is a collection such that the element in S_v satisfies $S(v, v_j) \geq M$. In this point of view, the essence of ontology mapping is to obtain a similarity function S and select a suitable parameter M . In our article, we focus on the technologies to yield an optimal similarity function S from dimensionality reduction standpoint. In fact, our approach for obtaining such similarity function is based on the linear mapping pair.

For ontology similarity measure, there are several effective learning tricks. Wang et al. [4] proposed to learn a score function which mapping each vertex to a real number, and the similarity between two vertices can

be measured according to the difference of real number they correspond to. Huang et al., [5] presented a fast ontology algorithm for calculating the ontology similarity in a short time. Gao and Liang [6] raised that the optimal ontology function can be determined by optimizing NDCG measure, and applied such idea in physics education. Gao and Gao [7] deduced the ontology function using the regression approach. Huang et al., [8] obtained ontology similarity function based on half transductive learning. Gao and Xu [9] explored the learning theory approach for ontology similarity computation using k -partite ranking method. Zhu and Gao [10] proposed a new criterion for ontology computation from AUC and multi-dividing standpoint. Gao et al., [11] presented a new ontology mapping algorithm using harmonic analysis and diffusion regularization on hypergraph. Very recently, Gao and Shi [12] proposed a new ontology similarity computation technology such that the new calculation model considers operational cost in the real implement.

In this paper, we determine the new ontology mapping algorithm based on dimensionality reduction idea and bilinear learning model. Using the optimization algorithm, we determine the linear mapping (L_1, L_2) to compute the similarity of vertices from two ontologies. The experiment is designed to show the efficiency of the algorithm.

2 Model and algorithm

For each vertex v , we use a vector to represent all its information. For two ontologies O_1 and O_2 , their structures can be determined by two ontology graphs

* Corresponding author e-mail: jzww@njnet.edu.cn

$G_1=(V_1,E_1)$ and $G_2=(V_2,E_2)$ respectively. Suppose that $V_1 \subset \mathbb{R}^{d_1}$ and $V_2 \subset \mathbb{R}^{d_2}$. That is to say, we use a vector with dimension d_1 to represent the information of vertex in V_1 and use a vector with dimension d_2 to represent the information of vertex in V_2 . For any $v_i \in V_1$, and $v_j \in V_2$, $S(v_i,v_j)=S_{ij}$ indicates the similarity between concepts corresponding to v_i and v_j . Our goal is to learn an optimal similarity function S based on the sample triple $D=\{(v_i,v_j,S_{ij})\}$, where $v_i \in V_1$, $v_j \in V_2$. For such triple D , let $V_1^D = \{v_i\} \subseteq V_1$, $V_2^D = \{v_j\} \subseteq V_2$, $n_1 = |V_1^D|$ and $n_2 = |V_2^D|$.

We are interested in searching a linear mapping pair (L_1,L_2) such that the corresponding images $L_1^T v_i$ and $L_2^T v_j$ are in the same d -dimensional latent space L with $d \ll \min\{d_1,d_2\}$ and the degree of similarity between ontology vertices $v_i \in V_1$ and $v_j \in V_2$ can be reduced to L 's dot product:

$$D_{L_1,L_2}(v_i,v_j) = v_i^T L_1 L_2^T v_j.$$

By virtue of the trick used in [13] for kernel learning, we aim to maximize the following expected version:

$$E_{v_i,v_j} \{S(v_i,v_j) D_{L_1,L_2}(v_i,v_j)\} = E_{v_i} E_{v_j|v_i} \{S(v_i,v_j) v_i^T L_1 L_2^T v_j\}. \tag{1}$$

The Equation (1) could be estimated as follows:

$$\frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} S_{ij} v_i^T L_1 L_2^T v_j.$$

Hence, the ontology mapping problem is boiled down to

$$\arg \max_{L_1,L_2} \frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} S_{ij} v_i^T L_1 L_2^T v_j, \tag{2}$$

s.t. $L_1 \in H_1$, $L_2 \in H_2$, where H_1 and H_2 are the hypothesis spaces for L_1 and L_2 respectively. Since the final computational model is linear in view of both ontology vertices v_i and v_j , learning model (2) is actually a bilinear model for calculating similarity in two spaces.

We apply l_1 norm and l_2 norm constraints on L_1 and L_2 . Let $|\cdot|$ and $\|\cdot\|$ be l_1 -norm and l_2 -norm respectively, and $l_{v_i,x}$ and $l_{v_j,y}$ be the x -th and y -th row of L_1 and L_2 . Specifically, we introduce two hypothesis spaces as:

$$H_1 = \{L_1 \mid |l_{v_i,x}| \leq \lambda_{v_i}, \|\cdot\| \leq \theta_{v_i}, x = 1, \dots, d_1\},$$

$$H_2 = \{L_2 \mid |l_{v_j,y}| \leq \lambda_{v_j}, \|\cdot\| \leq \theta_{v_j}, y = 1, \dots, d_2\},$$

where $\{\lambda_{v_i}, \theta_{v_i}, \lambda_{v_j}, \theta_{v_j}\}$ are parameters selected by

experts. Here the constraints relying on l_1 -norm will induce row-wise sparsity in L_1 and L_2 . Furthermore, the l_2 -norm on rows with regularization can avoid degenerative solutions. By virtue of the definition of H_1 and H_2 , we infer the following program:

$$\arg \max_{L_1,L_2} \frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} S_{ij} v_i^T L_1 L_2^T v_j, \tag{3}$$

s.t.:

$$|l_{v_i,x}| \leq \lambda_{v_i}, \|l_{v_i,x}\| \leq \theta_{v_i}, |l_{v_j,y}| \leq \lambda_{v_j}, \|l_{v_j,y}\| \leq \theta_{v_j}, \quad x \in \{1, \dots, d_1\}, \\ y \in \{1, \dots, d_2\}.$$

In practice reality, we solve the following variant version of Equation (3) for easier computation

$$\arg \min_{L_1,L_2} -\frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} S_{ij} v_i^T L_1 L_2^T v_j + \tag{4}$$

$$\beta \sum_{x=1}^{d_1} |l_{v_i,x}| + \gamma \sum_{y=1}^{d_2} |l_{v_j,y}|,$$

$$\text{s.t. } \|l_{v_i,x}\| \leq \theta_{v_i}, \|l_{v_j,y}\| \leq \theta_{v_j}, \quad x \in \{1, \dots, d_1\}, \quad y \in \{1, \dots, d_2\},$$

where $\beta > 0$ and $\gamma > 0$ are the balance parameters to control the trade-off between objective term and penalty term. For given L_2 , the objective mapping of Equation (4) can be re-represented as:

$$\sum_{x=1}^{d_1} \left\{ \left(-\frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} v_i^x S_{ij} L_2^T v_j \right)^T l_{v_i,x} + \beta |l_{v_i,x}| \right\}.$$

By using $\omega_x = [\omega_x^1, \omega_x^2, \dots, \omega_x^{d_1}]^T$ to represent the d -dimensional $\frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} v_i^x S_{ij} L_2^T v_j$, we infer the optimal $l_{v_i,x}$ as:

$$(l_{v_i,x}^k)^* = C_{v_i} \left(\max(|\omega_v^k| - \beta, 0) \text{sign}(\omega_v^k) \right), \quad k \in \{1, \dots, d\}, \tag{5}$$

where $l_{v_i,x}^k$ is the k -th element of $l_{v_i,x}$:

$$\text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

and C_{v_i} is a constant which makes $\|l_{v_i,x}^*\| = \theta_{v_i}$ if there exist non-zero elements in $l_{v_i,x}^*$, and $C_{v_i} = 0$ otherwise.

For given L_1 , the objective mapping of Equation (4) can similarly re-written as:

$$\sum_{y=1}^{d_2} \left\{ \left[-\frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} v_j^y S_{ij} L_1^T v_i \right]^T l_{v_j, y} + \gamma |l_{v_j, y}| \right\}.$$

In terms of the same fashion, we use $\eta_y = [\eta_y^1, \eta_y^2, \dots, \eta_y^d]^T$ to represent the d -dimensional

$\frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} v_j^y S_{ij} L_1^T v_i$, we yield the optimal $l_{v_j, y}$ as:
 $(l_{v_j, y}^k)^* = C_{v_j} (\max(|\eta_{v_j}^k| - \gamma, 0) \text{sign}(\eta_{v_j}^k))$, $k \in \{1, \dots, d\}$, (6)

where $l_{v_j, y}^k$ is the k -th element of $l_{v_j, y}$, and C_{v_j} is a constant which makes $\|l_{v_j, y}^*\| = \theta_{v_j}$ if there exist non-zero elements in $l_{v_j, y}^*$, and $C_{v_j} = 0$ otherwise.

Let:

$$w_{v_i, x} = \frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} v_i^x S_{ij} v_j$$

and

$$w_{v_j, y} = \frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} v_j^y S_{ij} v_i,$$

which does not depend on the change of L_1 and L_2 , and can be pre-calculated. It is easy to verify that:

$$\frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} v_i^x S_{ij} L_2^T v_j = L_2^T w_{v_i, x} =$$

and

$$\frac{1}{n_1 n_2} \sum_{v_i \in V_1^D} \sum_{v_j \in V_2^D} v_j^y S_{ij} L_1^T v_i = L_1^T w_{v_j, y}.$$

Let N_{v_i} be the average number of non-zeros in all v_i per dimension and N_{v_j} be the average number of non-zeros in all v_j per dimension, \tilde{n}_1 be the average number of related v_i samples per v_j and \tilde{n}_2 be the average number of related v_j samples per v_i , c_1 be the average number of non-zeros in each v_i sample and c_2 be the average number of non-zeros in each v_j sample. Now, we present the following two algorithms:

Algorithm 1. Calculating $w_{v_i, x}$ and $w_{v_j, y}$

Input $D = \{(v_i, v_j, S_{ij})\}$, where $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$.

For $x=1:d_1$, $w_{v_i, x} \leftarrow 0$; For $y=1:d_2$, $w_{v_j, y} \leftarrow 0$.

For $x=1:d_1$, $i=1:n_1, j=1:d_2$, $w_{v_i, x} \leftarrow w_{v_i, x} + \frac{1}{n_1 n_2} v_i^x S_{ij} v_j$.

For $y=1:d_2$, $i=1:n_1, j=1:d_2$, $w_{v_j, y} \leftarrow w_{v_j, y} + \frac{1}{n_1 n_2} v_j^y S_{ij} v_i$.

Output: $\{w_{v_i, x}\}_{x=1}^{d_1}$ and $\{w_{v_j, y}\}_{y=1}^{d_2}$.

Algorithm 2. Calculating L_1^t and L_2^t .

Input $\{w_{v_i, x}\}_{x=1}^{d_1}$, $\{w_{v_j, y}\}_{y=1}^{d_2}$, d , β , γ , θ_1 , θ_2 , and set L_1^0 and L_2^0 randomly $t \leftarrow 0$.

While $t \leq T$. For $x=1:d_1$, compute ω_x using $(L_1^t)^T w_{v_i, x}$ and determine $(l_{v_i, x})^*$ in terms of Equation (5), update L_1^{t+1} ;

For $y=1:d_2$, compute η_y using $(L_2^t)^T w_{v_j, y}$ and determine $(l_{v_j, y})^*$ in terms of Equation (6), update L_2^{t+1} ; $t \leftarrow t+1$.

Output L_1^t and L_2^t .

The complexities of Algorithm 1 and Algorithm 2 are $O(d_1 N_{v_i} \tilde{n}_2 c_2 + d_2 N_{v_j} \tilde{n}_1 c_1)$ and $O(d_1 W_1 d + d_2 W_2 d)$ respectively, where W_1 is the number of non-zeros for each $w_{v_i, x}$ on average and W_2 is the number of non-zeros for each $w_{v_j, y}$ on average.

After the similarity between vertices are determined by bilinear model, we select a strategy to derive finally ontology mapping. Following two strategies could be used for getting ontology mapping.

Strategy 1. For each $v \in V(G_i)$, $i=1,2$. Let $N \in \mathbb{N}$ be a parameter, and:

$$v_1 = \max_{v' \in V(G_i)} \{S(v, v')\},$$

$$v_2 = \max_{v' \in V(G_i), v' \neq v_1} \{S(v, v')\},$$

$$v_3 = \max_{v' \in V(G_i), v' \neq v_1, v' \neq v_2} \{S(v, v')\},$$

$$v_N = \max_{v' \in V(G_i), v' \neq v_1, v' \neq v_2, \dots, v' \neq v_{N-1}} \{S(v, v')\}.$$

Then, we deduce:

$$\text{map}(v) = \{v_1, v_2, \dots, v_N\}.$$

Strategy 2. For each $v \in V(G_i)$, $i=1,2$. Let $M \in \mathbb{R}^+$ be a parameter, and

$$\text{map}(v) = S\{(v, v') \geq M \mid v' \in V(G_i)\}.$$

3 Experiment

Experiment of relevance ontology mapping is designed below. In order to adjacent to the setting of ontology algorithm, we use a vector to express each vertex's information. Such vector contains the information of name, instance, attribute and structure of vertex. Here the instance of vertex refers to the set of its reachable vertex in the directed ontology graph.

We use physical education ontologies O_1 and O_2 (the structures of O_1 and O_2 are presented in Figures 1 and 2 respectively) for our experiment. The goal of this experiment is to determine ontology mapping between O_1

and O_2 via linear mapping L_1 and L_2 which is deduced by Algorithm 1 and Algorithm 2. $P@N$ criterion (Precision Ratio, see Craswell and Hawking, [14]) is applied to measure the equality of the experiment. We first give the closest N concepts for each vertex on the ontology graph with the help of experts, and then we obtain the first N concepts for every vertex on ontology graph by the algorithm and compute the precision ratio. Also, ontology algorithms in [11, 5] and [6] are employed to “physical education” ontology, and we compare the precision ratio which we get from four methods. Several experiment results refer to Table 1.

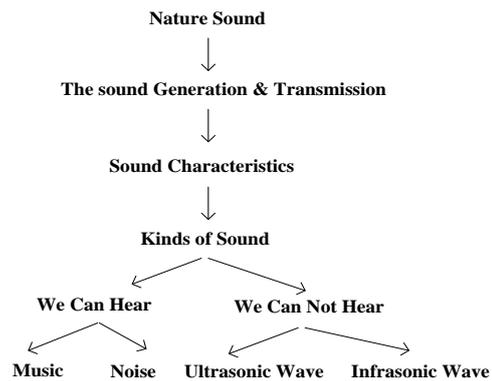


FIGURE 2 “Physical Education” Ontology O_2

The experiment results in Table 1 reveal that the precision ratio in terms of our algorithm higher than the precision ratio determined by algorithms proposed in [11, 5] and [6] by taking $N=1, 3$ or 5 . Specially, as N becomes large, such precision ratios in terms of our algorithm are increasing apparently. In this point of view, our algorithm is more efficient than algorithms raised in [11, 5] and [6] especially when N is sufficiently large.

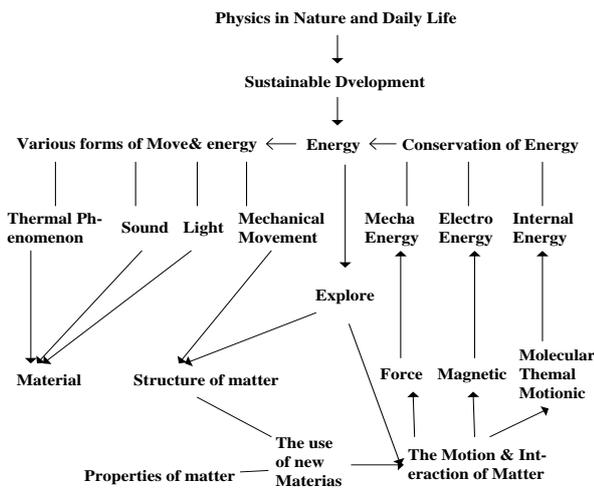


FIGURE 1 “Physical Education” Ontology O_1

TABLE 1 The experiment data of ontology mapping

	$P@1$ average precision ratio	$P@3$ average precision ratio	$P@5$ average precision ratio
Algorithm presented in our paper	70.97%	79.37%	90.48%
Algorithm presented in [11]	67.74%	77.42%	89.68%
Algorithm presented in [5]	61.29%	73.12%	79.35%
Algorithm presented in [6]	69.13%	75.56%	84.52%

4 Conclusions

In this paper, we propose a new computation model for ontology mapping application. The model is bilinear and the algorithm is essentially a kind of dimensionality reduction algorithm which maps the high-dimensional ontology space into low-dimensional. At last, simulation data shows that our new algorithm has high efficiency in physics education ontologies. The algorithm achieved in our paper illustrates the promising application prospects for ontology mapping. The technologies raised in our paper contribute to the state of the art.

References

[1] Bouzeghoub A, Elbyed A 2006 Ontology mapping for web-based educational systems interoperability *Journal of Interoperability in Business Information Systems* 1(1) 73-84
 [2] Hu B, Dasmahapatra S, Lewis P, Shadbolt N 2003 Ontology-based medical image annotation with description logics *Proceedings of*

Acknowledgment

First we thank the reviewers for their constructive comments in improving the quality of this paper. This work was supported in part by science and technology of Jiangsu province (BE2011173), and Key Laboratory of Computer Network and Information Integration founding in Southeast University. We also would like to thank the anonymous referees for providing us with constructive comments and suggestions.

the 15th IEEE International Conference on Tools with Artificial Intelligence California Sacramento USA 77-82
 [3] Fonseca F, Egenhofer E, Davis C, Camara G 2001 Semantic granularity in ontology-driven geographic information systems *AMAI Annals of Mathematics and Artificial Intelligence – Special Issue on Spatial and Temporal Granularity* 36 121-51

[4] Wang Y, Gao W, Zhang Y, Gao Y 2010 Ontology similarity computation use ranking learning Method *The 3rd International Conference on Computational Intelligence and Industrial Application* Wuhan China 20-2

[5] Huang X, Xu T, Gao W, Jia Z 2011 Ontology similarity measure and ontology mapping via fast ranking method *International Journal of Applied Physics and Mathematics* 1(1) 54-9

[6] Gao W, Liang L 2011 Ontology similarity measure by optimizing NDCG measure and application in physics education, *Future Communication, Computing, Control and Management* 142 415-21

[7] Gao Y, Gao W 2012 Ontology similarity measure and ontology mapping via learning optimization similarity function *International Journal of Machine Learning and Computing* 2(2) 107-12

[8] Huang X, Xu T, Gao W, Gong S 2011 Ontology similarity measure and ontology mapping using half transductive ranking *Proceedings of 4th IEEE international Conference on Computer Science and Information technology* Chengdu China 571-4

[9] Gao W, Xu T 2013 Stability analysis of learning algorithms for ontology similarity computation Abstract and Applied Analysis 2013 Article ID 174802

[10] Zhu L, Gao W Theoretical Analysis for New Multi-dividing Ontology Algorithm Based on AUC Criterion, Manuscript (in Chinese)

[11] Gao W, Gao Y, Liang L 2013 Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping *Journal of Chemical and Pharmaceutical Research* 5(9) 592-8

[12] Gao W, Shi L 2013 Ontology similarity measure algorithm with operational cost and application in biology science *An Indian Journal of BioTechnology (in Chinese)*

[13] Cristianini N, Shawe-taylor J, Elisseeff A, Kandola J 2011 On kernel-target alignment *Proceedings of the Neural Information Processing Systems NIPS'01* MIT Press 367-73

[14] Craswell N, Hawking D, Wilkinsin R, Wu M 2003 Overview of the TREC 2003 web track *Voorheer E (ed.) NIST Special Publication 500-255: The Twelfth Text REtrieval Conference (TREC 2003)* Gaithersburg MD 78-92

Authors	
	<p>Jian-Zhang Wu</p> <p>Current position, grades: Associate Professor at the School of Computer Science & Engineer (Southeast University). Scientific interests: cloud computing, information retrieval, machine learning.</p>
	<p>Yu Xiao</p> <p>Current position, grades: Lecturer at the School of Continuing Education (Southeast University). Scientific interests: web technology, cloud computing, educational technology.</p>
	<p>Wei Gao</p> <p>Current position, grades: Lecturer at School of School of Information Science and Technology (Yunnan Normal University). Scientific interests: information retrieval, graph theory, machine learning.</p>

The two-step motion compensation combined squint wavenumber domain algorithm based on fractional Fourier transform

Gewei Tan*, Wei Lin

School of information science and engineering, Huaqiao University, Xiamen, China

Received 1 June 2014, www.tsi.lv

Abstract

Fractional Fourier transform (FrFT) is a kind of generalized Fourier transform, which processes signals in the unified time-frequency domain and the linear frequency modulation signal can be well focused after FrFT. Motion error is an important factor affecting the SAR resolution, conventional wavenumber domain (CWD) algorithm is an ideal solution of SAR focusing problem as long as nominal straight flight track is given, especially in the case of high squint angles and long synthetic apertures, but it has certain limitation in processing airborne SAR data affected by motion error, so extended wavenumber domain algorithm (EWD) is presented. Pointing to the problem that the effect of error elimination is not obvious in processing non-stationary motion error using EWD algorithm, FrFT based the two-step motion compensation combined squint wavenumber domain algorithm is put forward in this paper, which is expected to eliminate the influence of motion error more effectively, so as to obtain high quality SAR images. The simulation results and the imaging results of real SAR data show that the proposed algorithm can eliminate the influence of motion error effectively. (the real SAR data provided by Institute of Electronics, Chinese Academy of Sciences).

Keywords: fractional Fourier transform, motion compensation, wavenumber domain algorithm, extended wavenumber domain algorithm, high resolution

1 Introduction

The system theory and related technology for SAR are established on keeping the flight path of radar platform in a straight line, but the atmospheric turbulence or other natural factors often make the aircraft deviated from the nominal track and generating motion error of antenna phase center (APC), which will cause the amplitude modulation and phase modulation of radar echo signal, result in the image blurring and geometric distortions. In order to obtain high quality SAR images, the motion error must be compensated [1,2].

Conventional wavenumber domain algorithm is an ideal solution of SAR focusing problem as long as nominal straight flight track is given, especially in the case of high squint angles and long synthetic apertures, but it has certain limitation in processing airborne SAR data affected by motion error, so EWD algorithm is presented, which can integrate with two-step motion compensation, but the compensation effect is not obvious when motion errors are non-stationary.

Fractional Fourier transform (FrFT) is proposed by Namias in 1980 [3], which is a new time-frequency analysis tool. Compared to the Fourier transform, FrFT has incomparable superiority in processing non-stationary signals. Chirp signal can be well focused after FrFT of specific rotation angle which provides a possibility to achieve high resolution and high accuracy

in SAR imaging. Especially when signal and interference source are coupled to each other, good separation effect in fractional domain can be obtained [4-6].

There has been literatures putting forward SAR imaging algorithm combining with FrFT. References [7,8] proposed an improving CS imaging algorithm based on FrFT. References [9,10] also proposed a combination of FrFT and RD imaging algorithm. These research results show that FrFT can contribute to the improvement of SAR resolution.

In order to eliminate the influence of motion error more effectively and improve the resolution, in the paper, FrFT is combined with the two step motion compensation technology and SAR imaging algorithm, through utilizing the superiority of FrFT in processing chirp signal and non-stationary signal, so as to obtain high quality SAR images.

2 Fractional Fourier transform

The fractional Fourier transform and its inverse transform for signal $x(t)$ is defined as:

$$X(u) = F_p [x(t)](u) = \int_{-\infty}^{+\infty} x(t) K_p(t, u) dt, \quad (1)$$

$$x(t) = F_{-p} [X(u)](t) = \int_{-\infty}^{+\infty} X(u) K_{-p}(t, u) du, \quad (2)$$

where:

* Corresponding author's e-mail: tangewei70@163.com

$$K_p(t, u) = \begin{cases} \sqrt{\frac{1-j\cot\alpha}{2\pi}} \exp\{j\pi((t^2+u^2)\cot\alpha - 2ut\csc\alpha)\} & \alpha \notin n\pi \\ \delta(t-u) & \alpha \in n \cdot 2\pi \\ \delta(t+u) & \alpha + \pi \in n \cdot 2\pi \end{cases} \quad (3)$$

and $p = \frac{2}{\pi}\alpha$, which is the order of FrFT when $p = 1$, FrFT is the Fourier transform.

The fractional Fourier transform for chirp signal $x(t) = e^{j2\pi kt^2}$ is as follows:

$$X(u) = F_p[x(t)](u) = \int_{-\infty}^{+\infty} x(t) K_p(t, u) dt = A e^{j\pi u^2 \cot\alpha} \int_{-\infty}^{+\infty} e^{j2\pi kt^2} e^{j\pi t^2 \cot\alpha - j2\pi ut \csc\alpha} dt$$

where $A = \sqrt{\frac{1-j\cot\alpha}{2\pi}}$, when $\cot\alpha = -2k$, $p = -\frac{2}{\pi} \arccot(2k) = p_{opt}$ is the optimal order of FrFT. Thus:

$$X(u) = F_p[e^{j2\pi kt^2}] = A e^{-j2\pi ku^2} \int_{-\infty}^{+\infty} e^{-j2\pi ut \csc\alpha} dt = 2\pi A e^{-j2\pi ku^2} \delta(2\pi \csc\alpha u) = \frac{A}{\csc\alpha} \delta(u) = C \delta(u) \quad (4)$$

So the optimal order FrFT for chirp signal is an impulse signal.

3 The motion error model of squint mode and SAR echo signal with motion error

Figures 1 and 2 are geometry models of SAR system under squint mode, in which: θ_{sq} is squint angle of antenna beam; r_{s0} is the range of antenna beam center without trajectory deviation; r_s is the range of antenna beam center with trajectory deviation; x_c is the APC azimuth coordinate; x_0 is the azimuth position of target; θ_s is antenna beam angle.

Based on law of cosines, there is:

$$R_0(x, x_0, r_0) = \sqrt{r_{s0}^2 + (x + x_c - x_0)^2 + 2r_{s0}(x + x_c - x_0)\sin\theta_{sq}},$$

$$R(x, x_0, r_0) = \sqrt{r_s^2 + (x + x_c - x_0)^2 + 2r_s(x + x_c - x_0)\sin\theta_{sq}},$$

where, $r_{s0} = \frac{r_0}{\cos\theta_{sq}}$, $r_s = \frac{r}{\cos\theta_{sq}}$, $r = r_0 + \delta r_0$,

$$r_s = r_{s0} + \delta r_s,$$

$$\delta R(x, x_0, r_0) = R(x, x_0, r_0) - R_0(x, x_0, r_0) \approx \delta r_s - \frac{(x + x_c - x_0)^2}{2r_{s0}^2} \delta r_s,$$

if $(x + x_c - x_0) \ll r_{s0}$,

$$\delta R \approx \delta r_s = \frac{\delta r_0(x, r_0)}{\cos\theta_{sq}}. \quad (5)$$

So motion compensation is determined by the APC position, the closest approach of target and squint angle, and has no relation with target position, therefore, the motion errors of all targets in the same range can be compensated together. Motion error can be divided into range-independent and range-dependent components [11], as following:

$$\delta R(x, r_0) = \delta R_r(x, r_m) + \delta R_{\Pi}(x, r_x).$$

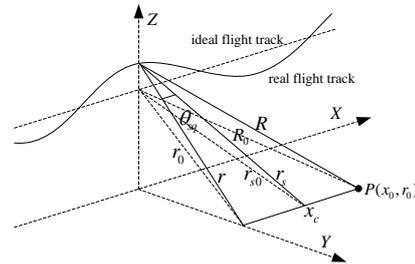


FIGURE1 Geometry model of SAR system under squint mode

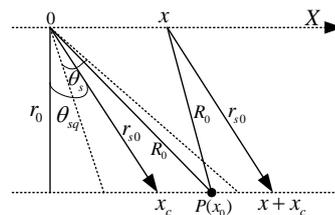


FIGURE2 Geometry relation in range plane

Suppose that SAR transmit the chirps to an observed scene, their echoes after the demodulation are:

$$sd(t, \tau; r_0) = \sigma(x_0, r_0) \text{rect}\left(\frac{\tau - \frac{2R}{c}}{T_p}\right) \exp\left\{-j\pi k \left(\tau - \frac{2R}{c}\right)^2\right\} \text{rect}\left(\frac{t - t_0}{T_s}\right) \exp\left\{-j\frac{4\pi R}{\lambda}\right\}, \quad (6)$$

where τ is fast time in the slant range direction, t is slow time along the radar flight path, c and λ are the speed of light and the radar wavelength respectively, and k is the chirp rate. $\text{rect}(\cdot)$ is the rectangle function, in which T_p and T_s are the pulse duration and synthetic aperture time. R is the instant range from the radar to a

point target in the observed scene, which includes motion error.

Ignore the azimuth motion error, the Fourier transform in the range direction first is performed by the stationary phase point, the result is:

$$sD(t, f_r; r_0) = C_1 \sigma(x_0, r_0) \text{rect}\left(\frac{f_r}{B_r}\right) \exp\left\{j\pi \frac{f_r^2}{k}\right\} \exp\left\{-j\frac{4\pi R_0}{c} f_r\right\} \exp\left\{-j\frac{4\pi(\delta R_l + \delta R_{II})}{c} f_r\right\} \cdot \text{rect}\left(\frac{t - t_0}{T_s}\right) \exp\left\{-j\frac{4\pi R_0}{\lambda}\right\} \exp\left\{-j\frac{4\pi(\delta R_l + \delta R_{II})}{\lambda}\right\} \quad (7)$$

Range-independent motion errors cause the echo envelope delay errors and azimuth phase errors, which can be compensated along with range compress. Range-dependent motion errors is related to slant range. In real SAR data, all point target are spread out in range and azimuth direction, so compensation to such error must be implemented after range compression and the range migration correction is completed and before azimuth compression.

4 Two-step motion compensation combined squint wavenumber domain algorithm based on the FrFT

Stolt interpolation is a key step in CWD algorithm, which not only increases the amount of calculation, but also brings additional error. Fractional Fourier transform can replace the interpolation operation and can transform signal into range-Doppler domain at the same time, thus improving the calculation efficiency and imaging resolution.

Processing flow for the two-step motion compensation combined squint wavenumber domain algorithm based on FrFT is as shown in Figure 3.

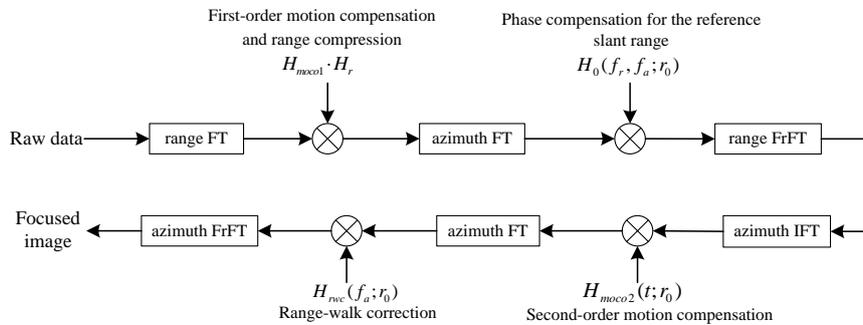


FIGURE 3 Processing flow for the two-step motion compensation combined squint wavenumber domain algorithm based on FrFT

First raw data is transformed into range frequency domain by Fourier transform on the range (shown as 7), then multiplying H_{moco1} and H_r to complete range focusing and first-order motion compensation.

$$H_{moco1} H_r = \exp\left\{j\frac{4\pi\delta R_l}{c}(f_r + f_c)\right\} \exp\left\{-\pi\frac{f_r^2}{K_r}\right\}. \quad (8)$$

Next performing azimuth FT, in the two-dimensional frequency domain, the signal is:

$$SD(f_a, f_r; r_0) = C_2 \sigma \text{rect}\left[\frac{f_r}{B_r}\right] \text{rect}\left[\frac{f_a}{B_d}\right] \exp\{j\psi(f_a, f_r; r_0)\}, \quad (9)$$

where $\psi(f_a, f_r; r_0) = \psi_0(f_a, f_r; r_m) + \psi_1(f_a, f_r; r_0)$,

$$\psi_0(f_a, f_r; r_m) = -\frac{4\pi r_m}{\lambda} \sqrt{\left(1 + \frac{f_r}{f_c}\right)^2 - \left(\frac{\lambda f_a}{2v}\right)^2}, \quad (10)$$

$$\psi_1(f_a, f_r; r_m) = -\frac{4\pi\delta r}{\lambda} \sqrt{\left(1 + \frac{f_r}{f_c}\right)^2 - \left(\frac{\lambda f_a}{2v}\right)^2},$$

where $r_0 = r_m + \delta r$, $\delta r \in \left[-\frac{w_r}{2}, \frac{w_r}{2}\right]$, r_m is the slant range of swath center, w_r is the swath width.

The Taylor series expansion around the Doppler frequency f_a for ψ_0 is:

$$\psi_0(f_a, f_r; r_m) = -\frac{4\pi r_m}{c} \left[f_c \gamma(f_a) + \frac{f_r}{\gamma(f_a)} - \frac{1 - \gamma(f_a)^2}{2f_c \gamma(f_a)^3} f_r^2 + \frac{1 - \gamma(f_a)^2}{2f_c \gamma(f_a)^5} f_r^3 + \dots \right] \quad (11)$$

First item of 11 corresponds to the azimuth compression, second to range cell migration, third to second range compression, the fourth is high-order coupling item of range and azimuth.

In squint mode, motion error is projected in the direction of antenna beam centre, the minimum

$$\psi_0(f_a, f_r; r_m) = -\frac{4\pi r_m}{\lambda} \left[\sqrt{\left(1 + \frac{f_r}{f_0}\right)^2 - \left(\frac{\lambda f_a}{2v}\right)^2} - \sqrt{1 - \left(\frac{\lambda f_a}{2v}\right)^2} - D(f_a - f_{DC})f_r \right],$$

$$\psi_1(f_a, f_r; r_m) = -\frac{4\pi \delta r}{\lambda} \sqrt{\left(1 + \frac{f_r}{f_0}\right)^2 - \left(\frac{\lambda f_a}{2v}\right)^2} - \frac{4\pi r_m}{\lambda} \left[\sqrt{1 - \left(\frac{\lambda f_a}{2v}\right)^2} + D(f_a - f_{DC})f_r \right],$$

where $D = \frac{\lambda \sin \theta_{sq}}{f_0 \cos^3 \theta_{sq}}$, f_{DC} is Doppler centroid,

$D(f_a - f_{DC})f_r$ is range-walk, which is obtained from the RCM expression in frequency domain [13]. Multiplying 9 with:

$$H_0(f_a, f_r; r_m) = \exp\{-j\psi_0(f_a, f_r; r_m)\} \quad ,$$

$$Sd(t, \tau; r_0) = \int_{-\infty}^{+\infty} [SD(f_a, f_r; r_0)H_0(f_a, f_r; r_m)]K_p(f_r, \tau)df_r =$$

$$C_3 \text{rect} \left[\frac{f_a}{B_d} \right] \int_{-\infty}^{+\infty} \text{rect} \left[\frac{f_r}{B_r} \right] \exp\{j\psi_1(f_a, f_r; r_0)\} \exp\{j\pi((f_r^2 + \tau^2) \cot \alpha - 2f_r \tau \csc \alpha)\} df_r = \quad (12)$$

$$C_4 \sigma \text{rect} \left[\frac{f_a}{B_d} \right] \exp\left\{-j \frac{4\pi r_0}{\lambda} \gamma(f_a)\right\} \exp\left\{-j \frac{4\pi r_0}{\lambda} D(f_a - f_{DC})\right\} \sin c\left(\tau - \frac{2r_0}{c}\right),$$

where $\alpha = -\cot^{-1}\left(\frac{2\delta r}{c} \frac{1 - \gamma(f_a)^2}{f_c \gamma(f_a)^3}\right)$, the optimal order

of FrFT is $p_{opt} = \frac{2}{\pi} \alpha$.

Next, performing azimuth IFT into two-dimensional time-domain, implementing second-order motion compensation, the corresponding phase compensation function is:

$$H_{moco2}(t, r_0) = \exp\left\{-j \frac{4\pi}{\lambda} \Delta R_H\right\}, \quad (13)$$

In range-Doppler domain, completing range-walk correction, and the compensation function is

approximate error of 5 is in the direction of range-walk (squint angle), so the second order motion compensation should be performed after correction of range-curve and before correction of range-walk [12], so Equation (10) is changed as:

can complete range-curve correction, second range compression and phase compensation for high-order coupling item of range and azimuth for the reference range. Then performing FrFT on the range so that completing residual range-curve correction and change the signal into range-Doppler domain:

$H_{rwc}(f_a, r_0) = \exp\left\{j \frac{4\pi r_0}{\lambda} D(f_a - f_{DC})\right\}$. In order to

obtain azimuth chirp signal, operating the Taylor series expansion around the Doppler frequency f_a for $\gamma(f_a)$,

$$\gamma(f_a) = \sqrt{1 - \left(\frac{\lambda f_a}{2v}\right)^2} \approx 1 - \frac{1}{2} \left(\frac{\lambda f_a}{2v}\right)^2.$$

Finally, performing FrFT on the azimuth by the optimal order $p_{opt} = \frac{2}{\pi} \beta = -\frac{2}{\pi} \text{arccot}\left(\frac{\lambda r_0}{2v^2}\right)$, the result

is:

$$sd(t, \tau; r_0) = \int_{-\infty}^{+\infty} [Sd(f_a, f_r; r_0) H_{rvc}(f_a, r_0)] K_p(f_a, \tau) df_a =$$

$$C_4 \sigma \sin c\left(\tau - \frac{2r_0}{c}\right) \int_{-\infty}^{+\infty} \text{rect}\left[\frac{f_a}{B_d}\right] \exp\left\{-j \frac{4\pi r_0}{\lambda} \gamma(f_a)\right\} \exp\left\{j\pi\left((f_a^2 + t^2) \cot \beta - 2f_a t \csc \beta\right)\right\} df_a =$$

$$C_5 \sigma \sin c\left(\frac{t}{\sin \beta}\right) \sin c\left(\tau - \frac{2r_0}{c}\right).$$
(14)

4 Simulations and SAR imaging results based on real data for the algorithm

Simulations of point target with motion error for the algorithm put forward in this paper is as follow, the simulation parameters as shown in Table 1.

The total error measured is non-stationary random motion error whose mean value is exponential function, so range-variant motion error is $\delta R_{11}(x, r_x) = \delta R(x, r_0) - \delta R_1(x, r_m)$.

TABLE 1 Simulation parameters

Parameter	value	Parameter	value
Carrier frequency	1.5GHz	Forward velocity	180m/s
Bandwidth for transmit signal	150MHz	Length of synthetic aperture	320m
Pulse duration of transmit signal	1.5μs	The number of azimuth sampling points	512
The number of range sampling points	1024	Range-invariant motion error	t ²

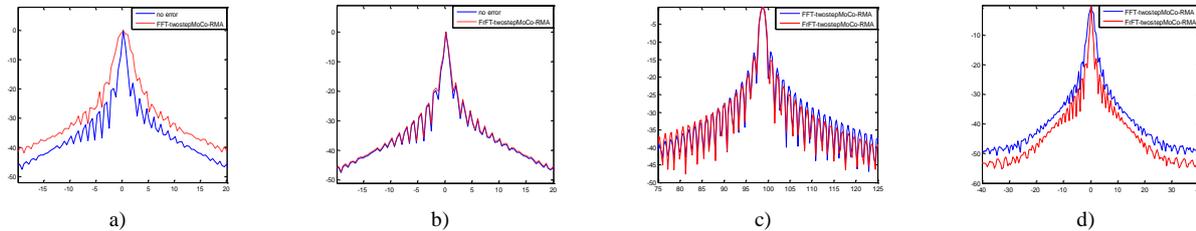


FIGURE 4 Comparing of azimuth impulse response between no error and with error after compensation

a) Comparison of azimuth impulse response with no error and after motion compensation using EWD algorithm; b) Comparison of a azimuth impulse response with no error and after motion compensation using the proposed algorithm; c) Comparison of range impulse response after motion compensation by the two algorithm; d) Comparison of azimuth impulse response after motion compensation by the two algorithm.

Simulation to point target with motion error using FFT based the EWD algorithm and FrFT based the two-step motion compensation combined squint wavenumber domain algorithm are shown in Figure 4. Figure 4b and d) shows that the main lobe of impulse response in azimuth direction is narrower and the influence of quadratic phase error is eliminated virtually. The wider mainlobe of Figure 4a indicates the

influence of quadratic phase error still exists, which proves the processing effect of non-stationary motion error using FrFT is obvious than FFT, at the same time proves that the focusing effect of chirp signal using FrFT is better than FFT. The performance comparison of FFT-EWD and FrFT-EWD algorithm is shown in Table 2.

TABLE 2 performance comparison of impulse response in azimuth direction

Condition and processing algorithm	Main lobe	ISLR	PSLR
No error	1.10m	-26.012 dB	-31.432dB
FFT-EWD algorithm processing under motion error	4.864m	-24.681dB	-24.218dB
FrFT-EWD algorithm processing under motion error	1.125m	-25.589dB	-30.353dB

Figure 5a is the imaging results of the real SAR data with motion error, the blurring image shows there are obvious quadratic phase errors. Processing result for such SAR data using the traditional FFT based EWD algorithm is shown in Figure 5b, due to the elimination of most motion error, image resolution is improving significantly, but in some places with more details

(such as part of the circle line), the image is not clear and image resolution deteriorates because of the residual phase error. The processing result with the proposed FrFT based two step motion compensation squint wavenumber domain algorithm is shown in Figure 5c, where details information increases, the image resolution is further improved.

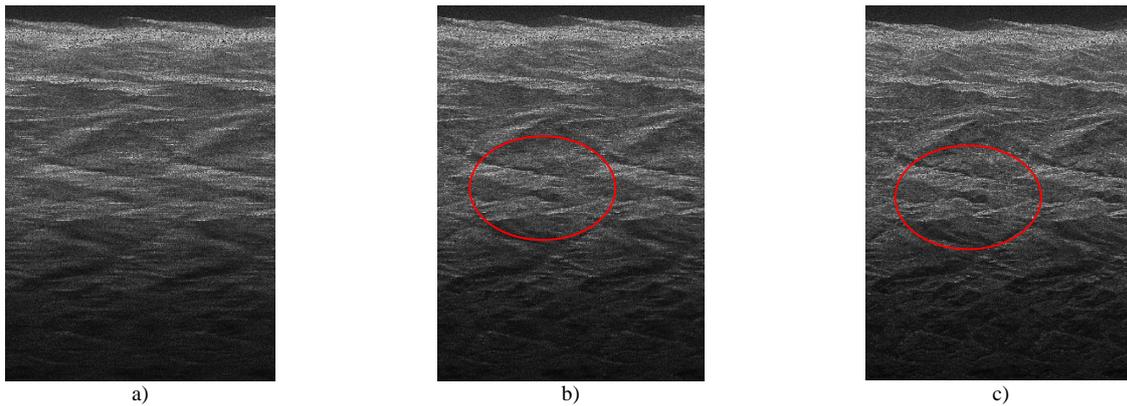


FIGURE 5 comparison of imaging results before and after motion compensation (resolution: $3m \times 3m$)
 a) SAR image with motion error; b) SAR image after FFT based the EWD algorithm processing; c) SAR image after FrFT based the two step motion compensation combined squint wavenumber domain algorithm processing

5 Conclusions

Motion error is a crucial factor to limit airborne SAR resolution improving. As modern SAR systems are continuously developing into the direction of higher spatial resolution, how to overcome the motion error caused by air turbulence is an urgent problem to radar

workers. The two-step motion compensation combined squint wavenumber domain algorithm based on fractional Fourier transform proposed in this paper can solve various motion errors effectively, especially can eliminate the image blurring caused by non-stationary motion errors, the research provides an effective solution scheme for squint SAR data with motion error.

References

- [1] Li Y, Liang X, Ding C, Zhou C, Chen L, Hong W 2012 A Motion Compensation Approach Integrated in the Omega-K Algorithm for Airborne SAR 2012 *IEEE International Conference on Imaging Systems and Techniques* 245-8
- [2] Mao Y, Xiang M, Wei L, Li Y, Hong W 2012 Error Analysis of SAR Motion Compensation 2012 *IEEE International Conference on Imaging Systems and Techniques* 377-80
- [3] Namias V 1980 The Fractional Fourier Transform and Its Application in Quantum Mechanics *IMA Journal of Applied Mathematics* 25(3) 241-65
- [4] Qi L, Tao R, Zhou S, Wang Y 2002 Adaptive Time-Varying Filter for Linear FM Signal in Fractional Fourier Domain *IEEE 6th International Conference on Signal Processing* 2 1425-8
- [5] Durak L, Aldirmaz S 2010 Adaptive Fractional Fourier Domain Filtering *Signal Processing* 90(4) 1188-96
- [6] Sejdić E, Djurović, Stanković 2011 Fractional Fourier Transform as a Signal Processing Tool: An Overview of Recent Developments *Signal Processing* 91(6) 1351-69
- [7] Amein A S, Soraghan J J 2005 *IEEE Signal Processing Letters* 12(10) 705-8
- [8] Amein A S, Soraghan J J 2006 *IEEE Transactions on Geosciences and Remote Sensing* 44(10) Part 2 2871-9
- [9] Clemente C, Soraghan J J 2012 Range Doppler and chirp scaling processing of synthetic aperture radar data using the fractional Fourier transform *IET Signal Processing* 6(5) 503-10
- [10] El-Mashed M G, Dessouky M I, El-Kordy M, Zahran O, Abd El-Samie F E 2012 Target Image Enhancement in Radar Imaging Using Fractional Fourier Transform Sensing and Imaging An international Journal 13(1) 37-53
- [11] Lanari R, Fornaro G 1997 *IEEE Transactions on Geoscience and Remote Sensing* 35(6) 1446-52
- [12] Fornaro G, Franceschetti G, Perna S 2004 Motion compensation of squinted airborne SAR raw data: role of processing geometry 2004 *IEEE International Geoscience and Remote Sensing Symposium* 2 1518-21
- [13] Tan G, Deng Y 2009 A kind of method of extended wavenumber domain algorithm for squint SAR motion compensation *Journal of electronics and information technology* 31(1) 156-9 (in Chinese)

Authors



Gewei Tan, born in January, 1970, Guiyang, China

Current position, grades: Doctor of Communication Engineering, Chair of Communication Engineering Department in Huaqiao University.
University studies: Bachelor's degree in physical electronics (1987-1991, Zhejiang University). Doctor's degree on communication & information system (2003-2008, Institute of Electronics, Chinese Academy of Sciences).

Scientific interest: SAR signal processing and motion compensation technology.

Publications: 15 Papers

Experience: Teaching experience of 9 years 10 scientific research projects



Wei Lin

Current position, grades: Master at communication engineering, lecturer at the communication engineering at Huaqiao University.

University studies: Bachelor's degree on microelectronics (2002-2005, University of Electronic Science and Technology of China). Master's degree on signal processing (2005-2008, University of Electronic Science and Technology of China).

Scientific interest: signal process and SAR motion compensation.

Publications: 8 Papers

Research on grid replacement technology and two-dimensional isothermal simulation on melt flow of blown film

Chao Wang*

College of Computer Science, South-central University for nationalities Hubei Wuhan, 430074, China

Received 1 June 2014, www.tsi.lv

Abstract

Established a melt flow model of two-dimensional isothermal simulation for spiral mandrel blow-film die, analysed die swell-effect at the die exit of blown film extrusion process by grid replacement technology, obtained the pressure distribution in the internal flow of blow-film die, and analysed the shape change of free surface at die extrusion section with different viscosity materials. The results showed that the resistance force of the melt flow under internal flow of blow-film die mainly from the spiral direction, with depth of spiral groove gradually shoaled, the direction of resistance force changed from spiral direction to axial direction. When high viscoelastic material entered blow-film die being extruded, the shape of free surface at die extrusion section would change significantly, meanwhile, the gradient of radial velocity of melt flow in the internal die became gradually reduced.

Keywords: spiral mandrel blow-film die, grid replacement, die swell-effect, free surface

1 Introduction

The technology of extrusion blown film from thermoplastic plastic began in the thirties. The first device of polyethylene tubular film was produced in the United States about in 1939. Today, this production method is widely used by people. This was inseparable with multifaceted applications of polyethylene. More than 90% of the low density polyethylene films between a few centimetres to several meters in width, 0.02 to 0.3 millimetres in thickness are produced by the blow mould. Because of the economical of this method, it can not only make the products have great change, but also obtained some finishing effect through the same process. This is particularly important for the use of thin films.

With the continuous development and industry technology upgrade as well as people continuously improve the diverse needs of packaging products, the late nineteen eighties, multi-layer co-extruded high barrier film was developed successful. This material has a permeability resistance, oil resistance, can retort resistance, heat sealing, etc., can be used in various types of food and cosmetics packaging. It greatly extended the product's shelf life. Thus, the plastic film for packaging had a rapid development trend towards the functional, lightweight and greenery [1-4].

However, due to the polymer molding process exhibit complex viscoelastic behaviours, the research of the melt flow in the die of the internal flow is relatively slow, using the traditional method of mathematical analysis and experimental methods have greater limitations. With the maturing of CAE technology, more and more scholars have conducted to basic research; the problems of its reliability and calculation effecting have been resolved.

In 1982 PLYFLOW was developed at the University of Louvain, Belgium. In 1997 was acquired by the world famous fluid analysis software company FLUENT. It is suitable for plastics, resins and other polymer materials, extrusion molding, blow molding, drawing, laminar mixing, the coating process of the flow and heat transfer and chemical reaction problems. It may also be used to simulate the polymer flow problems, such as a polymer melt, oil, ink and suspended solids flow simulation [5-7].

In this paper, the finite element method and polyflow software was applied to the planar structure of spiral mandrel die by two-dimensional numerical simulation, explored using grid replacement technical to analyse the internal flow of the melt in the die and the effect of die swell-effect to the morphology of the die extrusion section.

2 Calculation model

2.1 GEOMETRY

Analysis model adopt the spiral mandrel die as a basis which has been widely used in co-extrusion composite technology. As the complex shape of the internal structure, modelling and design would be more difficult, using the POLYFLOW software to do two-dimensional planar analysis for die profile. As shown in Figure 1.

Spiral mandrel die using central feeding method, can ensure the melt suffered shear force equal in the cross-section of each point, die gap uniformity, metl film speed stability.

* *Corresponding author's* e-mail: wangchaofly@126.com

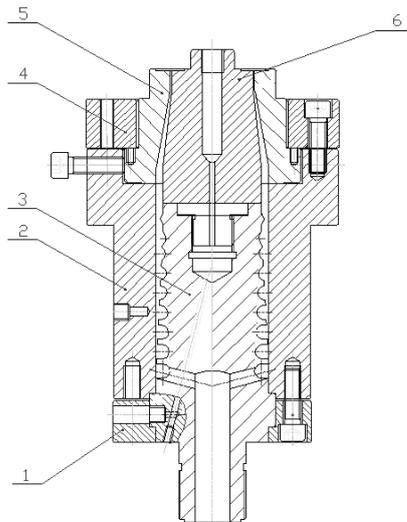


FIGURE 1 The internal flow channel of die head: 1. lower gland; 2. outer mold body; 3. spiral body; 4. upper gland; 5. adjust ring; 6. die

In the blown film process, the melt flow through spirochetes channel then being squeezed forcing through a die, so when to do the simulate, the outer mold body, the spiral body, the flow channel section and the die extrusion part (POLYFLOW set as the free surface) should deal with as two separate parts. Therefore, the thickness of extrusion section for free surface is taken of 2mm. The outer die body, upper and lower gland and adjustment ring can be neglected in simulation due to do not participate in the melt flow process.

2.2 MATHEMATICAL MODEL

In order to simplify the calculation, to do the assumptions of internal flow of melt in the die [8-10]. Assuming the melt in the die internal flow was isothermal flow, steady-state process and melt was incompressible, the continuity equation is:

$$\nabla v = 0. \tag{1}$$

Ignore the melt gravity and inertial force.

The momentum equation is:

$$\nabla \tau - \nabla P = 0, \tag{2}$$

where v is velocity vector, τ is stress tensor, P is pressure.

The polymer melt flow in spiral mandrel die, the changes of internal structure can cause the change of viscosity. Therefore, select Carlo model of generalized Newtonian isothermal flow as model. Carlo model uses five parameters model.

$$\frac{\eta - \eta_\infty}{\eta_0 - \eta_\infty} = \frac{1}{[1 + (\lambda\dot{\gamma})^a]^{\frac{1-n}{a}}}, \tag{3}$$

where, η_0 is zero shear viscosity; η_∞ is when $\dot{\gamma}$ tend to be very large the polymer shears thinning reach to another equilibrium viscosity, $\dot{\gamma}$ is shear rate, λ is

relaxation time, n as a parameter, λ and n were not changed with $\dot{\gamma}$. For many polymer fluid, when $\dot{\gamma}$ increases to a certain degree, the macromolecular chain prone to degradation, therefore η_∞ can take zero.

If $a = 1$, the equation can be simplified as Cross law model. Its expression is as follows

$$\eta = \frac{\eta_0}{1 + (\lambda\dot{\gamma})^m}. \tag{4}$$

When calculating, the value of n was 85000, λ was 0.2, m is cross law index, the value was 0.3.

3 Finite element simulation

3.1 MESHING

The limited release of extrusion will swell causing the surface position change, namely the die swell effect. In the die because of the internal restriction of mold wall, melt do no slip flow. After the extrusion, extrusion shape will be changed, causing the location of the free boundary changes and boundary mesh distortions. If still use the mesh form before, may affect the calculation accuracy, and even make the problem does not converge. In order to keep the grid still has good computational characteristics, need to adopt the grid replacement technology as the assistant to divide free surface mesh. When movement occurs in the free boundary, it would recalculate the distribution of nodes near the free boundary.

The grid replacement is the main method of treatment on free boundary node position, by continuously iteratively determining each node position to compose the final free surface shape.

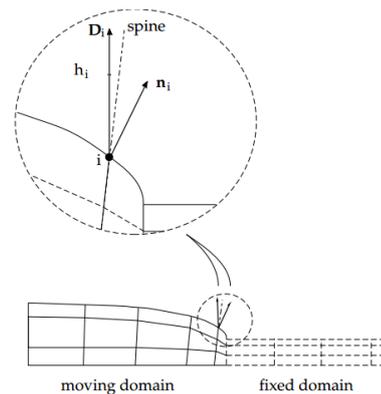


FIGURE 2 Vector relation between moving direction of free surface and other direction

The free movement of the boundary of node i can be determine by the outer normal vector n_i , direction vector D_i and the displacement of direction vector h_i , as shown in Figure 2. The outer normal vector is the vector perpendicular to the boundary. Direction vector refers to the actual direction of motion of vector node. Meanwhile, the kinematics equation set on the free boundary

constrains the normal displacement of nodes and the tangential displacement of nodes. Therefore, the core content of the grid replacement technology is to adjust the boundary tangential displacement and internal node position to minimize the grid deformation. By analysing and calculate these unknown quantity, can determine the position of node movement.

Due to the symmetry of the flow channel model, when computing select the 1/2 of the channel that between the outer die body and the spiral body to mesh. Use gambit software for modelling and meshing. Considering the complexity of the melt flow in the channel, creating quadrilateral mesh structure, and in the irregular parts in the channel was divided into multiple faces, in the specified edge at the interface to do the bilateral format classification, when meshing is completed then connect the face that belong to same sub-domain.

This two-dimensional model after meshing contains 282 nodes, 332 units, which contain 102 wire unit and 230 tetrahedral elements.

3.2 BOUNDARY CONDITIONS AND PHYSICAL PARAMETERS

The boundary in the model included the flow entrance, the flow exit, the wall boundary condition, the free surface boundary conditions, the rotating boundary conditions. Among them, assuming that the fluid is stationary in the wall, that is, fluid adhere to the wall surface and don't slip at the fluid solid interface. Specify the entrance volume flow rate was 18cm³/s. Export circumferential force was 0.

The solving order was used gambit software to pre-processing as modelling at first, then meshing, after that was sub- regional division, the die and the free surface of extrusion is divided into two independent sub region, finally specify the model boundary conditions and symmetry axis. After pre-processing the grid file were solved by polyflow software, the results were deal with flunt/post software to do post- processing.

Polyflow software provides a variety of mesh replacement method, can be applied to the 2D, 3D, blow mould extrusion and spinning of different flow problems. Here are solved by two-dimensional simulation which is more suitable for the spine, Its characteristic is the grid nodes along the spine to adjust distribution. Spinal method requires defining the initial spinal cord and the terminate spinal cord, the region that limited to use the mesh replacement method implement the normal direction slice to obtain the spinal cord for two-dimensional calculation area contains a free boundary. This method does not work on free boundary tangential adjustment, taking small amount of computation. The spinal cord is the node connecting line, does not require a straight line, as shown in Figure 3. There were one or two endpoints exist in free boundary.

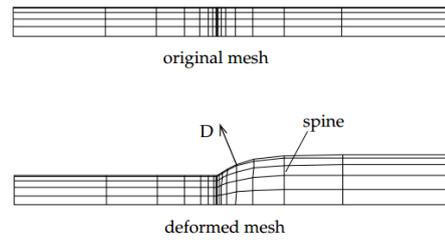


FIGURE 3 Meshing schematic by spine method

Consider a model of two spinal cord line consisting of point X₁ and X₂, when the free boundary position changed, the adjustment of internal nodes are as follow:

$$\delta x_i = w_{1i} \delta x_1 + w_{2i} \delta x_2, \tag{5}$$

$$w_{1i} = \frac{|x_i - x_2|}{|x_2 - x_1|}, \tag{6}$$

$$w_{2i} = \frac{|x_i - x_1|}{|x_2 - x_1|}. \tag{7}$$

Among them δx is the position variable.

4 Results and discussion

4.1 PRESSURE DISTRIBUTION

Figure 4 shows the contour map of melt flow pressure after mirroring the symmetry axis. The figure shows the main pressure drop of melt in the extrusion direction close to the die head.

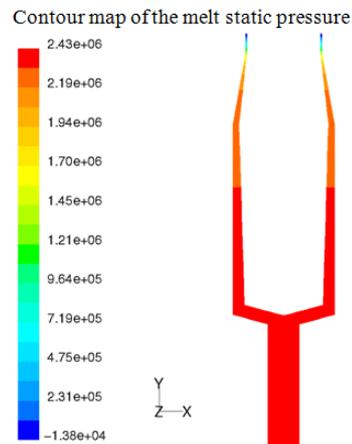


FIGURE 4 contour map of the melt static pressure

This phenomenon is mainly due to the design of spiral mandrel die flow of internal channel. In the process of melt flow, the melt flow along the helical direction in the spiral groove. At the same time by the effect of extrusion direction force and spiral tangent force, the flow resistance was large, but the volume of spiral groove in die head along the extrusion direction becomes gradually smaller, the thickness of annular gap increased larger. In the flow process the melt in the spiral groove continuous flow in the annular region. Finally, the spiral groove closed to the extrusion direction in the die head

completely disappeared, the melt was extruded along the extrusion direction, at this time the pressure at the die is smaller.

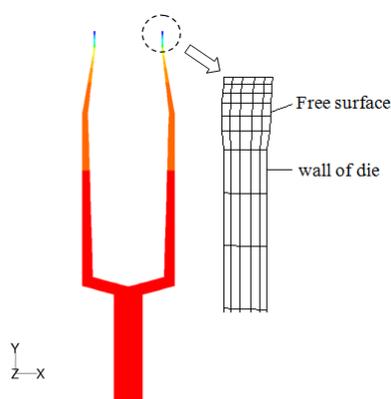


FIGURE 5 The shape of free surface in extrusion segment using the grid replacement method

Figure 5 shows when using the grid replacement method, the presentation of segment meshing of free surface and the die head exit. From the figure shows the shape of the free surface of the fluid as compared with the original shape has been changed. The extrusion segment remained 2mm thickness, but the shape of it had already apparent swell change.

4.2 EXTRUSION SWELL FLOW CHARACTERISTICS

For further study of melt viscoelasticity on outlet flow effect of extrusion swell form, requiring use the asymptotic method to analyse the model.

In the viscoelastic flow, due to the shear thinning and melt elasticity, the constitutive equations are highly nonlinear. Parameter asymptotic method is the simple and effective method to solve nonlinear problems. The parameter We (Weissenberg number) of viscoelastic fluid constitutive equation is the parameter to determine nonlinear problem. When We is higher, the nonlinear will stronger. When $We = 0$, the problem degenerates into Newton fluid linear problems. A typical test result of nonlinear problems can be used as the initial value of asymptotic function equation. After iteration computing converge the problem, the non-linear parameters will be obtained [10-13].

The previous simulation has proved that the meshing of Newtonian isothermal model was reasonable. Therefore, only changes the differential viscoelastic FENE-P as a mathematical model for fluid analysis. It would obtain the effect for viscoelasticity of fluid to die

swell morphology in different relaxation times near the die, as shown in Figure 6.

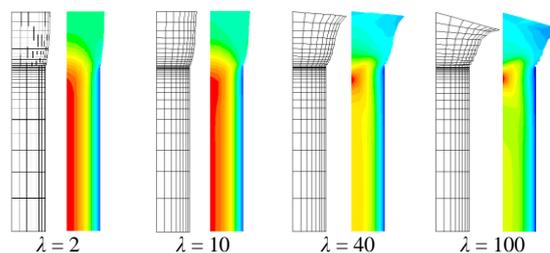


FIGURE 6 The contour map of meshing grid and Y-axis velocity in different relaxation times

For the analysis of die swell conditions in high viscoelastic state, change the relaxation time parameter in the algorithm. Draw the grid pattern of λ equal to 2, 10, 40, 100 and contour graphic of axial velocity in Y-axis direction. From the figure shows, when the value of λ was two, the fluid was approximated to the Newton fluid, the greater its value shows the fluid had stronger viscoelasticity. With the increased viscoelasticity of fluid, the flow state of free surface in extrusion segment had undergone significant change. The radial velocity gradient in the die decreased. Therefore, the effects of flow characteristics with viscoelasticity of different raw materials to the die extrusion segment are also different. In the film blowing process, by the blowing of wind ring and cooling control technique can ensure the forming film thickness are uniform of different raw materials.

5 Conclusion

Using the grid replacement technique to analyse the two-dimensional isothermal blown film extrusion process can simulate the swell effect of melt of extruded section, and to predict the shape changing of the free surface.

In the process of melt flow inside the die, since the volume of the spiral groove became gradually smaller, the thickness of the annular gap continuous increased, the suffered resistance of flow direction also gradually changed from spiral hoop direction to axial direction which along the extrusion direction. The melt pressure at the entrance of spiral groove was larger, then gradually became smaller along the extrusion direction, the spiral groove at the extrusion section were completely disappeared, the melt was extruded in the axial direction.

When the fluid which had strong viscoelastic was extruded, the shape of free surface in the extrusion segment would significantly change, the radial velocity gradient inside the die presented the gradual decreasing trend.

References

- [1] Callari J 1991 Tight layflat offered by bubble control system *Plastics World* 49(5) 27-8
- [2] LI S, Luo J, Gao L, Mao L 2011 Study on Processing and Property of Three-layer Co-extrusion Barrier films *China Plastics* 25(10) 20-3
- [3] Qi L, Liu H, Huang X The co-extrusion technology and its application *Engineering Plastics Application* 37(4) 83-7
- [4] Zhong Y, Wang Q, Xie P, Yang W 2011 Research Progress on preparation Methods of Barrier Polymer Composites *Plastics Science And Technology* 39(7) 103-6
- [5] Wei X, Wang X, Lin H 2008 The application progress in polymer molding by Polyflow software *Shanghai Plastics* 4 1-5

- [6] Zatloukal M, Tzoganak C, Perdikoulis J Saha P 2001 Numerical simulations of polymer flow in flat spiral dies *Polymer Engineering and Science* **41**(10) 1683-94
- [7] Miao L, Zhang Y, Xue P 2010 Structure Improvement and Development of Coextrusion Dies for Multi-layer Plastics-Films *China Plastics* **24**(2) 11-20
- [8] Gifford W A 2000 A three-dimensional Analysis of Coextrusion in a Single Manifold Flat Die *Polymer Engineering and Science* **40**(9) 2095-2100
- [9] Ma X, Li L 2009 Numerical Simulation on Melt Flow Characteristic of Flat Spiral Type Die for Blown Film *Plastics* **38**(3) 89-91
- [10] Elkoun S, Huneault M A, McCormick K, Puterbaugh F, Kale L 2005 LLDPE-Based Mono-and Multilayer Blown Films: Effect of Processing Parameters on Properties *Polymer Engineering and Science* **45**(9) 1214-20
- [11] Liu X 2011 Study on Preparation of Multi-layer Co-extruded PVDF cast Composite Film *Beijing: Beijing University of Chemical Technology*
- [12] Shi T, Wu D 2011 Foundation of polymer rheology *Beijing: Chemical Industry Press*
- [13] Verbeeten W, Peters G W M, Baaijens F P T 2002 Viscoelastic analysis of complex polymer melt flows using the extended pom-pom model *Journal of Non-Newtonian fluid mechanics* **108**(1-3) 301-26

Authors**Chao Wang**

Current position, grades: lector at the South-central University for nationalities China.

Scientific interest: automatic control technology, industrial intelligent control, computer simulation of polymer blow-molding process.

Analysis of cylindrical cam molded surface

Liyang Liu^{1*}, Chao Li², Qi Zhang³

¹School of Mechanical Engineering Shenyang Ligong University, Mechanical Engineering, Shenyang, China

²Shenyang Mint, Shenyang, China

³State Grid Yingkou Electric Supply Company

Received 12 June 2014, www.tsi.lv

Abstract

The mold surface of cylindrical cam is very complex. There are some questions in designing and using by unfolding picture of cylindrical cam. The thesis infers and builds the mathematical models of cylindrical cam through the theoretical analysis from the nature of space motion cylindrical cam. The models provide the theoretical reference of right projecting, manufacturing and examining.

Keywords: axial cam, spiral surface, equation

1 Introduction

Cam mechanism is a kind of typical facility. Owing to realizing expected motion with simply structure, cam is widely used in all kinds of machinery. Although cam is designed easily and moving perfectly, it is difficult to ensure the accurate process of cam molded surface, especially the cylindrical cam. The topic examines the designing and using of cylindrical cam surface.

2 The coordinates of the cutter and workpiece

Cam path of cylindrical roller includes arc, spiral line and transition arcs. The arc on cylindrical cam is a line perpendicular to cylindrical axis in unfolding picture. The spiral line on cylindrical cam is a line tilted to the cylindrical axis and the dip angle is spiral angle.

Cylindrical cam's molded surface is very complex as shown in the Figure 1. It consists of four arc paths and four spiral paths.

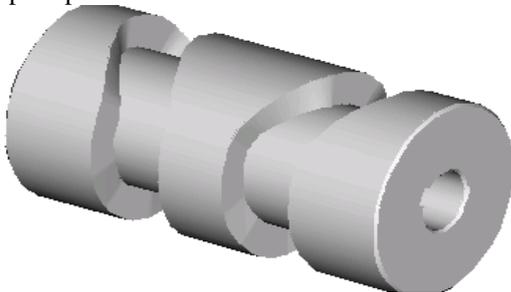


FIGURE 1 cylindrical cam

In order to design and analysis, the length of unfolding picture expanded from cylindrical cam excircle is $2\pi r$ (r is a radius of excircle). In expanded picture, transverse coordinate indicates axial location of cam path. The track of roller centre line is the theoretical line of

cam. The practical lines of cam path are two lines which are round of roller whose centre is on the graph. The Figure 2 is the unfolding picture of cylindrical cam. In this picture, the centre line of arc path which is horizontal line and the centre line of spiral path which is bias line. The completely cam path sees lines forming by cylindrical surface, when cylindrical roller is moving along the centre of unfolding picture.

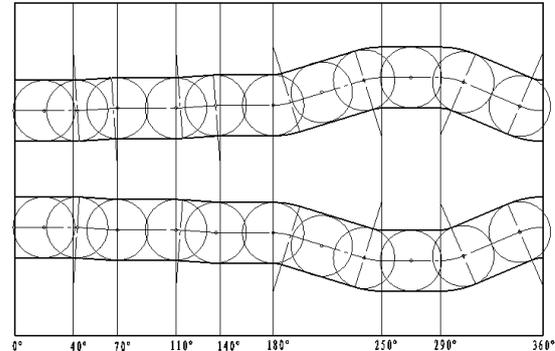


FIGURE 2 Expanded view along conductor rail of cylindrical cam

In order to provide a comprehensive analysis, it is necessary to establish two coordinate systems for cutter and workpiece to display space relation in the Figure 3.

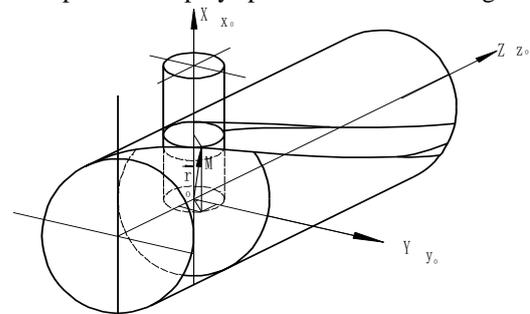


FIGURE 3 Coordinate of cutter and work (main body's relation in space)

* Corresponding author e-mail: liuliyang0806@163.com

Workpiece coordinate system is $O-X Y Z$, cutter coordinate system $O-X_0 Y_0 Z_0$.

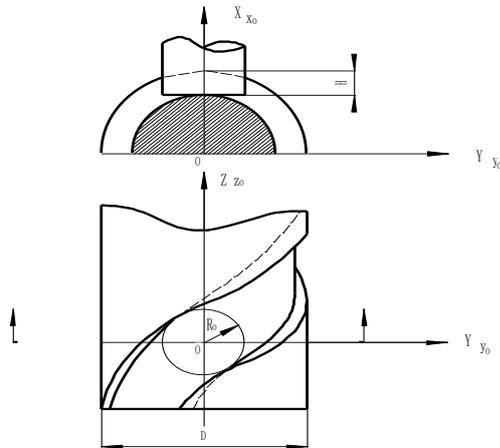


FIGURE 4 Relationship of plane projection of cutter and work

X_0 -axis of cutter is perpendicular to Z -axis of workpiece. The projection relationship of cutter and workpiece is displayed in the Figure 4. From the Figure 5, the equation of the cutter surface of revolution which is cylindrical is that:

$$\begin{cases} X_0 = R_0 \times c \tan \alpha \\ Y_0 = R_0 \times \cos \varphi \\ Z_0 = R_0 \times \sin \varphi \end{cases} \quad (1)$$

where R_0 -roller radius, α, φ - parameters. Having one α angle and one φ angle can confirm point M on the cylindrical surface. From Y_0 -axis, anticlockwise is positive and clockwise is negative.

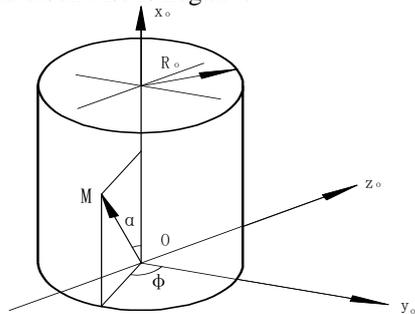


FIGURE 5 Revolution surface of cutter (in the same coordinates as Figure 3)

3 The contact line equation of cutter surface revolution and workpiece spiral surface

The contact condition of cutter revolution surface and workpiece spiral surface is displayed in Figure 3. The radius vector of point M which is relative to origin O of cutter and workpiece coordinate system is showed as follows.

To cutter:

$$OM = \vec{r} = x_0\vec{i} + y_0\vec{j} + z_0\vec{k}. \quad (2)$$

To workpiece:

$$OM = \vec{r} = x\vec{i} + y\vec{j} + z\vec{k}. \quad (3)$$

Now $x_0 = x, y_0 = y, z_0 = z, \vec{i}, \vec{j}, \vec{k}$ are the vectors of X -axis, Y -axis, Z -axis.

If rotation angular velocity of cutter revolution surface and workpiece are ω_0 and ω_1 , the line velocity of point M with cutter moving is showed as follows:

$$\vec{V}_0 = \omega_0(\vec{i} \times \vec{r}). \quad (4)$$

The line velocity of point M with workpiece moving is showed as follows:

$$\vec{V}_1 = \omega_1(\vec{k} \times \vec{r} + p \times \vec{k}). \quad (5)$$

P-spiral parameter $p = \frac{p_z}{2\pi}$ (p_z -helical pitch).

So the speed of relative movement is showed as follows:

$$\vec{V}_x = \vec{V}_1 - \vec{V}_0 = \omega_1(\vec{k} \times \vec{r} + p \times \vec{k}) - \omega_0(\vec{i} \times \vec{r}). \quad (6)$$

As the contact condition at point M of cutter revolution surface and workpiece spiral surface is $\vec{V}_x \perp \vec{n}$, then:

$$\vec{V}_x \times \vec{n} = 0. \quad (7)$$

$$\text{So } \omega_1(\vec{k} \times \vec{r} + p \times \vec{k}) \times \vec{n} - \omega_0(\vec{i} \times \vec{r}) \times \vec{n} = 0.$$

As cutter revolution surface is known and the line velocity \vec{V}_0 is vertical with normal \vec{n} , $\omega_0(\vec{i} \times \vec{r}) \times \vec{n} = 0$. Then the contact condition is showed as follows:

$$(\vec{k} \times \vec{r} + p \times \vec{k}) \times \vec{n} = 0. \quad (8)$$

Expressing by components, the contact condition is

$$\vec{k} \times \vec{r} = \vec{k} \times (x_0\vec{i} + y_0\vec{j} + z_0\vec{k}) = x_0\vec{j} - y_0\vec{i}. \quad (9)$$

As of $\vec{n} = \frac{\partial \vec{r}}{\partial \alpha} \times \frac{\partial \vec{r}}{\partial \varphi}$ so:

$$\vec{n} = -R_0^2(c \tan \alpha)' \sin \varphi \vec{k} - R_0^2(c \tan \alpha)' \cos \varphi \vec{j}. \quad (10)$$

Equations (9) and (10) lead to Equation (8). Cylindrical equation $x_0 = R_0 c \tan \alpha$ was introduced into. So $R_0 c \tan \alpha + p \tan \varphi = 0$, where $R_0 c \tan \alpha = a$, a is independent variable and radius of spiral line in

cylindrical cam which changes scope is $\frac{D}{2} \sim \frac{D}{2} - H$. (D -workpiece diameter; H -depth of spiral path.) So the contact condition is that:

$$a + p \tan \varphi = 0. \tag{11}$$

So the contact equation is:

$$\begin{cases} a + p \tan \varphi = 0 \\ x_0 = \sqrt{R^2 - (R_0 \cos \varphi)^2} \\ y_0 = R_0 \cos \varphi \\ z_0 = R_0 \sin \varphi \end{cases} \tag{12}$$

4 The equation of workpiece spiral surface

When the contact line rotates round Z-axis moving spiral with known contact equation, workpiece spiral surface is found.

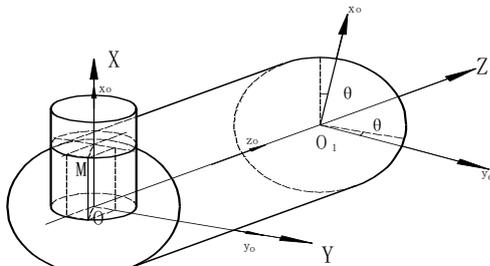


FIGURE 6 Cutter's coordinate played helical motion in work's coordinate

The contact line is obtained as $O-X, Y, Z$ coinciding with $O_1-X_0 Y_0 Z_0$. When cutter coordinate system moves from $O-X, Y, Z$ to $O_1-X_0 Y_0 Z_0$, it rotates θ angle and

References

[1] Zhao H 1993 CAM Mechanism Design Beijing. Higher Education Press
 [2] Shandong institute of technology Surface Mapping JiNan Shandong science and technology publishing house 1979
 [3] Volmer J 1976 CAM Mechanism Design. Beijing China Machine Press 1976
 [4] Hain K 1970 Challenge to Design Better Cams J. Mechanisms 5 283-6
 [5] Rees Jones J 1978 Cams and Cam Mechanisms Mechanical Engineering publications of the Institution of Mechanical Engineers
 [6] Zou H 1991 Modern Design of CAM mechanism Shanghai: Shanghai Jiao Tong University Press
 [7] Guan R, Tang C 1985 CAM and CAM Mechanism Beijing National Defence Industry Press
 [8] Lu J 1986 CAM Manufacturing Technology Beijing China Machine Press

moves P along Z -axis. Now the coordinate point M in is found in the Figure 5.

Equations (12) follows from the Equation (1), the spiral surface equation is:

$$\begin{cases} x = \sqrt{R^2 - (R_0 \cos \varphi)^2} \cos \theta - R_0 \sin \theta \cos \varphi \\ y = \sqrt{R^2 - (R_0 \cos \varphi)^2} \sin \theta + R_0 \cos \theta \cos \varphi \\ z = R_0 \sin \varphi + p \theta \end{cases} \tag{13}$$

From the equating, we know the contact line is not a line but a space curve and the spiral surface is also a space surface.

The spiral surface's shape and location is decided not only by the cylindrical diameter, cam path's depth and spiral line's helical pitch, but also by the diameter of roller. In addition, when finishing machining, the diameter of cylindrical cutter is equal to the roller diameter in order to ensure the accuracy. Otherwise, cylindrical roller can't move along the path and affect the moving stability.

5 Conclusions

The topic sets up the mathematical models of cylindrical cam surface and the course is clear and concise. The models grasp the nature of space motion of the facility. The equations have universal meaning, strong operability and good use for reference value for analysis of other cams.

Authors	
	<p>Liyang Liu, born in June, 1978, Shenyang, China</p> <p>Current position, grades: lecturer at the Northeastern Works at the School of Mechanical Engineering Shenyang Ligong University, Shenyang, China. University studies: Mechanical and Electronic Engineering at Shenyang University. Scientific interest: virtual reality and augmented reality applications in mechanical engineering. Publications: 4.</p>
	<p>Chao Li, born in January, 1980, Shenyang, China</p> <p>Current position, grades: engineer, works at Shenyang Mint, Shenyang, China. University studies: mechanical and electronic engineering in shenyang ligong university. Scientific interest: Maintenance of Equipment in Mechanical Engineering Publications: 3 Papers</p>
	<p>Qi Zhang, born in April, 1978, Shenyang, China</p> <p>Current position, engineer, State Grid Yingkou Electric Supply Company. University studies: Shenyang Agricultural University. Scientific interest: electrical equipment in mechanical engineering Publications: 1 papers</p>

Comparative study on prosthetic socket materials

Lifang Ma^{1, 2*}, Yaxin Wang¹, Yang Liu¹, Shizhong Zhang¹, Yu Chen¹

¹School of Materials Science & Engineering, Beijing Institute of Technology, Beijing, China

²National Research Center for Rehabilitation Technical Aids, Beijing, China

Received 22 July 2014, www.tsi.lv

Abstract

Prosthetic socket materials must exhibit a good processing performance so that a variety of desirable shapes can easily be formed, thereby enabling more controllable and adjustable socket production process and socket compatibility with the human anatomy and movement mechanics. Additionally, they need to feature high strength and light weight, and must be comfortable for the patient to use. In this study, the properties and application potential of current prosthetic socket materials, such as thermoplastic sheets, low-temperature thermoplastic sheets, silicone-based materials, and resin-based composite materials, were compared. Additionally, the matrix used in resin-based composite materials was investigated by infrared spectroscopy.

Keywords: prosthesis, socket, thermoplastic sheet, resin-based composite material

1 Introduction

A prosthesis is an artificial limb that is produced and assembled to restore body form and function, and to compensate for the disability caused by limb amputation. For the >200 million amputees in China, receiving a prosthesis is an important step towards their rehabilitation and reintegration into society. The prosthetic socket that connects the stump to the prosthesis is a critical part of the prosthesis and determines its performance. The prosthetic socket enables transfer of force between the stump and the prosthesis [1-5]. The basic requirements of a prosthetic socket are its ability to bear loads, control the prosthesis, and suspend the prosthesis. The prosthetic socket material must exhibit a good processing performance to allow the facile production of a variety of desirable shapes, thus enabling more controllable and adjustable socket production process and socket compatibility with the human anatomy and movement mechanics. Additionally, it must possess high strength and light weight, and be comfortable for the patient to use [6-17]. However, studies on prosthetic socket materials are rare. In this paper, we compare the properties and application potential of four types of existing prosthetic socket materials.

TABLE 1 Properties and performance of different thermoplastic sheet materials used for prosthesis socket fabrication

		LDPE	HDPE	PP-H	PP-C	PETG
Processing performance	Melting temperature of the crystalline part (°C)	100–120	125–138	160–175	150–175	>150
	Melt index (g/10 min)	0.25–27		0.4–100	0.6–100	
	Elongation at break (%)	100–650	350–525	100–600	200–500	110
Mechanical properties	Bending strength (breaking or yield strength) (psi)			6000–8000	5000–7000	10200
	Tensile modulus (×10 ³ psi)	25–41	130–150	165–225	130–180	
	Shore D hardness	44–50	61–63	76	70–73	78
Thermal properties	Linear expansion coefficient (10 ⁻⁶ /°C)	100–220	130–200	81–100	68–95	
	Thermal conductivity	8	11–12	2.8	3.5–4	
Physical properties	Relative density	0.917–0.932	0.94	0.9–0.91	0.89–0.905	1.27

* Corresponding author e-mail: malifang@sohu.com

2 Thermoplastic sheets

Thermoplastic sheets are widely applied in the fabrication of prosthetic sockets including temporary sockets, transparent sockets for experimental purposes, and long-term prosthetic sockets. Among the polymers used as raw materials for thermoplastic sheet production are polyethylene (PE), polypropylene (PP), and modified polyesters, e.g., polyethylene terephthalate glycolate (PETG).

Table 1 presents the properties of various thermoplastic sheet materials used for prosthesis socket fabrication. Low-density polyethylene (LDPE), synthesized via a high-pressure polymerization process, consists of long and branched chains, and has a low crystallinity, density, and strength, but good toughness. It can be used to manufacture flexible sockets. High-density polyethylene (HDPE), synthesized via a low-pressure method, exhibits less branching, a higher crystallinity and strength, a good processing performance, and excellent toughness. It is therefore used to produce temporary prosthesis sockets.

PP has also been used to fabricate prosthetic sockets (Figure 1) because of its higher melting point, strength, rigidity, and resistance to bending and fatigue when compared with PE. However, PP is more brittle than PE, especially at low temperatures. To reduce the brittleness of PP, the latter can be copolymerized with ethylene to obtain a polypropylene copolymer (PP-C). Despite the lower resulting strength and rigidity of PP-C when compared with those of the polypropylene homopolymer (PP-H), the brittleness of the PP material is greatly reduced. Thus, PP-C is usually selected for low-temperature applications. However, because the viscosity of the PP melt is sensitive to changes in temperature, the processing temperature needs to be strictly controlled and maintained at $\sim 185^{\circ}\text{C}$ during socket fabrication. Furthermore, because of the lower thermal conductivity of PP relative to that of PE, the heat generated during the grinding process cannot be easily released, thereby leading to the adhesion of debris to the polishing head. Therefore, low revolution speeds during the grinding process and a metallic grinding head for coarse grinding are necessary to ensure efficient heat dissipation.



FIGURE 1 PP prosthesis socket

Polyethylene terephthalate (PET) can crystallize into a configuration that reduces the transparency of the resulting product. To circumvent this issue, the resin can be modified accordingly to inhibit PET crystallization, thus generating products with good transparency. For example, modification of PET with cyclohexanediol affords non-crystalline PETG with good transparency. The resulting PETG can then be used for the fabrication of transparent prosthetic sockets for experimental purposes (Figure 2). As shown in Figure 3, the elastic modulus of PETG is higher than that of other thermoplastic sheets.



FIGURE 2 Transparent PETG prosthesis socket

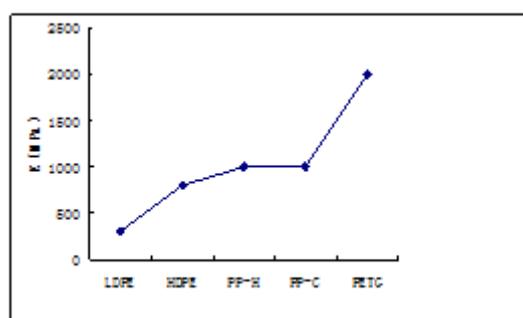


FIGURE 3 Elastic modulus of different thermoplastic sheet materials used for prosthesis socket fabrication

3 Low-temperature thermoplastic sheets

Plastic sheets have been widely used in the field of prosthetic products. However, socket production from ordinary plastic sheets is a complex process that includes the production and filling of the plaster female mold, modification of the plaster male mold, thermoplastic molding, and trimming of the products. In contrast, low-temperature thermoplastic materials can be shaped directly on the patients' bodies, and the product can be used directly after trimming, which greatly reduces the working and processing time. Low-temperature thermoplastic materials have a relatively low softening temperature and can be shaped in the temperature range from 55 to 75°C that human skin can withstand. Currently, the most frequently used low-temperature thermoplastic materials are polycaprolactone (PCL) and transpolyisoprene (TPI).

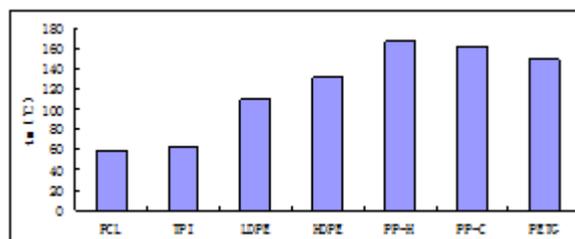


FIGURE 4 Melting temperature of different thermoplastic sheet materials used for prosthesis socket fabrication

PCL is a semi-crystalline polymer that can be synthesized via ring-opening polymerization of ϵ -caprolactone in the presence of a catalyst. PCL has a crystalline melting temperature of $\sim 60^{\circ}\text{C}$ and a glass transition temperature of $\sim 60^{\circ}\text{C}$. Its mechanical properties are similar to those of polyolefin, with a tensile strength of 13 – 30MPa and an elongation at break between 300 and 600% . TPI has a similar chemical composition, but a different configuration from that of natural rubber. Additionally, the good regularity of the molecular chains of TPI affords arrangement of the molecular chains with a high degree of ordering, resulting in a crystalline state. As a result, TPI loses its elasticity in the temperature region around room temperature, and can be used as a plastic. TPI has a low melting point of $\sim 64^{\circ}\text{C}$; hence, above this temperature, TPI has a soft, viscous

flow appearance. Tables 1 and 2 compare the properties and performance of different low-temperature thermoplastic materials and ordinary plastic materials. Based on the melting temperatures given in Tables 1 and 2, the low-temperature thermoplastic sheets melt and deform more easily than ordinary plastic sheets at high

temperatures. Hence, their application potential in the production of prosthetic sockets is limited. To better satisfy the requirements for successful application in the production of prosthetic sockets, some modifications are necessary to improve the mechanical properties of low-temperature thermoplastic sheets.

TABLE 2 Properties of low-temperature thermoplastic sheet materials

	PCL	TPI	LDPE	HDPE	PP-H	PP-C	PETG
Melting temperature (°C)	58–61	64	100–120	125–138	160–175	150–175	>150
Shore D hardness	50–55	50	44–50	61–63	76	70–73	78–80
Density (g/cm ³)	1.145	0.96	0.917–0.932	0.94	0.9–0.91	0.89–0.905	1.27

4 Silicone rubber and silicone gel

Silicone rubbers and silicone gels have been widely used for the production of prosthetic socket inner liners because of their excellent biocompatibility, high temperature tolerance and chemical resistance, and lack of colour and odor, additionally, they do not promote bacterial growth or induce tarnish and corrosion of other materials [18, 19]. Their three main functions are as follows. First, they exert a skin-protecting effect, especially under poor skin conditions involving scars that are prone to damage owing to compression and friction. These types of wounds are difficult to heal and may even form ulcers. To prevent scar damage, an oily lubricant or softening scar drugs can be smeared on the skin surface of the stump to reduce friction between the scar and

socket. Compression pressure can also be applied to the scar to soften the scar and prevent scar damage. Second, silicone rubbers and silicone gels exert a suspension effect. The silicone rubber stump sleeve can be rolled over and sleeved on the stump without the need for application of a lubricant on the skin. Additionally, silicone rubber sleeves generally show good adhesion to the skin, which can reduce both skin friction and shear force between the skin and internal wall of the socket, thus increasing the stump capability for prosthetic suspension. Third, silicone rubbers and silicone gels can improve the load-bearing capacity of the stump. Owing to the soft texture of silicone rubbers, they can adjust accordingly to protuberances of the stump bone and improve the load-bearing ability of the stump.

TABLE 3 Physical, mechanical, and chemical properties of silicone-based prosthesis products

Properties and performance	Index	Properties and performance	Index
Tensile strength (MPa)	≥7.00	Change in pH	≤1.5
Rupture strength (kN/m)	≥14.00	Heavy metal content (g/mL)	≤1.0
Elongation at break (%)	≥250	Residual content after evaporation (mg/mL)	≤0.05
Permanent deformation at break (%)	≤8	Consumption of KMnO ₄ /mL	≤6.5
Shore A hardness	45–80	UV absorbance at 220 nm	≤0.3
Thermal aging	70°C for 72h		
Change of tensile strength (%)	≥-15		
Change of elongation at break (%)	≥-25		

Table 3 lists the physical, mechanical, and chemical properties of silicone-based prosthesis products. Because silicone-based prostheses cannot transfer force between the stump and prosthesis owing to the low material strength, a silicone rubber or a silicone gel prosthetic inner liner must always be used in combination with prosthetic socket materials.

5 Resin-based composite materials

Resin-based composites used for prosthesis socket fabrication are typically fibre-reinforced plastic materials. Among the most frequently used reinforcing fibres are glass, carbon, and aramid fibres. Commonly used base materials are, for instance, epoxy resin, unsaturated polyester resin, and poly (methyl methacrylate) (PMMA). The main features of resin-based composite materials are their high strength and light weight. Table 4 compares the

performance of different resin-based composite materials and metallic materials.

The most commonly used prosthetic socket material is the modified PMMA-based fibre composite material. It is non-toxic and therefore does not induce strong skin irritation effects that are common for additives used in epoxy resin and unsaturated polyester. Additionally, the cured fibre-PMMA composite material is a thermoplastic and the shape of the products can be modified by annealing and subsequent processing. The composite material is typically prepared via a two-step process. The first step involves the formation of a prepolymer via bulk polymerization of methyl methacrylate monomers; the prepolymer is further modified to form the matrix of the prosthetic socket material. The second step involves the manufacture of the prosthetic socket using the base material and the fibre material via a composite processing technology.

TABLE 4 Performance of different resin-based composite materials and metallic materials

	Aluminium alloy	Titanium alloy	Steel	Glass fibre-reinforced composites	Carbon fibre-reinforced composites
Elastic modulus (MPa)	75	110	210	30	88
Tensile strength (MPa)	350	800	1100	720	900
Specific strength	125	178	141	343	600
Specific stiffness	27	24	27	14	59
Density (g/cm ³)	1.145	0.96	0.917–0.932	0.94	0.9–0.91



FIGURE 5 Carbon fibre-reinforced plastic prosthetic socket



FIGURE 6 Shaping and molding process of prosthetic sockets based on carbon fibre-reinforced plastic materials.

Figures 5 and 6 show an example of a carbon fibre-reinforced socket and its molding process, respectively. Prosthetic sockets and orthoses produced via this process have a thin texture, light weight, and high mechanical

strength. When resin-based composite materials are used to produce prosthetics and orthoses, the packing of the fibre material must not be too loose because only a tight fibre packing can ensure material capability to effectively withstand the desired load. Besides, the fibres should be oriented parallel to the direction of maximum loading so that the fibre material can withstand maximum loading. Furthermore, the high-strength fibre materials, e.g. carbon fibres, should have sufficient length and need to be laid on the weakest part of the composites to ensure effective material reinforcement. Processing and socket production technologies using resin-based composite materials are widely applied; however, studies on such materials have been rarely reported to date. In this paper, the base material of resin-based prosthetic sockets was investigated by infrared spectroscopy, and the results are shown in Figure 7. The peaks observed at 1730cm⁻¹ corresponding to C=O stretching vibration and at 1150, 1190, 1240, and 1268cm⁻¹ corresponding to C–C–O–C stretching vibrations are in good agreement with the characteristic bands of methyl methacrylate. Two characteristic peaks corresponding to amines and ether groups appeared at 1450 and 1600cm⁻¹, respectively, and can be attributed to the modification of the PMMA prosthetic socket [20-29]. Figure 8 shows the molecular weight distributions of the base materials used for resin-based prosthetic socket fabrication. The preparation of the most stable base material will be reported in due course in the following study.

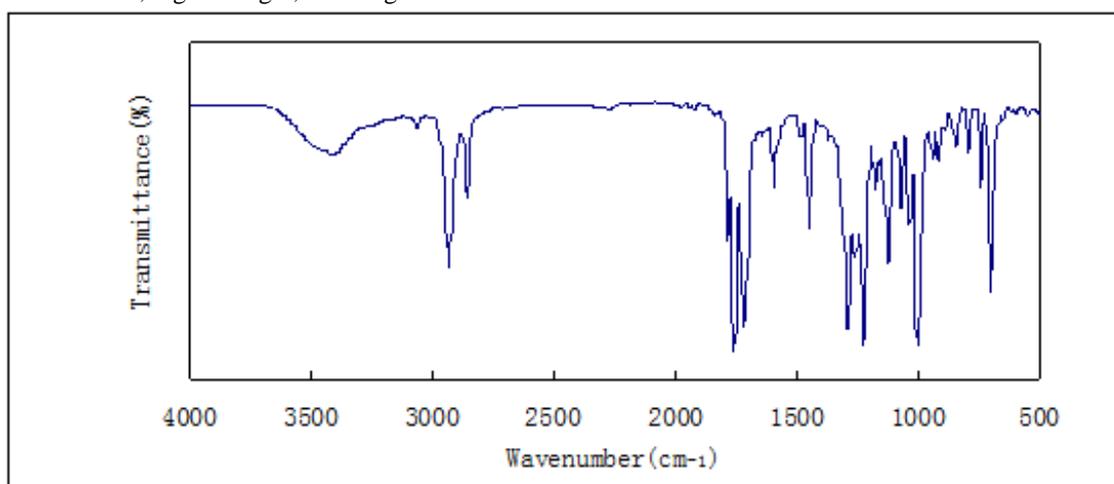


FIGURE 7 Infrared spectrum of the base material used for resin-based prosthetic socket fabrication

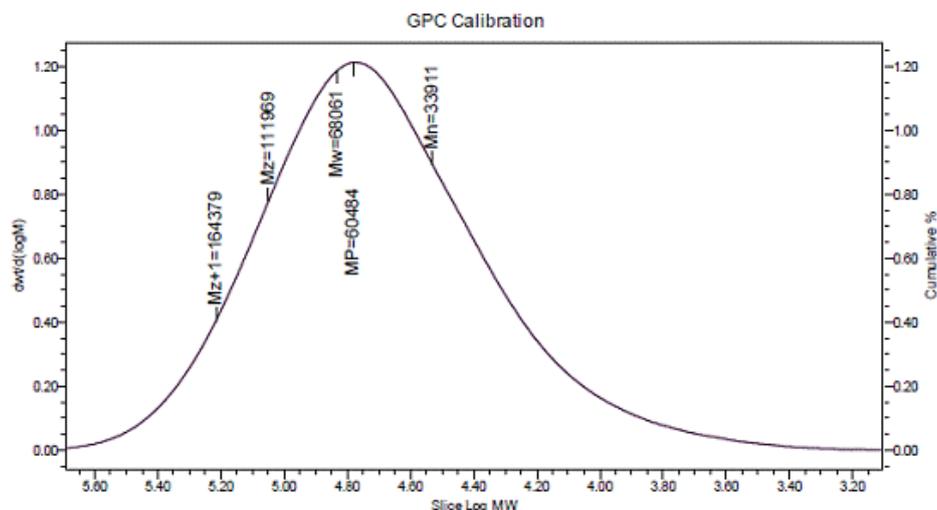


FIGURE 8. Molecular weight distribution of the base materials used for resin-based prosthetic socket fabrication

6 Conclusion

Currently, there are several gaps in the research of prosthetic socket materials and their modification for application. More specifically, several fundamental aspects relating to resin-based composite materials need to be studied including the reinforcement of the resin

matrix, preparation of the prepolymer for resin matrix fabrication, effective shaping and molding of both the resin matrix and carbon fibres, and the curing process at room temperature. In-depth research of these aspects will assist standardization of the production process of prosthetic sockets and further improve the overall quality of prosthetic sockets.

References

- [1] Biddiss E, Chau T 2007 Upper-limb prosthetics critical factors in device abandonment *American Journal of Physical Medicine & Rehabilitation* **86**(12) 977-87
- [2] Randall D A, Williams T W, Albuquerque M J, Altobelli David E 2011 Prosthetic sockets stabilized by alternating areas of tissue compression and release *Journal of Rehabilitation Research and Development* **48**(6) 679-96
- [3] Gerzeli S, Torbica A, Fattore G 2009 Cost utility analysis of knee prosthesis with complete microprocessor control (C-leg) compared with mechanical technology in trans-femoral amputees *Eur J Health Econ* **10**(1) 47-55
- [4] Silver-Thorn B, Current T, Kuhse B 2012 Preliminary investigation of residual limb plantarflexion and dorsiflexion muscle activity during treadmill walking for trans-tibial amputees *Prosthet Orthot Int* **36**(4) 435-42
- [5] Fey N P, Klute G K Neptune R R 2012 Optimization of prosthetic foot stiffness to reduce metabolic cost and intact knee loading during below-knee amputee walking: a theoretical study *J Biomech Eng* **134**(11) 111005
- [6] Postema K, Hermens H J, de Vries J, Koopman H F J M, Eisma W H 1997 Energy storage and release of prosthetic feet Part 1: Biomechanical analysis related to user benefits *Prosthet Orthot Int* **21**(1) 17-27
- [7] Rouse E J, Hargrove L J, Peshkin M A, Kuiken T A 2011 Design and validation of a platform robot for determination of ankle impedance during ambulation *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society 2011* 8179-82
- [8] Ciobanu O 2012 The use of CAD/CAM and rapid fabrication technologies in prosthesis and orthotics manufacturing *Rev Med Chir Soc Med Nat Ias* **116**(2), 642-648
- [9] Barr J B, Wutzke C J, Threlkeld A J 2012 Longitudinal gait analysis of a person with a transfemoral amputation using three different prosthetic knee/foot pairs *Physiother Theory Pract* **28**(5) 407-11
- [10] McNealy L L, Gard S A 2008 Effect of prosthetic ankle units on the gait of persons with bilateral trans-femoral amputations *Prosthet Orthot Int* **32**(1) 111-26
- [11] Yeung L F, Leung A K, Zhang M, Lee W C 2012 Long-distance walking effects on trans-tibial amputees compensatory gait patterns and implications on prosthetic designs and training *Gait Posture* **35**(2) 328-33
- [12] Pailler D, Sautreuil P, Piera J B, Genty M, Goujon H 2004 Evolution in prostheses for sprinters with lower-limb amputation *Ann Readapt Med Phys* **47**(6) 374-81
- [13] Hirons R 2012 Preparing our Paralympians: research and development at Ossur, UK Interview by Sarah A Curran *Prosthet Orthot Int* **36**(3) 366-9
- [14] Hafner B J, Sanders J E, Czerniecki J M, Ferguson J 2002 Transtibial energy-storage-and-return prosthetic devices: a review of energy concepts and a proposed nomenclature *J Rehabil Res Dev* **39**(1) 1-11
- [15] Gauthier M A, Zhang Z, Zhu X X 2009 New dental composites containing multimethacrylate derivatives of bileacids: A comparative study with commercial monomers *ACS Appl Mater Interf* **1**(4) 824-32
- [16] Winkler P A, Stummer W, Linke R, Krishnan K G, Tatsch K 2000 Influence of cranioplasty on postural blood flow regulation, cerebrovascular reserve capacity and cerebral glucose metabolism *Childs Nerv Syst* **16** 247-52
- [17] Greenwald R. M, Deanw R. C, Board J 2003 Volume management: Smart variable geometry socket (SVGS) technology for lower-limb prostheses *JPO: Journal of Prosthetics and Orthotics* **15**(3) 107-12
- [18] Rotaru H, Baciut M, Stan H, Bran S, Chezan H, Iosif A, Tomescu M, Kim S G, Rotaru A, Baciut G 2006 Silicone rubber mould cast poly ethyl methacrylate-hydroxyapatite plate used for repairing a large skull defect *J Craniomaxillofac Surg* **34**(4) 242-6
- [19] Min Cai, Yan Jing, Hongyan Shi, Renguo Song, Hongwen Zhang, Mengchao Zhao 2013 Polymethylmethacrylate Grafted onto

- Surfaces of Silicone Rubber via Atom Transfer Radical Polymerization *Polymer Materials Science and Engineering* **29** (5) 23-32
- [20] Barszczewska-Rybarek I M 2009 Structure property relationships in dimethacrylate networks based on Bis-GMA, UDMA and TEGDMA *Dental Mater* **25**(9) 1082-9
- [21] Jin Z, Liu C, Zhang W 2006 Nano Carbon-poly(Methyl Methacrylate) Composition Materials *Acta Polymerica Sinica* **2**(2) 320-4 (in Chinese)
- [22] Li Y, Liu P, Wei Z, Wang Y, Yu H 2014 Preparation and Oil-Absorbing Properties of Poly(Lauryl Methacrylate) *Polymer Materials Science and Engineering* **29**(2) 5-8 (in Chinese)
- [23] Mravljak M, Sernek M 2011 The influence of curing temperature on rheological properties of epoxy adhesives *Drvna Industrija* **62**(1) 19-25
- [24] Jaruchattada J, Fuongfuchat A, Pattamaprom C 2012 Rheological investigation of cure kinetics and adhesive strength of polyurethane acrylate adhesive *Journal of Applied Polymer Science* **123**(4) 2344-50
- [25] Du Q, Du Z, Li M, Sun X, Zhang C, Zou W 2013 Rheological Properties of Polyurethane Adhesive in Curing Process *Polymer Materials Science and Engineering* **29**(3) 31-4
- [26] Huanqin Chen, Chunbao Huang, Huifang Shen, Kai Zhang 2014 Emulsion Grafting Polymerization of Chloroprene Latex with Methyl Methacrylate *Polymer Materials Science and Engineering* **29** (2) 141-44
- [27] Bai R, Wei Z B, Zhang F A 2013 MMA solution polymerization in the channel of modified SBA-15 *Acta Polymerica Sinica* **2013**(7), 849-55 (in Chinese)
- [28] Cai J, Chen T, Wang G. Z, et al 2012 The study on the reverse atom transfer radical polymerization of MMA catalyzed by acetylacetonate cobalt(II) complex supported by ionic liquid *Adv Mater Res* **2012**(1) 476-8
- [29] Sun F, Wei Z, Zhang F 2014 Effect of Initiator Content on Methyl Methacrylate Solution Polymerization in the Channels of Mesoporous SBA-15 *Polymer Materials Science and Engineering* **30**(6) 29-32

Authors	
	<p>Lifang Ma, born in November, 1981, Beijing, P.R. China</p> <p>Current position, grades: PhD student in material engineering at Beijing Institute of Technology China, National Research Centre for Rehabilitation Technical Aids.</p> <p>University studies: Master's degree from Beijing Institute of Technology in China.</p> <p>Scientific interest: materials, rehabilitation, modelling,</p> <p>Publications: more than 10 papers.</p> <p>Experience: working experience of 8 years in rehabilitation, 2 scientific research projects, 3 patents</p>
	<p>Yaxin Wang, born in October, 1987, Beijing, P.R. China</p> <p>Current position, grades: Graduate student at the School of Chemistry, Beijing Institute of Technology, China.</p> <p>University studies: B.Sc. in chemistry at Hebei Normal University</p> <p>Scientific interest: polyurethane-silicone composite materials.</p>
	<p>Yang Liu, born in August, 1990, Beijing, P.R. China</p> <p>Current position, grades: the graduate student at the school of material science and engineering of Beijing Institute of Technology, China.</p> <p>University studies: B.E. in Polymer Science and Engineering at Anhui University of Anhui in China.</p> <p>Scientific interest: biomedical polymer material, natural polymer materials.</p> <p>Publications: 1</p> <p>Experience: 1 scientific research project.</p>
	<p>Shizhong Zhang, born in June, 1994, Beijing, P.R. China</p> <p>Current position, grades: Bachelor degree student in material engineering at Beijing Institute of Technology.</p> <p>Scientific interest: materials and chemistry.</p>
	<p>Yu Chen, born in April 1979, Beijing, P.R. China</p> <p>Current position, grades: Associate Professor of School of Materials Science and Engineering Beijing Institute of Technology.</p> <p>University studies: PhD at Beijing Institute of Technology in China.</p> <p>Scientific interest: development and application of biological medical dressings with antibacterial, hemostatic function and energetic materials research.</p> <p>Publications: more than 40 papers, 16 national invention patents, 10 authorized national invention patents.</p> <p>Experience: more than 10 projects.</p>

Extraction of soil salinization information from the Manas river basin based on TM Images

Ling Wang*, Peng Guo, Lin Liu

Geography Department of Shihezi University, Shihezi 832000, China

Received 26 May 2014, www.tsi.lv

Abstract

Remote sensing technology is widely used in real-time observations. In this study, therefore, salinization information was extracted from Thematic Mapper (TM) images. In particular, remote sensing information regarding saline soil was obtained to analyse its dynamic changes. This soil collected from the Manas River Basin in Xinjiang Province China was selected as the research area. Data from Landsat TM remote sensing images with seven bands were obtained in August 2010 as inputs, and salinization information was extracted using the e-Cognition system. As per this information, high-salinity soil is mainly distributed outside the oasis. The data were analysed further through the normalized differential vegetation, normalized difference water, and remote sensing image indices. Analysis results show that overall classification accuracy can reach 83.7%, thus demonstrating that the automatic extraction of information regarding saline soil is highly accurate. Furthermore, this information can be automatically and precisely extracted using an object-oriented method.

Keywords: Manas river, river basin, soil salinization, NDVI

1 Introduction

Soil salinization predominantly limits sustainable agricultural development. The estimated salinization level is 9.55×10^8 hm², which accounts for 7.26% of the land surface on the Earth. Hence, salinization has become an international problem. Specifically, salinization is severe in the Manas River and has stilted the utilization efficiency of its water resources. As a result, sustainable agricultural development is restricted in this region. All aspects of saline alkali soil (properties, range, geographical distribution, and saline degree) must therefore be effectively determined. These factors may help facilitate the monitoring and control of soil salinization in the Manas River Basin.

Soil salinization was first monitored through satellite remote sensing in the 1970s. In the early 1980s, multi-band and temporal remote sensing were widely used to assess saline soil [1, 2]. In the 1990s, however, visual interpretation was developed as an important method of monitoring soil salinity [3], and it has remained significant since then. Researchers report that comprehensive analysis and image feature analysis methods can eliminate the interference of objects with foreign bodies in a single spectrum [4-7]. Zeng first presented the concept of a "geographical control system", which considered the soil and the landscape area as a whole [8]. Zhang obtained meteorological data from the National Oceanic and Atmospheric Administration to establish a regression model between soil salinization and daily minimum and maximum temperatures [9-11].

In the current study, processing technology for remote sensing images was adopted to determine soil salinity in

the Manas River. This technology may clarify the distribution of soil salinization consistently and can guide the dynamic monitoring of this salinization in the Manas River Basin. We also develop a reasonable salinity control solution by analysing the progression of salinization. The results of this study promote the development of agricultural production in the Manas River Basin.

2 Research area

The Manas River Basin (longitude 85°00' – 86°30', north latitude 43°30' – 45°40') is located at the northern foot of Tianshan Mountain in Xinjiang, China. Administratively, this basin covers Manas County, Shawan County, and the reclaimed Shi He-zi area for a total area of 26500 km². The research area also includes six rivers, namely, the Taxi River, Manas River, Ning River, Gold River, South River, and Bayin River. upstream to downstream, and the constitution of soil salinity has shifted from sulfate to chloride salt accordingly [12].

3 Data sources and methods

3.1 DATA SOURCES

A Landsat Thematic Mapper (TM) remote sensing image was obtained for this research. The image contains seven bands (Table 1) and has a spatial resolution of 30m×30m upon resampling. The research area was pinpointed according to its coordinates, and the pixel array was 7904 (row)×11479 (line). This study also utilized land use, stream, and meteorological data.

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

TABLE 1 Bands of the remote sensing image (TM)

Band NO.	Range (μm)	Main function
TM1	Blue waveband (0.45–0.52)	The blue waveband is used to Distinguish soil and vegetation, as well as artificial objects
TM2	Green waveband (0.52–0.60)	The red waveband is used to detect the healthy of the plants and reflect the reflectivity of the plant.
TM3	Red waveband (0.62–0.69)	The red waveband is used to measure the pigments of vegetation and distinguish the artificial surface feature.
TM4	Near-infrared waveband (0.76–0.90)	The near-infrared waveband is used to determine the condition of the crop, drawing water boundary and detect the soil humidity.
TM5	Middle-infrared waveband (1.55–1.75)	The middle-infrared waveband is used to detect the soil moisture and the water content, distinguish the cloud and snow.
TM6	Infrared waveband (1.04–1.25)	The infrared waveband is used to detect the thermal radiation of the earth surface's object.
TM7	Middle-infrared waveband (2.08–2.35)	The middle-infrared waveband is used to monitor the radiation source.

3.2 EXTRACTION OF SOIL SALINIZATION INFORMATION

Soil salinization information can be extracted by classifying the remote sensing images. This method is based on various combinations of spectral data [13] and is object-oriented. Since the 1970s, it has been widely integrated into the classification of the remote sensing images at the interpretation stage [14, 15]. Object-oriented image analysis mainly involves two procedures: image segmentation and information extraction. Both processes are interactional and cyclic. We can then extract class information according to the dimensions of the image object [16].

4 Spectral analysis of salinized soil, variable selection, and establishment of a taxonomy system

4.1 SPECTRAL ANALYSIS OF SALINIZED SOIL

During the dry season of universal salt, a salt crust is formed on the surface of salinized soil, and the spectral

reflectance of this soil is greater than those of other soil types. Furthermore, the colour of the salinized soil image is thinnest relative to those of the images of other soil types regardless of the visibility of the spectrum or of the near-infrared band. Soil salinity is mainly induced by white crystal; thus, we can determine the extent of soil salinization according to the white marks in the image during spectral analysis.

4.2 VARIABLE SELECTION

The *NDWI* is the normalized ratio of the green waveband and the near-infrared band. Its formula can be described as follows:

$$NDWI = (Green - NIR) / (Green + NIR), \quad (1)$$

where *Green* represents the green waves and *NIR* signifies the infrared waves. *Green* and *NIR* denote the second and fourth bands, respectively, in the Landsat TM remote sensing image.

The *NDSI* is a quantitative index used to observe ice. It is the core of *SNOMAP* arithmetic. In remote optical sensing, this index is the universal method of extracting accumulated snow. It can not only recognize accumulated snow as its primary function, but it can also accurately determine snow-clouds. Thus, it may enhance the sensitivity of soil monitoring.

The *NDVI* is a remote sensing indicator that can reflect on remote sensing data regarding the research area, as displayed in Table 2.

4.3 CLASSIFICATION SYSTEM OF LAND USE

We established a classification system of land use based the land cover conditions. This index can be defined as follows:

$$NDVI = (NIR - R) / (NIR + R), \quad (2)$$

where *NIR* is the reflected value of the near-infrared band and *R* is the reflected value of the red wave band. The *NDVI* can detect vegetation progression and coverage; hence, it effectively extracts vegetation information [17].

TABLE 2 Classification system of land use

No.	Land use	Definition
1	Tillage	Crop land
2	Forestry	Forest land (arbor, bush, bamboo)
3	Grassland	Herbaceous plants (covers the degree below 5% of the entire area)
4	Stream	Natural or artificial stream
5	Lake	Natural ponds
6	Reservoir	Artificial ponds
7	Glacier	Lands covered with glaciers and snow
8	Residential land	Urban settlements, traffic paths
9	Severe salinization	Salinity of the surface soil $\geq 75 \text{ g kg}^{-1}$; Vegetation coverage is 0%—1%
10	Moderate salinization	Salinity of the surface soil is 45 g kg^{-1} — 75 g kg^{-1} ; Vegetation coverage is 5%
11	Mild salinization	Salinity of the surface soil is 15 g kg^{-1} — 45 g kg^{-1} ; Vegetation coverage is 15%
12	Bare soil	Soil texture cover; < 5% of the ground is covered with vegetation
13	Bare rock	Rock covers the ground; > 5% of the ground is covered with rocks

5 Extraction and analysis of remote sensing information on soil salinization

In this research, e-Cognition image processing software was adopted to extract non-salinization information. Prior to classification, the image was pre-processed to facilitate image segmentation and information extraction.

5.1 MULTI-SCALE IMAGE SEGMENTATION

The multi-scale partition method used in this study is the most commonly used partition approach. It defines a

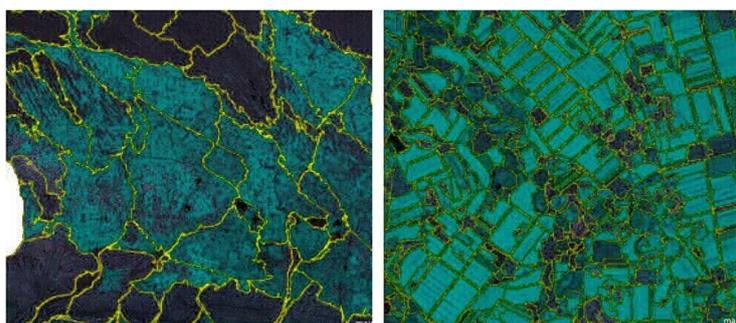


FIGURE 1 Image split at scales of 600 and 15

When the parameters of the segmented image are high in value, the image is actually poorly split. The image split at a scale of 15 may accurately segment the various ground features of soil into different polygons. By contrast, the image segmented at a scale set of 600 displays a low segmentation accuracy. Hence, we selected the image split at a scale of 15 as the base map of soil salinization in this study.

5.2 POLYGON CONSTRUCTION

Polygons may be constructed within the split image. Prior to this procedure, however, we must verify the image features (e.g., boundary, shape, area, length, and width) and the mean values and standard deviations of these features to classify image treatment. If image segmentation is poor, the scale parameter of the image must be readjusted. As per an analysis of the feature of the image divided at a set scale of 15, the studied image meets the requirements of the research.

5.3 HIERARCHY OF LOADING CLASS

We segmented the remote sensing image of the research area into the following categories based on surface feature type: Tillage, forest land, meadow, stream, lake, reservoir, glacier, residential land, severe salinization, moderate salinization, mild salinization, bare soil, and bare rock. The different surface features are presented in different colours.

specific scale for the polygon of the target image and highly optimizes image segmentation. In multi-scale segmentation, we first identified the compiled layers. We then measured each compiled layer under different weights. We considered every band to be coverage, and the compiled layers correspond to the band number. Therefore, we can determine the weights of all compiled layers. The multi-scale data were segmented based on different characteristics and classifications. In this study, the segment dimensions were set at 600 and 15 to divide the remote sensing images and to generate differently scaled images, as shown in Figure 1.

5.4 IMAGE CLASSIFICATION

5.4.1 Classification of non-salinized land according to land type

In this study, the two characteristic variables NDWI and NDSI were regarded as the base map. We obtained different interval values by trial and error. The wave and the glacier were extracted using optimum threshold value methods, whereas vegetation was detected by the eCognition processing system for remote sensing images. Briefly, the Classification step is as follows: Feature view → Object features → Customized → NDVI. Once we generated the NDVI, the values of this index could be adjusted to validate the optimal ranges of the categories tillage, forest land, and grassland. The land type between the bare rock and bare soil could be identified by adopting the interval value of NDVI obtained from a reiteration of the test.

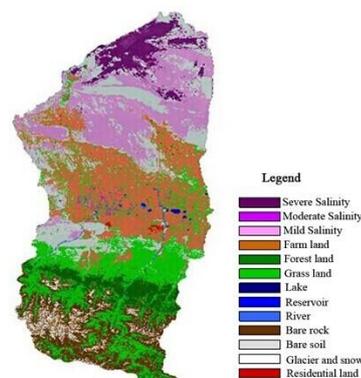


FIGURE 2 Classification of degrees of salinity at the Manas Basin

5.4.2 Classification of the extent of salinization

Based on brightness and the NDVI values of the bands, computers can automatically derive the extent of salinization. As the degree of salinity in an area increases, land cover decreases. We therefore set an appropriate NDVI threshold value to distinguish saline soil from non-saline soil. The brightness values of the severely saline soil and the sandy soil from TM1 are close, and the brightness values of the clay and the sandy soils are maximized. Hence, we can determine severe salinity by setting threshold values for both TM1 and TM7.

In this study, visual interpretation methods were used to determine the NDVI values at various extents of salinity (moderate and mild salinization alone were considered). Consequently, the e-Cognition system extracted these values through logical calculus.

5.5 ANALYSIS OF SALINIZATION INFORMATION

As illustrated in Figure 2, the soil of the research area displays different degrees of salinity. The distribution of TABLE 3 Proportion of salinized areas in the research region

Salinity degree	Mild	Moderate	Severe	Non-salinized
Area (m ²)	6390855	168259	2028581	28608394
Proportion (%)	17.18	0.46	5.45	76.91

5.6 EVALUATION OF SALINIZATION CLASSIFICATION ACCURACY

Once the shadowgraph of this study was sorted, we validated its accuracy based on the four classification results of the e-Cognition processing system for remote

TABLE 4 Accuracy of soil salinity classification

Type	Producer accuracy (%)	User accuracy (%)	Mean (%)
Mild salinity	80	83.72	81.85
Moderate salinity	71.11	100	85.55
Severe salinity	100	100	100
Total: 83.7%			

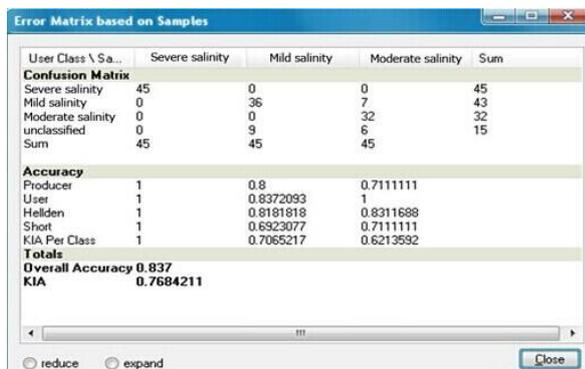


FIGURE 3 e-Cognition system results

6 Results and conclusions

1) The extracted information revealed that severely saline soil is mainly distributed outside the oasis. Moderately saline soil is highly distributed around the central tillage

salt concentration is presented in different areas and shapes. Moreover, the moderately saline soil is spread across the surrounding tillage and grassland. The mildly saline soil is distributed across the phases of tillage. The degree of salinization of the soil in the Manas River Basin is accentuated from upstream to downstream and from south to north. Furthermore, soil salinization is concentrated and distributed in the upstream of the alluvial plain and around the reservoir. Land is fertile and water resources are abundant within the oasis, and this region is not salinized because of the lack of ground water. The distribution of saline soil within the tillage portion is generally slight. In addition, the mildly saline soil is widely distributed outside the oasis. We calculated the proportion of saline soil according to the attributes of salinization classification (Table 3).

Table 3 reveals that the minimum and maximum proportions of the Manas River Basin are composed of areas with moderately and mildly salinized soils, respectively. The salinized areas accounts for 23.09% of the entire research region.

sensing images. Moreover, sample accuracy was verified using the confusion and error matrices. The result of the accuracy test is 0.837. However, soil salinization was difficult to distinguish in the remote sensing image obtained in August 2010. Table 4 and Figure 3 presents the evaluation results.

portion of the Manas River, whereas the mildly saline soil is distributed in the bare soil area.

- 2) According to the proportion of the degree of salinization, the areas with mildly and moderately saline soil account for 17.18% and 0.46% of the total area, respectively.
- 3) In this study, NDVI, NDWI, and NDSI were used to analyze soil salinity. The analysis results show that overall classification accuracy can reach 83.7% and that the automatic extraction of soil salinity information is highly accurate.
- 4) The remote sensing image, which was obtained during a period of active vegetation, influenced classification accuracy slightly.
- 5) The remote sensing image was classified using object-oriented methods. The results suggest that these methods may extract actual geographies automatically. Hence, this research may promote the development of both remote sensing and geographic information systems.

Acknowledgements

The study was supported by the Research Foundation for

Advanced Talents of Shihezi University (RCZX201130), the Natural Science Foundation of Shihezi University (ZR KXYB-05) and the National Natural Science Foundation of China (41361073).

References

- [1] Wang Z, Zhu S, Yu R 1993 Salt-affected soils of China *Beijing: Beijing Science Press* 400-515 (in Chinese)
- [2] Huo D, Zhang J, Zhang J 2001 Using the cbers-1 satellite data to study the thematic information extraction of saline dry land *Remote Sensing for Land & Resources* 18(2) 48-52.
- [3] Singh, A N, Baumgardner M F, Kristoff S J 1977 Delineating Salt-affected Soils in Part of the Ganges Plain by Digital Analysis of Landsat Data. *Technical Report, Laboratory for Applications of Remote Sensing Purdue University West Lafayette Indiana USA*
- [4] Zeng Z 1984 Automatic Recognition and mapping of soil types using Landsat images: Computer classification and the spectroscopical and Geographical Analyses of the Results *Acta Pedologica Sinica* 21(2) 183-93 (in Chinese)
- [5] Dai C, Yang Y, Shi X 1986 Inventory of the low productive soils in Huang-huai-hai plain by using remote sensing *Remote Sensing of Environment* 1(2) 81-91 (in Chinese)
- [6] Wang X 1992 Analyses of the salt-affected soils in Yu-dong plain by using remote sensing *Remote Sensing Information* 10(4) 24-7 (in Chinese)
- [7] Zhang H, Shang S 1992 Application of NOAA AVNRR data in degree of salinity *Remote Sensing Information* 10(1) 24-6 (in Chinese)
- [8] Liu Q, Luo J, Liu G 2000 Primary explorations of the applied potentials of ZY-1 satellite data at the Yellow River Delta *Geo-Information Science* 2(2) 56-7 (in Chinese)
- [9] Jiang H 2007 Research on the soil-salinization spatial-temporal evolution based on 3S technique in arid area. *Urimqi: Xinjiang University (in Chinese)*
- [10] Liu K, Zheng X, Xie Y 2005 Agricultural water-Saving potential analysis in Manas river valley *Journal of Shihezi University (Natural Science)* 23(2) 237-9 (in Chinese)
- [11] Fan Z, Ma Y, Ma Y 2001 Salinized soils and their improvement and utilization in West China *Arid Zone Research* 18(3) 1-6 (in Chinese)
- [12] Li N, Wu L, Wang S Xia J, Zhu H 2011 Analysis on relationships between soil salinization and spectra in Manas River Valley *Acta Agriculturae Universitatis Jiangxiensis* 33(6) 1242-7 (in Chinese)
- [13] Lilles R M, Kiefer R W 2003 *Remote Sensing and Image Interpretation (Fourth Edition)*. Beijing: Electronic Industry Press
- [14] Tang G, Zhang B, Sun Y 2005 Object oriented image analysis and automatic separation of the woodland and area in scanned color relief map *Surveying and Mapping* 28(4) 171-6 (in Chinese)
- [15] Ludig M, Cârdei P, Muraru V, Mihaolv N 2014 Ways to optimize the electromagnetic waves applications in agriculture and food industry *Journal INMATEH – Agricultural Engineering* 42(1) 75-82
- [16] Hua H 2014 Research on ontology construction and information extraction technology based on wordnet *Journal of Digital Information Management* 12(2) 114-9
- [17] He L, Meng X, Du J, Wang Y, Cai Y, Tang X 2014 Study on the Ecological Characteristics and Change Analysis of Xin Jiang Junggar Basin Based on the NDVI *Nature Environment and Pollution Technology* 13(1) 191-6

Authors



Ling Wang, born in September, 1974, Shihezi, P.R. China

Current position, grades: associate Professor, Shihezi University, China.
University studies: Dr. Degree in Crop Cultivation and Farming System at Shihezi University in China.
Scientific interest: 3S technology application.



Peng Guo, born in June, 1981, Shihezi, P.R. China

Current position, grades: on-the-job Doctor; lecturer at Shihezi University, China.
University studies: Master degree in Geographical Information Systems at Xinjiang University.
Scientific interest: remote sensing application.



Lin Liu, born in June, 1981, Shihezi, P.R. China

Current position, grades: lecturer at Shihezi University, China.
University studies: Master degree in Geographical Information Systems at Northwest Normal University.
Scientific interest: GIS Technology application in resources and environment.

Volume integral equation-based electromagnetic inversion of 3D complex resistivity bodies

Jianping Li^{1*}, Tongxiao Shang², Yixiao Guan³

¹State Key Laboratory of Mining Disaster Prevention and Control, Shandong University of Science and Technology, Qingdao, 266590, China

²Key Laboratory of Earth Fissures Geological Disaster of MLR, Nanjing, 210018, China

³Geological Survey of Jiangsu Province, Nanjing 210018, Jiangsu, China

Received 2 July 2014, www.tsi.lv

Abstract

This study proposed the volume integral equation method to invert the complex resistivity parameters of 3D bodies in homogeneous half space. The partial derivative matrix of 3D bodies to complex resistivity was obtained with the use of electromagnetic field partial derivatives to the real resistivity combined with the complex resistivity spectrum partial derivatives of the Cole–Cole model. The classical damped least squares inversion method was applied. Results show that with at least four frequency data points, true complex resistivity parameters, as well as geometric parameters, can be inverted. The four parameters in the Cole–Cole model are interdependent. Chargeability and time constant are highly correlated, such that either one of these factors should be fixed prior to inversion. Furthermore, the proposed inversion algorithm we proposed has high efficiency because it only works in a split unit.

Keywords: volume integral equation, complex resistivity, 3D EM, damped least square method

1 Introduction

Spectral induced polarization (SIP), also known as complex resistivity (CR), was developed in the 1970s as a method of induced polarization. This approach uses the conventional resistivity method with a dipole-dipole device. In this method, the current passes through the power supply electrode to the underground in several frequencies, such that the difference in potential between the two electrodes can be observed. Apparent resistivity can then be calculated by using the traditional formula [1, 2] on the basis of a large number of measurement results for rock, ore, and outcrop samples. Pelton considered that the CR of rock and ore caused by IP effect rate varies with frequency (CR spectrum), which can be represented by the Cole–Cole model [3] as follows:

$$\rho(i\omega) = \rho_0 \left\{ 1 - m \left[1 - \frac{1}{1 + i\omega\tau^c} \right] \right\}, \quad (1)$$

where $\rho(i\omega)$ is the CR, ρ_0 is the direct resistivity, m is the chargeability, τ is the time constant, and c is the frequency dependence. These parameters are collectively known as the Cole–Cole model spectrum parameters.

In the SIP forward simulation, Hohmann calculated the 3D electromagnetic field of the CR body in homogeneous half-space by using an integral equation method [4]. Soinen simulated the CR spectrum of a rectangular solid with dipole and middle gradient arrangements in polarized and non-polarized earth [5, 6]. Weller studied the induced

polarization response of a complex conductivity model by using the finite differential method [7].

As regards SIP inversion, Luo presented an analytical method in which the true spectrum is derived from the apparent spectrum [8] of the polarization body. Zhang proposed that the apparent spectrum can directly invert the true spectrum of the polarization body [9]. Wang introduced an algorithm through which the apparent spectrum directly inverts the true spectrum [10]. Liu reported that only through a plurality of apparent IP spectra measured in different positions can the true Cole–Cole parameters of a polarization body be determined. Therefore, inversion should also include a Cole–Cole parameter, true rock resistivity, geometry parameters, excitation parameters, measuring location, and other relevant parameters of the mathematical and physical field to achieve the real spectral parameter inversion [11, 12].

This paper proposes and implements a direct inversion of 3D CR parameter algorithm. First, the electromagnetic field of the 3D CR body is calculated. Then, the partial derivative matrix of the electromagnetic field to 3D CR body is deduced according to the partial derivative equation given by Eaton [13], combined with the partial derivative of CR spectrum on the parameters of the Cole–Cole model. Finally, the damped least square method is used to directly inverse each Cole–Cole model parameter.

2 Basic theory of forward integral equation

Forward calculation is the foundation of inverse calculation. Thus, we first introduce forward theory.

* Corresponding author e-mail: wsljp2000@qq.com

According to the theory of integral equation, the tensor function of electromagnetic fields introduced by Green and the equations used by Maxwell are used to obtain an integral equation of a 3D CR body in a homogeneous earth [14].

$$\mathbf{F}(\mathbf{r}) = \mathbf{F}_1(\mathbf{r}) + \int_V \Delta\sigma \cdot \tilde{G}^F(\mathbf{r}; \mathbf{r}') \cdot \mathbf{E}(\mathbf{r}') dv', \quad (2)$$

where $\mathbf{F}(\mathbf{r})$ denotes total electromagnetic field at \mathbf{r} , $\mathbf{F}_1(\mathbf{r})$ denotes primary electromagnetic field at \mathbf{r} , $\Delta\sigma = \sigma - \sigma_0$, σ denotes complex conductivity of 3D body, σ_0 denotes conductivity of homogeneous earth, $\tilde{G}^F(\mathbf{r}; \mathbf{r}')$ denotes electromagnetic the tensor function, and $\mathbf{E}(\mathbf{r}')$ denotes the electric field inside the 3D body. For convenience of numerical calculation, the 3D CR body is divided into N cubic cells. Assuming that resistivity is uniformly distributed in each cell, Equation (2) then becomes:

$$\mathbf{F}(\mathbf{r}) = \mathbf{F}_1(\mathbf{r}) + \sum_{n=1}^N \Delta\sigma_n \int_{V_n} \tilde{G}^F(\mathbf{r}; \mathbf{r}') \cdot \mathbf{E}(\mathbf{r}_n) dv'. \quad (3)$$

From the above Equation (3), we can obtain the electric field in each cell of 3D CR body, namely:

$$\mathbf{E}(\mathbf{r}_m) = \mathbf{E}_1(\mathbf{r}_m) + \sum_{n=1}^N \Delta\sigma_n \cdot \left(\int_{V_n} \tilde{G}^E(\mathbf{r}_m; \mathbf{r}') dv' \right) \cdot \mathbf{E}(\mathbf{r}_n). \quad (4)$$

First, we obtain $\mathbf{E}(\mathbf{r}_m)$ by solving Equation (4).

3 Theory of inversion

In inversion, we use the damped least squares method to fit the forward simulation data with the measured data. The CR parameters are then steadily changed. Finally, the best fitting of data is achieved. The inversion procedure can be easily expressed as follows: \mathbf{f}_i represents the measured field, $\mathbf{F}_1(\mathbf{r})$ represents forward simulation data, and \bar{X} represents CR parameters of each cell. The fitting degree of the field between forward simulation and that measured by the relative deviation of $\delta_i(\bar{X})$ is given by:

$$\delta_i(\bar{X}) = [\mathbf{f}_{si} - \mathbf{F}_1(\mathbf{r})] / \mathbf{f}_{si}. \quad (5)$$

Therefore, the inversion fitting error $\phi(\bar{X})$ is:

$$\phi(\bar{X}) = \sum_{i=1}^k [\delta_i(\bar{X})]^2. \quad (6)$$

In Equation (6), the measured and simulation fields both have three components, and the equation $i=1,2,\dots,l$ represents every observation location or frequency point.

The forward and deviation functions, as well as the fitting error, are nonlinear. Thus, approximate linearization processing should be applied to the deviation function to overcome the difficulties in solving

nonlinear equations. Given a set of initial values \bar{X}^0 of model parameters, Taylor expansion at \bar{X}^0 is used, and each order over second-order partial derivatives is omitted, i.e.:

$$\delta_i(\bar{X}) \approx \delta_i(\bar{X}^0) + \sum_{k=1}^n \frac{\partial \delta_i(\bar{X}^0)}{\partial x_k} \Delta x_k, \quad (7)$$

where k is the k -th model parameter, and $\Delta x_k = x_k - x_k^0$ is the model modification. We then let $p_{ik} = \frac{\partial \delta_i(\bar{X}^0)}{\partial x_k}$.

Substituting Equation (7) into Equation (6) yields

$$\phi(\bar{X}) = \sum_{i=1}^l [\delta_i(\bar{X}^0) + \sum_{k=1}^n p_{ik} \cdot \Delta x_k]^2, \quad (8)$$

$\phi(\bar{X})$ is expressed as a multiple function of the conductivity model variables $\Delta x_1, \Delta x_2, \dots, \Delta x_n$, with the minimum condition given by:

$$\frac{\partial \phi(\bar{X})}{\Delta x_j} = 2 \sum_{i=1}^l \left[\delta_i(\bar{X}^0) + \sum_{k=1}^n p_{ik} \cdot \Delta x_k \right] \cdot p_{ij} = 0. \quad (9)$$

Equation (9) is further derived as:

$$\sum_{i=1}^l \sum_{k=1}^n p_{ij} p_{ik} \Delta x_k = - \sum_{i=1}^l \delta_i(\bar{X}^0) \cdot p_{ij}. \quad (10)$$

Subsequently, we can derive the following equation with the use of $j = 1, 2, \dots, n$:

$$(P^T P + \lambda) \cdot \Delta X = S, \quad (11)$$

where P is a Jacobian matrix with elements p_{ik} ; P^T is the transpose matrix of P ; λ is the damping factor, which is a positive constant; $\Delta X = (\Delta x_1, \Delta x_2, \dots, \Delta x_n)^T$; and S is the right-side vector with elements given by

$$s_j = - \sum_{i=1}^m \delta_i(\bar{X}) \cdot p_{ij}.$$

We use Equation (11) to calculate the model modification value ΔX and then take $\bar{X} = \bar{X}^0 + \Delta X$ as a new parameter. The fitting error is then recalculated. Numerous iterations are required to achieve a fitting error that is less than the small positive number ε , with \bar{X} as the most appropriate inversion result.

4 Partial derivative equation of CR parameters

\bar{X} is set as an array of complex coal-coal model parameters, with element x_{kl} denoting the l -th ($l = 1, 2, 3, 4$) CR parameter of the k -th ($k = 1, 2, \dots, N$) cell. $\Delta \bar{\sigma}$ is an array of complex conductivity differences, with element $\Delta \sigma_k$ denoting the differences between the complex conductivity of the k -th cell and the conductivity

of the surrounding rock. $\bar{\rho}$ is an array of CR, and its element ρ_k denotes CR of the k -th cell.

According to the foregoing expression, the Jacobian matrix element is

$$P_{ikl} = \frac{\partial \delta_i(\bar{X}^0)}{\partial x_{kl}} \tag{12}$$

Substituting (5) into (12) yields

$$P_{ikl} = -\frac{1}{f_{si}} \frac{\partial \mathbf{F}_i}{\partial x_{kl}} \tag{13}$$

According to the derivative rule of compound functions,

$$\frac{\partial \mathbf{F}_i}{\partial x_{kl}} = \frac{\partial \mathbf{F}_i}{\partial \Delta \sigma_k} \cdot \frac{\partial \Delta \sigma_k}{\partial x_{kl}} = \frac{\partial \mathbf{F}_i}{\partial \Delta \sigma_k} \cdot \frac{\partial \Delta \sigma_k}{\partial \rho_k} \cdot \frac{\partial \rho_k}{\partial x_{kl}} \tag{14}$$

we simplify the above equation as

$$\frac{\partial \mathbf{F}_i}{\partial x_{kl}} = \frac{\partial \mathbf{F}_i}{\partial \Delta \sigma_k} \cdot \frac{-1}{\rho_k^2} \cdot \frac{\partial \rho_k}{\partial x_{kl}} = -\Delta \sigma_k^{-2} \cdot \frac{\partial \mathbf{F}_i}{\partial \Delta \sigma_k} \cdot \frac{\partial \rho_k}{\partial x_{kl}} \tag{15}$$

where $\frac{\partial \rho_k}{\partial x_{kl}}$ is the partial derivative of CR to the l -th CR parameter of the k -th cell, which can be calculated by using Equation (1) [15].

Equation (1) is decomposed into the real and imaginary parts:

$$\rho(i\omega) = \rho_0 \cdot \left[1 - m + \frac{mR}{R^2 + I^2} - i \frac{mI}{R^2 + I^2} \right], \tag{16}$$

where

$$R = 1 + \omega \tau^c \cos \frac{\pi c}{2}, \tag{17}$$

$$I = \omega \tau^c \sin \frac{\pi c}{2}, \tag{18}$$

$$\frac{\partial \rho}{\partial m} = \rho_0 \cdot \left[\frac{R}{R^2 + I^2} - 1 - i \frac{I}{R^2 + I^2} \right], \tag{19}$$

$$\frac{\partial \rho}{\partial c} = m \cdot \rho_0 \cdot \left\{ \left[\frac{I^2 - R^2}{R^2 + I^2} \cdot \frac{\partial R}{\partial c} - \frac{2RI}{R^2 + I^2} \cdot \frac{\partial I}{\partial c} \right] + i \cdot \left[\frac{2RI}{R^2 + I^2} \cdot \frac{\partial R}{\partial c} + \frac{I^2 - R^2}{R^2 + I^2} \cdot \frac{\partial I}{\partial c} \right] \right\}, \tag{20}$$

$$\frac{\partial \rho}{\partial \tau} = m \cdot \rho_0 \cdot \left\{ \left[\frac{I^2 - R^2}{R^2 + I^2} \cdot \frac{\partial R}{\partial \tau} - \frac{2RI}{R^2 + I^2} \cdot \frac{\partial I}{\partial \tau} \right] + i \cdot \left[\frac{2RI}{R^2 + I^2} \cdot \frac{\partial R}{\partial \tau} + \frac{I^2 - R^2}{R^2 + I^2} \cdot \frac{\partial I}{\partial \tau} \right] \right\}, \tag{21}$$

$$\frac{\partial \rho}{\partial \rho_0} = 1 - m + \frac{mR}{R^2 + I^2} - i \frac{mI}{R^2 + I^2}, \tag{22}$$

$$\left. \begin{aligned} \frac{\partial R}{\partial c} &= \omega \tau^c \left[\ln \omega \tau \cdot \cos \frac{\pi c}{2} - \frac{\pi}{2} \sin \frac{\pi c}{2} \right] \\ \frac{\partial I}{\partial c} &= \omega \tau^c \left[\ln \omega \tau \cdot \sin \frac{\pi c}{2} + \frac{\pi}{2} \cos \frac{\pi c}{2} \right] \\ \frac{\partial R}{\partial \tau} &= \frac{c}{\tau} \omega \tau^c \cdot \cos \frac{\pi c}{2} \\ \frac{\partial I}{\partial \tau} &= \frac{c}{\tau} \omega \tau^c \cdot \sin \frac{\pi c}{2} \end{aligned} \right\} \tag{23}$$

We then solve the partial derivative $\frac{\partial \mathbf{F}_i}{\partial \Delta \sigma_k}$, which is the

total field on earth relative to the conductivity difference of each cell of the underground abnormal body. Assuming that background resistivity is known, we take $\mathbf{F}_1(r)$ as a constant in Equation (3) relative to CR. That is:

$$\frac{\partial \mathbf{F}_i}{\partial \Delta \sigma_k} = \frac{1}{\partial \Delta \sigma_k} \left(\sum_{n=1}^N \Delta \sigma_n \int_{V_n} \mathcal{G}^{ff}(r, r') \cdot \mathbf{E}(r_n) \right). \tag{24}$$

We let $p_{in}^f = \int_{V_n} \mathcal{G}^{ff}(r, r') \cdot \mathbf{E}(r_n)$, where \mathcal{G}^{ff} is the electromagnetic tensor function of the n -th cell of an abnormal body on the ground observation points. Through further derivation, we obtain:

$$\frac{\partial \mathbf{F}_i}{\partial \Delta \sigma_k} = p_{ik}^f + \left(\sum_{n=1}^N \Delta \sigma_n \frac{\partial p_{in}^f}{\partial \Delta \sigma_k} \right). \tag{25}$$

Two steps are required to solve Equation (25). First, the partial derivatives of the electric field of each cell relative to the conductivity difference are solved. Second, the partial derivatives of the total field relative to the conductivity difference are determined by superimposing the sum. \mathbf{F}_i has three components, namely these are x , y , and z . Taking the x -component as an example, we derive:

$$\frac{\partial (p_{in}^f)_x}{\partial \Delta \sigma_k} = \frac{\partial E_x(r_n)}{\partial \Delta \sigma_k} \Gamma_{xixn}^f + \frac{\partial E_y(r_n)}{\partial \Delta \sigma_k} \Gamma_{xiyn}^f + \frac{\partial E_z(r_n)}{\partial \Delta \sigma_k} \Gamma_{xizn}^f, \tag{26}$$

where $\frac{\partial E_x(r_n)}{\partial \Delta \sigma_k}, \frac{\partial E_y(r_n)}{\partial \Delta \sigma_k}, \frac{\partial E_z(r_n)}{\partial \Delta \sigma_k}$ are partial derivatives of the electric field of the n -th cell relative to the conductivity difference of the abnormal k -th cell. We also consider $\Gamma_{xim}^f = \int_{V_n} \mathcal{G}_{xi'}^f(r_i, r') dV', \Gamma_{xym}^f = \int_{V_n} \mathcal{G}_{xy'}^f(r_i, r') dV'$, and $\Gamma_{xizn}^f = \int_{V_n} \mathcal{G}_{xi'z'}^f(r_i, r') dV'$.

We can obtain $\frac{\partial E_x(r_n)}{\partial \Delta \sigma_k}, \frac{\partial E_y(r_n)}{\partial \Delta \sigma_k}, \frac{\partial E_z(r_n)}{\partial \Delta \sigma_k}$ by using the same method employed to solve Equation (4).

After discretization, matrix equations can be expressed as:

$$\frac{\partial E(r_m)}{\partial \Delta \sigma_k} = \int_{V_k} \mathcal{G}^f(r_m, r') E(r_k) + \sum_{n=1}^N \Delta \sigma_n \int_{V_n} \mathcal{G}^f(r_m, r') dv' \cdot \frac{\partial E(r_n)}{\partial \Delta \sigma_k} \quad (27)$$

Therefore, by solving n times n linear equations, we derive

$$\frac{\partial E_x(r_n)}{\partial \Delta \sigma_k}, \frac{\partial E_y(r_n)}{\partial \Delta \sigma_k}, \frac{\partial E_z(r_n)}{\partial \Delta \sigma_k}, (n = 1, 2, \dots, N; k = 1, 2, \dots, N).$$

Substituting these parameters into Equation (26) and then substituting Equation (26) into Equation (25), we can obtain the partial derivative matrix of the observation field on the earth relative to the conductivity difference of each cell of the abnormal body. $\frac{\partial \mathbf{F}_i}{\partial \Delta \sigma_k}$ is substituted by solving Equation (25) and the results were obtained by solving Equations (19)–(22) into Equation (15). Thus, we obtain the partial derivative matrix of total observation field on the earth to the CR parameters of Cole-Cole model.

5 Inversion examples of 3D CR body

As shown in Figure 1, a geoelectric model is designed for the inversion of CR parameters. We find 3D CR anomalies in homogeneous earth. These anomalies have the following parameters: centre coordinate of (0, 0, 200); size of 200×200×50m; excitation source coordinate of (800, 0, 0); frequencies of 0.1, 1, 10, and 100 Hz; electric dipole moment of 1 A·m; non-polarization of surrounding

rock resistivity of 100 Ω·m ; and CR parameters of $\rho_0=10\Omega \cdot m, m=0.6, c=0.1, \tau=10s$. The abnormal body is divided into 4×4×1 cells. Resistivity is assumed to be uniformly distributed in each cell and is equal to the value of the centre.

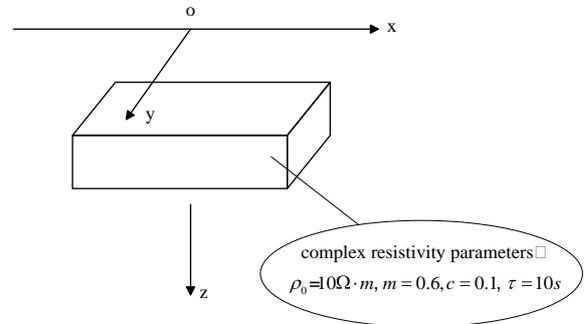


FIGURE 1 3D complex resistivity model in homogeneous earth

The four parameters of CR are inverted. The initial CR parameters of inversion are $\rho_0 = 30 \Omega \cdot m, \tau = 100s, c = 0.25$ and $m = 0.3$. The damping factor is 1, whereas its scaling factor is 5. This factor is calculated by using the inversion algorithm. Results after eight iterations are shown in Figure 2. The figures show that ρ_0, c and m converge close to the true value. However, the results of time constant τ is poor and is far from the true value. Moreover, fitting error decreases as iterations increases. As shown in Figure 3, fitting error is 1.40×10^{-3} , which is significantly greater than the given threshold 1×10^{-7} . Thus, this inversion process has failed.

According to analysis, accurate results cannot be obtained if the four parameters of each cell are inverted simultaneously because of the very strong correlation between chargeability m and time constant τ [16]. m is one of the key parameters of CR inversion, whereas τ can be determined through physical measurements or estimated from prior work experience. Therefore, we fix the time constant τ of each cell and then insert the other three parameters. In the subsequent inversion, we fix τ as a true value 10, and the initial value of the other three parameters are the same as described above. Results after 16 iterations are shown in Figure 4. For this inversion, the fitting error is 1.40×10^{-3} (Figure 5), which is less than the given threshold 1×10^{-7} . Thus, the inversion process is successful.

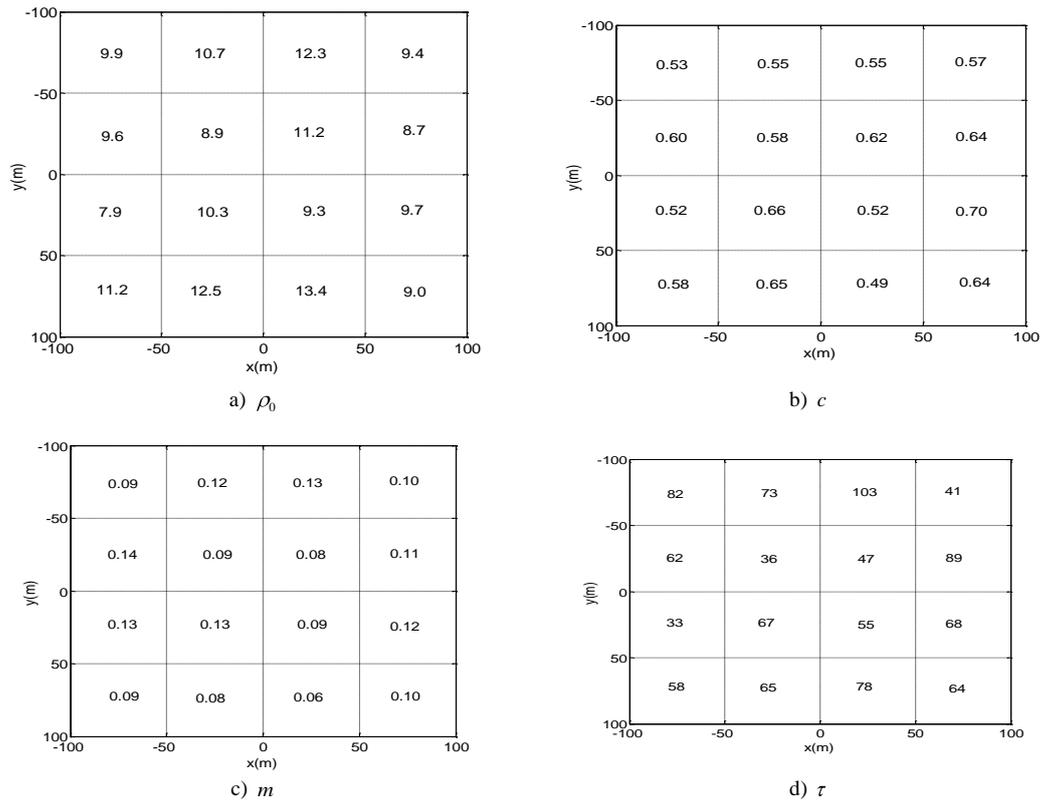


FIGURE 2 Inversion results of complex resistivity 200 m underground

As shown in Figure 4, although the difference between the initial and true values is significant, the CR parameters converge to the true value after inversion. This result indicates that the algorithm is correct and effective.

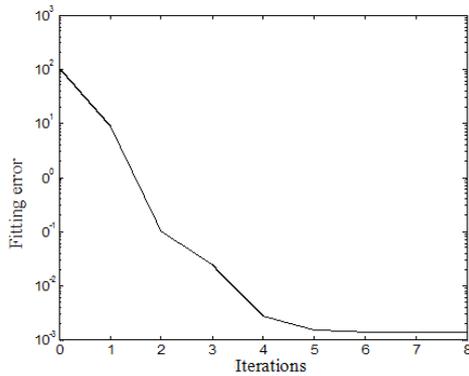


FIGURE 3 Relative error curve with the number of iterations

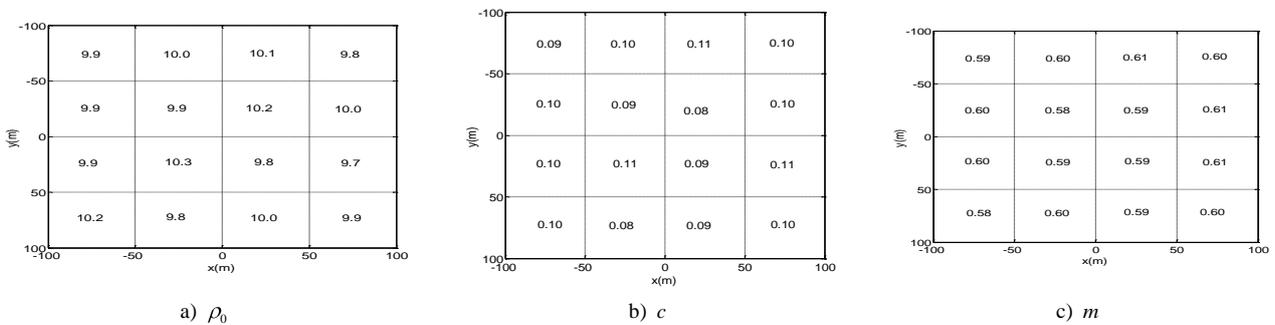


FIGURE 4 Inversion results of complex resistivity 200 m underground at a fixed time constant

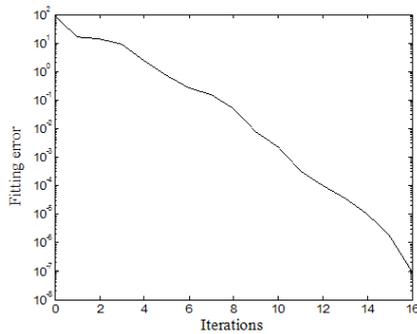


FIGURE 5 Relative error curve with the number of iterations at a fixed time constant

6 Conclusion

This study proposed and implemented a CR inversion algorithm. Through model simulation, we obtain the following conclusions:

References

- [1] Pelton W H, Ward S H, Hallof P G, Sill W R, Nelson P H 1978 Mineral discrimination and removal of inductive coupling with multifrequency IP *Geophysics* **43**(3) 588-609
- [2] Carlson N R, Hughes L J, Zonge K L 1981 Hydrocarbon exploration using induced polarization, apparent resistivity and electromagnetic scattering *Technical papers 51st Annual International SEG Meeting and Exposition Los Angeles* **3** 1339-58
- [3] Cole K S, Cole R H 1941 Dispersion and absorption in dielectrics, I. Alternating current characteristics *J.Chem Phys* **9** 341-51
- [4] Hohmann G W 1975 Three dimensional induced polarization and electromagnetic modeling *Geophysics* **40**(2) 309-24
- [5] Soininen H 1984 The behavior of the apparent resistivity phase spectrum in the case of a polarizable prism in an unpolarizable half-space *Geophysics* **49**(9) 1534-40
- [6] Soininen H 1985 The behavior of the apparent resistivity phase spectrum in the case of two polarizable media *Geophysics* **50**(5) 810-9
- [7] Weller A, Seichter M, Kampke A 1996 Induced-polarization modelling using complex electrical conductivities *Geophysical Journal International* **127**(2) 387-98
- [8] Luo Y Z, Fang S 1986 An approximate inversion of the apparent complex resistivity spectrum *Earth Science-Journal of Wuhan College of Geology* **11**(1) 93-102 (in Chinese)
- [9] Zhang G Q, Cui X W, Luo Y Z 1987 The determination of intrinsic parameters by inverting IP apparent spectrum *Geology and prospecting* **23**(4) 48-54 (in Chinese)
- [10] Wang T L, Zhang S Z 1990 A new algorithm for determination of intrinsic induced-polarization spectra of the earth *Acta geophysica sinica* **33**(6) 712-21 (in Chinese)
- [11] Liu S, Guan S Y, Gao P F 1994 Joint sip inversion for estimation of intrinsic cole-cole parameters of a polarizable ellipsoid *Acta geophysica sinica* **37**(S1) 542-51 (in Chinese)
- [12] Liu S, Xue J, Xu J H 2000 The evaluation of real cole-cole parameters of underground polarized bodies *Geophysical & Geochemical exploration* **24**(1) 51-61 (in Chinese)
- [13] Eaton P A 1989 3D electromagnetic inversion using integral Equation *Geophysical prospecting* **37**(4) 407-26
- [14] Hohmann G W 1975 Three-dimensional induced polarization and
- [15] electromagnetic modeling *Geophysics* **40**(2) 309-24
- [16] Luo Y Z, Zhang G Q 1998 Theory and Application of Spectral Induced Polarization *Society of Exploration Geophysicists :Tulsa USA*
- [17] Zhang H, Li T L, Dong R X 2006 Modeling electromagnetic responses of complex resistivity 3-D body using volume integral equation method *Coal geology & exploration* **34**(1) 70-4 (in Chinese)

Authors



Jianping Li, born in July, 1980, Weifang, Shandong, China

Current position, grades: the Lecturer of Shandong University of Science and Technology, China.

University studies: B.Sc. in Applied Geophysics, M.Sc. and PHD in Earth exploration and Information Technology from Jilin University in China.

Scientific interest: geophysical prospecting and numerical simulation of geophysical field

Publications: 15 papers

Experience: teaching experience of 7 years, 5 scientific research projects.



Tongxiao Shang, born in March, 1984, Fenyang, Shanxi, China

Current position, grades: the Engineer of Geological Survey of Jiangsu Province, China.

University studies: B.Sc. in applied geophysics from Jilin University in China, M.Sc. in Earth exploration and Information Technology from Jilin University in China.

Scientific interest: electromagnetic method.

Publications: 3 papers.



Yixiao Guan, born in April, 1983, Biyang, Henan, China

Current position, grades: the Engineer of Geological Survey of Jiangsu Province, China.

University studies: B.Sc. in Applied Geophysics, M.Sc. in Earth exploration and Information Technology from Jilin University in China.

Scientific interest: gravity, magnetism, electricity

Publications: 2 papers.

Authors' index					
An Xinguo	73	Li Jinxu	63	Wang Ying	108
An Yi	231	Li Juan	223, 231	Wang Yiqiang	131
Cai Ming	91	Li Meng	176	Wang Yujing	192
Chen Chunling	254	Li Ran	320, 327	Wang Zhan-ao	242
Chen Dongyan	176	Li Shouying	50	Wang Zhongmin	108
Chen Guangsheng	7	Li Yihui	363	Wang Ziya	320
Chen Lei	292	Li Zhang	143	Wen Bin	63
Chen Liwan	14	Lin Wei	373	Wu Jian-Zhang	368
Chen Qiang	14	Liu H	39	Wu Yujian	216
Chen Yaoting	205	Liu Jian	45	Wu Zhengren	347
Chen Yu	387	Liu Lin	393	Xiao Linjing	57
Cheng Binghua	122	Liu Liyang	384	Xiao Yu	39, 368
Chu Jian	333	Liu Mei	347	Xie Hualong	339
Ci Jiaojin	170	Liu Peikun	57	Xie Hui	14
Cui Xi-min	249	Liu Xiaoqing	306	Xin Gao	143
Dai Junyu	205	Liu Xinran	115	Xing Zhijie	210
Deng Li-jun	45	Liu Yang	387	Xu Hao	199
Dong Huimin	85	Liu Yan-Liang	236	Xu Jian	80
Du Xinling	216	Liu Yaqiu	22	Xu Jing	269
Fu Liping	223	Long Hai-Fei	68	Xu Lei	186
Gan Jianhou	63	Lu Xiaoyong	31	Xu Ming-xing	280
Gao Cai-yun	249	Ma Lifang	387	Xu Shan	333
Gao Wei	368	Meng Lingyi	216	Xu Tongyu	254
Gao Yan	127	Niu Xiaowei	14	Yang Jian	158
Gu Yan	131	Peng Jinshuan	186	Yang Junru	57
Guan Yixiao	398	Pu Zongshan	292	Yang Mei	312
Guo Peng	393	Qi Jiawei	254	Yang Ting	7
Han Chang-jun	148	Qian Chong	242	Yang Xinghua	57
Hao Shengang	143	Qin Tong	115	Yang Yao	306
Hao Xiaohui	127	Shang Tongxiao	398	Yang Yong	100, 358
He Nan	339	Shao Fei	122	Yao Guanxin	269
Hong X G	39	Sheng Jing	68	Yin Lei	363
Hou Ling	176	Shi Hao-ran	306	Yu Chao	292
Hu Yangcheng	275	Shi Tao	210	Yu Yue	280
Hu Yong-shi	280	Sun Zeyu	50	Yuan Xiuhua	131
Jia Tian	316	Sun Zhenyu	137	Zhai Sheping	108
Jiang Yicheng	192	Suo Chenxia	100	Zhang Bangcheng	131
Jiao Shuhong	137	Suo Juanjuan	85	Zhang Dongwen	73
Jilin Zhang	259	Tan Gewei	373	Zhang Hongjuan	127
Jin Baoquan	127	Tong Chunya	153	Zhang Huaping	164
Jing Lixian	231	Wang Chao	379	Zhang Qi	384
Jing Weipeng	22	Wang Danling	353	Zhang Shizhong	387
Kang Shouqiang	192	Wang Guoyu	333	Zhang Xiaonan	31
Kong Xianguang	363	Wang Jiajing	137	Zhang Yan	286
Lei Zhang	298	Wang Ling	393	Zhang Yuekan	57
Li Chao	384	Wang Long	254	Zheng Zuting	223
Li Dan	7	Wang Songling	347	Zhong Shun-Hong	91
Li Fei	339	Wang Xiaowen	363	Zhou Xiaoqin	131
Li Hongbing	14	Wang Yanfei	353	Zhou Zheng-Mao	91
Li Jianping	398	Wang Yaxin	387	Zhu Weiwei	358

Cumulative Index

Mathematical and Computer Modelling

Dan Li, Ting Yang, Guangsheng Chen A AHP-based method to solve contradiction matrix with multiple engineering parameters

Computer Modelling & New Technologies 2014 18(8) 7-13

Currently, in the use of TRIZ contradiction matrix table, users need to manually find optimization parameters and deterioration parameters of the invention for the corresponding inventive principles. When many parameters are queried, the user is hard to get the statistics, which are most likely to correspond to the invention and have to rely on tedious accumulative calculation to predict the most likely corresponding inventive principle. In this paper, we aimed to apply the analytic hierarchy process (AHP) to predict the inventive principle; on the basis of the successful cases data, we can take advantage of AHP for the statistics and projections of 40 invention principle through the optimization parameters and deterioration parameters chosen. In this way, we ranked 40 invention principles by the use of probability to give users inventive principles of efficient prediction results and provide the user with a practical guide at the same time.

Keywords: TRIZ, analytic hierarchy process, contradiction matrix table, multiple engineering parameters

Xiaowei Niu, Liwan Chen, Qiang Chen, Hui Xie, Hongbing Li An improved eigenstructure method for estimating DOA in the presence of parameters uncertainties

Computer Modelling & New Technologies 2014 18(8) 14-21

This paper presents an improved eigenstructure-based method for estimating the direction of arrival (DOA) of received signal in uniform circular-array, in the presence of sensor gain and phase uncertainties. A simple sensor gain and phase uncertainties calibration method, which does not require any prior knowledge of the DOAs, but also being capable of eliminating the DOA estimation ambiguity, is proposed. The performance of the proposed method is demonstrated by some representative computer simulation.

Keywords: eigenstructure, gain and phase uncertainty calibration, DOA estimation

Weipeng Jing, Yaqiu Liu Multiple DAGs reliability model and fault-tolerant scheduling algorithm in cloud computing system

Computer Modelling & New Technologies 2014 18(8) 22-30

In this paper, in order to provide the reliable scientific workflow scheduling problem for cloud computing, a dynamic of RANK-Hierarchical algorithm is put forward, which taking account of communication contention as well as supporting task dependencies (CCRH). A communication contention model is first defined, as soon as the earliest completion of the primary and backup task is deduced, besides the executive processor is limited, use dynamic hierarchical method and calculate of each DAG unfair degree factor for multiple DAGs scientific workflow. It can deal with the problem that multiple DAGs workflow comes at different time and have various kinds of structure. Both the theory and experiments have proved the algorithm not only improve the scheduling fairness of multiple DAGs workflow but also shorten the average execution Makespan effectively while meeting reliability constraints and meanwhile the produce well robustness.

Keywords: cloud computing, multiple DAGs, RANK-Hierarchical, reliability

Xiaonan Zhang, Xiaoyong Lu Multi-state system reliability assessment based on Bayesian networks

Computer Modelling & New Technologies 2014 18(8) 31-38

This paper considers a problem of multi-state system reliability modelling and assessment. By using the advantages of uncertainty reasoning and figurative expression of Bayesian network, a new method of modelling and assessment of multi-state system reliability based on BN is proposed to determine the nodes of BN and the multiple states of components of system, and to give the probability of each state and then utilizing conditional probability distributing (CPD) to describe the relationship among the component states, so as to express the states of correlated nodes and build a BN model of multi-state system. The model can clearly express the multiple states of system and component and the state probability, and also call directly calculate the system reliability on the basis of multiple state probabilities of component, thereby carrying out qualitative analysis and quantitative assessment of multi-state system reliability. By means of an example of multi-state radar system, we give the detailed multi-state system reliability analysis process based on BN. This paper not only proves the effectiveness of assessment of multi-state system reliability based on BN,

but contributes to good help of complex system reliability, safety analysis.

Keywords: multi-state system, Bayesian networks, reliability assessment

X G Hong, H Liu, Y Xiao Fast fractional-pel interpolation algorithm of H.264 based on CUDA

Computer Modelling & New Technologies 2014 18(8) 39-44

H.264 video standard introduces fractional pixel motion compensation technology to obtain a more precise motion vector and a higher compression ratio. But, it increases the complexity of the motion compensation process at the same time. In order to solve the difficulties, we analysis the procedure of fractional-pel interpolation in H.264 and propose a fast fractional-pel interpolation algorithm based on CUDA. Experimental results show that the fast algorithm enables locating fractional pixel effectively and improves the speed of fractional pixel motion estimation. Compared with the CPU serial algorithm, the fast algorithm can significantly improve encoding rate almost four times in processing high-resolution video sequences.

Keywords: H.264, CUDA, interpolation algorithm, fractional-pel

Li-jun Deng, Jian Liu New approach for ventilation network graph drawing based on Sugiyama method and GA-SA algorithm

Computer Modelling & New Technologies 2014 18(8) 45-49

Ventilation network graph has an important place in the management of a coal mine. In that case, aesthetics plays a major role for generating readable and understandable layouts. Besides, the drawing is required to be oval. The traditional longest path method for drawing ventilation network graph is inefficient and cannot effectively reduce the number of arc crossings because of the geometric intersection method. In this paper, we developed a new approach to draw ventilation network graph, consist of Sugiyama method framework, the longest path method and GA-SA algorithm. The longest path method was employed to rank nodes, and long arcs were removed by solving integer programming problem to minimize the sum length of ventilation network. Then genetic algorithm and simulated annealing algorithm were adopted to optimize the node order on reducing the number of arc crossings. In order to make the drawing be oval, a modified version of the longest path method was made to calculate node coordinates and arc shapes, which is called the longest parallel path method. Finally, computational experiments were carried out on two test ventilation network with our new approach.

Keywords: ventilation network graph, longest path method, integer programming, Sugiyama method, simulated annealing-genetic algorithm

Zeyu Sun, Shouying Li Multicast routing algorithm based on cloud computing strategy in wireless sensor networks

Computer Modelling & New Technologies 2014 18(8) 50-56

In the paper, the existing wireless sensor network routing protocol classification and comparative studies, then A targeted selection of typical clustering routing protocol Low Energy Adaptive Clustering Hierarchy (LEACH) protocol for the study, points Insufficient analysis of LEACH, and on this basis to select the optimal number of cluster head LEACH protocol, Selected on the basis of cluster head node, communication between clusters, cluster head node in a distributed fashion to improve "change After the agreement into improving the selection criteria LEACH cluster head, so that the cluster head node distributed more evenly, avoiding Free a single node excessive energy consumption. Finally, the improved LEACH protocol is added by way of randomly selected based on the energy value of closed Residual nodes on MATLAB simulation platform, the simulation results show that adding the residual section Point effectively extending the network life time improves the efficiency of the network.

Keywords: wireless sensor network, routing protocol, network lifetime, sensor nodes

Yuekan Zhang, Peikun Liu, Linjing Xiao, Xinghua Yang, Junru Yang Computational simulation of the effects of vortex finder diameter on the air core in a hydrocyclone separator

Computer Modelling & New Technologies 2014 18(8) 57-62

Air core is inherent to the solid-liquid hydrocyclone, the air core dimension plays a significant role in the separation efficiency. However, the formation mechanism of the air core in hydrocyclone has not arrived at an agreement. To further understand the flow behaviour of air core in hydrocyclones, in this paper, the volume of fluid (VOF) multiphase model and the Reynolds Stress Model (RSM) were adopted in this study to simulate the flow fields inside a hydrocyclone. The effect of the varying vortex finder diameter on the formation and development of air core was

analysed, and the generation and development of the air core were investigated. The results showed that air core could be generated in shorter time and more stable state with larger vortex finder diameter. In addition, the diameter of the air core increased with the vortex finder diameter.

Keywords: hydrocyclone, computational fluid dynamics, VOF multiphase, diameter of vortex finder, air core

Jianhou Gan, Bin Wen, Jinxu Li Research on ontology mapping of tourism information resources based on description logic

Computer Modelling & New Technologies 2014 18(8) 63-67

In this paper, the Ontology mapping of tourism information resources is discussed. Ontology mapping tries to find the corresponding relationships between two entities, and to achieve interoperability and information sharing in heterogeneous Ontology. We researched the Ontology mapping of tourism information resources and introduced the Description Logics to solve the Ontology mapping. The key relationships and determining relationships among Ontologies in description logic based Ontology mapping is first described. Then a description logic based mapping model is proposed, which can solve the problem of Ontology semantic heterogeneity.

Keywords: tourism information resources, resource integration, ontology isomerism, ontology mapping, description logic

Jing Sheng, Hai-Fei Long Parametric modelling and simulation on oblique cutting based on MSC.Marc

Computer Modelling & New Technologies 2014 18(8) 68-72

The finite element simulation of oblique cutting is a complex and professional process. It is necessary to build a system to construct a model in order to obtain simulation values more conveniently and rapidly. The key techniques of 3-D parametric modelling with MSC.Marc software metal oblique cutting simulation was presented in this investigate. The modelling rule based on the process was carried out. The system, designed using C++ Builder, can access data, which includes the geometrical angles and dimensions of tool, the sizes of workpiece, the relative position between tool and workpiece. Meanwhile their properties and cutting conditions, etc. were stored. The procedure file modelling in the MSC.Marc environment automatically is generated by the program. So the parametric modelling of simulation is completed by calling the procedure file. Further, an example was given and the simulation model was also verified. Therefore, the parametric modelling is a kind of effective way for metal cutting simulation.

Keywords: oblique cutting, numerical simulation, parametric modelling, interface design

Dongwen Zhang, Xinguo An H_∞ fault-tolerant control for nonlinear singular system via a fault diagnosis observer

Computer Modelling & New Technologies 2014 18(8) 73-79

An H_∞ fault-tolerant control scheme based on fault diagnosis observer was developed for a class of nonlinear singular systems with external disturbances and actuator faults. A fault diagnosis observer was designed to estimate the system states and the actuator faults and a sufficient condition for the existence of this observer was presented in the form of feasibility problem of a linear matrix inequality. Based on linear matrix inequality (LMI) technique and the estimates of the states and faults, an H_∞ fault-tolerant control scheme was worked out. The H_∞ fault-tolerant control system via a state feedback controller can be made solvable, impulse free, asymptotically stable, and the effect of external disturbances on the system was attenuated in terms of the prescribed H_∞ performance index. Finally, a simulation example was given to illustrate the procedure of designing the fault diagnosis observer and the state feedback controller, and the simulation result showed the effectiveness of the proposed method.

Keywords: singular systems, H_∞ control, fault-tolerant control, observer, linear matrix inequality (LMI)

Jian Xu Equilibrium distributions of the queue length in M/M/c queuing system

Computer Modelling & New Technologies 2014 18(8) 80-84

In this paper, An M/M/c queuing system with multiple working vacations and vacation interruption is considered. All servers work at a lower rate rather than completely stop during a vacation period. Meanwhile, we introduce another vacation policy: vacation interruption. Otherwise using matrix-geometric solution method, we obtain steady-state distribution for queue length.

Keywords: M/M/c; working vacation; matrix-geometric solution

Juanjuan Suo, Huimin Dong Novel method for quality assessment of computational translation

Computer Modelling & New Technologies 2014 18(8) 85-90

To overcome the shortcomings that there are few feasible methods and models in the comprehensive assessment on the quality of the computational translation, a novel mathematical tool, the unascertained measure was introduced. After the introduction of the basic knowledge of the Unascertained Sets, the unascertained measure was defined and the comprehensive assessment model was set up. Then the method was introduced to the quality assessment of the machine translation. Engineering practices shows that the method can complete the assessment systematically and scientifically without any assumption.

Keywords: computational translation, quality assessment, unascertained measure, model

Zheng-Mao Zhou, Shun-Hong Zhong, Ming Cai An efficient and flexible modelling approach for multi-DSP system

Computer Modelling & New Technologies 2014 18(8) 91-99

With the development of the information technology, single digital signal processor (DSP) cannot meet the requirements of massive data processing. Multi-DSP parallel processing mode has been commonly used in real-time information processing system. New technology is also making it much easier to integrate multiple DSPs into a single silicon chip. However, designers of a new multi-DSP system and software are confronted with problems such as short product life-time. Meanwhile, product verification is indispensable before launching into the market. In this paper, a multi-DSP simulation platform is developed to solve these problems. The designed multi-DSP platform is based on an ISS-SystemC structure and has three common interconnect interfaces. An AMBA bus-shared memory model is designed for the expansion of the simulation system. A thread-agent method is proposed to optimize the performance of SystemC thread and the experiment results show that the multi-DSP parallel processing mode can improve the processing performance of the system significantly.

Keywords: simulation platform, multi-DSP, ISS-SystemC, SystemC optimization

Chenxia Suo, Yong Yang A study on application of judgment matrix intelligent correction method in satisfaction evaluation

Computer Modelling & New Technologies 2014 18(8) 100-107

In order to improve rural folk house renovation in the satisfaction evaluation accurately, this paper puts forward a model for rural folk house renovation in the satisfaction evaluation based on intelligent expert judgement matrix adjustment method(AGA-LCAHP).By means of extracting the offset degree resulting in the inconsistency of AHP judgment matrix, this paper puts forward the new method of using accelerating genetic algorithm to locate the element of judgment matrix and calculate the AHP element ranking weight. This algorithm takes offset information as the foundation of correcting judgment matrix to avoid the subjectivity of correction; at the same time, it reserves and extracts the consistency information of judgment matrix, with the consistency index as the orientation of optimization. The case study result shows that the AGA-LCAHP method features high computational accuracy and stable calculation result and also has the popularization and application value in other comprehensive assessment.

Keywords: analytic hierarchy process, judgment matrix, consistency, offset degree, genetic algorithm

Computer and Information Technologies

Ying Wang, Zhongmin Wang, Sheping Zhai Spectral colour calibration for multi-ink printer

Computer Modelling & New Technologies 2014 18(8) 108-114

To implement colour calibration during outputting multispectral images on a multi-ink printer, a new spectral colour calibration method is proposed. Firstly, by uniform sampling in the multi-ink printer colour space, measuring the spectral reflectance of the samples and then transforming the reflectance data to a low dimension spectral space, a forward look-up table is created. Then by sampling in the low dimension spectral space and using a nonlinear optimization to calculate the mapping points of these samples in the printer colour space, a backward look-up table is established. Meanwhile, to improve the optimization accuracy and shorten the computing time, an algorithm is designed to determine the optimization parameters based on the samples. Finally, a multi-linear interpolation method is carried out on the forward and backward look-up table to achieve the spectral colour calibration of the multi-ink printer. Experiments show that the new method not only takes advantage of the high calibration precision and less time-consuming of the look-up table, but also solves the problem brought by the high dimension of the spectral data to the look-up table method by utilizing the nonlinear optimization and dimension reduction. Compared with the spectral colour calibration model methods, the new method improves the colorimetric and spectral precision obviously. It also

raises the time efficiency of the inverse calibration significantly.

Keywords: spectral colour calibration, multispectral image, multi-ink printing, look-up table

Tong Qin, Xinran Liu A trust-based resource selection algorithm in Cloud Computing

Computer Modelling & New Technologies 2014 18(8) 115-121

From the point of the safety in Cloud Computing and the virtual resource under the Cloud, here proposed a cognitive trust model of Cloud virtual resource, based on Bayesian, and the model supports the effective resource selection as a basis. Furthermore, on the basis of the description of process and question in virtual resource selection, here comes a Trust and Resource Scheduling oriented Cloud resource selection model which takes QoS, trust, resource scheduling and other indexes into account. After applying the improved CHC-TSSM Genetic Algorithm in the model, the simulation experiment confirms the feasibility and efficiency of the algorithm, which can resolve the trust and scheduling problem in resource selection effectively.

Keywords: resource selection, trust, Cloud Computing

Binghua Cheng, Fei Shao Optimal routing strategy on weighted networks

Computer Modelling & New Technologies 2014 18(8) 122-126

It is of great importance to improve the transfer capacity of the weighted networks. In this paper, the traffic dynamics on weighted networks is investigated based on global information. It is shown by simulations that the weighted network transfer capacity depends strongly on the tuneable parameter in three different node delivery capability schemes: constant, proportional to node degree and proportional to node strength. Furthermore simulations on both computer-generated and real world networks show that different tuneable parameter is suitable for different node delivery capability scheme.

Keywords: weighted network, BBV network, routing strategy, transfer capacity

Baoquan Jin, Xiaohui Hao, Hongjuan Zhang, Yan Gao The power supply design software model for mine based on VBA technology

Computer Modelling & New Technologies 2014 18(8) 127-130

Coal mine power supply system is increasingly complex, resulting in difficulties for power computing and tuning. A software design program is proposed by using VBA development tools embedded in AutoCAD secondary development, adopt object-oriented program, ActiveX automation technology and visual programming techniques to achieve human-computer interface and database interaction, design one of the mine power supply design software with function mapping, statistical calculations, protection setting, design reports automatically output, data maintenance. The study shows that the software can greatly simplify the power supply design process, improve efficiency, and can effectively improve the coal mine management level.

Keywords: VBA, secondary development, ActiveX automation technology, visual programming

Yan Gu, Yiqiang Wang, Xiuhua Yuan, Xiaoqin Zhou, Bangcheng Zhang Software reliability allocation model of CNC system based on software architecture

Computer Modelling & New Technologies 2014 18(8) 131-136

In order to guarantee the implementation of reliability target, software reliability allocation model of CNC system was established based on software architecture. Software architecture of CNC system was set up, which decomposed CNC system into the functional units, reliability indexes of the system can be distributed into each component from top to bottom. The relative weight of software element in each level of the architecture was determined with analytic hierarchy process (AHP) method. The software reliability allocation model was built by taking the maximum practicability of CNC system as the target function, the reliability and cost function of component as the constraints. The reliability of each component was calculated through culture algorithm (CA). According to the result, the reliability allocation worked out is reasonable and feasible, and during the development of allocation model, the practicability of CNC system was guaranteed and development cost was also saved effectively.

Keywords: software reliability of CNC system, software architecture, reliability allocation, analytic hierarchy process, cultural algorithm

Jiajing Wang, Shuhong Jiao, Zhenyu Sun A novel unsupervised segmentation for remote sensing image using MRF

Computer Modelling & New Technologies 2014 18(8) 137-142

The image segmentation is the basis of image interpretation in remote sensing applications and plays vital role in image analysis. The Markov Random Field (MRF) approach is widely studied for use in segmentation of remote sensing image, which is an important extraction technique in recognition problems. This paper presents an unsupervised segmentation method for remote sensing image using the MRF. A novel neighbourhood system for the energy function has been proposed, the segmentation of remote sensing image and the optimization process of the parameters are performed simultaneously for the unsupervised segmentation in iterative condition. The experimental results on Synthetic Aperture Radar (SAR) images show that the proposed method performs better than the conventional Bayesian segmentation methods.

Keywords: remote sensing image, MRF, unsupervised segmentation, parameter estimation, SAR

Shengang Hao, Zhang Li, Gao Xin Formal resource request representation for remote environment control system

Computer Modelling & New Technologies 2014 18(8) 143-147

Many open distributed systems across Internet such as those in grid computing and RECS (remote environment control system) involve the requesting, allocation and maintenance of sorts of resources. The discovery of large amount of resources in different sites is an important issue for the design of these systems. The booming semantic Web technology provides a suitable infrastructure for the publishing, requesting and matchmaking of resources. This paper presents a generic representation for quantified resource requesting with Semantic Web. It allows the representation of complex resource descriptions such as containment hierarchies and disjoint constraints between them. A model-theoretic semantics for matchmaking with countable resources is given for this representation. A constraint-based technique for the matchmaking check with such representation is designed to ensure the correctness for remote environment control system.

Keywords: quantified resource, RECS (remote environment control system), resource matchmaking, service-oriented architecture

Chang-jun Han Study on load capacity-based cascading failure model in the computer network Study on load capacity-based cascading failure model in the computer network

Computer Modelling & New Technologies 2014 18(8) 148-152

Studies on the cascading failure process and characteristics in the computer network are beneficial to guiding the system construction and improving the performance. Based on the load initialization capacity of the computer system, routing control strategies and node forwarding rate, this paper constructed one cascading failure model considering service performance in the computer network specific to the influence of cascading failures on the service performance of the computer system network. This model considered multiple influence parameters and effectively measured the variable values of influence parameters of cascading failures on the service performance of the computer system network. Through comprehensive analyses, this model can effectively provide practical guiding significance for the prevention and control of cascading failures in the network.

Keywords: load capacity, computer network, service performance, cascading failure

Chunya Tong A novel edge detection method based on 2-D Gabor wavelet

Computer Modelling & New Technologies 2014 18(8) 153-157

With the features of substantial data and complex landmark, remote sensing images need a higher requirement for edge detection operator. Due to the limitations of grads operator and Canny operator in edge detection, this paper presents an edge detection method based on 2-D Gabor wavelet real part and the experimental analysis shows this method was better on edge detection.

Keywords: edge detection, 2-D Gabor wavelet, real part, remote sensing images

Jian Yang A strategy for fault management in LDC wireless sensor network

Computer Modelling & New Technologies 2014 18(8) 158-163

With rapid development of hardware, wireless sensor networks (WSN) have been applied in a wide range of fields. However, energy cost constrains putting WSN into use. To reduce energy cost, extending life time, WSN in low-duty-cycle (LDC) draws researchers' attention. In general, work time of a node only occupies 0.1%-10% in a cycle. This model certainly reduces the energy for idle listening. On the other hand, it makes the probability of congestion very

high due to a node that can only receive packets when waking up. This paper proposes a new LDCWSN model to solve the congestion from duty schedule. With the model, we show a strategy for WSN fault averting, diagnosing and recovery based on congestion in nodes. We include some attributes of LDC WSN in our strategy, i.e. probability of congestion, scheduler, and link quality. By improving the selection of nodes on every level, we get a low rate for network's fault appearance, low E2E delay and long lifetime. The simulation's result shows that our strategy has a better performance in packet loss, energy cost and time delay than proposed WSN fault management.

Keywords: low-duty-cycle, congestion control, congestion recovery, fault management, wireless sensor network

Operation research and decision making

Huaping Zhang A performance evaluation model of green supply chain based on fuzzy analysis method of multi-attribute decision-making

Computer Modelling & New Technologies 2014 18(8) 164-169

This paper proposes a performance evaluation model of green supply chain based on fuzzy analysis method of multi-attribute decision-making. In this model, an evaluation index system is established with economic profit, environment protection, business process and customer service taken into consideration. Fuzzy analysis method of multi-attribute decision-making is introduced to get the fuzzy incidence degree of different performance evaluation indicators. Analysis of performance evaluation of green supply chain is based on the fuzzy incidence degree. Finally, the model and the algorithm are proved to be scientific and feasible through case study.

Keywords: green supply chain, performance evaluation, multi attribute decision making, fuzzy theory, model

Jiaojin Ci A performance evaluation model of supply Chain based on extension correlation function

Computer Modelling & New Technologies 2014 18(8) 170-175

Performance evaluation of supply chain is complex and of uncertainty, and is influenced by factors of multiple levels. Directing at the features mentioned above, this paper studied the performance evaluation of supply chain of multiple attributes, and put forward a performance evaluation model and algorithm of supply chain based on extension correlation function. Via analysis of relative factors in the process of performance evaluation of supply chain, the model gave out an index system of performance evaluation of supply chain and by standardization of different evaluation indexes, it built an improved extension correlation function between evaluation indexes of enterprise supply chain performance and the ideal range of supply chain performance. Thus, the comprehensive weighted extension goodness between enterprise supply chain performance and the ideal range of supply chain performance could be obtained. According to the value of comprehensive weighted extension goodness, the supply chain with the optimal implementation effect could be selected, so as to offer effective support for the follow-up implementation of supply chain. Finally, in order to offer a scientific method for improving the supply chain and the enterprise competitiveness, the model and algorithm was tested an actual case.

Keywords: supply chain, performance evaluation, extension correlation function, model

Ling Hou, Meng Li, Dongyan Chen Sourcing and pricing strategy research of competition supply chain under supply disruption

Computer Modelling & New Technologies 2014 18(8) 176-185

Under the environment of supply disruption, it is significant to study decision-making, because sourcing strategies of retailers impact the profit of the supply chain while the pricing strategies of suppliers affect all aspects of the supply chain. In this paper, the demand distribution function of each supply chain is obtained, which is based on the total demand of two supply chains with given distribution function, and the sourcing and pricing problems are obtained in supply chain network under the environment of supply disruption. In order to decompose the total demand with the given distribution function, customer choice theory is adopted to acquire the demand of each supply chain. By game theory and optimization theory, we obtain the sourcing strategies of two retailers and the pricing strategies of two suppliers in this system. Finally based on the assumption of a uniform demand distribution, the outcomes of the proposed models are demonstrated with a numerical example. The results show that when disruption probability or delivery cost are high, retailers will only order from the spot market although the spot market wholesale prices are a little high; but when the disruption probability is moderate or low, the retailer would rather place orders from suppliers. Specific purchasing method depends on the competition ability between suppliers.

Keywords: supply chain network, supply disruption, sourcing strategies, pricing strategies

Jinshuan Peng, Lei Xu Lane changing intent identification based on logistic regression model*Computer Modelling & New Technologies 2014 18(8) 186-191*

To reduce the risk of the lane changing behaviors, based on integrated collection platform, the research group conducts experiments under real road environment for the purpose of studying drivers' lane changing intent identification. On the basis of the drivers' fixation characteristics of the rearview mirrors before changing lanes, the length of lane changing intent time window is determined. Based upon differential analysis of visual characteristics between lane keeping and lane changing intent stages, saccade numbers, visual search extent, saccade amplitude, standard deviation of head rotation angles in the horizontal direction are selected as the characteristic indexes of the identification. The logistic model is built according to feature extraction of the leaning samples, then applied to the identification process after the validity test. Results show that the identification success rate may reach 90.42%, thus verifying the feasibility and effectiveness of the logistic model to identify drivers' lane changing intent.

Keywords: lane change, intent identification, logistic model, index system

Yujing Wang, Yicheng Jiang, Shouqiang Kang The application of time domain and frequency domain statistical factors on rolling bearing performance degradation assessment*Computer Modelling & New Technologies 2014 18(8) 192-198*

Rolling bearing performance degradation assessment is a predict and prevent technology. In order to assess the performance degradation degree of the rolling bearing, and make the time domain and frequency domain statistical factors be applied more effectively in rolling bearing performance degradation assessment, a comprehensive analysis method is proposed based on time domain and frequency domain statistical factors. Time domain and frequency domain statistical factors are calculated and analysed for the life cycle data of the rolling bearing. Outer raceway moderate fault and severe fault of the rolling bearing can be distinguished well by peak-to-peak level, the root-mean-square (RMS) value, and kurtosis value of time domain factors; normal state and mild fault can be distinguished better by frequency centroid, F_3 , F_4 and F_5 of frequency domain factors than each time domain factor. The outer raceway performance degradation condition of the rolling bearing can be monitored well by using the proposed comprehensive analysis method, which uses partly frequency domain factors to analyse mild fault and partly time domain factors to analyse moderate fault and severe fault.

Keywords: rolling bearing, life cycle, statistical factor, performance degradation assessment

Hao Xu Evolvement of the Nanjing urban green land based on GIS analysis*Computer Modelling & New Technologies 2014 18(8) 199-204*

Nanjing is one of the important central cities in the Yangtze River Delta, which goes through dramatic urbanization development in recent 40 years. In this research, green land distribution data in 1966, 1981 and 2004 were extracted from topographic maps, aerial imagery, and so on. Time series analysis has been conducted on the GIS platform; characteristics of the evolution of urban green land from 1966 to 2004 were summarized from the perspective of the scale, function, pattern analysis. The results reveal that an environmental function plays an important role, status of recreational function began to rise, the spatial pattern of green land has a tendency of specialization and complication, the downward of its total size accelerates. Woodland which maintain the basic pattern of the recent 40 years green land system in Nanjing remains stable structure.

Keywords: green land, evolvement, GIS, analysis, Nanjing

Yaoting Chen, Junyu Dai A comparative analysis on circulation efficiency of different banana circulation modes in Zhangzhou city based on DEA model*Computer Modelling & New Technologies 2014 18(8) 205-209*

In the paper, the software DEAP was used to evaluate relative efficiency of 8 kinds of banana circulation modes in Zhangzhou City. CCR model and BCC model were employed. The related index included input (unit circulation cost, circulation time, and circulation loss rate) and output (net profit). The result showed that, the comprehensive efficiency, pure technology efficiency, scale efficiency in Mode 1 (farmers-third party logistics-supermarket-consumers) and Mode 8 (farmers-banana sales stalls-consumers) were relatively efficient. Therefore, it was judged that the main factors affecting the circulation efficiency were the compression of circulation level and the ascension of scale and professional level.

Keywords: DEA, circulation mode, circulation efficiency

Tao Shi, Zhijie Xing A Comprehensive FAHP evaluation model on domestic sports economy development mode

Computer Modelling & New Technologies 2014 18(8) 210-215

Sports economy refers to rational allocation of various resources to develop sports related functions and economy. This study identified the bottleneck factors constraining the development of sports economy. A fuzzy hierarchical evaluation model was developed using FAHP. The results indicated internal competitiveness, development of market-oriented economy and social development played important role to sports economy. Thus, policy recommendations on sustainable and sound sports economic development were proposed. The future prospect and development mode were analysed.

Keywords: sports economy, fuzzy analytic hierarchy process, evaluation model

Xinling Du, Lingyi Meng, Yujian Wu A study on the determinants of e-commerce customer satisfaction

Computer Modelling & New Technologies 2014 18(8) 216-222

Based on existing research, a model and evaluation index systems are developed for assessing e-commerce satisfaction of customers. We centralizes associated questionnaires for six topics, namely, convenient operation for online shopping, product information, distribution service, safety and reliability of the system, handling of customer complaints, and staff services, that are administered among college students in Guangdong Province of China who have experienced purchasing items online. The proposed model explores the factors that affect customer satisfaction. The effects of these factors are then analysed using descriptive statistics and factor analysis method to verify the accuracy of the model and to draw Safety and reliability of the system has the highest correlation coefficient with customer satisfaction, different gender and grade factors significantly affect customer satisfaction.

Keywords: e-commerce, customer satisfaction, factor analysis model, evaluation index system

Liping Fu, Juan Li, Zutong Zheng A study on the efficiency of public culture service based on DEA cross evaluation

Computer Modelling & New Technologies 2014 18(8) 223-230

This paper analysed the inputs and outputs of public cultural services in 31 provinces of China in 2012, including municipalities and autonomous regions, based on the CCR model of DEA and the DEA cross-evaluation model. The DEAP2.1 software was also used for this empirical analysis to probe the performance and problems of the inputs and outputs of public cultural services. From analysis, this paper reached some conclusions, including the development of cultural services should be based on the increase of industrial inputs, as well as the optimization of resource allocation, so as to achieve the optimum state of inputs and outputs. The research results can provide a reference for the further improvement of the quality and inputs and outputs efficiency of public cultural services in all regions.

Keywords: Public Culture Services, DEA Model, Inputs and Outputs, Performance Evaluation

Lixian Jing, Juan Li, Yi An Application of CCD-model-based DEA analysis method in research on agricultural economic growth

Computer Modelling & New Technologies 2014 18(8) 231-235

This paper presents research into the statistical data related to China's economic growth from 1990 to 2011, obtained by application of the DEA analysis method to the data in the China Statistical Yearbook 2012. First, a brief introduction is given to the principle structure of the Data Envelopment Analysis mathematical model. Second, the CCD model is established, based on modelling of the decision-making unit DMU_j in the Data Envelopment Analysis model; and a detailed discussion is made of the corresponding linear program. Finally, the optimum value of this program is obtained and the results evaluated through study of the data collected by the mathematical model established in the paper. The main conclusions are as follows: the per capita income of China's farmers is closely related to the national total agricultural economic output and the total power of agricultural machinery, but is not that closely related to the crop planting area, area of affected crops or crops disaster area. Further, China's agricultural development shall lead to an advanced level of huge mechanization.

Keywords: CCD model, DEA analysis method, programming model, non-Archimedean infinitesimal

Yan-Liang Liu Impact analysis of trade openness on China's economic growth

Computer Modelling & New Technologies 2014 18(8) 236-241

Trade openness plays can affect China's economy significantly. In this paper, effect of trade openness on China's economic growth was discussed by using the error correction model, Granger causality test and impulse-response function. Moreover, an in-depth analysis on strategies that can facilitate China's economic development was carried out. The empirical research results demonstrated that: 1) there's a short-run equilibrium relationship rather than a long-run one between trade openness and economic growth; 2) the import trade is related with the export trade. Due to the hedging, their collaborative effect is smaller than their independent effects; 3) trade openness can facilitate economic growth within a certain time period, which presents a "U-shaped" influence mode.

Keywords: trade openness, economic growth, effect model

Chong Qian, Zhan-ao Wang Influence of China's trade imbalance on economy in the background of great nation

Computer Modelling & New Technologies 2014 18(8) 242-248

The developed countries have been accusing China for its trade surplus for a long time. In order to verify the influence of trade imbalance on the economy, our research applied techniques like error correction function, Granger test and impulse response function. The research outcomes manifest that: the influence of trade imbalance on the economy is not significant. On the contrary, the economy imposes certain impact on trade imbalance. Both imports and exports significantly promote the development of economy. Furthermore, imports remarkably facilitate exports. This can also be attributed as the reason of China's trade deficit before 1992.

Keywords: great nation, trade imbalance, error correction model

Cai-yun Gao, Xi-min Cui Nonlinear time series of deformation forecasting using improved BP neural networks

Computer Modelling & New Technologies 2014 18(8) 249-253

Although the back propagation neural network has been successfully employed in various fields and demonstrated promising results, literatures show its performance still could be improved. Therefore, we present a comprehensive comparison study on the application of different BP algorithm in time series of deformation forecasting. Four types of typical improved BP algorithm, namely, momentum, conjugate gradient, Quasi-Newton and Levenberg-marquardt algorithms, are investigated. An illustrative example of high-rise building settlement deformation is adopted for demonstration. Results show that the improved BP algorithms can increase the prediction accuracy and have faster convergence speed.

Keywords: artificial neural networks, back propagation, deformation forecasting, learning rate, convergence, improved

Chunling Chen, Long Wang, Tongyu Xu, Jiawei Qi Relative humidity prediction of northern greenhouse environmental factors on the basis of a radial basis function neural network

Computer Modelling & New Technologies 2014 18(8) 254-258

With its advantages of abundant resource, popularity, and efficiency, solar greenhouse is the only type of greenhouse that is widely used in Northern China. This study proposes a simulation prediction model that is based on a radial basis function artificial neural network. This model is suitable for dealing with humidity in northern solar greenhouses. We select 600 groups of training data to establish the network model and to verify its accuracy. We then randomly select 80 groups for validation. With a 7.35% average error rate, the prediction model shows satisfactory performance. Thus, the results can be used to predict the relative humidity curve in a greenhouse, as well as provide a scientific basis for reasonable regulation and control of a greenhouse environment.

Keywords: solar greenhouse, relative humidity, predict model, radial basis function neural network

Zhang Jilin Research on stock analysis methods based on fluid mechanics

Computer Modelling & New Technologies 2014 18(8) 259-268

In this study, the leading indicator is considered as a new technical analysis method for evaluating stocks. This paper is written using the special characteristics and features of the principles of fluid motion, which can be applied to microscopic and macroscopic aspects of stock prices. Microscopic aspect: the stock that changes fastest is considered the leading indicator, which can be easily found by auto-filtering in Excel after working out the speed rate via program Matlab. The trend of other stocks will be predicted by analysing the leading indicator. Macroscopic aspect: determine the leading indicator, which are the top 10 stocks listed in the speed analysis results in the microscopic aspect. The trend of A-shares in a certain time period will be predicted according to the trend of the leading indicator. In order to

verify the viability of the leading indicator analysis method in microscopic and macroscopic aspects, we do the following research. We collect 4 days of stock price data to find out the leading indicator via the microscopic or macroscopic leading indicator method. When the corresponding leading indicator is found, then we predict the trend of other stocks and A-shares in a certain time period, so that the feasibility of the leading indicator methods can be proved.

Keywords: financial physics, leading indicator, stock analysis, Matlab

Jing Xu, Guanxin Yao Rural logistics service providers pricing and competition-cooperation research considering the 3PL accessibility

Computer Modelling & New Technologies 2014 18(8) 269-274

Rural logistics service providers pricing and competition-cooperation problems are researched in this article considering the 3PL accessibility in which the post logistics dominants. In the competition stage a model of master-slave Stackelberg game is to be established, and in the stage of cooperation, a cooperation model under different circumstances and the optimal pricing scheme without government intervention are discussed. This study could provide some economic phenomena existing explanations, as well as a reference and theoretical support for the decision-making act of Postal Logistics and other rural logistics-related subject.

Keywords: agricultural products logistics, post logistics, pricing, game theory, rural logistics

Yangcheng Hu Stakeholder orientation and financial performance of NPOs: development and testing of a mediating model

Computer Modelling & New Technologies 2014 18(8) 275-279

Literature examining how stakeholder-oriented strategy helps non-profit organizations (NPOs) cope with environmental uncertainty is limited. Thus, this study builds and tests a mediating model to investigate the relationships among stakeholder orientation, social performance and financial performance of NPOs. Data from 32 NPOs were collected in east China. Multiple regression analysis was employed to examine the relationships among the variables. The results revealed that stakeholder orientation has a positive and significant relationship with both social performance and financial performance. Meanwhile, it was discovered that social performance is significantly related to financial performance. Further, the results provided empirical evidence that the relationship between stakeholder orientation and financial performance was fully mediated by social performance. The limitations and future research directions were also discussed. This paper not only contributes to the theory of stakeholder orientation but also provides practical implications for NPOs on how to improve organizational performance.

Keywords: stakeholder orientation, NPOs, social performance, financial performance

Yong-shi Hu, Yue Yu, Ming-xing Xu Study on distribution and economic growth based on feder-model: evidence from Fujian province in China

Computer Modelling & New Technologies 18(8) 280-285

The effect of distribution expansion on economic growth is examined as one of the sources of growth. The research available mostly focused on the direct effect of distribution and ignored the externality effect of distribution. The paper tested the direct effect of distribution on economic growth in Fujian province, China with the indicators such as direct contribution and direct contribution rate, and an analytical framework is developed, incorporating the possibility that marginal factor productivities are not equal in the distribution and non-distribution sectors of the economy. Econometric analysis utilizing this framework in the period 1978-2011 indicates that the existence of externality effect generated by distribution sector, which is much larger than the direct effect, significantly affected the economic growth in Fujian province, China. However, the results indicate the marginal productivity of distribution sector is lower than non-distribution sector in Fujian province, which constrains the externality effect of distribution sector. We should take action to improve the level of modernization and informationization of distribution sector and turn distribution sector from labour-intensive to technology-intensive as soon as possible.

Keywords: distribution sector, direct effect, externality effect, feder-model

Yan Zhang Study on the land desertification early-warning system of Xinjiang in China

Computer Modelling & New Technologies 2014 18(8) 286-291

Land desertification is one of the greatest disasters in the world because of its wide-range influence, long duration, and enormous loss. The desertification early-warning system is very important to decrease the disaster occurrence and the

reduction of disaster loss. The grid accumulation desertification early-warning model can provide this. Therefore, Land Desertification Early-warning System of Xinjiang (LDES_XJ) was built in Browser/Server (B/S) structure based on the grid accumulation land desertification early-warning model. The functions include the publishing of land desertification warning information sub-system, the desertification data management, the map browsing and query have been realized in the LDES_XJ. The system is a fundamental platform for the land desertification prevention and control as well as a convenient communication platform for the government, which makes it more convenient for relevant people to obtain the desertification early-warning information and also plays an important role in the desertification prevention and control in Xinjiang, China.

Keywords: desertification prevention, early-warning model, management information system, B/S structure

Chao Yu, Zongshan Pu, Lei Chen Study on the location choice of Chinese outward foreign direct investment

Computer Modelling & New Technologies 2014 18(8) 292-297

With the steady implementation of the “Go Out” strategy, China has been developing rapidly in the outward foreign direct investment. This paper, based on the 2003-2012 panel data of Chinese outward foreign direct investment, applies systematic GMM model and variable intercept model to study the location choice of Chinese outward foreign direct investment. Research findings show that, the resource endowment and bilateral trade of the host country have significantly positive impacts on Chinese outward foreign direct investment, whereas the market size and technical merit of the host country bring about significant negative impacts. Therefore, domestic corporate outward foreign direct investment should still focus mainly on resource acquisition and trade complementarity rather than acquiring technologies.

Keywords: outward foreign direct investment, location choice, technical merit, GMM model

Zhang Lei The collaborative optimization of uncertain supply chain network under multi-generation co-existence

Computer Modelling & New Technologies 2014 18(8) 298-305

This paper studies the collaborative optimization of multi-generation co-existence supply chain network of single manufacturer, multi-distribution centre and market. Firstly, we considered the effects of prices, time and substitutability on the market demand of both new and old products and the whole production and marketing decision, and then subdivide the products demand of each generation. Secondly, considering that as the lifecycle stage of each generation differs, there exist differences in demand characteristics, which make the structure of supply chain network obviously different, we built WCVaR risk optimization model of production-distribution network under co-existence, which subjects to scattering distribution constraints and solve the model by Lingo11.0 and then made simulation analysis. Lastly, we verified the validity of the model with the optimization results of numerical simulation.

Keywords: multi-generation co-existence, supply chain network, collaborative optimization, WCVaR

Hao-ran Shi, Xiaoqing Liu, Yao Yang On dynamic iterative algorithm and the loss of newsvendor problem

Computer Modelling & New Technologies 2014 18(8) 306-311

Based on the thinking and method of dynamic programming, this paper calculates the expected profit of every selling cycle of newsvendor by historical data and calculates the expected profit of the second selling cycle by that of the first selling cycle. We will get the optimal purchasing quantity as the expected profit begins to fall. Besides, this paper also discusses the stock loss and gets the optimal purchasing quantity by empirical examples.

Keywords: newsvendor problem, marginal income, dynamic iterative algorithm, optimal purchasing quantity

Mei Yang A market segmentation model of enterprise marketing based on an improved grey correlation analysis methods

Computer Modelling & New Technologies 2014 18(8) 312-315

In order to make out quick and effective enterprise sales strategy, this paper studies market segmentation for key brand products and proposes a market segmentation model based on an improved grey correlation analysis methods. Key indicators about market segmentation are analysed first. These indicators are then subject to standardization to be in the unified measurement. On this basis, classical grey correlation analysis methods get impTabled and Hamming distance is adopted to cope with fuzzy information. Grey correlation coefficient and grey correlation degree of market segmentation are acquired. AHP is introduced to assign weight to indicators of market segmentation in order to get

weighed comprehensive grey correlation degree. Finally, test is given to the effectiveness, scientific nature and feasibility of the model and algorithm through case study of a brand product.

Keywords: market segmentation, marketing strategy, grey correlation analysis, correlation degree, model

Tian Jia Study on enterprise extension marketing model based on extension engineering methods

Computer Modelling & New Technologies 2014 18(8) 316-319

This paper studies the marketing strategy of enterprises and proposes an enterprise extension marketing model based on extension engineering methods under market economy. The matter-element model for enterprise marketing is established on extension theory, and the matter-element features, values and discourse domain are studied to produce more marketing strategies. These strategies are subject to extension priority-degree evaluation for convergence analysis. Compatibility analysis is also carried out. Empirical studies prove the model to be feasible and effective.

Keywords: extension marketing, extension engineering, marketing strategy, market, model

Ziya Wang, Ran Li An improved Grey prediction model and the application in college sports information management system

Computer Modelling & New Technologies 2014 18(8) 320-326

The sports elective course is an important means for college students to exercise their physical education and quality education. Because the students can choose the elective courses voluntarily online, teachers and colleges cannot learn the number of the students who choose the sports elective courses before the deadline. In this paper, we propose an improved GM (1, 1) prediction model to forecast the number of the students who take part in the sports elective courses and apply this model in the college sports information management system. Firstly, this paper puts forward the Grey prediction model with time parameter. Then, this paper studies the constructing mechanism and the modelling characteristics of this model. At last, we use the improved GM (1, 1) model to forecast the number of students who select the college sports elective courses in college sports information management system. High precision of the fitting and forecasting are obtained in the experiment while the result verifies the validity of the model.

Keywords: Grey precision model, college sports information management system, sports elective course

Ran Li An improved method of controlling bullwhip effect and the analysis of the bullwhip effect

Computer Modelling & New Technologies 2014 18(8) 327-332

The bullwhip effect is an important parameter to measure whether the logistics management is good or not. The bullwhip effect affects the production, inventory, transport efficiency in logistics management seriously. In this paper, we establish the structural model of supply chain with multi distribution centre and apply control method to inhibition the bullwhip effect. We analyse the control mechanism of the bullwhip effect and present the control arithmetic to control the bullwhip effect. All the processes are under the circumstance that the demand is worst according to the control theory. At last we processed a stochastic control simulation experiment to control the bullwhip effect. The result shows that the bullwhip effect is inhibited and the bullwhip effect is reduced and stable. The first part of this paper is the related problem description. The second part is basic model and quantitative description of bullwhip effect. The third is control method. The last part is a simulation example.

Keywords: bullwhip effect, multi distribution centre, logistics management

NATURE PHENOMENA AND INNOVATIVE ENGINEERING

Jian Chu, Guoyu Wang, Shan Xu An extension evaluation model of the operation state of aero engine

Computer Modelling & New Technologies 2014 18(8) 333-338

Traditional fault diagnosis with single parameter fails to evaluate the operation state of aero engine. This paper analyzes the extension evaluation based on supervision information of engine's performance and works out an evaluation system as well as an extension evaluation model and algorithm based on incidence function in extension theory. Through the model, parameters under the operation state of aero engine are studied. The incidence function in extension theory between classic domain and section domain of all attributes and parameters is established in the corresponding evaluation system. This makes it possible to acquire the state level of aero engine according to incidence in extension theory.

Keywords: aero engine, operation state, incidence function, evaluation, model

Hualong Xie, Nan He, Fei Li The bionic design, virtual prototype modelling and motion simulation of biped

robot with heterogeneous legs*Computer Modelling & New Technologies 2014 18(8) 339-346*

The research of humanoid robot and intelligent prosthesis is integrated and a new-style humanoid robot named biped robot with heterogeneous legs (BRHL) is proposed to provide an ideal test-bed for intelligent bionic leg (IBL). The research background and concept of BRHL are introduced. The existed problems of common humanoid robot leg are analysed in detail. Based on bionics, the joint structure and driving scheme of artificial leg with pneumatic muscle actuator and bionic leg with magneto-rheological damper are designed. Using Pro/E, the virtual prototypes of artificial leg, bionic leg and BRHL are established. The bionic characteristics of artificial leg and bionic leg are analysed and the motion simulation of BRHL using ADAMS is done. The simulation indicates that BRHL can simulate amputees with IBL well and is an ideal test-bed for IBL.

Keywords: biped robot with heterogeneous legs, intelligent bionic leg, artificial leg, pneumatic muscle actuator, motion simulation

Mei Liu, Songling Wang, Zhengren Wu Influence of liquid physical properties on liquid film flow characteristics of uneven wall*Computer Modelling & New Technologies 2014 18(8) 347-352*

Two-dimensional model of inclined uneven wall was established based on VOF method to numerically simulate the flow characteristics of liquid film. The impact of physical property on the flow field is studied. Water, acetone and ethyl alcohol were selected as the medium. The results indicate that liquid film thickness increases with liquid viscosity, while the phase difference between free surface and the uneven wall has no change. The continuous uniform film is easy to form while considering the surface tension. Furthermore, the phase difference and liquid film thickness both increase when taking the surface tension into account.

Keywords: uneven wall, liquid film, flow characteristics, physical properties, VOF method

Danling Wang, Yanfei Wang Multifractal analysis on gene and PPI networks*Computer Modelling & New Technologies 2014 18(8) 353-357*

Multifractal analysis is a useful way to systematically describe the spatial heterogeneity of both theoretical and experimental fractal patterns. In this paper, we introduce a new box-covering algorithm to compute the generalized fractal dimensions of complex networks. We apply our method on networks built on disease-related gene microarray data and PPI networks. For each microarray data, we compare the difference of multifractal behaviour between gene networks that reconstructed from patients and normal microrarrays. The result suggests that multifractality exists in all the gene networks we generated and the differences in the shape of the D_q curves are obvious for all microarray data sets. Meanwhile, multifractal analysis could provide a potentially useful tool for gene clustering and identification between healthy people and patients. For the analysis of PPI networks, the results support that the algorithm is a suitable and effective tool to perform multifractal analysis of complex networks, and this method can be a useful tool to cluster and classify real PPI networks of organisms.

Keywords: multifractal analysis, self-similarity, gene networks, PPI networks

Yong Yang, Weiwei Zhu Calculation of microstress in machining distortion of titanium alloy monolithic component based on x-ray diffraction experiment*Computer Modelling & New Technologies 2014 18(8) 358-362*

Machining distortion of titanium alloy monolithic component is closely related to the internal stress of material. In this paper, the various factors causing width effects of X-ray diffraction line were analysed, and the effect of stacking fault on diffraction spectrum was excluded according to the result of TEM experiment. The true diffraction spectrum, which can reflect inside information of Ti6Al4V titanium alloy, was determined using MDI JADE peak shape analysis method. Further, the calculation model of microstress was constructed. Combining with the result of X-ray diffraction experiment, microstress in machining distortion of titanium alloy monolithic component was calculated. This work establishes the foundation for investigating the mechanism of machining distortion of titanium alloy monolithic component.

Keywords: titanium alloy, machining distortion, microstress, diffraction line width effect

Xianguang Kong, Yihui Li, Lei Yin, Xiaowen Wang Research and implement of CATIA parametric modelling-based cutter information integration in VERICUT

Computer Modelling & New Technologies 2014 18(8) 363-367

In order to achieve the integration of the VERICUT tool library and tool management software, this paper studies the VERICUT tool integration technology based on CATIA parametric modelling. As the tool modelling ability is limited, VERICUT can neither achieve complex tool modelling nor generate the tool model by parameterization. Therefore, the secondary development based on VERICUT is unable to meet the integration requirements of tool information. This paper proposes a new integration approach, which transforms the integration of VERICUT and tool management software into VERICUT and CATIA tool information. Meanwhile, a feasible integrated development process is put forward. First, the tool model is parametrically driven by CATIA. Then, the integration of the CATIA tool library and tool management software is realized based on CAA. Finally, the transformation of CATIA and VERICUT tool model is accomplished, which indirectly realizes the integration of the VERICUT tool library and tool management software. During the NC machining simulation, VERICUT inherits the tool information and model generated by CATIA NC programming, which can ensure consistency with the tool information in tool management software. This paper solves the problem of integration of tool management software and VERICUT in the context of a digital manufacturing project; successful application of the proposed approach has greatly improved the efficiency of NC programming.

Keywords: VERICUT, CATIA, tool management software, parametric modelling, TLS, CAA

Jian-Zhang Wu, Yu Xiao, Wei Gao Bilinear model for ontology mapping*Computer Modelling & New Technologies 2014 18(8) 368-372*

As a model of concept representation, ontology has widely applied to various disciplines. Ontology mapping is used to create the link between different ontologies. In this paper, we present a new ontology mapping algorithm by virtue of bilinear model. The linear mapping pair is given by the iterative procedure. Two strategies are manifested to obtain the finally ontology mapping. The simulation experimental results show that the proposed new technologies have high accuracy and efficiency on ontology mapping in certain applications.

Keywords: ontology, ontology mapping, linear mapping, bilinear model, dimensionality reduction

Gewei Tan, Wei Lin The two-step motion compensation combined squint wavenumber domain algorithm based on fractional Fourier transform*Computer Modelling & New Technologies 2014 18(8) 373-378*

Fractional Fourier transform (FrFT) is a kind of generalized Fourier transform, which processes signals in the unified time-frequency domain and the linear frequency modulation signal can be well focused after FrFT. Motion error is an important factor affecting the SAR resolution, conventional wavenumber domain (CWD) algorithm is an ideal solution of SAR focusing problem as long as nominal straight flight track is given, especially in the case of high squint angles and long synthetic apertures, but it has certain limitation in processing airborne SAR data affected by motion error, so extended wavenumber domain algorithm (EWD) is presented. Pointing to the problem that the effect of error elimination is not obvious in processing non-stationary motion error using EWD algorithm, FrFT based the two-step motion compensation combined squint wavenumber domain algorithm is put forward in this paper, which is expected to eliminate the influence of motion error more effectively, so as to obtain high quality SAR images. The simulation results and the imaging results of real SAR data show that the proposed algorithm can eliminate the influence of motion error effectively. (the real SAR data provided by Institute of Electronics, Chinese Academy of Sciences).

Keywords: fractional Fourier transform, motion compensation, wavenumber domain algorithm, extended wavenumber domain algorithm, high resolution

Chao Wang Research on grid replacement technology and two-dimensional isothermal simulation on melt flow of blown film*Computer Modelling & New Technologies 2014 18(8) 379-383*

Established a melt flow model of two-dimensional isothermal simulation for spiral mandrel blow-film die, analysed die swell-effect at the die exit of blown film extrusion process by grid replacement technology, obtained the pressure distribution in the internal flow of blow-film die, and analysed the shape change of free surface at die extrusion section with different viscosity materials. The results showed that the resistance force of the melt flow under internal flow of blow-film die mainly from the spiral direction, with depth of spiral groove gradually shoaled, the direction of resistance force changed from spiral direction to axial direction. When high viscoelastic material entered blow-film die being extruded, the shape of free surface at die extrusion section would change significantly, meanwhile, the gradient of radial velocity of melt flow in the internal die became gradually reduced.

Keywords: spiral mandrel blow-film die, grid replacement, die swell-effect, free surface

Liyang Liu, Chao Li, Qi Zhang Analysis of cylindrical cam molded surface

Computer Modelling & New Technologies 2014 18(8) 384-386

The mold surface of cylindrical cam is very complex. There are some questions in designing and using by unfolding picture of cylindrical cam. The thesis infers and builds the mathematical models of cylindrical cam through the theoretical analysis from the nature of space motion cylindrical cam. The models provide the theoretical reference of right projecting, manufacturing and examining.

Keywords: axial cam, spiral surface, equation

Lifang Ma, Yaxin Wang, Yang Liu, Shizhong Zhang, Yu Chen Comparative study on prosthetic socket materials

Computer Modelling & New Technologies 2014 18(8) 387-392

Prosthetic socket materials must exhibit a good processing performance so that a variety of desirable shapes can easily be formed, thereby enabling more controllable and adjustable socket production process and socket compatibility with the human anatomy and movement mechanics. Additionally, they need to feature high strength and light weight, and must be comfortable for the patient to use. In this study, the properties and application potential of current prosthetic socket materials, such as thermoplastic sheets, low-temperature thermoplastic sheets, silicone-based materials, and resin-based composite materials, were compared. Additionally, the matrix used in resin-based composite materials was investigated by infrared spectroscopy.

Keywords: prosthesis, socket, thermoplastic sheet, resin-based composite material

Ling Wang, Peng Guo, Lin Liu Extraction of soil salinization information from the Manas river basin based on TM Images

Computer Modelling & New Technologies 2014 18(8) 393-397

Remote sensing technology is widely used in real-time observations. In this study, therefore, salinization information was extracted from Thematic Mapper (TM) images. In particular, remote sensing information regarding saline soil was obtained to analyse its dynamic changes. This soil collected from the Manas River Basin in Xinjiang Province China was selected as the research area. Data from Landsat TM remote sensing images with seven bands were obtained in August 2010 as inputs, and salinization information was extracted using the e-Cognition system. As per this information, high-salinity soil is mainly distributed outside the oasis. The data were analysed further through the normalized differential vegetation, normalized difference water, and remote sensing image indices. Analysis results show that overall classification accuracy can reach 83.7%, thus demonstrating that the automatic extraction of information regarding saline soil is highly accurate. Furthermore, this information can be automatically and precisely extracted using an object-oriented method.

Keywords: Manas river, river basin, soil salinization, NDVI

Jianping Li, Tongxiao Shang, Yixiao Guan Volume integral equation-based electromagnetic inversion of 3D complex resistivity bodies

Computer Modelling & New Technologies 2014 18(8) 398-403

This study proposed the volume integral equation method to invert the complex resistivity parameters of 3D bodies in homogeneous half space. The partial derivative matrix of 3D bodies to complex resistivity was obtained with the use of electromagnetic field partial derivatives to the real resistivity combined with the complex resistivity spectrum partial derivatives of the Cole–Cole model. The classical damped least squares inversion method was applied. Results show that with at least four frequency data points, true complex resistivity parameters, as well as geometric parameters, can be inverted. The four parameters in the Cole–Cole model are interdependent. Chargeability and time constant are highly correlated, such that either one of these factors should be fixed prior to inversion. Furthermore, the proposed inversion algorithm we proposed has high efficiency because it only works in a split unit.

Keywords: volume integral equation, complex resistivity, 3D EM, damped least square method