



COMPUTER MODELLING
AND
NEW TECHNOLOGIES

2015
VOLUME 19 NO 6

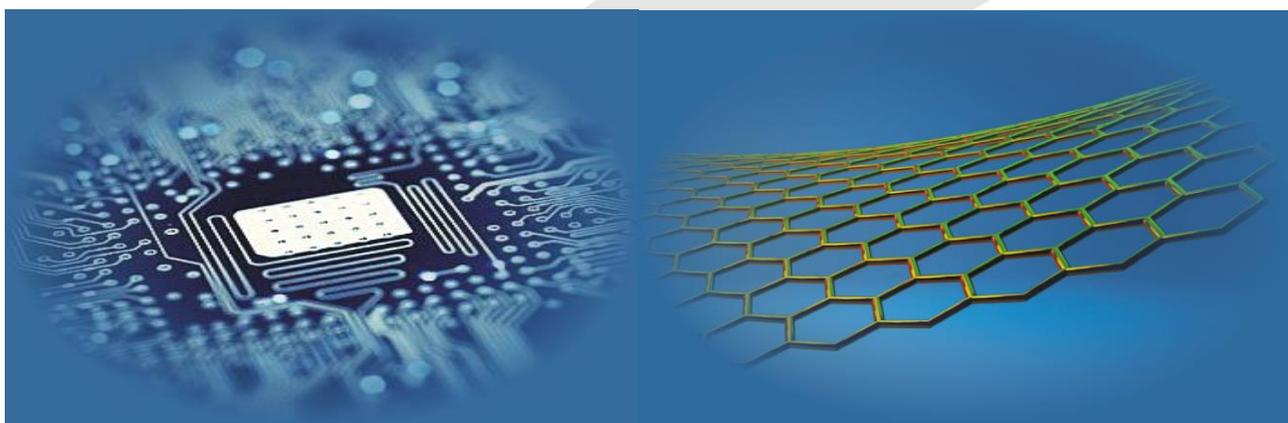
ISSN 1407-5806 ISSN 1407-5814 on-line

Latvian Transport Development and Education Association

Computer Modelling and New Technologies

2015 Volume 19 No 6

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



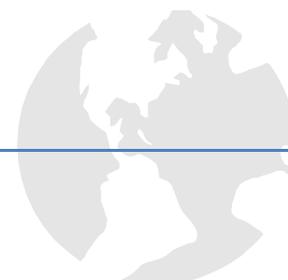
Riga – 2015

EDITORIAL BOARD

Prof. Igor Kabashkin	Chairman of the Board , <i>Transport & Telecommunication Institute, Latvia</i>
Prof. Yuri Shunin	Editor-in-Chief , <i>ISMA University, 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>ISMA University, 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>ISMA University, 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, 2015, Vol. 19, No.6 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 Compendex



Editors' Remarks

Lost Time

by Rabindranath Tagore

On many an idle day have I grieved over
lost time. But it is never lost, my lord.
Thou hast taken every moment of my life in
thine own hands.

I was tired and sleeping on my idle bed
and imagined all work had ceased.

In the morning I woke up and found my
garden full with wonders of flowers.

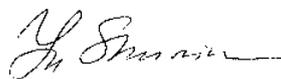
Hidden in the heart of things thou art
nourishing seeds into sprouts, buds into
blossoms, and ripening flowers into
fruitfulness.

Rabindranath Tagore (1861-1941)*

This 19th volume No.6 includes two review papers and several research papers on **Computer and Information Technologies, Operation Research and Decision Making and Nature Phenomena and Innovative Engineering.**

Our journal policy is directed to fundamental and applied scientific researches, innovative technologies and industry, which is the fundamentals of the full-scale multi-disciplinary modelling and simulation. This edition is the continuation of our publishing activities. We hope our journal will be of interest for research community and professionals. We are open for collaboration both in the research field and publishing. We hope that the journal's contributors will consider collaboration with the Editorial Board as useful and constructive.

EDITORS



Yuri Shunin



Igor Kabashkin

* **Rabindranath Tagore (7 May 1861 – 7 August 1941)**, was a Bengali poet, novelist, musician, painter and playwright who reshaped Bengali literature and music. As author of Gitanjali with its "profoundly sensitive, fresh and beautiful verse", he was the first non-European and the only Indian to be awarded the Nobel Prize for Literature in 1913. His poetry in translation was viewed as spiritual, and this together with his mesmerizing persona gave him a prophet-like aura in the west. His "elegant prose and magical poetry" still remain largely unknown outside the confines of Bengal.



Content

REVIEWS		
D Fink, A Kiv, Y Shunin, N Mykytenko, T Lobanova-Shunina, A Mansharipova, T Koycheva, R Muhamediev, V Gopeyenko, N Burlutskaya, Y Zhukovskii, S Bellucci	The nature of oscillations of ion currents in the ion track electronics	7
R I Muhamedyev	Machine learning methods: An overview	14
MATHEMATICAL AND COMPUTER MODELLING		
A Mrochko	Modelling of the subsystem of estimation of navigational parameters in automatic vehicle control systems	30
I Ter-Saakova	Optimisation of coordination's selection by innovation and investment projects	35
Jinge Luo, Xiaofei Wang, Junhui Luo	Earth as a building material for reproduction of ancient buildings in China Meishan cultural park	40
Authors' Index		45
Cumulative Index		46



The nature of oscillations of ion currents in the ion track electronics

**D Fink¹, A Kiv^{2*}, Y Shunin³, N Mykytenko⁴, T Lobanova-Shunina⁶,
A Mansharipova⁵, T Koycheva⁴, R Muhamediev⁵, V Gopeyenko⁷,
N Burlutskaya⁷, Y Zhukovskii³, S Bellucci⁸**

¹Departamento de Física, Universidad Autónoma Metropolitana-Iztapalapa, PO Box 55-534, 09340 México, D.F., México

²Ben-Gurion University, PO Box 653, Beer-Sheva 84105, Israel

³Institute of Solid State Physics, Latvian University, 8 Kengaraga Str., LV-1063 Riga, Latvia

⁴South-Ukrainian National Pedagogical University, 26020 Odessa, Ukraine

⁵Almaty University, Kazakhstan

⁶Riga Technical University, Faculty of Mechanical Engineering, Transport and Aeronautics, Latvia

⁷ISMA University, 1 Lomonosova Str., Bld 6, LV-1019, Riga, Latvia

⁸INFN-Laboratori Nazionali di Frascati, Via Enrico Fermi 40, I-00044, Frascati-Rome, Italy

*Corresponding author's e-mail: kiv@bgu.ac.il

Abstract

The paper contains the description of the main features of ion current pulsations in track devices. The conditions under which the pulsations arise are discussed. We describe different approaches that are used for interpretation of the effect of ion current pulsations. In particular the generalized model of current spikes in track devices is considered. To create this model a special modification of the classical molecular dynamics was developed. The results of application of this model coincide with the main experimental data concerning the ion current pulsations in track devices.

Keywords: track devices ion current pulsations nanoporous membranes

1 Introduction

2 Collective Interaction in Ion Track Electronics

3 Oscillating currents through nanoporous membranes in electrolytes

4 Generalized model of current spikes in track devices

4.1 MODIFICATION OF CLASSICAL MOLECULAR DYNAMICS

4.2 DESCRIPTION OF THE MODEL AND RESULTS OF SIMULATION

5 Conclusion

References

1 Introduction

Since the sixties of the past century it is known that energetic (with tens of MeV or more) heavy (with atomic masses being usually larger than that of Ar) ion irradiation ("swift heavy ions", SHI) introduces very narrow (~ some nm) but long (typically 10-100 μm) parallel trails of damage in irradiated polymer foils, the so-called latent ion tracks. The damage shows up primarily by the formation of radiochemical reaction products. Whereas the smaller ones readily escape from the irradiated zone thus leaving behind them nanoscopic voids, the larger ones tend to aggregate towards carbonaceous clusters. Thus emerging structural disorder along the tracks modifies their electronic behavior.

The newly created intrinsic free volume enables electrolytes to penetrate into the polymer, thus forming parallel liquid nanowires. In case that the tracks penetrate through all the foil the conducting connections emerge between the foil front and back sides.

Upon proper design, the irradiated polymer foils may

exhibit electronic properties that mimic bioelectronic functionalities, as they resemble somewhat biological membranes which also contain a number of parallel electrolyte-filled nanopores.

The ion track technology is, in particular, directed towards biosensing applications. In this case the ion tracks are functionalized directly by attaching organic or bioactive compounds (such as enzymes) to their walls. The recent advances in this field allow monitoring and tracking biomolecules in areas such as environment, food quality and health. The presently developed ion track-based nanosensors provide high sensitivity, low power and low cost [1].

The creation of new biosensors and their further improvement requires a careful study of the mechanisms of electrolytes passage through the tracks.

2 Collective interaction in ion track electronics

In [2] the capability of a multitude of parallel electroactive nanostructures in a given substrate to show collective interactions has been considered. Two types of electroactive nanostructures are described: (a) electrolyte-filled current

spike-emitting latent ion tracks in thin polymer foils, (b) metal cluster-filled etched ion tracks in TEMPOS structures (i.e. in thin SiO₂ layers on Si substrates). Both these electroactive nanostructures can be operated either by applying a constant voltage to them, or by application of a sinusoidal voltage at low frequency, and in both cases current spike emission can be triggered. An electroactive nanostructure can influence the performance of neighboring nanostructures by modifying the entrance or exit potentials (or both) of the latter one, via lateral charge exchange through the common front or backside conductors or contacts. For these two cases, this leads to two different effects: The collective interaction of many current spike-emitting latent tracks in electrolytic ambient leads to pulse-locked synchronization similarly to its representation in Neural Network theory.

In TEMPOS structures with etched tracks in SiO₂ on Si and with metal nanoclusters coverage of the oxide layer and the track walls (with at least two contacts on the oxide surface and one on the Si substrate), the collective track interaction can induce negative differential resistance [3]. This is the consequence of a chain reaction triggered by spontaneous opening of previously closed (or closing of previously open) neighbored tracks. Periodic repetition of such opening/closing processes leads to self-pulsating devices.

3 Oscillating currents through nanoporous membranes in electrolytes

The effect of current pulsations when ions pass through the tracks is of great interest for creation of new biosensing devices. [4-7]. There are different attempts to explain the phenomenon of current oscillations in track-containing foils embedded in electrolytes.

To describe the properties of nanopores (in particular the ion transport in tracks) different models and mathematical methods are used. For example, the current through nanotracks is described by stationary Poisson Nernst Planck equations [8, 9]. Molecular dynamics simulation is used in [10, 12] to describe the ion current rectification.

One of the models connects these oscillations with carbonaceous clusters that form along the latent tracks [7]. These clusters might behave as obstacles for the smooth ionic current passage along them, upon application of a DC or low frequency AC voltage across the track-containing polymer foil. As a result, charges may pile up in front of them until the intrinsic electric field across them exceeds the breakthrough field strength. At that moment current spikes eventually associated with negative differential resistances emerge. As the spike height is decreased by eventual surface adsorption layers, pulsating tracks can also be exploited for biosensing [13]. Foils with current spike emitting tracks are thought to mimic neurons. In a multitude of such tracks, the individual randomly emitted spikes synchronize themselves towards phaselocked oscillations [14, 15], similarly as they occur for neurons in the human brain, where their interaction results in the formation of brain waves. The frequency of these collective track pulsations is around 0.1...30 Hz [2]. Hence in order of magnitude which is similar to brain waves. The presently available neural network theory describes the behavior of pulsating tracks at least qualitatively well.

In [16] it was suggested that the current oscillation mechanism is linked to the competition of two processes in

the tracks: a) The adsorption of charged ions at the internal track walls and b) The ion spike when the number of accumulated ions reaches some critical value. The negative charge of the pore walls [17] causes the adsorption of positive ions. Also chemical reactions may influence this adsorption. During accumulation of the adsorbed ions at the inner surfaces of the tracks the repulsive forces begin to prevent the penetration of new positive ions into the tracks, hence the ion current decreases. At some threshold voltage applied to the track and for some critical number of adsorbed ions an ion spike emerges which leads to desorption of the ions accumulated inside the track, hence to an increase of the ion current. The current maximum corresponds to the open track with a minimum of adsorbed ions. Thereafter newly adsorbed ions can be accumulated so that the process repeats. Hence an oscillation current emerges with a frequency $\nu = 1/\tau_1$, (τ_1 being the period of oscillations) that is determined by the rate of accumulation of adsorbed ions and by the probability that an ion spike occurs, the latter depending on the applied voltage. In order that ions can penetrate at all through the track, the track radius must exceed a certain threshold value r_{min} , which is determined by the thickness of the ion adsorption layer. On other hand, for too large track radii r_{max} the adsorption layer does not control the ion penetration through the track. Thus, current oscillations will occur only for track radii r with $r_{min} < r < r_{max}$.

The above model does not take into account the possible interaction of nanopores.

Hence this refers to the case of one track per membrane only. In the case of many tracks per membrane the statistical interaction between them modifies the conditions that determine the oscillation frequencies. Here, the ion current minimum corresponds to the situation when most tracks are closed simultaneously. Such a situation is realized as a result of statistical interaction processes and demands some time (τ_2). This interaction may be ionic charge equilibration processes between neighboring tracks via currents running from one charge carrier cloud near a track entrance and/or exit to the other, or via the (more improbable) ionic diffusion across the tracks, and by different adsorption conditions in different tracks. A realization of the situation when almost all tracks are open demands also a given time (τ_3).

It is clear that we have $\tau_1 \ll \tau_2 + \tau_3$; in the case of many tracks the frequency of current oscillations is much less than in the case of one individual track. In the case of many tracks per membrane the current oscillations do not drop to zero as in the case of individual tracks because in the first case some tracks are open always.

4 Generalized model of current spikes in track devices

For ion current pulsations in track-containing foils the following features are established [18, 19]:

- Spike emission depends on the amplitude and frequency of the applied voltage;
- The high ion track density (higher than some threshold density) is necessary to obtain the effect of current spikes;
- Current spikes preferentially occur at pronounced, rather equidistant applied voltages;
- Maximal spike heights do not seem to be affected markedly by the frequency of the applied voltage;

- Current spike spectra are not always reproducible though their principle features remain the same;
- Current spike emission appears to vanish rapidly with the frequency of applied voltage. Its decrease indicates the existence of a threshold frequency for spike emission.

The model described in [20] allows studying the general case of the ion current pulsations in the track-containing polymer foils embedded in electrolyte. A schematic representation of the appropriate structure can be seen in Fig. 1. To construct the model the classical method of Molecular Dynamics (MD) was modified so that a new MD approach allows describing subthreshold radiation effects [21].

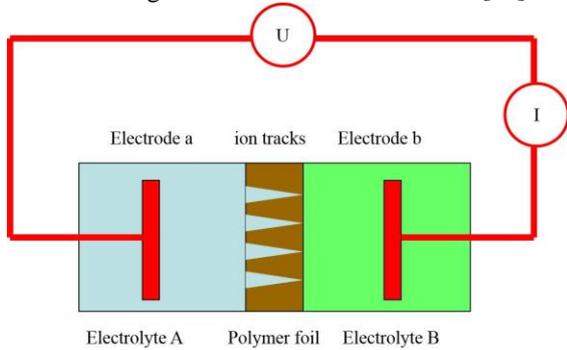


FIGURE 1 Principle arrangement of experimental setup to study current spike emission in ion track-containing foils embedded in electrolytes (current/voltage measurements)

4.1 MODIFICATION OF CLASSICAL MOLECULAR DYNAMICS

To create the model of current spikes in track devices a modification of classical MD was performed in [22-24]. In the MD method the classical equations of motion with an appropriate potential of the interaction between particles are solved. The Verlet algorithm [25] is usually used to solve these equations.

To simulate the effect of atomic collisions in the modified MD method the “shock function” (FSH) is introduced. Then in the Verlet algorithm the total force acting on atom *i* is presented by expression:

$$\vec{F}_i^n = -\sum \frac{\partial \Phi(\vec{r}_{ij})}{\partial x_i} + \vec{F}_{SH}, \quad (1)$$

where $\Phi(\vec{r}_{ij})$ is the interatomic potential and $r_{ij} = |\vec{r}_i - \vec{r}_j|$ is the distance between atoms *i* and *j*.

The pulses that are transferred to lattice atoms during irradiation are characterized by special random function (RF). This function shows which atom in the irradiated sample is knocked, what energy value is chosen from the selected energy interval, and what the direction of hit is. RF inserted to the computer program performs three tasks:

- Selects an atom in the target lattice which gets a hit;
- Selects an energy value from the interval ($\varepsilon_1, \varepsilon_2$);
- Selects an orientation of the pulse transferred to the target atom.

A linear congruent generator for RF is an algorithm that yields a sequence of pseudo-randomized numbers calculated with a discontinuous piecewise linear equation. This generator presents one of the best-known pseudorandom number generator algorithms that are easily implemented and fast, especially on computer hardware which can

provide modulo arithmetic by storage-bit truncation. These generators based on linear congruent method are especially useful for non-cryptographic applications, such as modeling. They are effective and most used in empirical tests and show good statistical characteristics.

The generator is defined by the recurrence relation:

$$X_{n+1} = (aX_n + c) \pmod{m}, \quad (2)$$

where X_0 is an initial value. In our model the parameters are: $m = 232$, $a = 214013$, $c = 2531011$.

In order to determine the kinetic energy transferred to the lattice atom the scaling of the forces FSH should be implemented. As a reference point we used the energy that is necessary for irreversible displacement of lattice atom to interstitial position in elastic collision (Ed). Then the following relation may be written:

$$\frac{(F_{SH}t)^2}{2M} = E_d, \quad (3)$$

where *t* is duration of the action of the force in the process of one hit.

In computer experiment the value of the force FSH was gradually increased to the point where the atoms begin to move irreversibly to interstitial positions. This magnitude of the force FSH corresponds to $E_d \approx 25 - 30$ eV.

To apply the proposed approach a model crystal with a cubic lattice that consists of 8000 atoms is constructed. Lenard-Jones potential [26] was used. These parameters were slightly varied to stabilize the lattice of the model crystal.

The proposed approach allows studying so called subthreshold radiation effects in solids [21]. Such mechanisms are used in the model.

4.2 DESCRIPTION OF THE MODEL AND RESULTS OF SIMULATION

In the model the track structure is represented as a two-dimensional system of currents. Really, in any plane that intersects the tracks one can fix some current value in each track (see Fig. 2). Therefore in the indicated plane we have a set of currents that perform oscillations in the appropriate conditions. The model describes a system of pulsating currents regardless of the material, in which the tracks are created, and does not include the description of real material and tracks themselves. Current pulsations are simulated by oscillations of model particles. Their masses are varied in a large range of values (five orders) in order to get a wide variety of oscillation spectra and compare their qualitative view with views of experimental pulsations of currents. To simulate the real experimental conditions the value of SF and its frequency were varied.

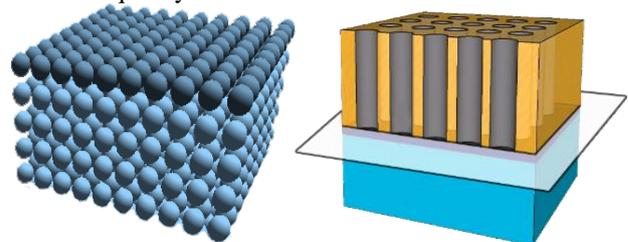


FIGURE 2 The model crystal (on the left) and the model for pulsating ion currents (on the right)

In the simulation experiment it was assumed that the average amplitude of the model particles oscillations correspond to the average spike heights, the value of SF – to the value of applied voltage, and the frequency of SF action – to the frequency of applied voltage. The values of SF were chosen so that model particles (MP) did not leave irreversibly their nodes under the action of SF. The directions of SF action

were determined by the random function and change within the upper hemisphere which corresponds to the directions of motion of the ions in the track. As a result of action of SF, regular small oscillations of MP as well as large amplitudes arise as shown in Figure 3. This figure demonstrates the similarity of oscillation spectra in real experiment and as a result of computer modelling.

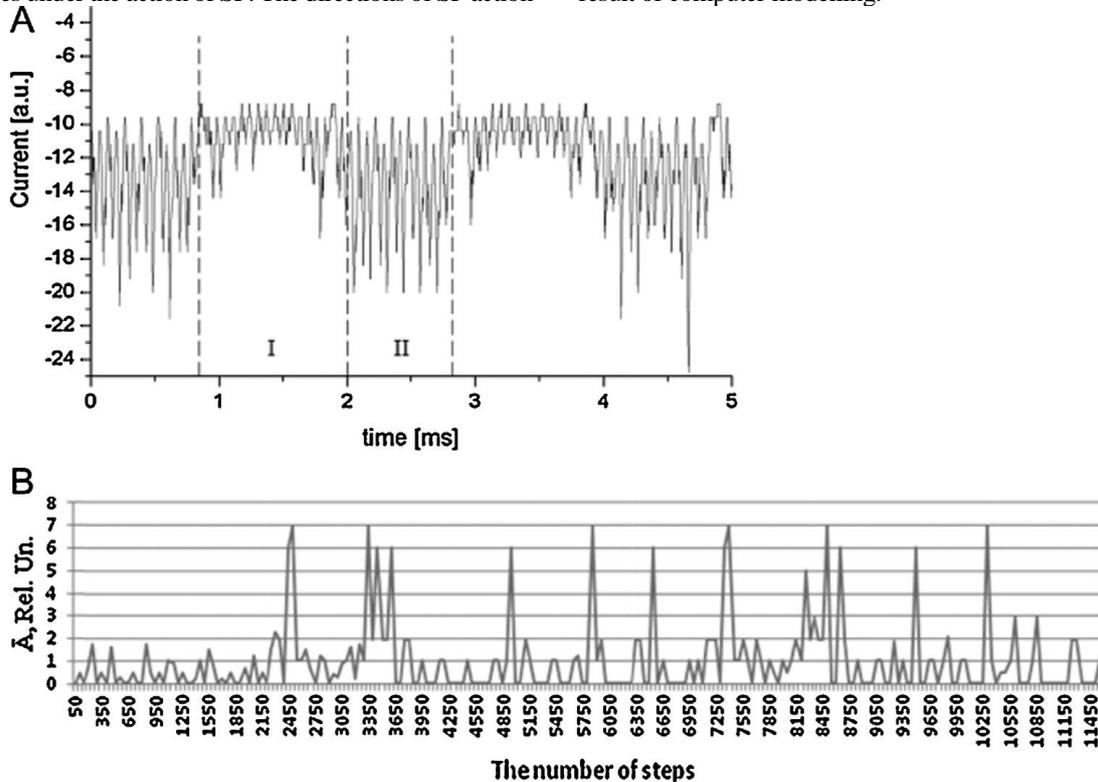


FIGURE 3 Current spikes in conditions of real experiment (on top). Region I corresponds to regular oscillations, region II corresponds to current spikes. Illustration of model spikes in the model experiment (down). At the horizontal axis is the average amplitude of MP oscillations.

The dependence of the average amplitude A of MP oscillations on the value of SF is displayed in Figure 4 (on top). This dependence is in good agreement with the dependence of the average spike height on the value of applied voltage shown in Figure 4 (down). In Figure 5 (on top) the dependence of the average amplitude of MP oscillations on the frequency of SF action is shown. The corresponding dependence of the average spike height on the frequency of applied voltage is presented in Figure 5 (down).

The dependence in Figure 5 can be explained by a “memory effect” which manifests itself in the mechanism of spikes formation: each next spike “remembers” the information about previous spikes. As a result of this effect according to the model there is some optimum frequency of the external exciting factor that provides a maximum average spike height. It means that at higher frequencies of the SF action the model lattice is too disordered after the previous spikes, and the conditions for the synchronization of individual spikes are less favorable. On the other hand, at too low frequencies the model lattice is completely restored after the previous spikes (“forgetting about the previous disordering”) and this worsens the conditions for the formation of new integrated spikes.

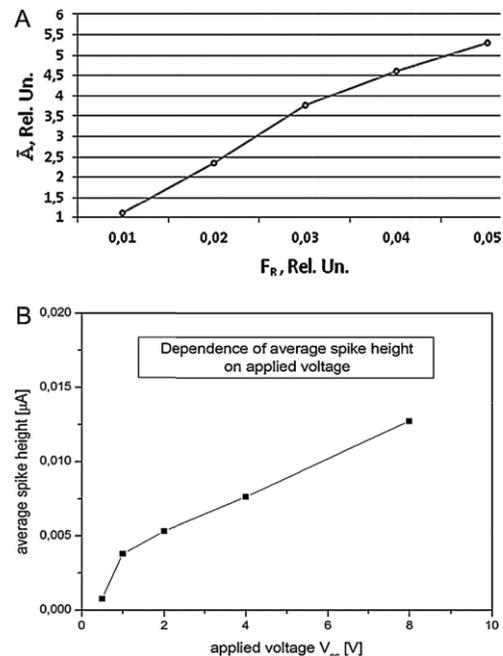


FIGURE 4 Dependence of the average amplitude of MP oscillations on the value of SF (on top); Dependence of the average spike height on the value of the applied voltage (down)

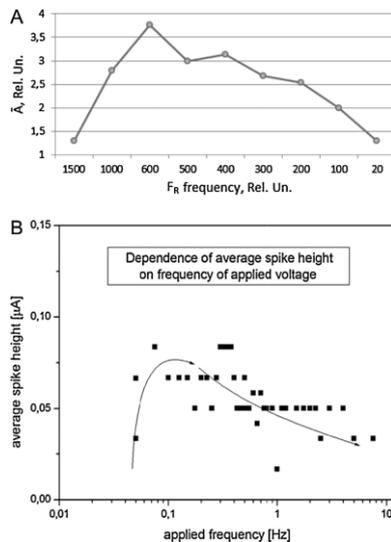


FIGURE 5 Dependence of the average amplitude of MP oscillations on the frequency of SF action (on top); Dependence of the average spike height on the frequency of applied voltage (down), here the curve is drawn by the method of least squares.

References

- [1] Alfonta L, Bukelman O, Chandra A, Fahrner W R, Fink D, Fuks D, Golovanov V, Hnatowicz V, Hoppe K, Kiv A, Klinkovich I, Landau M, Morante J R, Tkachenko N V, Vacik J, Valden M 2009 Strategies towards advanced ion track-based biosensors *Rad. Effects and Defects in Solids* **164** 431–7
- [2] Fink D, Gerardo Munoz H, Alfonta L, Mandabi Y, Dias J F, de Souza C T, Bacakova L E, Vacik J, Hnatowicz V, Kiv A, Fuks D, Papaleo R M Status and Perspectives of Ion Track Electronics for Advanced Biosensing, (Y.N. Shunin and A.E. Kiv eds.) *Nanodevices and Nanomaterials for Ecological Security NATO Science for Peace and Security Series B: Physics and Biophysics* DOI 10.1007/978-94-007-4119-5_24
- [3] Fink D, Petrov A, Hoppe K, Fahrner W R 2003 Characterization of “TEMPOS” structures *Proceedings of MRS meeting* 1–5
- [4] Powell M R, Sullivan M, Vlasiouk I, Constantin D, Sudre O, Martens C C, Eisenberg R S, Siwy Z S 2008 Nanoprecipitation-assisted ion current oscillations *Nat. Nanotech* **3** 51–9
- [5] Fink D, Cruz S, Munoz G, Kiv A 2010 Current spikes in polymeric latent and funnel-type ion tracks *Rad. Effects and Defects in Solids* **166** 373–88
- [6] Fink D, Munoz G H, Vacik J, Alfonta L 2011 Pulsed biosensing *IEEE Sens.* **11** 1084–9
- [7] Fink D, Kiv A, Fuks D, Tabacnic M, Rizutto A, Silva A, Chandra A 2007 Irradiation induced pulsations of reverse biased metal oxide/silicon structures *Appl. Phys. Lett* **91** 83512–20
- [8] Cervera J, Schiedt B, Ramirez P 2005 Poisson - Nernst–Planck model for ionic transport through synthetic conical nanopores *Europhysics. Lett* **71** 35–41
- [9] Cervera J, Schiedt B, Neumann R, Mafe S, Ramirez P 2006 Ionic conduction, rectification and selectivity in single conical nanopores *Chem. Phys* **124** 104706–10
- [10] Cruz-Chu E R, Aksimentiev A, Schulten K 2009 Ionic current rectification through silica nanopores *Phys. Chem* **113** 1850–62
- [11] Cruz-Chu E R, Ritz T, Siwy Z S, Schulten K 2009 Molecular control of ionic conduction in polymer nanopores *Faraday Discuss* **143** 47–62
- [12] Siwy Z, Apel P, Dobrev D, Neumann R, Spoor R, Trautmann C, Voss K 2003 *Nucl. Instr. Meth* **208** 143
- [13] Fink D, Vacik J, Garcia Arellano H, Muñoz G H, Alfonta L, Fahrner W R, Hoppe K, Kiv A 2014 Biosensors with Nuclear Tracks and Embedded Membranes *Key Engineering Materials* **605** 83-6
- [14] Fink D, Muñoz G H, Vacik J, Alfonta L 2011 Pulsed Biosensing *IEEE Sensors J.* **11**(4) 1084-7
- [15] Fink D, Kiv A, Fuks D, Saad A, Vacik J, Hnatowicz V, Chandra A 2010 Conducting Swift Heavy Ion Track Networks *Rad. Effects and Defects in Solids* **165** 227-44
- [16] Fink D, Kiv A, Fuks D, Saad A, Vacik J, Hnatowicz V, Chandra A 2010 Conducting Swift Heavy Ion Track Networks *Rad. Effects and Defects in Solids* **165** 227-44
- [17] Fink D, Hnatowicz V, Yu P 2004 *Apel Transport Processes in Tracks In: Transport Processes in Ion Irradiated Polymers* D. Fink Ed. *Springer Series in Materials Science* **65** 128-30 and references therein
- [18] Fink D, Kiv A, Fuks D, Munoz G, Cruz S A, Fahrner W R, Hoppe K 2011 Collective interaction in ion track electronics *CM&NT* **15** 7–18
- [19] Fink D, Cruz S, Munoz G, Kiv A 2011 Current spikes in polymeric latent and funnel-type ion tracks *Rad. Effects and Defects in Solids* **166** 373–88
- [20] Mykytenko N, Fink D, Kiv A 2015 Computer modeling of ion current pulsations in track-containing foils *Journal of Computational Science* **6** 34–9
- [21] Vavilov V S, Kiv A, Niyazova O 1981 *Formation and migration of defects in semiconductors*
- [22] Haile J M 1997 *Molecular Dynamics Simulation: Elementary Methods*
- [23] Griebel M, Knapek S, Zumbusch G 2007 *Numerical Simulation in Molecular Dynamics* Springer
- [24] Parviainen S, Djurabekova F, Timko H, Nordlund K 2011 Electronic processes in molecular dynamics simulations of nanoscale metal tips under electric fields *Comput. Mat. Sci* **50** 2075–9
- [25] Hairer E, Lubich Ch, Wanner G 2003 Geometric numerical integration illustrated by the Stormer-Verlet method *Acta Numerica* **1** 399–450
- [26] Mostafa Ghiaasiaan S 2011 *Convective Heat and Mass Transfer, Appendix K: Lennard Jones Potential Model Constants for Selected Molecules* Cambridge University Press

5 Conclusion

The effect of current spikes in track devices is described. Different explanations of this effect are discussed. A phenomenological model for ion current spikes in track-containing foils is considered in more detail.

Experimental results show that ion current spikes in track-containing foils arise for different forms of tracks and different types of materials and that the current spike effect is determined significantly by the mean distance between tracks. A computer experiment with the developed model for current spikes showed that this model reflects the main features of the ion current spikes in track structures. It is shown that taking into account only one factor (the interaction of currents in the system of tracks) leads to the result of an emergence of current spikes. The occurrence of current spikes takes place in a wide range of the potential parameters.

Authors	
	<p>Dietmar Fink</p> <p>Current position, grades: Dr., at present he is a guest Professor of the University of Mexico. Earlier he was a Head of laboratory in Hahn-Meitner institute (Berlin), the Editor-in-Chief of International journal "Radiation effects and defects in solids".</p> <p>Scientific interest: radiation effects in nanodevices</p> <p>Experience: one of the founders of track electronics</p>
	<p>Arnold Kiv</p> <p>Current position, grades: Professor-researcher of Department of materials engineering, Ben-Gurion University of the Negev, Head of Department of physical and mathematical modelling, South-Ukrainian national pedagogical university after K. D. Ushinskij</p> <p>Scientific interest: nanomaterials and nanoelectronics,</p> <p>Experience: He is an expert in the theory of real structure of solids and physical processes in electronic devices. In last decade he obtained significant results in the field of nanomaterials, track nanostructures and track electronics.</p>
	<p>Yuri Shunin</p> <p>Current position, grades: Professor and Vice-Rector on academic issues at Information Systems Management University and a leading researcher at the Institute of Solid State Physics, University of Latvia.</p> <p>University studies: PhD (physics and maths, solid state physics, 1982) at the Physics Institute of Latvian Academy of Sciences and Dr. Sc. Habil. (Physics and maths, solid state physics, 1992) at Saint-Petersburg Physical Technical Institute (Russia).</p> <p>Scientific interests: His current research activities concern nanophysics, nanoelectronics, nanodevices, nanomaterials, nanotechnologies, nanorisks, nanoeducation, and nanothinking</p> <p>Publications: over 470, 1 book with Springer</p> <p>Experience: director of NATO ARW "Nanodevices and Nanomaterials for Ecological Security," Riga, Latvia, 2011, a visiting researcher at Gesellschaft für Schwerionenforschung mbH, Darmstadt, Germany (1995), INFN—Laboratori Nazionali di Frascati, Frascati-Roma, Italy (2010 to 2015), participation in EU FP7 Projects CATHERINE (2008 to 2011) and CACOMEL (2010 to 2014), education practitioner in Higher Education from 1975 till nowadays</p>
	<p>Natalia Mykytenko</p> <p>Current position, grades: Post-graduate student of the Department of physical and mathematical modelling, South-Ukrainian national pedagogical university after K. D. Ushinskij</p> <p>University studies: South-Ukrainian national pedagogical university after K. D. Ushinskij.</p> <p>Scientific interest: She is a specialist in the field of neural networks and classical molecular dynamics. Her works are devoted to modelling of physical processes in track nanodevices.</p>
	<p>Tamara Lobanova-Shunina</p> <p>Current position, grades: Associate Professor at Riga Technical University, Aeronautics Institute, PhD, Dr.edu</p> <p>University studies: University of Latvia, She obtained her PhD (2009) on innovative education at South-Ukrainian National University</p> <p>Scientific interests: current research activities concern nanotechnologies, nanomanagement, nanoeducation, nanorisks, and nanothinking in the EU FP7 Project CACOMEL (2010 to 2014), special interests are connected with the systemic approach to nanosystems applications.</p> <p>Publications: 53 regular papers</p> <p>Experience: She was a member of the NATO ARW Local Organizing Committee 'Nanodevices and Nanomaterials for Ecological Security', Riga, Latvia, 2011. She has been working as a visiting researcher at INFN-Laboratori Nazionali di Frascati, Frascati-Roma, Italy (2010 to 2015). She was the Head of International Business Communications Department, Director of the study programme 'International Business Communications' at Information Systems Management University (till 2013). She is the Editorial Board member of the journal 'Innovative Information Technologies'. She is the author of over 53 scientific papers in international scientific journals.</p>
	<p>Almagul T Mansharipova, Kazakhstan, Almaty</p> <p>Current position, grades: Head of Department science KRMU, professor, member of Association aksam, Almaty, Kazakhstan</p> <p>University studies: Kazakh Russian medical university</p> <p>Scientific interest: Research in the field of medicine, cell biology, bioinformatics.</p> <p>Publications: 150 papers, 3 patents.</p>
	<p>Tatyana Koycheva</p> <p>Current position, grades: PhD in pedagogical sciences, As. Professor of the Department of Informatics and Applied Mathematics in South-Ukrainian National Pedagogical University after K. D. Ushinskij.</p> <p>Scientific interest: innovative education, distance education, influence of scientific communities on the corporate culture formation.</p> <p>Experience: Head of Scientific-Organizational departmentcommunities on the corporate culture formation.</p>
	<p>Ravil I Muhamediev, Kazakhstan, Almaty</p> <p>Current position, grades: Head of Department CSSE&T, professor, International IT University, Almaty, Kazakhstan</p> <p>University studies: International IT University, Almaty, Kazakhstan</p> <p>Scientific interest: stochastic simulation, machine learning, simulation of anisochronous systems</p> <p>Publications: 160 papers, 5 books</p>

	<p>Victor Gopeyenko</p> <p>Current position, grades: professor, Dr.Sc.Eng Vice-Rector on scientific issues at Information System Management University and the director of ISMU Computer Technologies Institute.</p> <p>University studies: Riga Civil Aviation Engineering Institute (Latvia) obtained his doctor's degree (Dr. Sc. Eng., 1987) at Riga Civil Aviation Engineering Institute (Latvia).</p> <p>Scientific interests: current research activities concern nanophysics, nanoelectronics, nanodevices, and nanotechnologies in the EU FP7 Project CACOMEL (2010 to 2014). His special interests concern carbon nanotubes and graphene systems applications and modeling.</p> <p>Publications: 80 regular papers in international scientific journals</p> <p>Experience: Was the member of local organizing committee of NATO ARW "Nanodevices and Nanomaterials for Ecological Security," Riga, Latvia, 2011, the editor-in-chief of the journal 'Information Technologies, Management and Society' and editorial board member of the journal 'Innovative Information Technologies'.</p>
	<p>Nataly Burlutskaya</p> <p>Current position, grades: a researcher at the Information Systems Management University and the Institute of Solid State Physics, University of Latvia</p> <p>University studies: Master degree in computer systems (2011) at Information Systems Management University, Riga, Latvia.</p> <p>Scientific interests: current research activities concern theoretical simulations of the electronic and electrical properties of carbon nanotubes and graphene nanoribbons in the EU FP7 Project CACOMEL (2010 to 2014).</p> <p>Publications: 30 regular papers</p> <p>Experience: Nataly Burlutskaya was the secretary of organizing committee of NATO ARW "Nanodevices and Nanomaterials for Ecological Security," Riga, Latvia, 2011.</p>
	<p>Yuri Zhukovskii</p> <p>Current position, grades: Dr.Chem., Head of Laboratory of Computer Modeling of Electronic Structure of Solids, Institute of Solid State Physics (University of Latvia).</p> <p>Publications: He is the author of over 120 regular and review papers in international scientific journals. His Hirsch index is 16.</p> <p>Experience: From 1977 until 1995 he was a researcher at the Institute of Inorganic Chemistry, Latvian Academy of Sciences. Since 1995 he has been a leading researcher at the Institute of Solid State Physics, University of Latvia. Within the last 20 years he has been granted several fellowships for collaboration, visiting activities and positions at seven universities and scientific centers of Canada, Finland, Germany, United Kingdom, and the United States. He has also been actively engaged in developing active collaboration with some scientific groups in Belarus, Italy, Russia, and Sweden. Simultaneously, he has been a contact person and participant in a number of collaboration projects under support of European Commission. His current research activities concern theoretical simulations on the atomic and electronic structure of crystalline solids (with 3D, 2D and 1D dimensionalities).</p>
	<p>Stefano Bellucci</p> <p>Current position, grades: PhD, Professor, currently coordinates all theoretical physics activities at INFN Laboratori Nazionali di Frascati (Italy).</p> <p>University studies: April 1982: Laurea in Physics (Magna cum Laude), University of Rome "La Sapienza" July 1984: Master in Physics of Elementary Particles, SISSA and University of Trieste, PhD in physics of elementary particles in 1986 at SISSA, Trieste, Italy.</p> <p>Publications: over 400 papers in peer-reviewed journals (with h = 40), and more than 10 invited book chapters, the editor of ten books with Springer</p> <p>Scientific interests: research interests include theoretical physics, condensed matter, nanoscience and nanotechnology, nanocarbon-based composites, and biomedical applications.</p> <p>Experience: Worked as a visiting researcher at the Brandeis University, Waltham, MA, USA (1983 to 1985); at the M.I.T., Cambridge, MA, USA (1985 to 1986); the University of Maryland, USA (1986 to 1987); at the University of California at Davis, USA (1987 to 1988). Editorial board member of the Springer Lecture Notes in Nanoscale Science and Technology, as well as the editorial board member of the Global Journal of Physics Express and the Journal of Physics & Astronomy</p>

Machine learning methods: An overview

Ravil I Muhamedyev

Institute of Problems of Information and Control, Ministry of Education and Science of the Republic of Kazakhstan. Pushkina 125, Almaty Kazakhstan

High School of management information systems ISMA, Lomonosov st. 1, Riga, Latvia

Corresponding author's e-mail: ravil.muhamedyev@gmail.com

Received 1 December 2015, www.cmnt.lv

Abstract

This review covers the vast field of machine learning (ML), and relates to weak artificial intelligence. It includes the taxonomy of ML algorithms, setup diagram of machine learning methods, the formal statement of ML and some frequently used algorithms (regressive, artificial neural networks, k-NN, SVN, LDAC, DLDA). It describes classification accuracy indicators, the use of "learning curves" for assessment of ML methods and data pre-processing methods, including methods of abnormal values elimination and normalization. It addresses issues of application of ML systems at the processing of big data and the approaches of their solution by methods of parallel computing, mapreduce and modification of gradient descent.

Keywords: machine learning learn ability preprocessing big data map reduce

1 Introduction

2 Machine learning techniques

2.1 TYPES OF MACHINE LEARNING ALGORITHMS

2.2 SETUP OF MACHINE LEARNING SYSTEMS

2.3 FORMAL DEFINITION OF A MACHINE LEARNING PROBLEM

2.4 REGRESSIVE ALGORITHMS AND DATA CLASSIFICATION ALGORITHMS

3 Quality assessment of machine learning systems

3.1 INDICATORS OF CLASSIFICATION ACCURACY ESTIMATES

3.2 "LEARN ABILITY" OF ALGORITHMS

4 Data preprocessing

4.1 ELIMINATION OF ABNORMAL VALUES

4.2 DATA NORMALIZING AND CENTERING METHODS

5 Machine learning in Big Data management

Conclusion

Acknowledgment

References

1 Introduction

The domain of Artificial Intelligence (AI) is quite extensive and includes numerous directions, from logics through methods of text tonality analysis. Traditionally the Strong AI and the Weak Ai are distinguished. The first one is oriented to creation of highly intelligent human-level problem systems, and eventually to creation of thinking machines. Such developments are financed by, DARPA [1], for instance. The book [2] also may be mentioned as one of the first to define the notion itself of the thought and to describe the principles of brain function in a popular way.

Weak AI is oriented to creation of applications that perform this or another intellectual ability of humans or animals. For instance, ability of safe across-country movement, beehive (ant) or distributed intelligence [3], systems of natural selection (referred to as genetic algorithms) etc. Frequently, the field of intelligent agents [4] and multi-agent systems, described in detail in the research papers by Gorodetsky V.I. [5, 6] are also included herein.

In the course of development, AI, being at the cutting-edge of scientific research, is gradually changing its matter as a science. If in the beginning of its development, its scope of interests included such problems as bio-identification, text recognition, etc., later on they transformed into the scope of technologies widely used in applied sciences, developments and industries [7]. One of successful directions of

artificial intelligence is machine learning (ML). ML is successfully applied to solve of many kinds of problems.

Nowadays ML is an established discipline in many ways. The methodology of application of ML considers selection of relevant data and its pre-processing, selection of adequate algorithms and solution quality assessment. Development of this vast domain includes search for optimal usage of accumulated potential of big and heterogeneous data, search for rapid learning methods and analysis of application features depending on the field of application.

The article describes consistently ML methods that found extensive practical application, methods of data preparation and indicators of classification quality assessment and algorithms. A certain part is dedicated to big data management. The article consists of the following parts.

Part 2 describes taxonomy of ML techniques, formal statement of a ML problem, some popular techniques and methodology of ML application to problems.

Part 3 describes classification quality indicators and methodology of learning curves application in assessment of ML systems.

Part 4 describes methods of data pre-processing.

Part 5 is dedicated to ML techniques application in big data management.

The conclusion briefly resumes the describe techniques and methodologies.

2 Machine learning techniques

Data interpretation is often connected to classification, when a certain object is to be related to one of the previously determined classes, to clustering, when objects are split into initially undetermined groups (clusters) and to forecasting, when by some volume of initial data describing the process background, for example, it is necessary to determine its future state in space or in time. In all the cases, when no strict formal techniques of classification or clustering are used, ML techniques are used extensively.

ML techniques include a vast class of algorithms, starting with solution trees, genetic algorithms, and metric techniques, such as k-NN, SVM, statistical methods, Bayesian networks and ending with artificial neural networks [4, 7, 8, 9]. On the merits this direction is meant to solve the central problem of an intelligent system, anticipatory to all other actions – assessment of a current object (situation).

One of practical applications, where ML techniques are used extensively since 1970's is extraction of commercial minerals. For instance, artificial neural networks are applied in petrography for log data analysis, in lithology for assessment of mineral resources base, in seismic sounding [10-17]. The research paper [18] is dedicated to application of neural networks, the research paper [18] is dedicated to decision of practical interpretation problems of oil production log data. The research papers [19-21] describe some results of feed-forward neural networks application for interpretation of well-log geophysical data during uranium exploration.

Still applications of ML are much more extensive. They include medicine [22-24], biology [25], robotics, municipal facilities and industry [26], service sector, ecology [27], innovative systems of communication [28], astronomy [29] etc.

Herewith the taxonomy of ML, key algorithms and features of their application are being considered.

2.1 TYPES OF MACHINE LEARNING ALGORITHMS

ML, as a part of a vast scientific direction called Artificial Intelligence (AI), on the merits is implementing the potential of AI idea. The main anticipation related to ML considers implementation of demand for flexible, adaptive, learning algorithms or methods of computation*. This results into new functions of systems and programs. ML opportunities, i.e. ability of learning and giving expert-level recommendations in a narrow application domain, provide algorithms, which are divided into two big groups, realizing the main property – ability of learning:

- Unsupervised learning (UL)
- Supervised learning (SL)

On top of that, specify

- Reinforcement learning [30] (RL)
- Semi-supervised learning [31] (SSL)

The main problem decided with use of ML algorithms involves relating the observed object to one or another class with the following automatic or manual decision generation. Such problems occur very frequently. As a specific illustration may be given the problems that arise in the process of mobile autonomous robot movement, and relating to

recognition of an expected path; recognition problems of face, facial gestures and emotions; analysis of user activity data from e-commerce systems, which allows to perform both interface optimization and planning of system's response. In broader terms, this is a data analysis in different information systems, letting forecast status and classification of objects. Solution approaches of this problem differ.

UL techniques solve the problem of clustering, when the range of initially undetermined objects is split into groups by means of automatic procedure based on the properties of these objects. Whereas the number of groups (clusters) may be given in advance or generated automatically. Such algorithms include the adaptive resonance theory -ART and self-organizing maps- SOM or Kohonen maps [32], and a vast group of clustering algorithms (k-means, mixture models, hierarchical clustering etc.) [33, 34].

SL decide the problem of classification, when finite groups of somewhat determined objects are distinguished in a potentially infinite aggregate of objects. As a rule, groups are formed by an expert. Furthermore the expert can explain or not the reasons of the initial classification.

The algorithm of classification should put the following undetermined objects to this or another expert-organized group by use of the initial classification based on the properties of these objects. SL include a big number of algorithms or a class of algorithms being often divided to linear and non-linear classifiers, depending on the shape (hyperplane or hypersurface) of object dividing classes. In a two-dimension case, linear classifiers divide classes with the only straight line, whereas non-linear classifiers – with the curve (Figure 1).

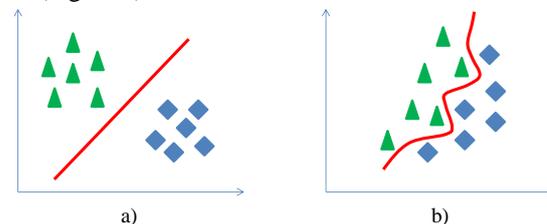


FIGURE 1 Linear (a) and non-linear (b) classifiers

Classification approaches for MO algorithms are presented in research papers [35, 36] in particular. Taxonomy of MO algorithms, not eligible for exhaustiveness, may be presented as the following below hierarchy structure (Figure 2):

Unsupervised learning (UL)

ART

ART1

ART2

ART3

SOM

Generative Topographic Map (GTM)

Cluster algorithms

k-means

K-Means++

K-Medoids

Fuzzy C-Means Clustering Algorithm (FCM)

Soft K-Means Clustering Algorithm (SKM)

* Methods of computation is the term that introduced by Donald Knuth to separate mathematically validated algorithms and empirical methods that frequently used in practice

K-Harmonic Means Clustering Algorithm (KHM)
 Kernel K-Means Clustering Algorithm (KKM)
 Spectral Clustering Algorithm (SCA)
 Density models (DM)
 Subspace models:
 mixture models (MM),
 hierarchical clustering (HC)
 Supervised learning (SL)
 Linear Classifiers
 Linear Discriminant Analysis Classifier (LDA)
 Logical regression (LR)
 Naive bayes Classifier (NBC)
 Perceptron (P)
 Non-linear Classifiers
 Quadratic Classifier (QC)
 Diagonal Linear Discriminant Analysis (DLDA)

Support Vector Classification (SVM) (Linear SVM и Non-linear SVM)
 Logistic regression (LogR)
 k-Nearest-Neighbor (k-NN)
 Decision Tree (DT)
 Random Forest (RF)
 Neural Networks (NN)
 Bayesian Networks (BN)
 Reinforcement learning (RL)
 Q-Learning
 Deterministic Q-Learning (DQL)
 Monte-Carlo Methods (MCM)
 Temporal Difference Methods (TDM)
 Sarsa
 Semi-supervised learning (SSL)

Taxonomy of ML algorithms

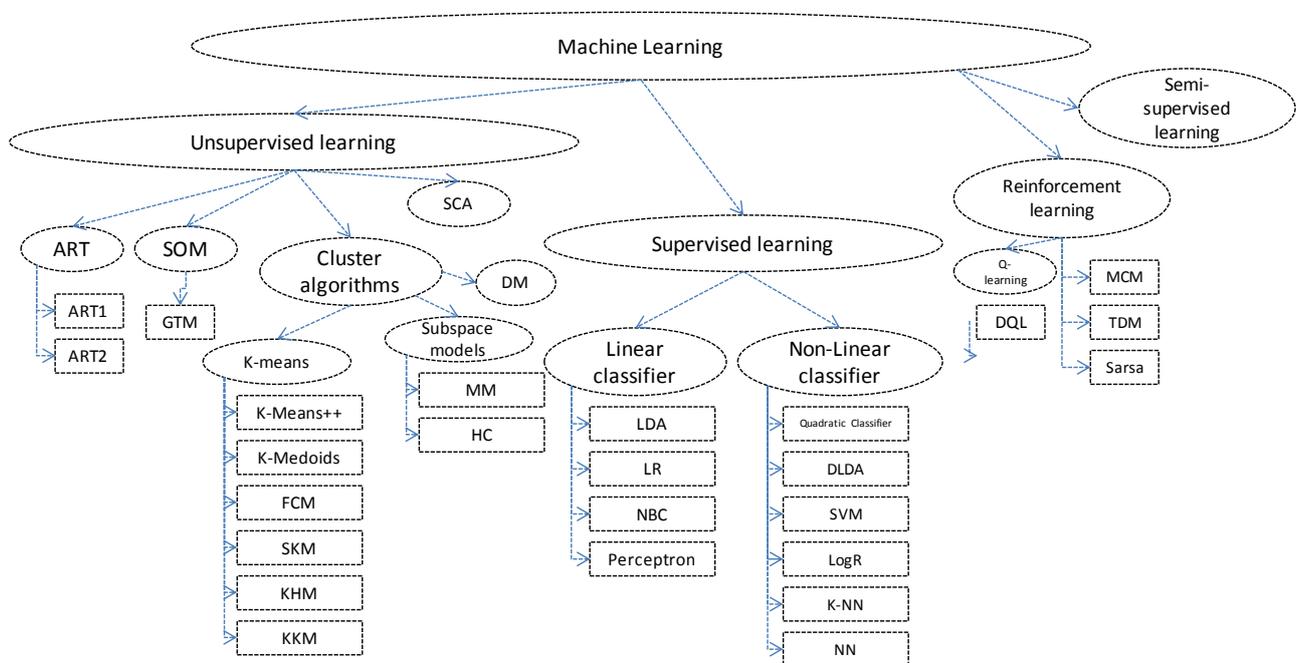


FIGURE 2 Taxonomy of machine learning algorithms

Each of the mentioned algorithms, in fact forms a kind of flock, modified according to these or another needs of programs and algorithms that often differ by computational complexity, difficulty of implementation and learning process automation, ability to classify only two types (binary classification) or several types of objects at once.

Classification and other kind heterogeneous data processing feature a range of peculiarities:

- 1) Data heterogeneity that will require introduction of different metrics for certain types of data in some cases
- 2) Frequent occurrence of big data volumes, when a problem must be solved relatively fast for incoming data portions, which obviously troubles the use of computationally complex learning techniques
- 3) Complexities of results integration.

2.2 SETUP OF MACHINE LEARNING SYSTEMS

Application of ML techniques in the problems that do not operate any pure mathematical models, only data and probably expert estimates, is often an optimal way of solution. A learning system, an artificial neural network in particular, is able to reproduce the behavior being difficult or impossible to formalize. In supervised learning problems it is often difficult to determine the quality of expert estimates. In a point of fact these problems include also the problems of disease risk detection, product quality assessment, speech recognition, prediction of share quotation level in financial markets, the problem of lithology type recognition, etc. Notwithstanding the experts give a list of relevant features, ranges of measured physical quantities, expert estimates may be contradictory or display errors.

Besides, classification data may contain abnormal values and mistakes, related to physical peculiarities of their

acquisition processes. Consequently a learning system can treat data with mistakes.

Analysis of ML techniques applicability, methods of data processing for application of the mentioned methods,

and comparison of algorithms against each other are a necessary condition for development of scientifically proven automatic data interpretation software system.

The general diagram of ML alignment by active problem is presented in Figure 3.

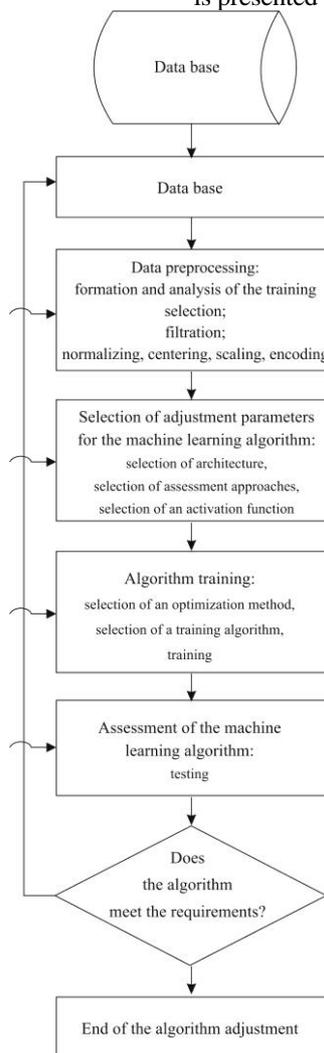


FIGURE 3 Generalized algorithm of a machine learning system alignment by active problem

The alignment technology is an interactive procedure, when correct functioning of the system requires multiple returns to the previous alignment blocks of parameters.

According to this diagram we need to specify the very problem of ML in formal form, to select algorithms to be used in the system, to define quality parameters of ML algorithms, to choose and to validate the methods of data pre-processing.

2.3 FORMAL DEFINITION OF A MACHINE LEARNING PROBLEM

A formal statement of a ML problem (an example-based learning problem or supervised learning problem) is known and to be as follows [37]. Let X and Y be two metric spaces: X (the space of admissible objects), Y (answer or label space) and a (target) function $y: X \rightarrow Y$, specified only in the finite aggregate of points (learning selection, sample sets): $y(x^1), \dots, y(x^m)$, i.e. the labels of objects x^1, \dots, x^m are known. It requires to build an algorithm A (or "instruct" an

algorithm), which calculated a $y(x)$ value by an object x (or „close enough” value, if an inaccurate solution is considered acceptable).

For a finite aggregate $Y = \{1, 2, \dots, l\}$, a problem is called a classification problem (per l non-intersecting classes). In this case it may be considered that a set X is split to classes K_1, \dots, K_l , where $K_i = \{x \in X \mid y(x) = i\}$ at $i \in \{1, 2, \dots, l\}$:

$$X = \bigcup_{i=1}^l K_i$$

When $Y = \{(\alpha_1, \dots, \alpha_l) \mid \alpha_1, \dots, \alpha_l \in \{0, 1\}\}$ speak of a classification problem per l intersecting classes. Here i-class – $K_i = \{x \in X \mid y(x) = (\alpha_1, \dots, \alpha_l), \alpha_i = 1\}$.

Frequently the loss or cost function $J(A(x), y(x))$ can be used in the problem to describe how “wrong” is our answer A(x) against a true answer y(x). In case of a classification problem it may be considered that.

$$J(A(x), y(x)) = \begin{cases} 1, A(x) \neq y(x) \\ 0, A(x) = y(x) \end{cases}$$

But in a regression problem –

$$J(A(x), y(x)) = |A(x) - y(x)|$$

or

$$J(A(x), y(x)) = (A(x) - y(x))^2.$$

A problem of learning using a set of training samples presents also an optimization task, which can be decided by search of the minimum value of a cost function $J(\theta)$ across all available examples, defined as the *Sum of Squared Differences* of “a forecasted” value and a real value of y through a wealth of examples m . Herein a hypothesis $h_\theta(x)$ “is profiled” to provide the minimum value $J(\theta)$ at a certain set of parameters $\theta_i \in \Theta$.

$$J(\theta) = \min \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2, \tag{1}$$

where m – is an aggregate of samples, h_θ – is a function of hypotheses that can be linear $h_\theta = \theta_0 + \theta_1 x$ or non-linear, for instance $h_\theta = \theta_0 + \theta_1 x + \theta_2 x^2$ with a varying set of parameters $\theta_i \in \Theta$.

Running ahead may be said that adjustments of parameters θ_i needs that parameters $x_j \in X$ (in a multi-dimensional case) were expressed in units of the same dimension and approximately of the same value. Most frequently normalization is used to render all the parameters in numbers of the range $0 \leq x \leq 1$ or $-1 \leq x \leq 1$. Basically, selection of a normalization function depends on a problem’s class. Furthermore, the process of data preprocessing may employ the methods that provide exclusion of anomalous values, exclusion of noises, for example, high-frequency, by the way of fitting, etc. Selection of such methods also depends on a problem class.

After the parameters have been normalized and data is properly framed, there is performed the search of the hypothesis function $h_\theta(x)$, which minimizes the cost $J(\theta)$. For solution of this problem a large number of algorithms is used, partly they are described below.

2.4 REGRESSIVE ALGORITHMS AND DATA CLASSIFICATION ALGORITHMS

2.4.1 Linear regression

A linear regression problem is stated as the search for the minimum cost function (see Formula 1) under the terms that a hypothesis function is a linear one $h_\theta = \theta_0 + \theta_1 x$. Obviously that such function realizes the linear classifier. The gradient descent algorithm is used to obtain an optimal function $h_\theta(x)$, its essence is about the consequent change of parameters θ_0, θ_1 from the following equation

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1), \tag{2}$$

where α is a learning parameter, but $\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ is a derivative of cost function by θ_j . The sign: = means assignment contrary to the sign of equality (=) in algebraic expressions.

Along with this the algorithm steps are arranged so that initially there occurs the simultaneous change of both parameters based on the Equation (2.2) and only afterwards the assignment of new values to them. Put it differently, the algorithmic sequence of one step of algorithm to be expressed in pseudo code will be the following in case of two parameters

$$temp0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1);$$

$$temp1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1);$$

$$\theta_0 := temp0;$$

$$\theta_1 := temp1;$$

Depending on the learning parameter α , the algorithm may reach the minimum (converge) or, at a too large α , not converge.

The simplest in realization, but not the most optimal in terms of time complexity the “Batch” Gradient Descent algorithm uses all learning examples through each step of the algorithm. In search for parameters θ instead of the gradient descent algorithm the matrix equation can be used $\Theta = (X^T X)^{-1} X^T y$, where Θ is a vector of parameters, $(X^T X)^{-1}$ is an inverse matrix of $X^T X$, X^T is a transpose matrix of X .

The advantage of matrix operations is that there is not necessary to try the parameter α and to iterate several times. The disadvantage refers to the necessity of getting the inverse matrix, which computational complexity is proportional to $O(n^3)$, and to impossibility of getting the inverse matrix in several cases.

2.4.2 Polynomial regression

Contrary to linear regression polynomial one operates non-linear function of the hypothesis $h_\theta = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \dots + \theta_n x^n$, which allows to plot extrapolated curves (hypersurfaces) of complex shape, but increases the number of parameters and computational complexity. Apart from that, the danger of “relearning” exists, when a curve becomes too complex it fits well the learning set, but it produces a large error across the test set.

In the case when a classifier loses its ability to generalize regularization is used to decrease the effect of high order values

$$J(\theta) = \min \frac{1}{2m} \left[\sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

Growth of the parameter λ results into extension of the ability to generalize the algorithm. The function of hypothesis turns into the straight line within limits at a very large value of the parameter.

2.4.3 Logistic regression

It is used if objects have to be divided into two classes, for instance, to "negative" and "positive". In this case the function of hypothesis requires fulfillment of the condition $0 \leq h_\theta(x) \leq 1$, which is achieved with use of a sigmoid (logistic) function

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

where Θ - a vector of parameters
Can be expressed also as

$$h_\theta(x) = g(\Theta^T x),$$

where $g(z)$ - a sigmoid function.

Let us remark that a sigmoid function also is widely used in neural networks as an activation function of neurons, as it is continuously differential and hence guarantees convergence of algorithms of neural network learning. The sample of a sigmoid is demonstrated in Figure 4.

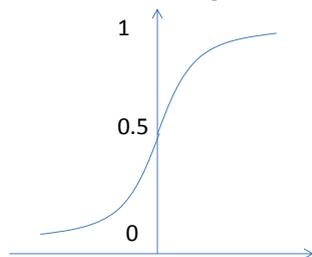


FIGURE 4 Sigmoid function

$h_\theta(x)$ can be considered as the likelihood that the object can be "positive" $h_\theta(x) \geq 0.5$ or "negative" $h_\theta(x) < 0.5$. In complex cases that requires a non-linear interface, for example in the shape of a circle (Figure 5) - $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$

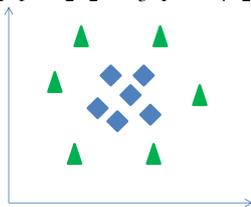


FIGURE 5 Non-linear interface between objects of different classes

Adjustment of parameters Θ , after selection of the function of hypothesis, is executed with regard to minimize the cost function as follows

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right]$$

Same as in case of a linear regression, minimization is reached with use of the gradient descent algorithms, though Conjugate gradient [38], BFGS, L-BFGS [39] is also applicable.

Logistics classifier can be used also for the case of several classes. Here, the classifier is adjusted separately for each class. The class, which a new object belongs to, is computed by value calculation of all functions of hypothesis and selection of maximum rated value among them $\max_i h_\theta^{(i)}(x)$, where i is a class number.

Put it differently, the object belongs to the class with the maximum function of hypothesis.

Same as in case of a linear regression, regularization is used to increase the generalizing ability of the algorithm

$$J(\theta) = \left[-\frac{1}{m} \sum_{i=1}^m y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

2.4.4 Artificial neural networks

Artificial neural networks (ANN) is an apparatus being actively researched since 1940's. NN, as a part of the theory of connectionism, has passed the prominent path from the epoch of overestimated expectations, afterwards, through the epoch of disappointments in 1970's, to a widely applicable technology at present. The link between biological neurons and opportunities of their modeling with use of logistic computations is stated in the paper Warren S. McCulloch, Walter Pitts [40], in the Rosenblatt's paper [41] the model of perceptron is described, in the books of Minsky M. and Papert S. [42-44] limitations of a single-layer perceptron are elicited. In 1974 Paul Werbos suggested the algorithms of back propagation [45, 46] applicable for learning of a multilayer perceptron or a neural network. The most popular and helpful ANN is a network of forward propagation, where non-linear elements (neurons) are represented as consequent layers, but information is distributed in one direction [47] (feed-forward neural networks). In 1989 the research papers by Cybenko G. [48] and Hornik K. etc. [49] demonstrated that such a network is able to approximate practically any function.

The theory of connectivism was contributed considerably by national scientists [50-53], who demonstrated the possibility to decide classical computational problems based on neural networks, thus putting a fundamental principle of neural computers generation.

Application of a neural network apparatus addresses decision of a wide range of computationally complex problems, such as optimization, signal processing, image recognition, forecasting and classification.

Let us consider application of feed-forward neural networks. An individual neuron represents a logistic element, consisting of input elements, an integrator, an activating elements and a single output (Figure 6).

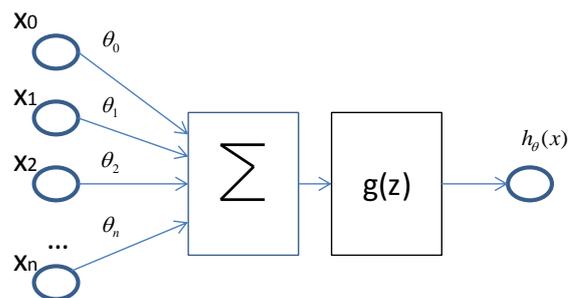


FIGURE 6 Diagram of a classical neuron

Neural output is determined with the formulas

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$$

$$h_\theta(x) = g(z),$$

where $g(z)$ – a sigmoid function.

To make the diagram simpler the integrator and the activating element are united, thus a multilayer network can look as in the Figure 7. The network contains three input neurons, three hidden layer neurons and one output neuron. In the Figure input neurons are marked with the symbol x , hidden layer neurons – with the symbols $a_1^{(2)}, a_2^{(2)}, a_3^{(2)}, a_0^{(2)}$, and output layer neurons – with the symbol $a_1^{(3)}$.

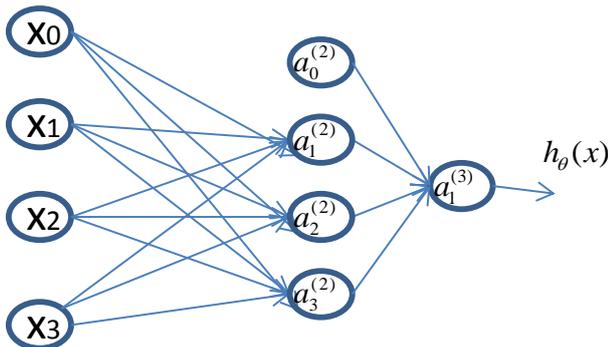


FIGURE 7 Diagram of a multilayer network

Output of each hidden layer neuron can be calculated same as the one for a single neuron:

$$a_1^{(2)} = g(\theta_{10}^{(1)} x_0 + \theta_{11}^{(1)} x_1 + \theta_{12}^{(1)} x_2 + \theta_{13}^{(1)} x_3)$$

$$a_2^{(2)} = g(\theta_{20}^{(1)} x_0 + \theta_{21}^{(1)} x_1 + \theta_{22}^{(1)} x_2 + \theta_{23}^{(1)} x_3)$$

$$a_3^{(2)} = g(\theta_{30}^{(1)} x_0 + \theta_{31}^{(1)} x_1 + \theta_{32}^{(1)} x_2 + \theta_{33}^{(1)} x_3)$$

The neural network output is determined from the equation:

$$h_\theta(x) = g(\theta_{10}^{(2)} a_0^{(2)} + \theta_{11}^{(2)} a_1^{(2)} + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)})$$

The benefit of the neural network is the opportunity to classify several classes at once. In this case the output layer contains the number of neurons equal to the number of classes. For example, as may be required to classify objects of two classes we will get a network (Figure 8).

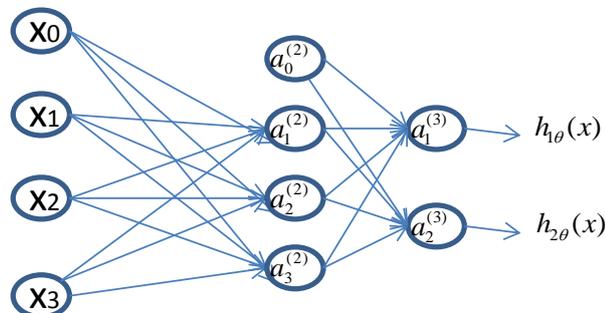


FIGURE 8 Diagram of a multilayer network with double output

To adjust the weights θ of the neural network (training of the network) they use the cost function, reminding the cost function for logistic regression.

$$J(\Theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log(h_\Theta(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - h_\Theta(x^{(i)}))_k \right] + \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{S_l} \sum_{j=1}^{S_{l+1}} (\Theta_{ji}^l)^2$$

where L – number of neural network layers,

S_l – number of neurons in the layer l ,

K – number of classes (equal to the number of neurons in the output layer),

Θ – a weight matrix.

To train a multilayer neural network they use the back propagation error algorithm and different modifications targeting to accelerate the training process.

2.4.5 *k-Nearest-Neighbor (k-NN) [54, 55] algorithm*

The algorithm is based on calculation of the number of objects in each class of the sphere (hypersphere) with the centre in the recognized object. The object belongs to the class, which objects dominate in this sphere. This technique supposes that weights have been chosen individually for every object.

If weights are not same, instead of calculation of the number of objects their weights can be added together. Thus, if the sphere around the recognized object contains 10 reference objects of class A with the weight 2 and 15 error/boundary objects of Class B with weights 1, the point will be referred to Class A.

Weights of objects in the sphere can be expressed as inversely proportional to their distance to the recognized object. Thus, the closer is an object, the more significant it is for this recognized object.

In total, a metric classifier can be described as:

$$a(u; X^l) = \arg \max_{y \in Y} \sum_{i=1}^l [y_u^{(i)} = y] w(i, u),$$

where $w(i, u)$ – a weight of an i -neighbour of an recognized object u , a $(u; X^l)$ – class of an object u , recognized by a selection X^l .

The radius of the hypersphere can be as constant, as dynamic. Moreover, in case of a dynamic radius, a radius of each point is adjusted so that the number of objects in each sphere is constant. Thus, for recognition in areas with varying density of selection, the number of „neighboring” objects (being actually used in recognition) will be the same. In this manner, there is excluded the situation when recognition suffers scarcity of data in low-density areas.

2.4.6 *Support Vector Classification (Linear SVM and Non-linear SVM) [56]*

This algorithms refers to the group of boundary techniques: it determines classes with use of area boundaries. The technique is based on the concept of solution planes. A solution plane divides objects of different classes. In high-dimension spaces instead of straight lines there should be considered hyperplanes of one dimension less than the one of the assumed space. In R^3 , for example, a hyperplane is a two-dimensional plane.

The support vector technique searches for samples, located on class boundaries (at least two), i.e. support vectors, and decides the problem of finding the division a set of objects to classes with use of a linear decision function. The support vector technique builds classification function $f(x)$ as follows:

$$f(x) = \text{sign}(\langle w, s \rangle + b),$$

where \langle, \rangle – scalar product,

w – normal (perpendicular) vector to a dividing hyperplane,

b – auxiliary parameter equal to the absolute module of a distance from a hyperplane to the origin of coordinates. If a parameter b equals zero, a hyperplane passes through the origin of coordinates.

Objects with $f(x)=1$ get into the first class, but objects with $f(x)=-1$ - into another one.

From the point of classification accuracy the best straight line to be selected is the one maximally distant from each class. Such straight line (in general — a hyperplane) is identified as an optimal dividing hyperplane. The problem consists in selection of w and b , maximizing this distance.

In case of non-linear division the approach of support vector machine adaptation is available. It is necessary to input the space of R_n indicators into the space H of a larger dimensionality using the function of the form: $\varphi: R_n \rightarrow H$.

Hence decision of the problem reduces to a linear divisible case, i.e. a dividing classification function is searched again as: $f(x) = \text{sign}(\langle w, \varphi(x) \rangle + b)$.

Also possible another option of data conversion – conversion to polar coordinates:

$$\begin{cases} x_1 = r \cos(\phi), \\ x_2 = r \sin(\phi). \end{cases}$$

2.4.7 Linear Discriminant Analysis Classifier (LDAC) [57]

In mathematical statistics classification is often identified as discrimination. Discriminatory analysis is the section of multivariate statistical analysis, which includes statistical classification methods of multivariate observations.

Linear discriminatory analysis is a statistical and ML method used in search for linear combination of variables eminently suitable for division of two and more classes of objects or events. LDA tries to express a dependant variable (class label) via a linear combination of other indicators or measurements. The acquired combination can be used as a linear classifier. The indicators used to distinguish one class (subset) from another are often identified as discriminatory variables.

Let the learning selection be expressed in matrixes X_1 and X_2 , having by I_1 and I_2 lines (objects). The number of columns (variables) is the same. Initial assumptions are as follows:

- Each class ($k=1$ or 2) represented with normal distribution;
- Covariance matrixes of these two classes are equal $\Sigma_1 = \Sigma_2 = \Sigma$.

The rule of LDA classification is the following: a new sample x is referred to that class, which it is closer to in Mahalanobis units:

$$d_k = (x - \mu_k) \sum^{-1} (x - \mu_k)^t, k = 1, 2$$

In practice unknown mathematical expectations and covariance matrix are replaced with their estimates

$$m_k = \frac{1}{I_k} \sum_{i=1}^{I_k} x_i,$$

$$S = \frac{1}{I_1 + I_2 - 2} (\tilde{X}_1^t \tilde{X}_1 + \tilde{X}_2^t \tilde{X}_2),$$

where \tilde{X}_k – a centered matrix X_k .

If we set equal a distance d_1 to a distance d_2 , $d_1=d_2$, in the mentioned above formula, it makes possible to find the formula of a class division curve. Along with this quadratic terms are reduced and the equation becomes linear

$$xw_1^t - v_1 = xw_2^t - v_2,$$

where

$$w_k = m_k S^{-1},$$

$$v_k = 0.5m_k S^{-1} m_k^t.$$

The values in different parts of the equation are identified as LDA-accounts, f_1 and f_2 . The sample refers to class 1, if $f_1 > f_2$, and, vice versa, to class 2, if $f_1 < f_2$.

2.4.8 Diagonal Linear Discriminant Analysis (DLDA)

Algorithm is a kind of the Linear Discriminatory analysis, described above. This algorithm was described in 2002 [58] as an optimized LDA modification for multivariate data (high-dimension problems). When all class densities have same diagonal covariance matrixes:

$$\Delta = \text{diag}(\sigma_1^2 \dots \sigma_G^2).$$

The classification rule will look as follows:

$$d_k = \sum_{i=1}^G \frac{(x_i - \mu_{ki})^2}{\sigma_i^2}.$$

3 Quality assessment of machine learning systems

In a view of abundance of algorithms selection of their use in a certain ML problem can become complicated. To compare algorithms and methods among themselves and probably with an expertise results they use indicators of algorithm classification accuracy and learning curves. The function of accuracy indicators is to give the estimate demonstrating how much the certain ML classifications or forecasts differ from the ones made by experts. Along with this frequently use the elementary indicator – percentage of correctly classified examples. For estimation of type I and II errors also use some important indicators: “precision”, “recall” and summarizing indicators - T1 Score and kappa. Their use is especially important in case of unequal by volume classes, whereas the number of objects of one type significantly exceeds the number of objects of another type. One more important indicator of ML method is its learning ability, i.e.

improving its accuracy indicators as the number of samples is increasing. At this point, it might happen that the method showing very good results with train set of examples will produce unsatisfactory results with a testing one, i.e. it does not feature the necessary degree of summarizing. The balance between summarizing ability and accuracy can be found with help of “learning curves”, which in a general case are able to show if this or another method can improve its result in a way that its accuracy indicators were approximately equal and satisfied the research domain requirements as for training as for testing setoff values.

3.1 INDICATORS OF CLASSIFICATION ACCURACY ESTIMATES

At present in the field of ML quality assessment most frequently apply:

Accuracy – fraction of correctly classified examples (percentage of correctly classified examples)

$$Ac = \frac{N_t}{N}$$

where N_t - number of correctly classified examples, N – total number of objects.

This indicator is essentially important, but if the number of objects in classes is substantially unequal, said to be uneven or “skewed” classes, it might happen that a very bad classifier would produce a large value Ac. For example, if type I objects make 90% of total number of objects, but type II objects only 10%, it will be sufficient for a classifier just to report on recognition of a type I object and its accuracy will reach 90%. Thus, even if the algorithm will never recognize correctly a type II object, still it will have a high indicator Ac. At this, if recognition of type II objects is of prime importance, an indicator Ac will just mislead. To avoid such an inadequate estimate some more important indicators are taken: “precision”, “recall” and a summarizing indicator - T1 Score, calculated from the following equations:

Precision: $P = \frac{T_p}{(T_p + F_p)}$,

Recall: $R = \frac{T_p}{(T_p + F_n)}$

T1 Score: $T1Score = \frac{2PR}{(P + R)}$

Let us explain the given equations.

Let us concern the case of classification of two classes (or one type I class and all other classes, which we assign the number 0) to. In this case the following situations in Table 1 are optional.

TABLE 1 Classification of two classes

	Actual class	
	1	0
Predicted class 1	True positive	False positive
Predicted class 0	False negative	True negative

Cases True positive (Tp) and True negative are the cases of correct operation of a classifier, appropriately, cases False

negative (Fn) and False positive (Fp) of incorrect operation. At this, Fn can be concerned as the sign of an excessively pessimistic (cautious) classifier, Fp – vice versa, as the sign of excessively optimistic or incautious classifier. Then

Precision: $P = \frac{T_p}{(T_p + F_p)}$

will show a part of correctly recognized objects of the specified class referring to the total number of objects taken by the classifier as objects of the specified class. Alternatively, **recall**

$$R = \frac{T_p}{(T_p + F_n)}$$

will show the ratio of correctly recognized objects to the total number of objects of the given class.

Both indicators and P and R show “confusion” of a classifier. Still P shows how optimistic is a classifier in its estimates, or how frequently it “likes” (low value of P) adding objects of other classes to the given one. While R shows how “pessimistic” is a classifier in its estimates, i.e. how frequently it neglects (low value of R) objects of the needed class.

Evidently, it is desirable that both of these indicators tend to 1. For some “average” estimate use

$$T1Score = \frac{2PR}{(P + R)}$$

which, as is clear from the equation, also tend to 1, if both indicators P and R are close to 1.

Let us notice that use of simple Average= (P+R)/2 can result into a wrong idea about properties of algorithm. For example, let us assume three algorithms showing the following Precision и Recall estimated (Table 2)

It is evident that a simple average (column Average) gives the highest estimate of an absolutely bad algorithm 3, which take all objects as required in error (P are very few). At the same time T1 Score shows the more correct result, giving the highest score to the algorithm 1, which shows close Precision and Recall estimates, and, consequently, will be weighed in their estimates.

TABLE 2

	Precision (P)	Recall (R)	Ave-rage	T1 Score
Algorithm 1	0.55	0.44	0.495	0.4888889
Algorithm 2	0.71	0.12	0.415	0.2053012
Algorithm 3	0.03	1	0.515	0.0582524

For a finer estimate of developed algorithms also use indicators of errors designed for parts of a selection: error on a control selection, error of cross-validation and validation techniques: fold validation, random subselection (sub-sampling) validation.

Moreover, for the problems with greatly differing numbers of class representatives, the cost function (errors) can be computed in a specific way, for example, in case of skewed classes:

$$J(\theta) = \min \sum_{k=1}^2 \frac{1}{|\{t | y(x_t) = k\}|} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2,$$

3.2 “LEARN ABILITY” OF ALGORITHMS

The Estimation of recognition algorithms by simple comparing the quality indexes (accuracy, precision, recall) has the disadvantage that makes it impossible to evaluate the algorithms in the dynamics of change in the volume of training sample. Particularly, in terms of neural networks, accuracy indicators are effected substantially by the number of hidden layers and the number of training examples. With the use of linear regression its order, for k-NN – is the nearest neighbours circle radius, etc. Along with this it is important to consider the algorithm’s ability of learning, overfitting or underfitting. The correct balance between underfit and overfit means the search of an algorithm and its parameters, which would be able to show consistent results for a testing set (or a cross validation set). An underfit algorithm will show equally inconsistent results both for test and train sets, while an overfit algorithm will demonstrate a high result for a train set and a low one for a testing set. Let us assume, in case of regression, the appropriate formulas of the curves, extrapolating distribution of training examples as shown below:

- a) $\Theta_0 + \Theta_1x$ - high bias(underfit)
- b) $\Theta_0 + \Theta_1x + \Theta_2x^2$ - just right
- c) $\Theta_0 + \Theta_1x + \Theta_2x^2 + \Theta_3x^3 + \Theta_4x^4$ - high variance (overfit)

Extrapolation results at some hypothetic distribution of train set objects are shown in Figure 9).

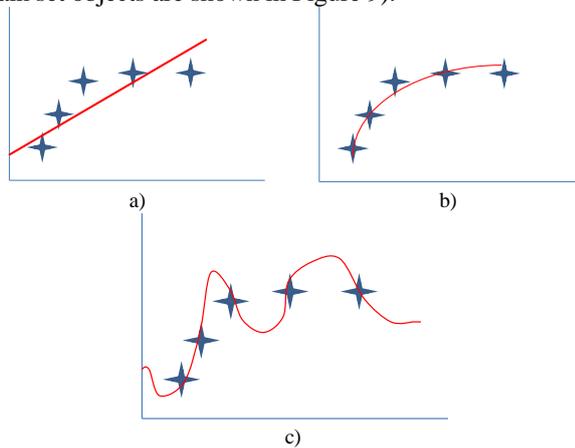


FIGURE 9 Chart of ML algorithm underfit and overfit
 a) too linear divisor (underfit) b) almost ideal case
 c) too many variables (overfit)

At this error indicators across train and test (cross validation – cv) sets are determined from identical equations (changes only a set of examples)

$$J_{train}(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2,$$

where m – train set of examples, m_{cv} - test set of examples (cross validation - cv),

h_{θ} – function of the hypothesis, which can be linear $h_{\theta} = \theta_0 + \theta_1x$ or non-linear, for example, $h_{\theta} = \theta_0 + \theta_1x + \theta_2x^2$ with different set of parameters $\theta_i \in \Theta$.

Model parameters that define the function $h(\Theta)$, are calculated with use of a train set, and validated by examples from a testing set.

Looking ahead, it might be stated that compensation of excess variables in case of overfit the regression model applies regularization, achieving that variables of higher order showed lower effect. The cost estimate formula with consideration of regularization is specified below.

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2,$$

where λ - a regularization parameter.

If a neural network is used reduction in the number of hidden layers of a neural network performs the similar function.

Application of regularization or reduction of the number of hidden layers increase summarizing ability of a ML algorithms and consequently decreases learning ability, in terms of flexibility of adjustment according to subtle differences between classes.

Let us notify that ML systems can be classified as high bias, having a comparatively low capability to generate complex interpolatory curves, and as high variance systems, being able to form curves (surfaces) of complex shape. Pattern of these algorithms (models) differs when the number of training examples increases. Te first ones, as a rule, summarize results ignoring frequently some, probably, essential differences between examples. The second ones, on the contrary, "trace" all nuances, probably random, but at the same summarize insufficiently. The first ones feature underfit, but the second ones – overfit (Figure 2.5) As a rule, it is impossible to evaluate model abilities during one experiment, as both first and second ones can produce close error indicators.

In this context, trainability of ML algorithms is estimated with help of so called learning curves, taking into account the following empirical patterns:

- In normal environment, with an effective algorithm, with increasing the number of training examples the error across the train set will slightly grow, but the error across the test set will decrease (Figure 10);
- If a system is comparatively linear (high bias) increase in the number of training examples will be of little use. The error both across the training and test sets will be approximately equal and large (Figure 11);
- In case of a high variance system, increase in the number of training examples will lead to reduction of the error value across the test set, but it will differ substantially from the error across the train set (Figure 12);

To decrease the test set error even more it is possible to increase essentially the training set (which is not always possible).

Thus, to estimate which of two groups the analyzed algorithm belongs to (too linear or too flexible) it is recommended to analyze the curve of learning errors with increasing body of the learning set, for example, if curves show convergence, but in parallel a high level of errors, this may point to a linear model (impossible to train it).

If unwilling properties of the algorithm have been detected it is possible to try to adjust it, to change the body of the learning set of examples or to choose additional properties of objects, by taking into account the following:

- increasing the size of the train set is helpful at high variance of the algorithms (many layers of neural network, high order of regression), when the program does not feature the needed degree of summarizing, and is adjusted mainly to the train set of examples and cannot normally classify examples from the test set (overfit error).
- Decrease in number of used properties or parameters is helpful at high variance of the algorithms (many layers of the neural network, high order of regression), i.e. once again for the cases of overfit, and when in parallel the number of training examples is impossible to be increased substantially;
- Use of additional properties is helpful in case of too linear algorithms (low order of regression, few neurons in hidden network layers, or few hidden layers), when the program will show the same bad results both on the test and train sets (underfit errors);
- Use of special synthesized (polynomial) properties, showing higher degrees and products of basic ($x_1^2, x_2^2, x_1 x_2, \dots$) is also helpful in case of too linear algorithms (low order of regression, few layers of the neural network) (underfit model).

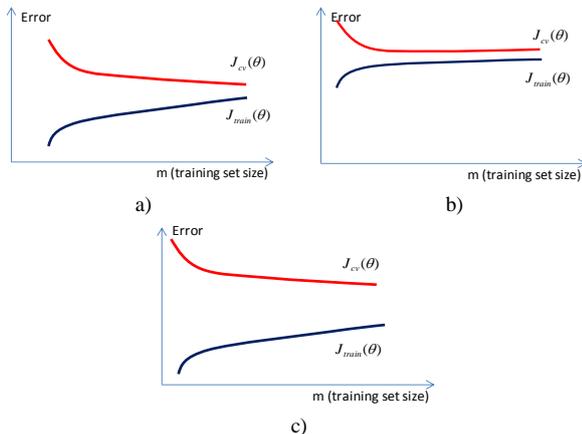


FIGURE 10 Learning curves

- a) normal situation, learning algorithm with good summarizing ability
- b) hard-trained algorithm, a high bias system
- c) high variance algorithm

4 Data preprocessing

ML method use require to render the processed data to a certain format that would allow to feed them to learning and analysis algorithms input. Along with this important cycles of processing are performed, including as a the rule,

cleaning of abnormal values; data normalizing; "fitting"; data reformatting; formatting of input data set, which for example can include formatting of a so-called dockable pane, needed for analysis of patterns of presented data sequences; data reconciliation, for example when one data set is time, distance, spectrum, etc. translated, with respect to others, etc. Let us notice that if the set of ML algorithms is known and the properties of these algorithms are mostly well studied, the processes of data preprocessing vary and depend directly on the object domain and quality of available data. Frequently, the specified above "classical" set is added with additional stages, letting finally raise the quality of data interpretation.

4.1 ELIMINATION OF ABNORMAL VALUES

4.1.1 Normal distribution-based algorithm of abnormal value detection

The normal distribution-based algorithm for abnormal activity detection is designed on the assumption that the whole set of "correct" objects forms Gaussian distribution (normal distribution), i.e. x values are distributed according to the normal law of distribution, defined by mathematical expectation - μ and mean square deviation - δ^2

$$x \sim N(\mu, \delta^2),$$

which can be presented graphically with Figure 11.

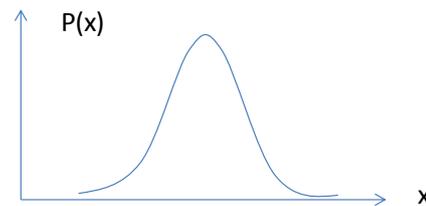


FIGURE 11 Approximate curve of x objects probability distribution subordinated to normal (Gaussian) law

Along with this probability of that or another value is calculated from the known formula

$$p(x; \mu, \delta^2) = \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(x-\mu)^2}{2\delta^2}\right).$$

If objects have several properties, then probabilities will subordinate to the multivariate Gaussian distribution.

At this,
$$p(x) = \prod_{j=1}^n p(x_j; \mu_j, \delta_j^2),$$
 where n – number

of object properties to be classified (number of parameters).

In this manner, the calculation algorithm of abnormal objects is as follows:

Step 1. Based on m train set examples the parameters of the multivariate Gaussian distribution are determined - $\{\mu_1, \dots, \mu_n; \delta_1^2, \dots, \delta_n^2\}$:

$$\mu_j = \frac{1}{m} \sum_{i=1}^m x_j^{(i)},$$

$$\delta^2_j = \frac{1}{m} \sum_{i=1}^m (x_j^{(i)} - \mu_j)^2,$$

where $x_j^{(i)}$ - a j-parameter of an x object in the training example I from the set of examples m

Step 2. For each new x item its probability is calculated

$$p(x) = \prod_{j=1}^n p(x_j; \mu_j, \delta_j^2) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\delta_j^2}} \exp\left(-\frac{(x_j - \mu_j)^2}{2\delta_j^2}\right).$$

Step 3. If the calculated probability is lower than some threshold values ε , then this object x is considered abnormal

If $p(x) < \varepsilon$ then x is abnormal.

In practice, training a system to detect abnormal values considers detection of the threshold value ε . Adjustment of this threshold can be performed based on available examples (similarly to supervised algorithms). After the series of experiments such a threshold ε is established so that all (or most part of) “wrong” objects would be detected as abnormal.

The applied approach of abnormal value detection is illustrated graphically with Figure 12

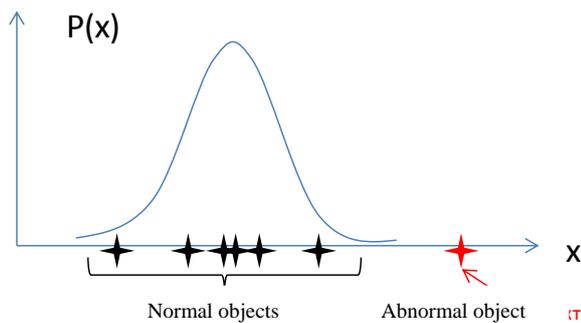


FIGURE 12 Exclusion of abnormal objects

Let us notice that if the distribution differs from normal, it can be normalized by calculating the logarithm of x or the degree of x values.

4.1.2 Use of ML for detection of abnormal values

Obvious, that supervised learning algorithms are quite feasible for detection of abnormal objects. In this case abnormal objects represent an individual class of objects, which can be detected with algorithms k-NN or logistical regression. In both cases adjustment of those object properties being essential in segregation of abnormal values.

4.2 DATA NORMALIZING AND CENTERING METHODS

4.2.1 Linear normalizing

In a view of cleaning the abnormalities data and operation comfort perform data normalizing and centering so that each component of the input vector is located on the segment from 0 to 1 or from -1 to 1. If knowing the change range of input variable there can be used the simplest type of conversion

$$p = \frac{(x - x_{\min})(b - a)}{(x_{\max} - x_{\min})} + a,$$

where [a,b] – the range of acceptable input signals; $[x_{\min}, x_{\max}]$ – the change range of input variable values; p – conversed input signal.

When using this method of conversion (linear normalizing) big change range of input variable can become a problem. In such cases it is possible to use conversion from equations of the sigmoid function or the hyperbolic tangent.

When coding quantitative variable in a general case it is necessary to consider the profound characteristic value, its location over the interval of values, accuracy of measurement. Conversion can be performed using two equations:

$$x_i = \frac{x_i - M(x_i)}{q(x_i)}$$

or

$$x_i = \frac{x_i - M(x_i)}{\max |x_i - M(x_i)|},$$

where x_i - an i-coordinate of an input vector X;

Sampling estimate of mathematical expectation x_i (mean value):

$$M(x_i) = 1/n \sum_{i=1}^n x_i$$

Sampling estimate of the mean square deviation:

$$q(x_i) = \sqrt{(1/n) \sum_{i=1}^n (x_i - M(x_i))^2}$$

To make minor changes of big values significant (for example, when the value of an input variable can reach 10000, but in terms of analysis significant is the value change by 1), three kinds of numerical data preprocessing are used – modular, positional and functional. For accounting of minor changes every value is coded with a vector formed by the rules specified below, instead of the only amount.

4.2.2 Modular preprocessing

There is given a set of positive numbers y_1, \dots, y_k . Calculate each component of the vector Z as follows

$$z_i = \frac{((x \bmod y_i) + y_i)(b - a)}{2y_i} + a,$$

where [a,b], as before, – the range of acceptable input signals.

Let us explain that when $x \bmod y$ is calculated the remainder on dividing x by y is returned, for example, if $x=5, y=3, x \bmod y=2$

4.2.3 Functional preprocessing

In a general case, transformation of an input characteristic x into a k-dimensional vector Z occurs as described below.

There are selected k numbers, satisfying the condition

$$x_{\min} < y_i < \dots < y_k < x_{\max},$$

The elements of the vector Z are computed

$$z_i = \frac{(\varphi(x - y_i) - \varphi_{\min})(b - a)}{\varphi_{\max} - \varphi_{\min}} + a,$$

where φ - the function, determined within the interval $[x_{\max} - y_k, x_{\min} - y_1]$, a $\varphi_{\max}, \varphi_{\min}$ - the maximal and the minimal value of the function within this interval.

4.2.4 Positional preprocessing [59]

The approach in this case is approximately the same as in case of building the positional representation. A positive y value is selected, meeting the condition $y^k \geq (x_{\max} - x_{\min})$. Let us shift an x parameter so that it had only positive values and zero. For calculation of the vector Z we use the following formulas

$$z_0 = (x - x_{\min}) \bmod y,$$

$$z_1 = ((x - x_{\min}) / y) \bmod y,$$

$$z_k = ((x - x_{\min}) / y^k) \bmod y.$$

Other methods of input value conversion – raising to the power, rooting, calculation of reciprocal values, exponential curves and logarithms, and also certain combinations of variables - products, quotients, etc. that can reduce the length of the input data vector.

5 Machine learning in Big Data management

The main problem of ML techniques application in classification and regression of big data are computational complexity of the cost function calculation (Equation (2.1)) and of appropriate parameters of the function of hypothesis (Equation (2.2)) due to a large number of examples.

For example, the standard gradient descent algorithm is an iterative procedure provided via comparatively large number of elementary steps.

Let us assume that we have 100 000 000 examples, 10 parameters and calculate parameters for 500 iterations of gradient descent at the average. The number of required calculations will make $5 \cdot 10^{17}$. Assuming that a computer is able to calculate 1 mln. iterations of gradient descent per minute, we get about 140 hours for calculation of only one set of parameters of the function of hypothesis. If it is needed to draft a learning curve the time of calculation increases proportionally to the number of curve point and can reach absolutely unsuitable values. Though the use of matrix operations (Equation (2.3)) eliminates iterations, still it becomes impossible due to great computational complexity of derivation of a reciprocal matrix.

To overcome “the curse of dimensionality” for the gradient descent algorithms two algorithms are suggested in this case: the one of Stochastic gradient descent – SGD [60] and the one of mini-batch gradient descent – MBGD [61].

The SGD algorithm functions as follows:

1. In the beginning examples from the train set are randomly rearranged.

2. Over all m examples from the learning set, for each of n parameters a new value of the parameter is calculated. In SGD pseudo-code this can be written as:

```

for iter:=1 to K
  for i:=1 to m
    for j:=1 to n
       $\theta_j := \theta_j - \alpha (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$ 
    end
  end
end

```

Number of iterations $K=1, \dots, 10$.

At the same time, the pseudo-code of the batch gradient descent- BGD algorithm looks as follows

```

Do
  for i:=1 to m
    for j:=1 to n
       $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$ 
    end
  end
enddo

```

Thus, at each of m steps of the BGD algorithm cumbersome summing is performed, while SGD uses only one example per each iteration. Assessment of computational complexity of SGD - $O(m)$ comparing to $O(m^2)$ for BGD under condition that the number of parameters n is much fewer than m. Still use of SGD results into approximate decision, not into the global minimum of the cost function. Besides, SGD, same as the described below MBGD also suffer the convergence problem. This means that instead of getting better their result can even get worse with increase in the number of processes examples.

The MBGD algorithm per each iteration uses only a part (b) of examples

```

for iter:=1 to K
  for i:=1 to m with step b
    for j:=1 to n
       $\theta_j := \theta_j - \alpha \frac{1}{b} \sum_{k=i}^{i+b} (h_{\theta}(x^{(k)}) - y^{(k)}) x_j^{(k)}$ 
    end
  end
end

```

Its computational complexity can be assessed as $O(bm)$, which in terms of $b \ll m$ can be traced to $O(m)$.

Except for the mentioned above SGD and MBGD paralleling of calculations can be applied to overcome the problem of large number of calculations [62, 63], which can allow to decide the problem of finding the parameters of the function of hypothesis with help of BGD during the acceptable period of time parallel to large number of independent processes.

The essence of the Map-reduce method is as follows. Calculation of the sum

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}.$$

Is performed on separate machines or processors. Taken that the number of processors is B pieces, the calculation can be performed like this:

$$\begin{aligned}
 Sum_{j0} &:= \sum_{i=1}^{m/B} (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} \\
 Sum_{j1} &:= \sum_{i=1}^{m/B} (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} \\
 &\dots \\
 Sum_{jk} &:= \sum_{i=k*(m/b)+1}^{m/B+k*(m/B)} (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} \\
 &\dots \\
 Sum_{jB-1} &:= \sum_{i=(B-1)*(m/b)+1}^{m/B+(B-1)*(m/B)} (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} \\
 \theta_j &:= \theta_j - \alpha \frac{1}{m} \left(\sum_{k=0}^{B-1} Sum_{jk} \right)
 \end{aligned}$$

For example, if the number of processors is 3, but the number of examples is 300000:

$$\begin{aligned}
 Sum_{j0} &:= \sum_{i=1}^{100000} (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} \\
 Sum_{j1} &:= \sum_{i=100001}^{200000} (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} \\
 Sum_{j3} &:= \sum_{i=200001}^{300000} (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} \\
 \theta_j &:= \theta_j - \alpha \frac{1}{300000} \left(\sum_{k=0}^3 Sum_{jk} \right)
 \end{aligned}$$

Obviously that computational complexity of the method can be rated as $O(\frac{m^2}{B})$. At small B values, growth of calculation speed will be negligent, but if the number of processors is comparable to the number of processors in modern supercomputer clusters (1000-1000000) the growth rate can reach up to several orders.

Application of ML in big data management has serious opportunities. Special languages and platforms are suggested for realization of ML potential in big data management. For example, in [64] the language (SystemML) is suggested and the ways of its application are described for realization of ML techniques in big clusters MapReduce. In [65] the

platform MLBase is described, which provides problem statement and use ML together with high-level operators.

6 Conclusion

The domain of Artificial Intelligence is very vast and includes many disciplines, beginning from logical reasoning to the methods of text tonality analysis. Traditionally recognize the so-called strong artificial intelligence and weak artificial intelligence. The first one is oriented at development of high intelligence human-centric decision systems, eventually at creation of intelligent machines. Weak artificial intelligence is oriented at development of applications, realizing this or another intellectual ability of humans or animals. The potential of weak artificial intelligence concept is realized using machine learning.

The review considers machine learning techniques as a part of weak artificial intelligence methods, applicable for analysis data including for big data processing. The taxonomy of machine learning techniques was proposed. This schema unites various kinds of supervised and unsupervised learning algorithms. At this, classification or other kind of heterogeneous data processing has a range of specific features: heterogeneity of data types, frequently occurring large massives of data, complications in data collation.

The classical diagram of adjustment for machine learning decision techniques is described. The ML problem is formally stated and some frequently used algorithms are described (linear regression, polynomial regression, logistical regression, artificial neural networks, algorithms k-NN, SVN, LDAC, DLDA). Assessment indicators of classification accuracy (accuracy, precision, recall) and the summarizing indicator (T1 Score) are considered in detail. The concept of learn ability of ML algorithms and its practical use (methods of learning curve interpretation) is described. Some data preprocessing methods are described in detail, including the methods of abnormal value elimination and normalization. Briefly considered the application of machine learning methods at the processing of big data and techniques of solving some specific problems with help of parallelized computing, mapreduce approach and modifications of the gradient descent algorithm.

Development of machine learning techniques proceeds in parallel with their practical application, which results into increasing number of applications, appearance special decision techniques of applied problems using ensembles of algorithms. Programming environments and languages including declarative ones are suggested now. Some of them are targeted to facilitate applications at big data. One of the significant task of future researches is development preprocessing methods that could be used at many kinds of data automatically or semi-automatically.

Acknowledgment

This work was supposed by grants 0168/GF4 and 2318/GF3 from the Ministry of Education and Science of the Republic of Kazakhstan.

References

- [1] Systems of Neuromorphic Adaptive Plastic Scalable Electronics [http://www.darpa.mil/Our_Work/DSO/Programs/Systems_of_Neuromorphic_Adaptive_Plastic_Scalable_Electronics_\(SYNAPSE\).aspx](http://www.darpa.mil/Our_Work/DSO/Programs/Systems_of_Neuromorphic_Adaptive_Plastic_Scalable_Electronics_(SYNAPSE).aspx) 10 Aug 2014
- [2] Блейкли С, Хокинс Д 2007 *Об интеллекте* 128
- [3] Weiß G 1999 *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence* 648
- [4] Russell S, Norvig P 2010 *Artificial Intelligence: A modern approach* 1078
- [5] Gorodetsky V I 2012 Self-organization and multi-agent systems *Proceedings of the Russian Academy of Sciences Theory and control systems* 2 92-120
- [6] Gorodetsky V I 2012 Self-organization and multi-agent systems Applications and technology development *Proceedings of the Russian Academy of Sciences Theory and control systems* 3 55-75
- [7] Jones M T 2008 *Artificial Intelligence: A Systems Approach New Delhi: INFINITY SCIENCE PRESS LLC* 500
- [8] Zhang G P 2000 Neural Networks for Classification: A Survey *IEEE Transactions on systems man and cybernetics—Part C: Applications and reviews* 30(4)
- [9] Kriesel D 2015 *A Brief Introduction to Neural Networks* http://www.dkriesel.com/en/science/neural_networks
- [10] Van der Baan M, Jutten C Neural networks in geophysical applications *Geophysics*. 65(4) 1032-47
- [11] Baldwin J L, Bateman R M, Wheatley C L 1990 Application of a neural network to the problem of mineral identification from well logs *The Log Analyst* 279-93
- [12] Benaouda B, Wadge G, Whitmark R B, Rothwell R G, MacLeod C 1999 Inferring the lithology of borehole rocks by applying neural network classifiers to downhole logs - an example from the Ocean Drilling Program *Geophysical Journal International* 136 477- 91
- [13] Saggaf M M, Nebrija E L 2003 Estimation of missing logs by regularized neural networks *AAPG Bulletin* 87(8) 1377-89
- [14] Тененёв В А, Якимович Б А, Сенилов М А, Паклин Н Б 2002 Интеллектуальные системы интерпретации геофизических исследований скважин *Искусственный интеллект* 3 338
- [15] Алёшин С П, Ляхов А Л 2011 Нейросетевая оценка минерально-сырьевой базы региона по данным геофизического мониторинг *Нові технології* 1(31) 39-43
- [16] Карпенко А Н, Булмасов О В 2015 Применение нейросетевых технологий при интерпретации данных геофизических исследований скважин. <http://oil-gas.platinov-s.com/index.php?name=articles&op=view&id=11&pag=3&num=1>
- [17] Rogers S J, Chen H C, Kopaska-Merkel D C, Fang J H 1995 Predicting permeability from porosity using artificial neural networks *AAPG Bulletin* 786-1797
- [18] Костиков Д В 2007 Инструментальные средства интерпретации геофизических исследований скважин на основе преобразованных каротажных диаграмм с помощью многослойной нейронной сети *Диссертация кандидата технических наук* 189
- [19] Muhamedyev R, Amirgaliev E, Iskakov S, Kuchin Y, Muhamedyeva E 2014 Integration of Results of Recognition Algorithms at the Uranium Deposits *Journal of ACSI* 18(3) 347-52
- [20] Амиргалиев Е Н, Исаков С Х, Кучин Я В, Мухамедиев Р И 2013 Интеграция алгоритмов распознавания литологических типов *Проблемы информатики* 4(21) 11-20
- [21] Amirgaliev E, Iskakov S, Kuchin Y, Muhamedyev R 2013 Machine-learning techniques in the pattern recognition of rock at uranium deposits *Proceedings of National Academy of Sciences of Kazakhstan* 3 82-8
- [22] Joseph A C, David S 2006 Wishart Applications of Machine Learning in Cancer Prediction and Prognosis *Cancer Informatics* 2 59-77
- [23] Shoeb A H, John V 2010 Gutttag Application of machine learning to epileptic seizure detection *Proceedings of the 27th International Conference on Machine Learning* 975-82
- [24] Mannini A, Angelo M S 2010 Machine learning methods for classifying human physical activity from on-body accelerometers *Sensors* 2 1154-75
- [25] Ballester P J, Mitchell J B 2010 A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking *Bioinformatics* 9 1169-75
- [26] Farrar C R, Worden K 2012 *Structural health monitoring: a machine learning perspective* 66
- [27] Recknagel F 2001 Application Of machine Learning To Ecological Modelling *Ecological Modelling* 303-10
- [28] Charles C, Hecker J, Stuntebeck E, Shea T O 2007 Applications of machine learning to cognitive radio networks *Wireless Communications IEEE* 4 47-52
- [29] Ball N M, Brunner R J 2010 Data mining and machine learning in astronomy *International Journal of Modern Physics* 7 1049-1106
- [30] Szepesvari C 2009 Algorithms for Reinforcement Learning Synthesis Lectures on Artificial Intelligence and Machine Learning series by Morgan & Claypool Publishers 98 <http://www.ualberta.ca/~szepesva/papers/RLAlgsInMDPs.pdf>
- [31] Zhu X 2008 Semi-Supervised Learning Literature Survey Computer Sciences http://pages.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf
- [32] Kohonen T 1982 Self-Organized Formation of Topologically Correct Feature Maps *Biological Cybernetics* 43(1) 59-69
- [33] Jain A K, Murty M N, Flynn P J 1999 Data Clustering: A Review *ACM Computing Surveys* 31(3)
- [34] Barbakh W A, Wu Y, Fyfe C 2009 Review of Clustering Algorithms. Non-Standard Parameter Adaptation for Exploratory *Data Analysis Studies in Computational Intelligence* 249 7-28
- [35] Ayodele T O 2010 Types of Machine Learning Algorithms *New Advances in Machine Learning* 19-48
- [36] Hamza A, Hamza I 2012 Taxonomy of Machine Learning Algorithms to classify realtime Interactive applications *International Journal of Computer Networks and Wireless Communications (IJCNWC)* 2(1) 69-73
- [37] Дьяконов А Г 2010 *Анализ данных, обучение по прецедентам, логические игры, системы WEKA, RapidMiner и MatLab (Практикум на ЭВМ кафедры математических методов прогнозирования)* 277
- [38] Möller M F A scaled conjugate gradient algorithm for fast supervised learning *Neural Networks* 6(4) 525-33
- [39] Dong C, Liu, J 1989 Nocedal On the limited memory BFGS method for large scale optimization *Mathematical Programming* 45(1-3) 503-28
- [40] McCulloch W S, Pitts W 1943 A logical calculus of the ideas immanent in nervous activity *The bulletin of mathematical biophysics* 5(4) 115-33
- [41] Rosenblatt F 2015 The perceptron: A probabilistic model for information storage and organization in the brain *Psychological Review* 65(6) 386-408 <http://dx.doi.org/10.1037/h0042519/>
- [42] Minsky M 1987 Seymour Papert *Perceptrons expanded edition* 308
- [43] Минский М 1971 Пейперт С *Перцептроны* 263
- [44] Minsky M 1987 Seymour Papert *Perceptrons expanded edition* 308
- [45] Werbos P 1974 Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences
- [46] Werbos P J 1988 Backpropagation: past and future *IEEE International Conference on Neural Networks* 1 343-53
- [47] Saad D 2009 1998 Introduction. On-Line Learning in Neural Networks *Cambridge University Press* 3-8
- [48] Cybenko G 1989 Approximation by superpositions of a sigmoidal function *Mathematics of Control Signals and Systems* 2(4) 304-14
- [49] Hornik K 1989 Multilayer feedforward networks are universal approximators *Neural Networks* 2 359-66
- [50] Галушкин А И 2006 Решение задач в нейросетевом логическом базисе *Нейрокомпьютеры: разработка, применение* 2 49-71
- [51] Галушкин А И 2010 Нейронные сети: основы теории *Горячая линия*
- [52] Ясницкий Л Н 2008 *Введение в искусственный интеллект: Уч.пос.для вузов* 176
- [53] *Нейрокомпьютеры: Учеб. Пособие для вузов* 2004 320
- [54] Dudani S A 1976 The distance-weighted k-nearest-neighbor rule

- Systems, Man and Cybernetics, IEEE Transactions* 4 325-7
- [55] K-nearest neighbor algorithm Support vector machine. http://en.wikipedia.org/wiki/Support_vector_machine 22.02.2012 http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm/ 5 Jun 2012.
- [56] Support vector machine. http://en.wikipedia.org/wiki/Support_vector_machine 22.02.2012 17 Mar 2013
- [57] Linear discriminant analysis. http://en.wikipedia.org/wiki/Linear_discriminant_analysis/ Nov2012
- [58] Dudoit S, Fridlyand J, Terence P 2002 Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data *Journal of the American Statistical Association* 97(457) 77-87
- [59] Миркес Е М 1998 *Нейрокомпьютер. Проект стандарта* <http://pca.narod.ru/MirkesNeurocomputer.htm> 19 Jan 2015.
- [60] Bottou L 2010 Large-Scale Machine Learning with Stochastic Gradient Descent *Proceedings of COMPSTAT'2010* 177-86
- [61] Bottou L 1998 Online learning and stochastic approximation. *On-Line Learning in Neural Networks Cambridge University Press* 9-43
- [62] Chu C-T 2006 Map-Reduce for Machine learning on multicore. *Advances in Neural Information Processing Systems Proceedings of the 2006 Conference* 281-310
- [63] Leskovec J, Rajaraman A, Ullman J D 2014 Mining of Massive Datasets *Cambridge University Press* 476
- [64] Amol G, Krishnamurthy R, Pednault E, Reinwald B, Sindhwani V, Tatikonda S, Tian Y, Vaithyanathan S 2011 SystemML Declarative machine learning on MapReduce *Data Engineering (ICDE) IEEE 27th International Conference* 231-42
- [65] Tim K, Talwalkar A, Duchi J C, Griffith R, Franklin M J, d Jordan M I 2013 MLbase: A Distributed Machine-learning System *Conference on Innovative Dta Systems Research (CIDR)* 7-9

Author



Ravil I Muhamedyev, Kazakhstan, Almaty

Current position, grades: Head of Department CSSE&T, professor, International IT University, Almaty, Kazakhstan

University studies: International IT University, Almaty, Kazakhstan

Scientific interest: stochastic simulation, machine learning, simulation of anisochronous systems

Publications: 160 papers, 5 books

Modeling of the subsystem of estimation of navigational parameters in automatic vehicle control systems

A Mrochko

ISMA University, 1 Lomonosova Str., Bld 6, LV-1019, Riga, Latvia

*Corresponding author's e-mail: aleksandrs.mrocko@isma.lv

Abstract

One of the most actively developing spheres of applying modern information technologies is transport. Divisions of different services, departments and organisations are actively introducing and employing the system of *Automatic Vehicle Location* (AVL). The information systems of Automatic Vehicle Location solve the task of controlling and guiding transport means. Employing modern telecommunication technologies along with satellite navigation systems (SNS) facilitates and improves controlling of the mobile objects (MO). Modelling of the work and analysis of these systems' efficiency indicators sufficiently reduces the periods and costs of their testing and introducing in a particular region.

Keywords: information technologies satellite navigation systems mobile objects controlling AVL systems

1 Introduction

Active development of transport systems all over the world and a great increase of the variety of the provided services have led to the formation of an applied complex area of transport-dispatch information technologies, which basics are the following:

- satellite navigation systems that measure the main navigation parameters of the MO (coordinates, speed and direction of movement);
- modern telecommunication systems that transmit the necessary information to the dispatch center and other traffic participants;
- cartographic and special software solving the problem of accumulation, conversion, storage and submission of information on board the mobile object and dispatch center;
- onboard sensors of information and information mapping equipment.

The main consumers of these services are air, road, rail and marine transport means. The system of monitoring of mobile objects in real time improves the efficiency of cargo and passenger transportation, as well as ensures the safety of transport means, cargo, passengers and crew.

The efficiency of employing transport means greatly depends on the efficiency of their informational support. All this explains the increased number of papers considering the methods of estimating the indicators of the efficiency of Automatic Vehicle Location Systems. The main indicators of the efficiency of the mobile objects surveillance systems are characteristics of accuracy and reliability (integrity, availability and continuity of service). The increased requirements to these systems' characteristics are achieved by means of employing, particularly, the technologies of global satellite navigation systems (SNS). Currently we are watching the accumulation of the world experience in applying global radio navigation systems (GPS and GLONASS) and technologies based on these systems.

At the end of the XX century began testing and introducing the above systems in different regions of the world. The aim of the research was to provide accurate positioning information for all transport means located in the system working zone independently from the meteorological and topographic conditions and with the required rate of the data renewal. Some experience in building and testing such systems was

received in realizing the CARD (CNS Applications Research and Development) project. The main research performed in the frame of this project is devoted to evaluating the characteristics of the data transmission line and developing the methods of estimating the efficiency indicators.

The first step in developing the methods of estimating the efficiency indicators should apparently be the analysis of errors of navigation – time determinations (NTD) of the SNS.

2 Analysis of errors of navigation – time determinations

The navigation task to be solved in the user's equipment (UE) SNS, in its simplest case, lies in defining space – time coordinates $P(t) = [x; y; z; W]^T$. In the latest samples of UE there is adopted a two stage procedure of processing information. At the stage of primary procession they perform those measurements of navigation parameters (distance - D , speed of distance change - \dot{D} , etc.), which are only functionally connected with the state vector $P(t)$. At the stage of secondary procession the received parameters are subjected to transformation based on navigation algorithms with the purpose of calculating vector $P(t)$.

The accuracy of determining by the SNS user the location coordinates ($x; y; z$), the speed (W) and other parameters is influenced by many factors. They are connected with the peculiarities of primary and secondary navigation measurements, with the characteristics of the used signals and the media of propagation. To facilitate the influence of different factors on the NTD quality at the primary stage of processing they introduce UERE (User Equivalent Range Error) and speed of its change UERRE (User Equivalent Range Rate Error), conditioned by the non-correlated constituents of measurement errors. Secondary navigation definitions are easily characterized with help of geometric factors indicated as various DOP (Dilution of Precision).

Consider the main sources of these measurements' errors in application to the adopted in GPS and GLONASS long distance method NTD. The expression of the measured distance to the i -satellite D_i in this case will look as follows:

$$D_i = D_{0i} + \delta D_{NS} + \delta D_{RL} + \delta D_{UE},$$

where D_{0i} – true value of the distance to the i -satellite; δD_{NS} – errors introduced in navigation satellites (NS) and control measuring set (CMS); δD_{RL} – errors introduced in the radio line "NS - user"; δD_{UE} – errors introduced by UE SNS.

Errors introduced in NS and CMS (δD_{NS}) are conditioned mainly by the insufficient frequency – temporal and ephemeris support of NS.

• **Errors of the frequency – temporal support** are caused mainly by insufficient procedures of verifying and storing the board time scale (board clock – BC) of NS. For typical caesium board frequency patterns the given errors between the correction moments may be approximated as follows:[1]

$$\sigma^2(t) = 2,5 \times 10^{-21}(t-t_c) + 5,76 \times 10^{-26}(t-t_c)^2,$$

where t – current time; t_c – time of correction BC.

CMS SNS is correcting the BC in such a way that $\sigma(t)$ of the BC shift would not exceed 10 nc. Besides, in the intervals between the apparatus corrections, the algorithmic correction of the BC of the given satellite is performed in UE. And here the unpredictable deviations of the BC of the given satellite in relation to the Time System Scale may reach 1nc (0.3 m) in one hour interval.

• **Errors of the ephemeris support** are caused by inaccurate definitions of orbit parameters NS in CMS and unpredictable shifts of NS in relation to the extra polar orbit. In SRNS GPS the average quadrant value of the ephemeris constituents UERE makes up about 1m [2].

• **Errors introduced in the radio line «NS – user» (δD_{RL})** are caused by insufficient knowledge of the ways of radio waves propagation in the Earth’s atmosphere (refraction of the satellite signals in the ionosphere and the troposphere).

• **Troposphere errors.**

Judging by the experimental data for GPS additional delays of the NS signal in the troposphere may reach 8...80 nc [2]. For average meteorological conditions (temperature, pressure and air humidity) the value of this delay is defined by the expression:

$$\Delta t_{TR} \approx \frac{K_T}{\sin \beta} \cdot \int_0^{S_T} (n-1)ds,$$

where K_T – parameter characterising the condition of the troposphere; β - angle of the NS place; n – coefficient of the radio waves refraction; S_T – length of the troposphere line sector.

Troposphere models used in SNS allow to reduce the troposphere errors up to nanosecond units. In compensating the troposphere refraction, the periodicity of the user corrections is determined by the speed of the corresponding delays’ change, which in normal circumstances does not exceed 10 m/h.

• **Ionosphere errors.**

The additional delay in the ionosphere Δt_{ion} of signal SNS GPS with frequency – f may be estimated as

$$\Delta t_{ion} = \frac{A}{f^2},$$

where A – the coefficient characterising the features of the propagation media.

The value of this delay changes widely depending on the

Earth region where the mobile object is located, the time of the day, the season of the year, the Solar and the geomagnetic activities, etc., and makes up 5-500 ns [4]. The average value of Δt_{ion} for GPS makes up 5-10ns at night and 30...50ns in daytime for the angles of place β reaching 90° . At $\beta < 15^\circ$ this delay increases by 2-3 times.

• **Errors caused by the pass variety (multi pass).**

These errors are mainly dependent on the mutual location of the satellite, the receiving antenna and the reflecting objects. Experimental research has shown wide range of values of the long distance errors due to the rays variety, which makes up the best 0,5-2m (using special antennas) and up to 100m the worst in urban high buildings conditions. In most unfavourable conditions failure of surveillance can be incurred.

• **Errors introduced by UE SNS (δD_{UE})** are caused by errors in the surveillance at the moment of the satellite signal incoming. Typical error of UE makes up about 1,5-10m – for the standard GPS accuracy code.

The analysis of errors of primary navigation parameters estimation using SNS has shown that their summarising value ($2\sigma_D$) can reach 100 m and exceeds the limits of acceptable values adopted for most applications of AVL systems. Substantial reduction of NTD errors (by up to ten times) can be possible by using the differential SNS work mode. The basis of the differential method is relative stability of the considerable part of the SNS error in time and space. The main slightly varying errors of defining distance in SNS are [1]:

- errors in NS synchronisation;
- errors caused by the faulty ephemeris support of NS;
- non-compensated ionosphere errors.

Errors in NS synchronisation are constant in space and quite stable in the considered temporal intervals. Fluctuation of board clock NS by about 10^{-14} or 10^{-13} in the time of up to 15 min results in distance errors from 3 mm to 3 m.

The error effect of the ephemeris information (ξ_i) can be characterised by the following model:

$$\xi_i < \frac{Ld_i}{D_i},$$

where d_i – error of the i - NS ephemeris (typical d_i value for GPS makes up 10 m); L , km - UE distance from the control point.

Calculations on this model show that space variation of distance measurement errors caused by insufficient ephemeris information is not substantial and at $d_i = 10$ m and $L < 200$ km does not exceed 10 sm and $L < 1000$ km - $\xi_i < 50$ sm. It is worth noting that with the data about the satellites constellation “aging” the errors of the ephemeris (d_i) also increase and, therefore, ξ_i increases.

Variation of ionosphere errors in time and space are characterised by the correlation function, which has times and space correlation radiuses at the level corresponding to several minute and thousands kilometres [4]. There is some of the experimental data of the temporal fluctuations of distance (D_i) errors caused by ionosphere [1], for example, in 1 min variation made up 0,1-0,2 m (σ), and in 6 min – 0,3-1,4m.

Differential work mode of SNS makes it possible to define and compensate the above errors. At the same time the main sources of errors in evaluating distances are noise constituents of UE, which in measuring curving delays make up metres and in measuring carrying delays –

centimetres and even millimetres. Another feasible source of errors is the property of multi pass.

• *Geometric factor in SNS.* Calculating the user's space – temporal coordinates is performed at the second stage of processing the NS signals. The ratio between the vector of errors in defining space – time coordinates $P(t)$ $\delta_M = |\delta_x \delta_y \delta_z \delta_D|$ and the vector of errors in measuring distances $\delta_M = |\delta_{D1} \delta_{D2} \delta_{D3} \delta_{D4}|^T$ depends on the geometry of the corresponding location of NS and the user. Due to some special peculiarities of the NS and the user's space locations, the measure of decreasing the accuracy of navigational definitions in SRNS is the geometric coefficient GDOP (Geometric Dilution of Precision). The most important characteristic of SNS is the precision of place location, therefore for a surface mobile object the Horizontal Dilution of Precision (HDOP) is more often used

$$HDOP = \frac{\sqrt{\sigma_x^2 + \sigma_y^2}}{\sigma_D}$$

The orbital configuration characteristics of satellites GPS provide, with the probability of 0.999 plus, the field of vision in a global working zone in any 24 hour interval of four satellites plus, the average value of HDOP making up 1.5 [2]. Increasing the number of visible satellites makes it possible to achieve a good gain in the accuracy of evaluating navigational parameters.

3 Reliability of navigation – time determinations

Besides accuracy properties of SRNS we should also regard reliability indicators of NTD as the indicators of efficiency of SRNS functioning. The main characteristic of reliability – system integrity – is defined as the ability to detect inadmissible system performance deterioration with the preset probability and time lag of informing the users thereabout [3].

The analysis of the factors, which influence the SNS integrity, makes it possible to divide them into two categories.

To the easily detected failures, there belong the following:

- signal fading from the NS;
- the distorted structure of the signal, which does not allow the user to come in synchronization with the satellite;
- presence of the sign in the navigation message of NS, which prohibits using its navigation information.

The user detects such situations without any additional equipment and without additional calculations.

To the difficultly detected failures, there belong:

- unpredictable shift of the board time scale (board clock) of the navigation satellite
- drift of frequency of the satellite reference generator;
- drift of the carrier frequency of the signal transmitted by the satellite;
- drift of the NS from the orbit;
- incorrect ephemerid information.

Such failures of navigation satellites lead to the errors of navigation-time determinations. Therefore, the problem of integrity control and development of algorithms and methods of detecting the above failures are of great practical interest. The analysis of SNS errors is the basis for the optimizing

methods of estimating the efficiency of the AVL systems.

4 Methods of estimating the radio navigation parameters (RNP)

The maximum exactness of determining the location of the MO can be achieved by a ranging measuring method or by a range-difference measuring one. The navigation parameters (NP) in this case are, correspondingly, either distance **D** or the difference between distances ΔD . Their values are received on the basis of estimating the vector of the radio navigation parameters $\vec{\lambda} = \{\lambda_U, \lambda_\psi, \lambda_\omega, \lambda_\tau\}$ of the received signal $S(t, \vec{\lambda})$

$$S(t, \vec{\lambda}) = U(t - \tau) \cdot \cos[\varpi \cdot t + \psi(t)]$$

The elements of the parameter $\vec{\lambda}$ may be amplitude $U(t)$, phase $\psi(t)$, frequency ω and the time of delay of signal τ .

Under real conditions of receiving signal $S(t, \vec{\lambda})$ the estimation of parameter $\vec{\lambda}$ is performed against the background of noises $n(t)$ and disturbances. The value of the radio navigation signal here is not known beforehand and changes randomly due to the noises.

Task setting. The receiving equipment of the radio navigation system (RNS) receives an additive mixture of a signal and a fluctuation noise

$$\xi(t) = S(t, \vec{\lambda}) + n(t),$$

where $\vec{\lambda}$ - the vector of the radio navigation parameters subjected to estimation; $n(t)$ – fluctuation white noise with the parameters $M[n(t)] = 0, k(\tau) = N_0 \delta(\tau)/2$.

It is suggested that parameters $\vec{\lambda}$ at the interval of observance $[0, T]$ are constant and the priori density of possibilities may be either unknown, or known partially, or known completely. It is needed, by the accepted realization of fluctuation $\xi(t)$, to estimate optimally at the interval of observance $[0, T]$ the values of parameters λ_α , where $\alpha = U, \psi, \omega, \tau$.

As the result of solving the task there should be received the algorithms and the structural scheme of the optimal parameters' estimation and calculated their exactness characteristics.

The optimal estimation of parameters $\vec{\lambda}^*$ depends on the chosen function of losses $Q(\vec{\lambda}, \vec{\lambda}^*)$, which form is determined by the physical essence of the task [5, 6]. While estimating the radio navigation parameters we most often use the quadric function of losses

$$Q(\vec{\lambda}, \vec{\lambda}^*) = (\vec{\lambda} - \vec{\lambda}^*)^T \cdot (\vec{\lambda} - \vec{\lambda}^*)$$

Among the most frequently used methods for estimating NP, we should point out: the Bayesian method, the maximum Likelihood method and the method of Least squares.

Bayesian method. NP estimation is performed on the basis of the posteriori probability density with priori information about the values of the parameters estimated

$$W_{PS}(\vec{\lambda}) = c \cdot L(\vec{\lambda}) \cdot W_{PR}(\vec{\lambda}),$$

where $W_{PS}(\vec{\lambda})$, $W_{PR}(\vec{\lambda})$ - are correspondingly the posteriori and priori densities of possibilities of the parameters estimated; $L(\vec{\lambda})$ - likelihood function; c - constant coefficient. For the quadratic loss function, estimation by the given method corresponds to the mathematical expectation of the posteriori distribution [5]

$$\vec{\lambda}^* = \int_{-\infty}^{\infty} \vec{\lambda} \cdot W_{PS}(\vec{\lambda}) d\vec{\lambda}.$$

Maximum Likelihood method. It is used to receive estimation $\vec{\lambda}^*$ without priori information about the characteristics of the parameters estimated. The optimal estimation $\vec{\lambda}^* = \vec{\lambda}_{opt}^*$ is the root of the equation

$$\frac{d \ln L(\vec{\lambda})}{d\vec{\lambda}} = 0$$

The above method of estimating the navigation parameter $\vec{\lambda}$ allows us getting only optimal algorithms of processing the received fluctuation $\xi(t)$. It does not give quantitative characteristics of the algorithm's estimation exactness. For their determination we can use the method of consecutive approaches along the smaller parameter or Cramer-Rao Inequality [5]. For the smaller parameter we use the relation signal/noise.

Method of Least squares. When the radio navigation parameter $\vec{\lambda}$ is the linear function of the measurements $\xi(t)$ we can use for its estimation the method of least squares. We can use it, for example, for estimating the amplitude of the signal and the time of its delay at the impulse method of measurements of the navigation parameter.

The analytical record of the fluctuation $\xi(t)$ can be presented in the form

$$\xi(t) = S\vec{\lambda} + \vec{V}, \tag{1}$$

where $\xi(t)$ - m -measure vector of measurements; $\vec{\lambda}$ - n -measure vector of the navigation parameters ($m \geq n$); S - the known matrix of observances of size $m \times n$; \vec{V} - vector of the noises of measurements.

In the equation (1) the error of measurements \vec{V} has zero mathematical expectation $M\{\vec{V}\} = 0$ and the covariance matrix $V = M\{\vec{V}, \vec{V}^T\}$. It is suggested that vectors \vec{V} and $\vec{\lambda}$ are non-correlated $M\{\vec{\lambda}, \vec{V}\} = 0$, other priori information on the parameter $\vec{\lambda}$ and the process $\xi(t)$ lacking.

According to the considered method, we need to determine such value of the estimation $\vec{\lambda}^*$ at which the sum of the squares of faults $\vec{\xi} = \xi - S\vec{\lambda}^*$ reaches its maximum

$$\|\vec{\xi}^2\| = \vec{\xi}^T \vec{\xi} = (\xi - S\vec{\lambda}^*)^T (\xi - S\vec{\lambda}^*) \rightarrow \min. \tag{2}$$

Leveling to zero the quotient derivatives on the parameter $\vec{\lambda}$ in the expression (2) we get the optimal estimation of the parameter

$$\vec{\lambda}^* = (S^T S)^{-1} S^T \vec{\xi} = p S^T \vec{\xi},$$

where $p = (S^T S)^{-1}$.

Comparing the above methods by their exactness shows that for larger relations the estimations signal/noise received by the maximum likelihood method and the Bayesian method appear to be pretty close [5]. Though, the maximum likelihood method is more preferable for estimating HP in the case of no priori information regarding the estimated parameters. It is caused by its relative simplicity and asymptotically efficient estimation at increasing the signal/noise ratio (S/N ratio).

It should be noted that during long time of observances the distribution of the $\vec{\lambda}^*$ assessment is normal. In this case, the maximum likelihood method and the method of least squares are identical and dispose of the least fault of estimation determined by the lower border of *Cramer-Rao inequality*.

5 Estimation of the radio impulse radio navigation parameters

On the basis of the maximum likelihood method, we get the estimation of phase (ϕ^*) and the temporary state of radio impulse (τ^*)

$$S(t, \tau) = A_0 \cdot g(t - \tau) \cdot \cos(\omega t + \phi).$$

The likelihood equation from which we deduct the algorithm of performance of the optimal phase measurer has the form [5]:

$$\frac{d}{d\phi} \ln F(\phi) = \frac{d}{d\phi} \left[\frac{2}{N} \int_0^T \xi(t) \cdot U(t - \tau) \cdot \cos(\omega t + \phi) dt \right]_{\phi^*} = 0, \tag{3}$$

where $F(\phi)$ - functional of the likelihood, N - noise spectral density.

From (3) there goes

$$\int_0^T \xi(t) \cdot \sin(\omega t + \phi^*) dt = 0. \tag{4}$$

The expression (4) is approximately modeled by the system phase automatic frequency.

The variance phase estimation, obtained by the method of successive approximations in the small parameter, equals

$$\sigma_{\phi^*}^2 = \frac{N}{2E},$$

where E - the signal energy.

Analogically, we can get the estimation of the temporary state of a known form impulse $U(t - \tau) = A_0 \cdot g(t - \tau)$ without the inter-impulse phase or frequency modulation. The likelihood functional for estimating parameter τ is

recorded in the form:

$$F(\tau) = \exp \left\{ \frac{2A_0}{N} \int_0^T \xi(t-\tau) \cdot g(t-\tau) dt \right\}.$$

An equation for estimating τ^* as

$$\int_0^T \xi(t) dt = \frac{1}{\beta^2 \cdot \frac{2E}{N}}.$$

The variance estimation of the temporal position is expressed as:

$$\sigma_{\tau^*}^2 = \frac{1}{\beta^2 \cdot \frac{2E}{N}},$$

where β^2 - value that specifies the width of the spectrum of the envelope radio impulse.

For the real rectangular pulse limited by spectrum band Δf , the value $\beta = \frac{2\Delta f}{t_{imp}}$, and the variance parameter τ^* estimation

$$\sigma_{\tau^*}^2 = \frac{t_{imp}}{2\Delta f \cdot \frac{2E}{N}}. \quad (5)$$

In [5, 6] we get the expressions for the dispersion of the estimation of the temporary state of different form radio impulses. The analysis shows that at the fixed energy of the signal the exactness of estimation of the temporary state of the radio impulse increases with the increase of the signal specter width and with the decrease of the impulse length.

Also, the exactness of the estimation of the radio impulse temporary state can be increased by applying the inter-impulse modulation, for example, by pseudo random succession. In this case, if the energy of signal E remains constant, the width of the radio impulse specter increases and the dispersion of the estimation decreases correspondingly (5).

On the other hand, the exactness of estimation of the radio impulse temporary state can be increased by multiple repletion of measuring parameter τ and further processing of the obtained selection by one of the above estimation methods. Multiple measurements can be obtained by applying the mode of the space recirculation of the signal.

References

- [1] *Спутниковые радионавигационные системы. Выпуск 1, 2* Под ред. М. С. Ярлыкова. – М: Издательство «Радиотехника» 2013
- [2] Understanding GPS: principles and application Elliotte Kaplan editor Artech House Publish Ers 1996
- [3] Durand J M, Michol N, Bouchard J 1990 GPS Availability part1 (11) Availability of Service Achievable for Different Categories of Civil

Optimal estimation of the distance by using the method of least squares on selecting volume κ is given by the expression (6)

$$D_k^* = p S^T \vec{D}_k, \quad (6)$$

where $p = (S^T S)^{-1}$.

With the account of the fact that the distance measurements are scale ones and the size of the matrix is $S^T \kappa \times 1$ estimation of the distance by the method of least squares coincides with the selection average

$$D_k^* = \frac{1}{k} \sum_{i=1}^k D_i$$

and is asymptotically efficient with the increase of the number of recirculation cycles.

Errors of distance estimation at the κ cycles of recirculation are determined by the covariant matrix

$$V_K = (S^T S)^{-1} S^T V S (S^T S)^{-1},$$

where V – covariant matrix of the measurement noises ΔD_i .

In case of scale measurement

$$V_K = \sigma_{D_k^*}^2 = \frac{\sigma_{\Delta D}^2}{k}. \quad (7)$$

Thus, the variance of the distance estimation decreases in reverse proportionality to the number of recirculation cycles. The lower border of the estimation variance is determined by the *Cramer-Rao inequality* [5, 6] and coincides, in this case, with the value (A.7).

It should be noted that the increase of the number of recirculation results in the decrease of the fluctuation measurement error, though it is accompanied with the increase of the dynamic error since the interval of averaging is increased. The analysis of the modeling results has confirmed our supposition about the existing optimal number of recirculations at which the errors of estimating the location of the MO would be minimal. Since the nature of the fluctuation and dynamic errors in measuring distance is different they can be considered independent and non-correlated ones. It gives possibility of analyzing the ways of decreasing the constituent errors independently for each other.

Analysis of sources of errors and estimation methods of navigation parameters sufficiently reduces the periods and costs of testing and introducing the *automatic vehicle control* systems in a particular region.

Author



Alexander Mrochko, 1957, Kiev, Ukraine

Current position, grades: Dr. Sc. Ing., Ass.prof

University studies: Information Systems Management Institute

Scientific interest:

Publications:

Experience:

Users Navigation 37 2(3)

[4] Черенкова Е Л, Чернышев О В 1984 Распространение радиоволн *Радио и связь*

[5] Тихонов В И 1982 Статистическая радиотехника *Радио и связь* 624

[6] Тихонов В И, Харисов В Н 1991 Статистический анализ и синтез радиотехнических устройств и систем *Радио и Связь*

Optimisation of coordination's selection by innovation and investment projects

Ilana Ter-Saakova*

Baltic International Academy, 4 Lomonosova Str., LV-1019, Riga, Latvia

**Corresponding author's e-mail: ilana@alida.lv*

Abstract

The steady onward progress of engineering developments, together with increasing competition, implies the need for the development of implementation for novelties and innovations. This involves a huge number of innovative projects emerging, thus consecutively defining the formation of investment criteria

Keywords: innovations investments project intensity probability model function

1 Introduction

The increasing competition in the goods and services market and the ongoing development of facilities require continuous improvement of the development and integration of innovations in various fields of science and technology. This results in the formation of a large number of innovative projects that cause, in turn, the need for appropriate funding, and therefore require the formation of investment projects.

This is especially evident in the preparation of investment plans in various industries intended for the development of a particular region. In this case, the investment appropriation requests exceed the financial capacity almost without exception.

2 Research

The ongoing development and the need for innovation development and integration set the problem of the comprehensive selection of innovative projects and their adequate funding at the expense of potential investment projects. In order to solve it, let us consider the following setting set up of the problem.

A set of innovative projects is formed {M}, each with sufficient definiteness of the result and measurability of their characteristics.

There is also a set of investment projects {N} (M C N) based on the above-mentioned investment projects.

There are time constraints on the implementation of innovation and investment projects, as well as cost and resource constraints on the project financing for the entire period and within its individual intervals.

One needs to select the best project versions, order them by preference in accordance with the given resource constraints and adopted criteria.

Let us introduce the following notation

X_{ik} is the i -th innovation project with k -th execution intensity;

Y_{lj} is the l -th investment project with j -th intensity version;

$X_{ik} \in (X); Y_{lj} \in (Y); (X)$ and (Y) are a set of Boolean variables.

$X_{ik} = 1$ mean that the i -th innovation project with k -th intensity is selected.

By project intensity we shall mean the project organisation version coordinated with the project executors and characterised by cost, duration, and expected completion.

In the course of project preparation, the probability of execution P_{ik} is possible and as well as non-execution possible $P_{ik} = 1 - P_{ik}$.

Within the project $X_{ik} = 1$ income B^L can be obtained with L -th efficiency criterion, where $L=1, A$ is a possible number of criteria.

Losses B^{-L} within the project are also possible.

In view of this notation, the problem of the coordinated selection of innovation and investment projects can be presented as a mathematical optimisation problem.

It is necessary to find the following:

$$\sum_{i=1}^L \sum_{k=1}^K (P_{ik} B^L X_{ik} Y_{in} - (1 - P_{ik}) B^{-L} X_{ik} Y_{in}) \rightarrow \max . \quad (1)$$

Since the cost and time resources are constrained, the equation (1) should be solved under the following constraints:

$$\sum_{i=1}^L \sum_{k=1}^K (C_{ik} X_{ik} + C_{in} Y_{in}) \leq C_{\max} , \quad (2)$$

$$\sum_{i=1}^L \sum_{k=1}^K (T_{ik} X_{ik} + T_{in} Y_{in}) \leq T_{\max} . \quad (3)$$

The duration of a particular i -th project can be divided into a number of discrete intervals, and as a rule can be equal to a month, a quarter or a year. In the case of the definition of each project by the univariate model network, the project parametric analysis procedure can be used for the given financial resources C and time T in a certain range.

When selecting projects we shall recognise that the result of the i -th innovative project must be equal to a similar investment project in content and characteristics.

The well-known approaches to solving vector optimisation problems can be used to find the optimal solution of the expression (1) on the selection of the best set of projects. The well-known approaches to solving vector optimisation problems can be used for the projects. Boolean integer programming can be used to find the numerical solution. Upon finding the solution, it is required to determine an optimal subset of projects to be developed, the start and end dates of

each of them, as well as the best version of the set of projects being formed in terms of the general optimality test.

The cost and the reliability of the project are functions of the project time. Determining the correlations between these parameters is an actual scientific and practical problem, since, as is known, the cost (C) and the duration (T) of the project are an important subject in the coordination of interests between the customer and the project developer under the relevant requirements on the project reliability.

Every project's range of characteristic parameters values (k range members) embodies its final result properties, its cost value C_k at k^{th} intensity variance and duration T_k , as its implementation time.

Being aware that, under a market economy, both the project cost (C) and time (T) are subject for agreement between the client and the contractor, immediately after the concluding of the contract, the project task supervisor proceeds to the elaboration of project implementation network model, seeking to provide the operative management of the works for the whole implementation period, accounting for possible interfering impacts that, in general, bear a random character. In the case the initiated project is defined with a univariate network model, each project work stage correlates to a monotonous decreasing time-dependent cost function $C = C(T)$, whose type and the existence interval $[T_{min}, T_{max}]$ can be found. Under the precondition of an a-priori accurate and reliable project implementation assessment, the dependence $C(T)$ forms a convex function (Figure 1).

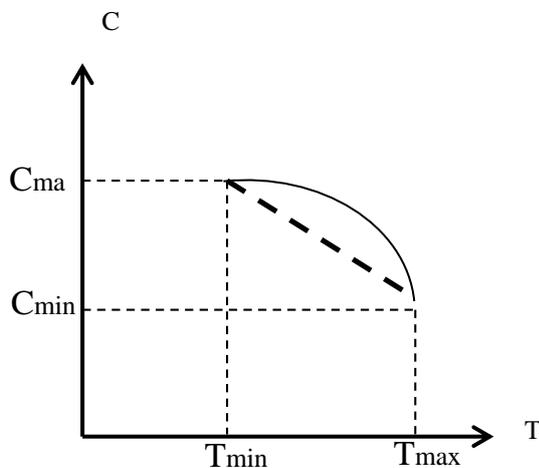


FIGURE 1 Time-dependent cost function $C = C(T)$

Let us consider the generalized reliable project implementation assessment characteristic P , i.e. the probability of the project's successful completion within scheduled delays as a T -dependent function. According to the leading experts in the field, this characteristic is revealed as a monotonous increasing one, while the T -dependent, has a concave shape (Figure 2); here, the project reliable implementation parameter slightly decreases when there exists an insignificant difference between T and T_{max} , and, when the T value is further diminished, the P parameter goes down more abruptly.

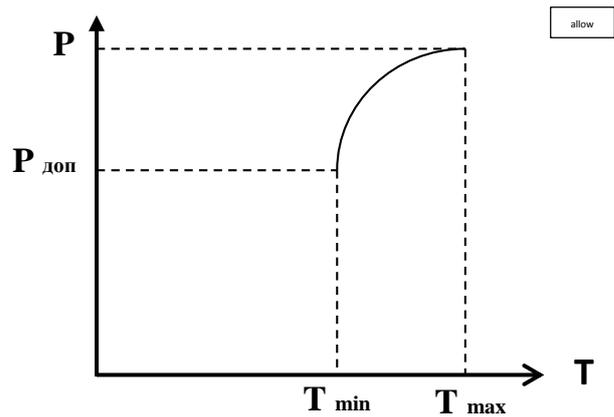


FIGURE 2 Time-dependent project reliable implementation assessment characteristic

On one hand, the T_{max} is determined with the maximum allowed (Client side) project development delay and, on the another, with sufficiently high project implementation reliability degree characteristic for the specified allowed delay (as assessed by the Contractor's experts).

The T_{min} , value is determined with the criterion that the probability of the project's successful implementation for every $T \in [T_{min}, T_{max}]$ should never be less that the limiting allowed value $P_{allow} < P(T_{min})$ and, consecutively $P = P(T) > P_{allow}$ within interval $T \in [T_{min}, T_{max}]$.

Determining the T_{min} value, other considerations include: availability of client finances, degree of the client's motivation to speed up receiving the results of the final project.

Taking into account the dependence of $P = P(T)$, we can determine $M(C)$ mathematical expectation as the project cost function $f(T)$:

$$f(T) = C(T) \times P(T). \tag{4}$$

To proceed with the first approximation, according to Figure 1, we choose a linear dependency:

$$C = C(T) = -KT + R, \tag{5}$$

where

$$K > 0, R > 0, (-KT + R) > 0.$$

And, examining Figure 2, we observe $P'(T) > 0$ and $P''(T) < 0$, which means that the $P(T)$ function is positive, monotonously increasing and convex.

Upon differentiation of the project cost mathematical expectation

$$f(T) = C(T) \times P(T) \text{ we obtain:}$$

$$f'(T) = (-KT + R) \times P'(T) - KP'(T). \tag{6}$$

The necessary precondition satisfying the $f(T)$ monotonously decreases within all intervals $[T_{min}, T_{max}]$ refers to the $f'(T) < 0$, criterion, that is:

$$(-KT + R) \times P'(T) < KP'(T), T \in [T_{min}, T_{max}]. \tag{7}$$

Consecutively, both K and P assigned values should satisfy conditions (7), either definition of region T should be

positioned to the right of the maximum of the $f(T)$ function.

The $f(T)$ function is convex regardless of which K and P values are assigned, as always, the conditions are satisfied above, and this will be $f'' = (-KT + R) * P''(T) - 2KP'(T) < 0$. Respectively, the time-dependent cost function will appear as shown in Figure 1.

During the discussion and definition of the project's implementation conditions, we must consider the variety of variance from the viewpoint of the final result criteria. The implementation conditions are established with the final product parameters (features) limit values. It is certain that these features should reflect not only the technical, but also the economical characteristics.

The provided interval of whole values of every parameter variance, range within relevant limits, and is divided into non-crossing subpopulations (either subpopulations with essentially insignificant crossing areas) arranged by the increase in the level of consumer-demanded parameter quality, we can find the L index represents the number of project implementation variances, every one individually complying with the integrity of requirements to the final result. These criteria for innovative projects can be formulated as parameter values of the final scientific technical product required to satisfy, within the evolved subpopulation limits in respect of every above-considered characteristic.

Every project implementation variance correlates, as aforementioned, with the values: $(T_{max} C_{min})$, $(T_{min} C_{max})$, and $C = C(T)$ function, found within these intervals.

In such a way, the target project can correlate with the family $\{C = C(T)\}$

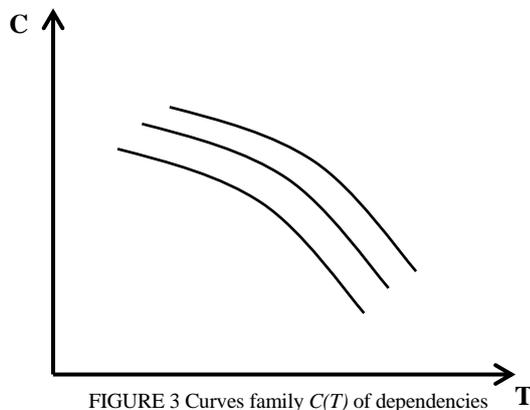


FIGURE 3 Curves family $C(T)$ of dependencies

Both the innovative project cost and its implementation delay (beginning from the scientific research and experimental design development), are significantly influenced by the decisions taken at the earliest stages of development. Several foreign experts estimate the situation as follows: as early as the completion of the project draft, 75% of such new project costs are defined, nevertheless namely this specific stage costs amount nearly 5% of the project's total cost. Thus, at the stage of conceptual design embodiment and pre-production prototype manufacturing, the real possibility to reduce the newly designed product cost is shortened to about 20%, and during the serial production stage (general expenses reaching the maximum), this possibility never exceeds 5%.

3 Results

These ciphers evidently show that time and money saving during the early development stage of the innovative project obligatorily involves the resulting innovative product increased cost.

Such an approach serves as a stimulus, not only for the project's accelerated implementation, but also to achieve the highest possible characteristics of the final scientific and engineering product.

In the process of developing the project's contract conditions (requirements to the newly-created engineering product inclusive) agreeing with the client, the contractor can elaborate a univariance (no alternatives) network model of project implementation. The potential contractors for the works establish every work stage and its respective network model parametric evaluation system. Every i^{th} work within such a network model obtains the following characteristics:

\overline{R}_i - material expenses vector;

\overline{Q}_i - labour expenses vector (staff number by the required specializations);

t_i - work implementation time

These assessment parameters are concurrently elaborated by the potential contractors on the basis of the desired reliability level (probability of successful work completion within t_i period), equal to 1 ($P_i = 1$).

Knowing the project's overall cost (labor inclusive) and the $C = C(T)$ function, the contractors are motivated to assign all mentioned network model's parameters criteria in a manner that the increase in money gained by every contractor per time unit is in direct relation to the increase of labor productivity.

The client's interest in reducing the input of material resources is due to the competition between possible contractors departing from the income distribution after expenses are covered.

When input established assessment criteria for resources (for a given contractor's team) defined are the following: the project implementation overall time T , critical works stages, and required time reserves for every j^{th} variance of the network model (Δt_j) and every work (Δt_j).

Provided $T \leq T_{max} - \Delta T$, where ΔT - scheduled time reserve to secure the project implementation reliability reserve, the given team can develop the project with a degree of high reliability with respect to complying with scheduled delays. The $\Delta T > 0$ selection is due to the possible inaccuracy of the contractor's estimations. The ΔT is convenient to be linked with the assessment of the established accuracy of the work's parameters. This accuracy being high, we can assume $\Delta T \approx 0$.

Let us consider the case of $T > T_{max} - \Delta T$. Hypothetically, the probability of P project successful completion and the same for every separate work in its implementation network model (P_i) is directly proportional to the expected labour input or, a fixed team member, equivalent to the work duration. Therefore, we can conclude that, when works can be fulfilled by a fixed team during T_i time with the probability P_i , then it can be performed during T_i / P_i time with the

probability equal to one unit.

In this case, the function $P_i(T_i)$ can be represented as:

$$P_i = \frac{t_i}{t_i^{(1)}}$$

where $t_i^{(1)}$ – work duration when the fulfilment probability is equal to one.

Admitting the model as the time-dependent function $P_i(T_i)$ is related to a specific character and the peculiarities of some considered work stage. In general, such a dependency model can be expressed with the function of the following kind:

$$P_i = \left(\frac{t_i}{t_i^{(1)}} \right)^n \tag{8}$$

where the order index “n” depends on the work duration time correlated to the probability of successful completion of the works.

Provided the works successful completion probability increases slower than its length, i.e. if the function $P_i(T_i)$ is monotonously and convexly increasing, then the order index “n” (8) is $n < 1$.

To illustrate, we expose the example of $P_i(T_i)$ function type:

$$P_i = \left(\frac{t_i}{t_i^{(1)}} \right)^{\frac{1}{2}}$$

When the works successful completion probability increases faster than its length, i.e. if the function $P_i(T_i)$ is monotonously and concavely increasing, then the order index “n” (8) is $n > 1$.

To illustrate, we expose the example of this case:

$$P_i = \left(\frac{t_i}{t_i^{(1)}} \right)^{\frac{3}{2}}$$

In general cases, we should consider “reliability for the duration”, dependencies differentiating them to be the class of works or type, and for the whole project.

The contractors must specify and give grounds to which $P_i(T_i)$ function types the given work corresponds. In other words, every work type should obtain a substantiated value of “n” order index. Provided, according to the contractors, any of the above-proposed dependencies does not represent an adequate description of some work’s character, they are invited to substantiate and propose their own $P_i(T_i)$ function type for agreeing with the client and further feasibility analysis.

We suppose an approximation of the project reliable fulfilment linear dependency onto the duration both for separate works and the whole project. To achieve the project implementation length equal to the maximum allowed value $T_{max} - \Delta T$, it is sufficient to evaluate time for every part of the work for the network model, including the reserves multiplied by the:

$$P = \frac{T_{max} - \Delta T}{T} \tag{9}$$

Now, we compare the t_i and $(t_i + t_i) \times P$. Provided $t_i \leq (t_i + t_i) \times P$, then the fulfillment of these works is also characterized by their completion reliability within the scheduled period, equal to one unit, and the new time reserve shall be found from:

$$t_i(P - 1) + Pt_i \geq 0 \tag{10}$$

Let us find the minimum allowed value P_{allow} for the reliable project implementation by a given team and the same parameter for every separate work within the given network model. Supposing $P > P_{allow}$. Now we define works characterized by the reliability index, equal to one unit. From expression (9) we conclude that the initial time reserves of the works should satisfy the condition:

$$\Delta t_i \geq t_i \left(\frac{1}{P_{allow}} - 1 \right) \tag{11}$$

For example, when $P_{allow} = 0,9$, then

$$t_i \geq 0,11t_i \tag{12}$$

In other words, when some work time reserve is 11% above the assigned expected completion time, their implementation reliability is equal to one.

The works not complying with (12) are characterized by the P_i probability of successful completion within the allowed time $(t_i + t_i) \times P$.

There, from some i^{th} work implementation probability P_i at the assigned linear approximations will be found as:

$$P_i = \frac{t_i + \Delta t_i}{t_i} P \tag{13}$$

For that, the following condition is required to be satisfied:

$$\frac{(t_i + \Delta t_i)P}{t_i} \leq P_i \leq 1 \tag{14}$$

From (13) we conclude that the lesser number is some i^{th} work completion time reserve Δt_i respective to t_i , the lower is the lower P_i limit, reaching the minimum value of P_{allow} for the critical stage works. i.e. when $\Delta t_i \rightarrow 0$, at $P_{i\ min} = P$ and $P_{i\ max} \rightarrow 1$. For example, at $P = P_{allow} = 0,9$ and $\Delta t_i = 0,05t_i$, using (13), we find that $P_i = 0,945$.

$$\Delta t_i > 0,11t_i \tag{15}$$

i.e. for the reserve work time which is 11% above the assigned expected completion time, their implementation reliability is equal to one.

The works not satisfying (15) are characterized by the P_i probability of successful completion within the allowed time

$$(t_i + \Delta t_i) P \tag{16}$$

$$P_i = \frac{(t_i + \Delta t_i)P}{t_i}.$$

And they comply with the condition:

$$\frac{(t_i + \Delta t_i)P}{t_i} \leq P_i \leq 1. \quad (17)$$

From (12) we conclude, that the lesser is Δt_i , respective to t_i , then the lower is the lower limit for P_i , reaching its minimum P_{allow} for the critical stage works, i.e. when $\Delta t_i \rightarrow 0$, this function is linear at the proportionality

coefficient of $\frac{\Delta t_i^t}{t_i}$, at that $P_{i\ min} = P$ and $P_{i\ max} \rightarrow 1$.

For example, at $P = P_{allow} = 0,9$ and $\Delta t_i = 0,05t_i$
 $P_i = 1,005 \times 0,9 = 0,9045$.

References

- [1] Schumpeter J A 1982 *Theory of economic development*
- [2] Kolobov A, Kochetkov V, Omelchenko I 2007 *Knowledge-intensive industries innovative activity economics*
- [3] Nokes S, Major I, Greenwood M, Goodman M 2004 *The Definitive Guide to Project Management: Every executives fast-track to delivering on time and on budget*
- [4] Meredith J R, Mantel S J 2011 *Project Management: A Managerial Approach*
- [5] Mantel S J, Meredith J R, Shafer S M, Sutton M M 2013 *Project Management in Practice*
- [6] Kerzner H 2013 *A System Approach to Planning scheduling and controlling*
- [7] Yourdon E 2001 *Managing High-Intensity Projects 1st Edition*

Author



Ilana Ter-Saakova, 1975, Riga, Latvia

Current position, grades: Chairman of a Board AS Alida Turs
University studies: PhD BSI
Publications: 4

4 Conclusions

An expression has been obtained for the project duration defined by the maximum permissible project time with regard to the selected reserve.

The minimum permissible value of the probability P_{add} was defined for the project reliability, as well as the values of the time reserves that ensure that the project reliability is close to one.

An equation has been obtained for solving the problem of the coordinated selection of innovative and investment projects, which optimises the selection of the best set of projects with the given financial and time constraints.

As the cost and the duration of the project are an important subject in the coordination of interests between the customer and the project developer, the obtained dependences of these functions with account of the project reliability can be the basis of the coordination of the parties at the stage of concluding an engineering contract.

Earth as a building material for reproduction of ancient buildings in China Meishan cultural park

Jingde Luo^{1*}, Xiaofei Wang², Junhui Luo³

¹School of Arts and Design and School of Architecture, Hunan University, Changsha, 410082 China

²School of Civil Engineering, Dalian University of Technology, Dalian 116024 China

³Institute of Geotechnical Engineering, Southeast University, Nanjing 210096, China

*Corresponding author's e-mail: jingde.luo@gmail.com

Received 1 March 2015, www.cmnt.lv

Abstract

Rammed earth was vastly used for ancient buildings in the Meishan region in the central south part of China. However, these rammed earth buildings are rarely found nowadays. During the construction of Meishan Cultural Park, a park that reproduces the ancient architectures and cultures of Meishan region, the rammed earth architectures need to be reconstructed. In this paper, preliminary designs of these buildings were investigated. Specifically, the procedures to build the wall and to select the soils were introduced first. Then the seismic resistance behaviours of a typical rammed earth building in the ancient Meishan region was investigated using finite element modelling. The designed rammed earth walls were found to be able to satisfy the strength requirements of modern building standards. This paper also shows that the modelling of rammed earth buildings can be geotechnical in nature, as opposed to simply structural analysis.

Keywords: Rammed earth, Soil selection, Ancient buildings, Cultural park, Finite element modelling

1 Introduction

The rammed earth buildings have been constructed widely in the rural area of Human, China because of the availability of these construction materials and their good performances of maintaining indoor weather including temperature and humidity etc.. Air humidity in a room has a significant impact on the well-being of inhabitants and earth is one of the best material to balance indoor humidity (Pacheco-Torgal and Jalali 2012). This is because that earth material is a porous material that has the capacity to absorb and desorb from air and thus is able to balance the indoor humidity. The effectiveness of this balancing depends on their speed of absorbing and desorbing. In other words, rammed earth is a green material with high energy efficiency.

Two key issues regarding the reproduction and potential revival of this type of buildings are that: 1) selection of suitable soil types that can be used for rammed earth construction; 2) such structures that made of local soil are often susceptible to earthquake damage because of their heavy weight and low strength (Liu et al. 2015).

This paper firstly proposed a building procedure of using local earth to build the walls. The processes of the local earth and the mix design of the wall material were investigated. Secondly, a typical rammed earth structure in the cultural park under seismic loading was simulated using ANSYS (ANSYS Inc, Canonsburg, PA, USA) and ABAQUS (Dassault Systemes, France) and the deformations of the building were analysed. These guidelines, written in the style of a submission CM&NT, discuss how to prepare your paper using Microsoft Word. In addition to the usual guidance on style/formatting, there are notes and links to assist in using some of Word's features such as inserting graphics, formatting equations and so forth.

2 Methods and Materials

2.1 THE BUILDING PROCEDURES OF RAMMED EARTH WALL

Every building has many components such as the walls, the roof and the floor amongst others. But in the context of the rammed earth building, the walls are the most important structure that peoples are interested in.

The supporting base (i.e., foundation) for traditional rammed walls normally are constructed using stones gathered from the locality of the building site. These supporting bases not only provide support for the walls but also prevent the wall from being affected by the adverse effects of ground water. In this design, stones were used keep in line with the traditional design methods. Of note, this kind of supporting base was proved to be strong enough to bear the loading of the whole building.



FIGURE 1 The forms and rammer used to build rammed earth walls in southern China

During the building of the wall, each form is filled first, then another form is placed above it, and the process begins again. This process is repeated until the wall finished. The lower forms can be taken off as soon as the form above is begun. Pneumatic rammers were used to compact the earth within the forms. Figure 1 shows the forms and the rammer used to build the rammed earth wall in south part of China.

2.2 SELECTION OF SOIL

The composition of soils in different regions varies considerably due to the origin and the climatic conditions. It was shown that rammed earth allows a wide variety of soils with different composition. However, careful selection of soil is still necessary to control its quality.

In China, there is no official code that provides guidelines for the selection of soil for rammed earth. In this project, we used the maximum and minimum values of content of clay, silt, sand and gravel that are recommended to aid selection by (Shang 2005, Steve 2010). Extensive experiments have shown that as long as each substance is in the recommended range, the strength and durability of the earth wall are likely to meet the requirements. In this project, sandy loam sub-soils directly taken near the construction site were used. Topsoil is not used because it contains too much organic matters and it is highly compressible. The substances designed for this project are: clay (20%), silt (20%), sand (30%) and gravel (30%). A detailed study carried out at University of Bath indicates that the compressive strength of about 1.0–3.0 MPa can be obtained with unstabilized rammed earth (Maniatidis and Walker 2003) easily.

The mechanical behaviour of rammed earth were affected by various factors. Compaction is the most important factor that influences its stiffness. One significant factor that influences the compaction is the moisture content. If moisture is not enough, the wall may not have sufficient green strength to be able to withstand the mechanical disturbances mainly during the installation or uninstallation processes of forms. If there is too much water in the soils, the soil may become very sticky and may hamper the compaction process. The water content can also affect the density to a certain extent. The water content used in this project is 10% according to previous research (Bahar et al. 2004). The water content can be tested through a simple drop test in situ, which is demonstrated very useful (Morel et al. 2001). In the drop test, a ball of soil is made in the palm and then the ball is dropped on to a floor, from about one meter height. If the ball breaks into four to five pieces, the water content is good. If it crumbles away, the soil is too dry or if it stays as one pat, it is too wet.

In summary, this project will use the proposed mix design of clay, silt, sand and gravel and the water content of 10%.

3 Finite element modelling of a typical rammed earth building under seismic loading

A typical rammed earth building that is going to be built in China Meishan Cultural Park was used for this study. The building is a one-story building which is a widely used building type in ancient Meishan region. These kinds of buildings were mainly used for the ancient tea industry.

In this study, response spectrum analysis and time history analysis were performed to evaluate the structural behaviours of the rammed earth building under seismic loadings. Response spectrum analysis is fast and is recommended by the design code of China. However, time history method can give more accuracy information including all sources of the nonlinear and time dependant material geometry effect although it is very computational expensive. Therefore, it is better to use the two methods combined to investigate the design of the rammed earth building. In this section, finite element modelling (FEM) was used to perform these studies.

3.1 GEOMETRY

The height of the building is 8 m and the wall thickness is 240mm. One door (900mm×2100mm) and four windows (400mm×600mm) were placed on the front wall. The architectural design of the building is shown in Figure 2. In the finite element analysis, only the rammed earth building was considered. The adjacent lower building was not considered.



FIGURE 2 Geometry of the rammed earth building (left)

3.2 MATERIAL PROPERTIES

The rammed earth material is treated as a homogeneous and isotropic material. This hypothesis has been proved to be adequate to model the rammed earth in a previous study (Bui et al. 2009).

Rammed earth can be modelled using the Mohr-Coulomb failure criterion (Nowamooz and Chazallon 2011). Mohr-Coulomb model is an elastic-perfectly plastic model, which is often used to model soil behaviour in general and serves as a first-order model. The stress-strain curve in the elastic range is linear with two parameters: Young's modulus, E and Poisson's ratio, ν .

In the Mohr-Coulomb failure criterion, the material failure is controlled by the maximum shear stress, which depends on the normal stress. The Mohr-Coulomb criterion can be written as:

$$\tau = c + \sigma \tan \phi, \quad (1)$$

where τ is the shear stress, σ is the normal stress, c is the cohesion of the material, and ϕ is the material angle of friction.

The material constants used for the rammed earth are: Young's modulus $E = 60\text{MPa}$, Poisson's ratio $= 0.3$, cohesion $c = 150\text{kPa}$, friction angle $= 45$ degrees, angle of dilation $= 0$ (Jaquin et al. 2006).

The roof was mainly made of wood. Therefore, the

material properties of wood were used for the roof with a elastic modulus of 100 MPa and Poisons' ratio of 0.3. The density of the roof was 710 kg/m3 (Tankut et al. 2014).

3.3 MESHING

The geometry of the building used to simulate the seismic behaviours were discretised into 8,461 8-node hexahedral elements (Figure 3). All the parts in the finite element model were connected by sharing their boundary nodes. The mesh density was proved to be fine enough to provide reasonable accurate results.

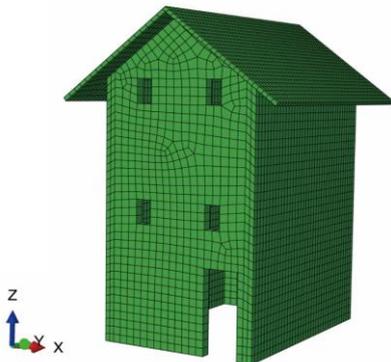


FIGURE 3 Finite element mesh of the rammed earth walls and roof

Two simulations were performed to investigate the seismic behaviours of this rammed earth building. Firstly, a model analysis was used to determine the frequency of the building. Response spectrum analysis, which is recommended by the China design code for normal buildings (GB5001 2001), was performed based on the model analysis results. Secondly, ground acceleration data from a rare earthquake was applied to the structure directly. The detailed loading scenarios were introduced below.

3.4 RESPONSE SPECTRUM ANALYSIS

Response spectrum analysis is a simplified method to evaluate the rammed earth building under seismic loading. There are two steps in running a response spectrum analysis. First a modal analysis was needed to obtain the modes of the structure. Secondly response spectrum analysis was performed using the models obtained in the model analysis as inputs. ANSYS Workbench (Ansys Inc., Canonsburg, PA, USA) was used to obtain the natural frequency and mode shapes of this building.

In response spectrum analysis, the key step is to make a response spectrum. In this paper, a standard response spectrum recommended by The China code for seismic design of buildings (GB5001 2001) was used. The response spectrum is shown in Figure 4.

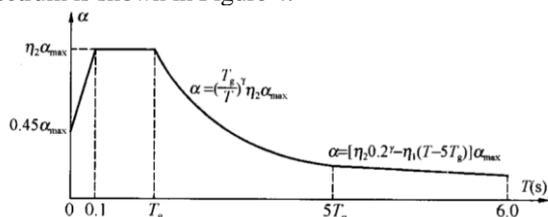


FIGURE 4 The standard response spectrum recommended by China design code (GB5001 2001)

The parameters used in the response spectrum curve are determined by the site class and building class etc.. These parameters can be determined from China code for seismic design of buildings (GB5001 2001) and seismic ground motion parameter zonation map of China (GB18306 2008). The parameters are listed in Table 1. The final response spectrum (Figure 5) was obtained based on these parameters.

TABLE 1 The parameters used to determine the response spectrum

Parameters	Values
α_{max}	0.04
η_1	0.02
γ	0.9
T_g	0.35
η_2	1

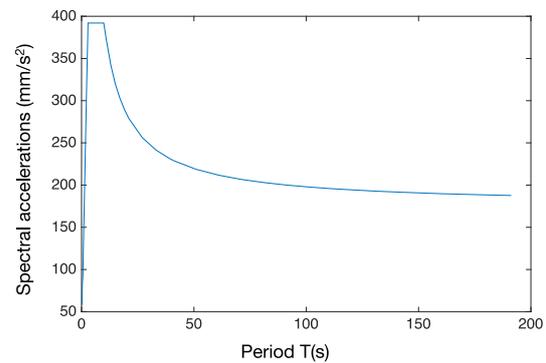


FIGURE 5 Response spectrum of seismic acceleration according to China design code(GB5001 2001)

3.5 TIME HISTORY ANALYSIS

In the response spectrum analysis, only the elastic behaviors can be included. In order to see the rammed earth building's plastic deformations and to see the building's strengths under very rare earthquakes, a time history analysis was performed. In the time history analysis, ground motion recordings of the 2008 Wenchuan Ms8.0 earthquake was applied on the building to examine the buildings behaviors under extreme loading conditions. The ground data used in this research was observed at a station that is about 100 km from the epicenter as illustrated in Figure 6.

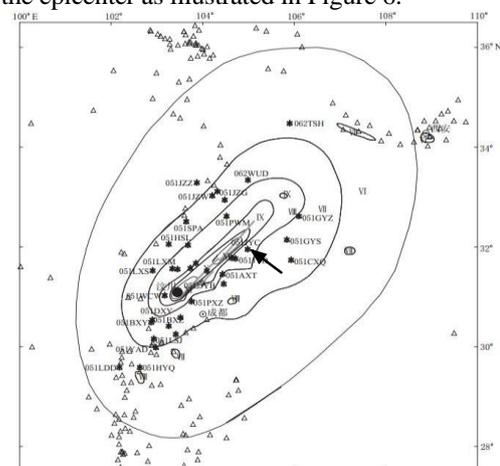


FIGURE 6. The seismic ground acceleration data used in this study was observed at the Jiangyou station as indicated by the black arrow.

The original ground motion data (Figure 7) were smoothed using a Butterworth low pass filter and then resampled to reduce the data points to allow a reasonable computational time in the simulation. Only the most detrimental part (with high acceleration; 36 seconds) of the ground data was applied to the structure in the simulation (Figure 8).

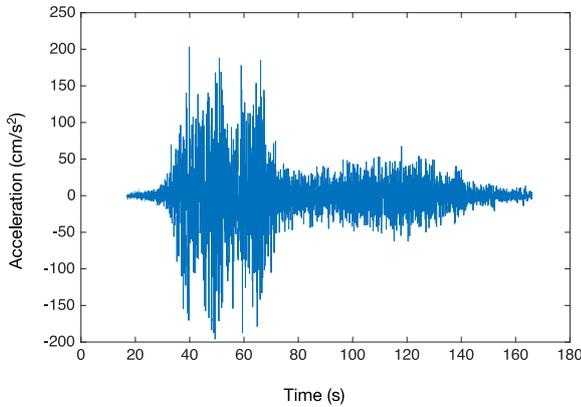


FIGURE 7 The original ground motion recordings of the 2008 Wenchuan Ms8.0 earthquake

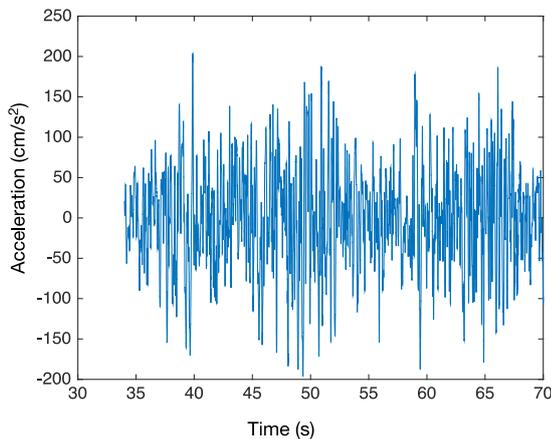


FIGURE 8 The ground acceleration data used in this analysis. Note that only the most detrimental section of the original data was used

Before any seismic loading, an initial stress state corresponding to the gravity load was applied to the structure. The ground acceleration was applied to the horizontal direction of the building (x direction in Figure 3).

4 Results

4.1 RESPONSE SPECTRUM ANALYSIS

Because the ground acceleration is applied in the x-direction in the response spectrum analysis, one needs to make sure that the effective mass in the x-direction is higher than 90% of the total mass. In the mode analysis, the results show that by requesting 8 modes there was 100% participating mass in the x-direction, which meets the requirements for the response spectrum analysis. Moreover it is found that 1 of the modes (mode 1) are contributing with 90% of the effective mass and consequently it can be expected that the earthquake response will be dominated by the first mode.

The frequencies and are listed in Table 2. The first three vibration modes are illustrated in Figure 9.

TABLE 2. The frequencies of the rammed earth building.

Mode	1	2	3	4	5	6	7	8
Frequency	1.4146	1.4619	1.9694	2.2163	2.8516	2.8934	2.9247	2.9792

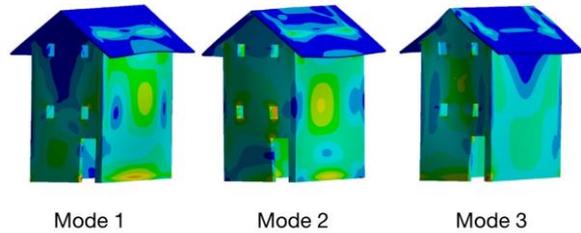


FIGURE 9 The first three vibration modes of the rammed earth building

The results of the x direction (horizontal) displacement of the rammed earth walls are presented in Figure 10. The middle part of the sidewall experienced the largest deformation. However, these deformations are relative small and will not affect the safety of the building. The stress levels were all way smaller than the yield stress. Therefore, the rammed earth walls meet the requirements of the China design code for normal buildings (GB5001 2001).

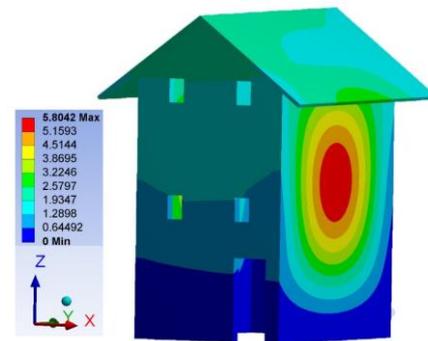


FIGURE 10 The x displacement of the rammed earth building (Unit: mm)

4.2 TIME HISTORY ANALYSIS

Because the ground accelerations of an extreme rare earthquake were applied on this structure, plastic deformations start to appear in this rammed earth wall after 36 seconds of seismic loading. The plastic strain distribution of this building under rare earthquake loadings is displayed in Figure 11. The von Mises stress distributions were also shown in Figure 12. The plastic strain is a good indicator to understand where plasticity yield occurs. One can observe, at the end of the step (36s), that plastic strains appear near the center and bottom of the sidewall.

The equivalent stresses were larger at the sidewall and the front wall, especially near the door and the windows. The plastic strain was not observed at the back wall in this building. Therefore, under extreme loading conditions, the building will experience plastic deformations but plastic deformations only present at a very small part of the walls. It should be noted that in this time history analysis, the seismic loadings were from an extreme severe earthquake and thus were way larger than those of the response spectrum analysis. Therefore, their results should be different.

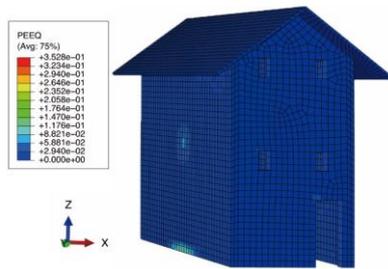


Figure 11 Plastic strain distribution in the rammed earth building under a rare earthquake

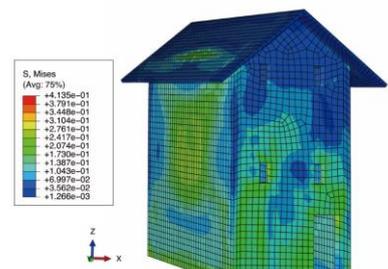


Figure 12 Von Mises stress distribution after a rare earthquake loading

5 Conclusions

This paper proposed a preliminary design of rammed earth buildings in China Meishan region. The proposed design can use the affordable, locally sourced sustainable construction materials. The proposed design of using rammed earth is of useful for rescuing the heritage and also it will promote the use of a rediscovered environmentally friendly building material.

A typical rammed earth building under seismic loadings was simulated through response spectrum analysis and time history analysis. The numerical modelling was performed with ANSYS and Abaqus finite element code. The designed rammed earth building is safe under the most possible earthquake conditions in that region and it is also safe under an extreme rare earthquake.

In this paper, the rammed earth was treated as a homogeneous single-phase material. Mohr-Coulomb failure criterion was used in the time history analysis. However, the rammed earth wall can be two-phase (soil and water). To become useful in the building field, rammed earth requires standards which will have to take into account the complex hydro-mechanical behaviour of this material. Therefore, the analysis of rammed earth building can be geotechnical in nature, rather than simple structure analysis.

References

- [1] Bahar R, Benazzoug M, Kenai S 2004 *Cement and Concrete Composites* **26**(7) 811-20
- [2] Bui Q B, Morel J C, Hans S, Meunier N 2009 *Materials and Structures* **42**(8) 1101-16
- [3] GB5001 2001 *Construction Ministry of PRC: Code for Seismic Design of Buildings* China Architecture & Building Press
- [4] GB18306 2008 *General Administration of Quality and Technology Supervision of the People's Republic of China: Seismic ground motion parameter zonation map of China* Beijing: China Standards Publishing House
- [5] Hall M, Djerbib Y 2004 *Construction and Building Materials* **18**(4) 281-6
- [6] Jaquin P, Augarde C, Gerrard C 2006 *Structural Analysis of Historical Constructions* P Lourenco New Delhi
- [7] Keefe L 2005 *Earth Building: Methods and Materials Repair and Conservation* Taylor & Francis
- [8] Liu K, Wang M, Wang Y 2015 *Construction and Building Materials* **100** 91-101
- [9] Maniatidis V, Walker P 2003 *A Review of Rammed Earth Construction* University of Bath
- [10] Morel J C, Mesbah A, Oggero M, Walker P 2001 *Building and Environment* **36**(10) 1119-26
- [11] Nowamooz H, Chazallon C 2011 *Construction and Building Materials* **25**(4) 2112-21
- [12] Pacheco-Torgal F, Jalali S 2012 *Construction and Building Materials* **29** 512-9
- [13] Shang J 2005 *A Study of Optimization of the Ecological Building Material System of Traditional Rammed Earth Dwellings* PhD Xi'an University of Architecture and Technology
- [14] Steve B 2010 *Journal of Green Building* **5**(1) 101-14
- [15] Tankut N, Tankut A N, Zor M 2014 *Drvna industrija* **65**(2) 159-71
- [16] Venkatarama R, Jagadish K S 2003 *Energy and Buildings* **35**(2) 129-37
- [17] Williams-Ellis C, Eastwick-Field J C, Eastwick-Field E 1947 *Building in cob, pisé, and stabilized earth* Country Life

Authors

	<p>Jinge Luo, 1986, China</p> <p>Current position, grades: PhD candidate University studies: Architecture Scientific interest: Architecture, Rural Buildings, Construction Technologies Publications: 3 Experience: PhD candidate in Human University, China, research filed is rural buildings in Hunan province, China, experience in designing of modern and ancient buildings.</p>
	<p>Xiaofei Wang, 1987, China</p> <p>Current position, grades: Master student University studies: Structural Engineering Scientific interest: Concrete Structure, New Construction Materials Publications: 4 Experience: a bachelor degree of Civil Engineering, experience in computational simulation and experimental research on novel concrete structure and rammed earth structures.</p>
	<p>Junhui Luo, 1984, China</p> <p>Current position, grades: PhD candidate University studies: Geotechnical Engineering Scientific interest: Soil Mechanics Publications: 4 Experience: research on soil and rock mechanics for many years, computational and experimental experiences of research on soil structures.</p>

AUTHORS' INDEX	
Bellucci S	7
Burlutskaya N	7
Fink D	7
Gopeyenko V	7
Kiv A	7
Koycheva T	7
Lobanova-Shunina T	7
Luo Jinge	40
Luo Junhui	40
Mansharipova A	7
Mrochko A	30
Muhamediev R	7, 14
Mykytenko N	7
Shunin Yu	14
Ter-Saakova I	35
Wang Xiaofei	40
Zhukovskii Yu	7

REVIEWS

The nature of oscillations of ion currents in the ion track electronics

D Fink, A Kiv, Y Shunin, N Mykytenko, T Lobanova-Shunina, A Mansharipova, T Koycheva,
R Muhamediev, V Gopeyenko, N Burlutskaya, Y Zhukovskii, S Bellucci

Computer Modelling & New Technologies 2015 19(6) 7-13

The paper contains the description of the main features of ion current pulsations in track devices. The conditions under which the pulsations arise are discussed. We describe different approaches that are used for interpretation of the effect of ion current pulsations. In particular the generalized model of current spikes in track devices is considered. To create this model a special modification of the classical molecular dynamics was developed. The results of application of this model coincide with the main experimental data concerning the ion current pulsations in track devices.

Keywords: track devices ion current pulsations nanoporous membranes

Machine learning methods: An overview

Ravil I Muhamedyev

Computer Modelling & New Technologies 2015 19(6) 14-29

This review covers the vast field of machine learning (ML), and relates to weak artificial intelligence. It includes the taxonomy of ML algorithms, setup diagram of machine learning methods, the formal statement of ML and some frequently used algorithms (regressive, artificial neural networks, k-NN, SVN, LDAC, DLDA). It describes classification accuracy indicators, the use of "learning curves" for assessment of ML methods and data pre-processing methods, including methods of abnormal values elimination and normalization. It addresses issues of application of ML systems at the processing of big data and the approaches of their solution by methods of parallel computing, mapreduce and modification of gradient descent.

Keywords: machine learning learn ability preprocessing big data map reduce

MATHEMATICAL AND COMPUTER MODELLING

Modelling of the subsystem of estimation of navigational parameters in automatic vehicle control systems

A Mrochko

Computer Modelling & New Technologies 2015 19(6) 30-34

One of the most actively developing spheres of applying modern information technologies is transport. Divisions of different services, departments and organisations are actively introducing and employing the system of *Automatic Vehicle Location (AVL)*. The information systems of Automatic Vehicle Location solve the task of controlling and guiding transport means. Employing modern telecommunication technologies along with satellite navigation systems (SNS) facilitates and improves controlling of the mobile objects (MO). Modelling of the work and analysis of these systems' efficiency indicators sufficiently reduces the periods and costs of their testing and introducing in a particular region.

Keywords: information technologies satellite navigation systems mobile objects controlling AVL systems

Optimisation of coordination's selection by innovation and investment projects

Iana Ter-Saakova

Computer Modelling & New Technologies 2015 19(6) 35-39

The steady onward progress of engineering developments, together with increasing competition, implies the need for the development of implementation for novelties and innovations. This involves a huge number of innovative projects emerging, thus consecutively defining the formation of investment criteria

Keywords: innovations investments project intensity probability model function

Earth as a building material for reproduction of ancient buildings in China Meishan cultural park

Jinge Luo, Xiaofei Wang, Junhui Luo

Computer Modelling & New Technologies 2015 19(6) 40-44

Rammed earth was vastly used for ancient buildings in the Meishan region in the central south part of China. However, these rammed earth buildings are rarely found nowadays. During the construction of Meishan Cultural Park, a park that reproduces the ancient architectures and cultures of Meishan region, the rammed earth architectures need to be reconstructed. In this paper, preliminary designs of these buildings were investigated. Specifically, the procedures to build the wall and to select the soils were introduced first. Then the seismic resistance behaviours of a typical rammed earth building in the ancient Meishan region was investigated using finite element modelling. The designed rammed earth walls were found to be able to satisfy the strength requirements of modern building standards. This paper also shows that the modelling of rammed earth buildings can be geotechnical in nature, as opposed to simply structural analysis.

Keywords: Rammed earth, Soil selection, Ancient buildings, Cultural park, Finite element modelling